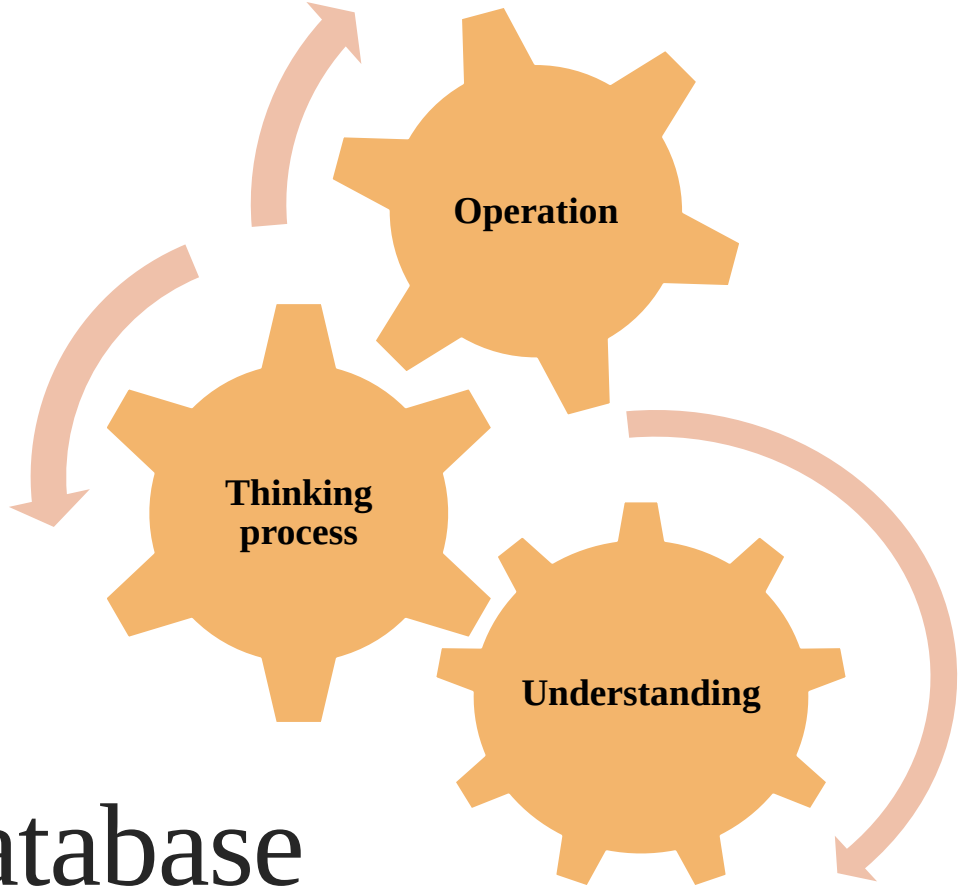


SAS Programming:

National Health Insurance Database

20190530

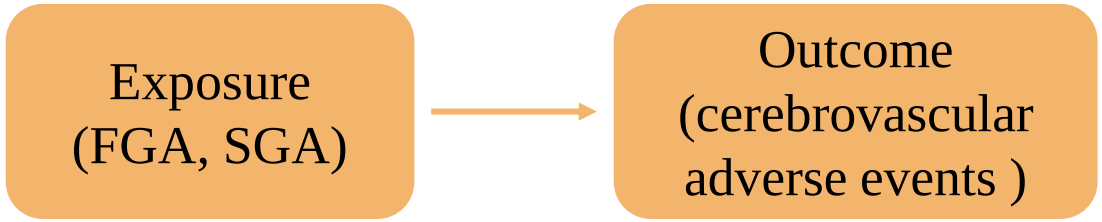
臨床藥學與藥物科技研究所

陳儒伶

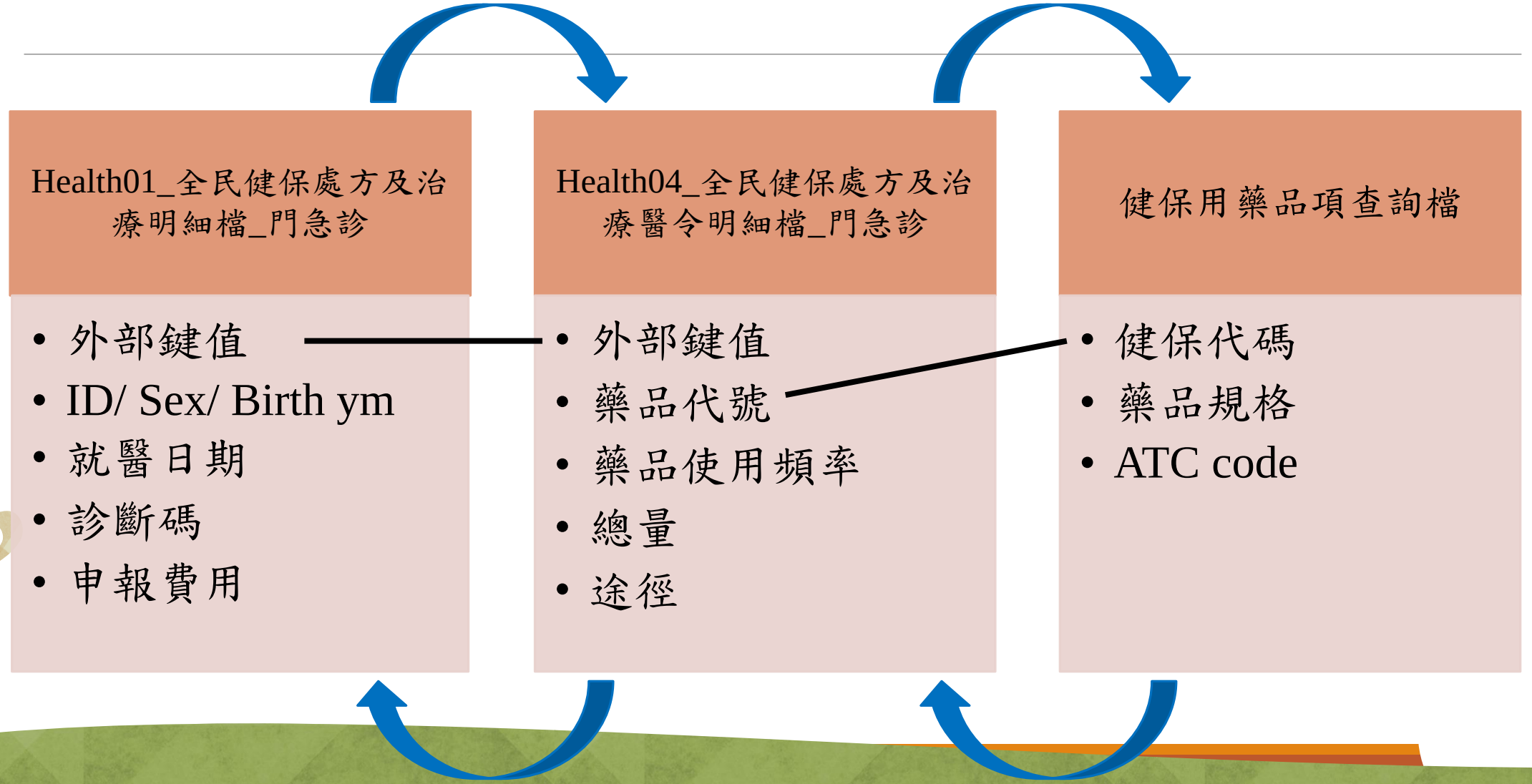
Research question

Is the risk of cerebrovascular adverse events with second-generation antipsychotic (SGA) users lower than those who taking first-generation antipsychotics (FGA) in older adults?

- Retrospective cohort study



1st step – extract antipsychotic user



基本語法 – if … then…

條件限制

配合substr function 進行條件篩選，篩選出antipsychotic agents (N05A) 之健保代碼。

```
data n05a; set a.nhi_atc;
        if substr(atc_7,1,4)='N05A';  /*selection of antipsychotics*/
        if substr(atc_7,1,5)='N05AN' then delete; /*exclude lithium*/
run;
```

基本語法 – if … then…else

將資料進行分組，分為first or second generation antipsychotic agents

```
data n05a; set a.nhi_atc;
        if substr(atc_7,1,7) in ('N05AE04', 'N05AH02', 'N05AH03',
        'N05AH04' , 'N05AH06','N05AL05' , 'N05AX08' ,'N05AX09',
        'N05AX11' ,'N05AX12','N05AX13')
        then index_class='SGA'; else index_class='FGA';  /**grouping**/
        index_agent=atc_7;
run;
```

條件設定應互斥且周延

資料合併 merge

將資料做橫向連結 (增加變數variable)

```
proc sort data=c.n05a; by drug_code; run;
proc sort data=a.H_nhi_opdto9501_10; by drug_no; run;
```

注意：使用MERGE時，要記得先將資料依據"**鍵值**"排序(PROC SORT)。

```
data opdto_9501;
   merge a.H_nhi_opdto9501_10 (in=x)  c.n05a(in=y rename= (drug_code=drug_no));
   by drug_no;
   if x and y;
run;
```

x,y 可以自訂，但不可與dataset中的variable名稱相同

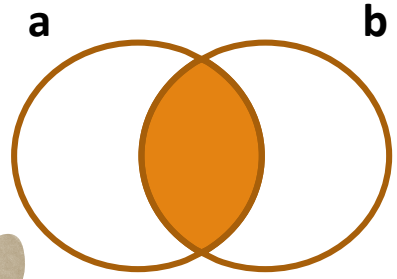
依據"**鍵值**"連結

兩個資料表間取"**交集**"

資料合併

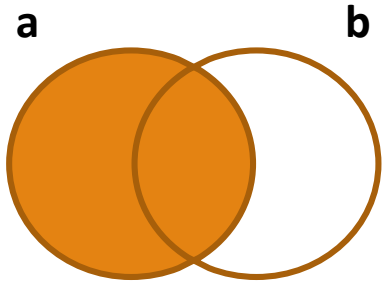
SAS語法為例：
DATA xxx ; MERGE a b;
If.....

交集



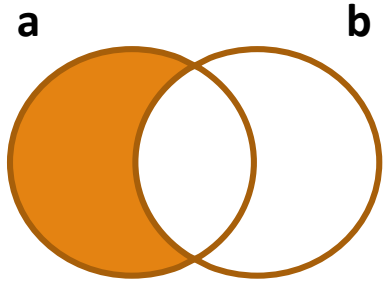
if a and b;

$a \cup (a \cap b)$



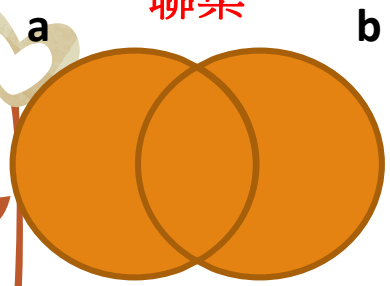
if a;

$a \cap b'$



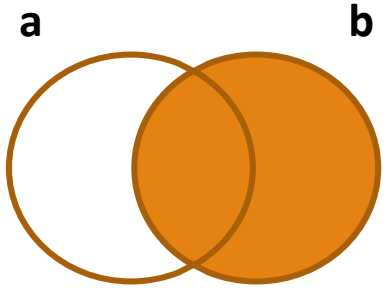
if a and not b;

聯集



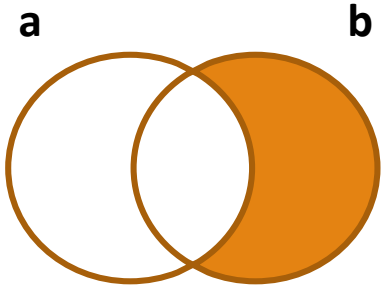
無需用 if

$b \cup (a \cap b)$



if b;

$a' \cap b$



if b and not a;

Proc sql

資料合併前不需要sort，參照的欄位名稱可以不相同

```
proc sql;
   create table b.master1 as select
   a.*, b.eligibility, b.out_date
   from b.master a
   left join idfile3 b on a.id=b.id
;
quit;
```

新資料集名稱

要留下的variable，彼此之間以逗號間隔。
若要保留原dataset所有欄位，則以*表示:如a.*

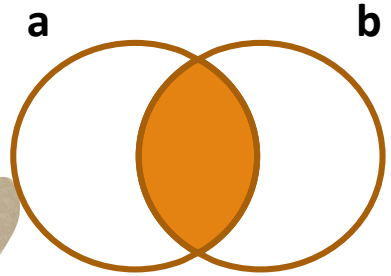
要合併的資料集，on要參照的var 。若要依據多個var連結資料，以and串聯；如：
on (a.id=b.id and a.sex=b.gender)

Proc sql 需以quit 結尾

資料合併 (Join)

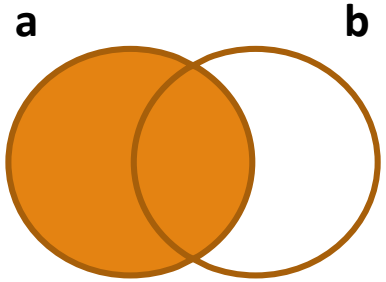
SAS與法 SQL 程序：
PROC SQL;
CAREAT TABLE xxx as
SELECT* FROM yy as **a** …JOIN… zz as **b** …
on **a**.key=**b**.key ……

交集



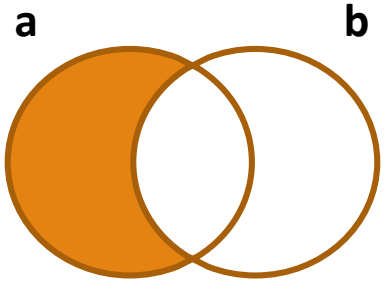
inner join

$a \cup (a \cap b)$



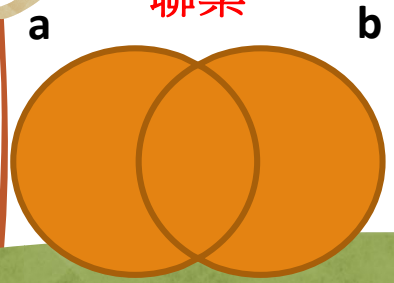
left join

$a \cap b'$



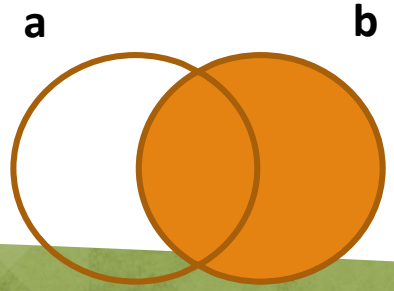
left join…
…where
b.key is NULL

聯集



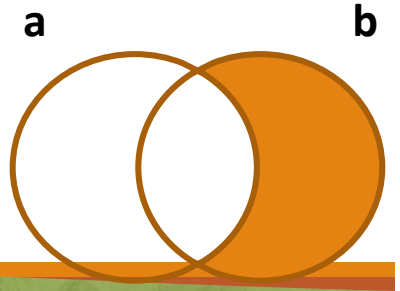
full join

$b \cup (a \cap b)$



right join

$a' \cap b$



right join…
…where
a.key is NULL

資料合併 set

將資料做縱向連結 (增加observation)

```
data car6;
set cars3 cars4;
run;
```

VIEWTABLE: Work.Cars3

	make	model	mpg	weight	price
1	AMC	Concord	22	2930	4099
2	AMC	Pacer	17	3350	4749
3	AMC	Spirit	22	2640	3799

VIEWTABLE: Work.Cars4

	make	model	mpg	weight	price
1	AMC	Concord	22	2930	4099
2	AMC	Pacer	17	3350	4749
3	AMC	Spirit	22	2640	3799
4	Buick	Century	20	3250	4816
5	Buick	Electra	15	4080	7827

VIEWTABLE: Work.Car6

	make	model	mpg	weight	price
1	AMC	Concord	22	2930	4099
2	AMC	Pacer	17	3350	4749
3	AMC	Spirit	22	2640	3799
4	AMC	Concord	22	2930	4099
5	AMC	Pacer	17	3350	4749
6	AMC	Spirit	22	2640	3799
7	Buick	Century	20	3250	4816
8	Buick	Electra	15	4080	7827

➢ 欄位名稱與格式需相同

	make	model	mpg	weight	price	var1	var2	var3	var4
1	AMC	Concord	22	2930	4099				
2	AMC	Pacer	17	3350	4749				
3	AMC	Spirit	22	2640	3799				
4			.	.	.	200101	1	0b64808157449e660c2b767d8bf2cce	200102
5			.	.	.	200101	1	7ead6cd0b8d12f97fc285ae5f8c68dc2	200102
6			.	.	.	200101	1	a78800b3ace1fdd28af8505332f5d41f	200102
7			.	.	.	200105	1	1a71b6fddc9c366d9d48f5e356eae18	200106
8			.	.	.	200104	1	6f0db153524d2b1babda6030b0fa42b	200105

```
1061  data test2;
1062  set cc test1;
ERROR: Variable func_ym has been defined as both character and numeric.
1063  run;
```

%MACRO Statement-1

使用於重複執行的程序

```
data n05_95;
    set opdte_9501-opdte_9512;
run;
data n05_96;
    set opdte_9601-opdte_9612;
run;
……
```

給定此macro名稱

要改變的variable名稱

```
%macro yr(x);
data n05_&x.;
set opdte_&x.01-opdte_&x.12;
run;
%mend;
%yr (95);
%yr (96);
%yr (97);
%yr (98);
```

在macro內，要改變的variable 需以 **&x.** 呈現

%macro開始，%mend結尾。

呼叫macro執行內容

%MACRO Statement-2

```
%macro yr(y1, y2);
%do x=&y1. %to &y2.;
data n05_&x.;
    set opdte_&x.01-opdte_&x.12;
run;
%end;
%mend;
%yr (95, 98);
```

```
%macro yr;
%do x=95 %to 98;
data n05_&x.;
    set opdte_&x.01-opdte_&x.12;
run;
%end;
%mend;
%yr ;
```

Do loop 迴圈

1. IF .THEN DO; END;（根據某條件做很多事）像是設定虛擬變數

```
IF Edu=1 THEN DO ; D1 = 0; D2 = 0 ; D3 = 0; END;
IF Edu=2 THEN DO ; D1 = 1; D2 = 0 ; D3 = 0; END;
IF Edu=3 THEN DO ; D1 = 0; D2 = 1 ; D3 = 0; END;
IF Edu=4 THEN DO ; D1 = 0; D2 = 0 ; D3 = 1; END;
RUN;
```

2. 搭配macro進行反覆運算 (%do)

   %DO *macro-variable=start* %TO *stop* ;

   text and macro language statements

   %END;

Find first / last record

```
proc sort data=n05_index1;
   by id index_date;run;
proc sort data=n05_index1 nodupkey
out=b.master; by id; run;
```

日期由先至後排，去重複後留下最早日期的那一筆

```
proc sort data=n05_index1;
   by id descending index_date;run;
proc sort data=n05_index1
nodupkey out=b.master; by id; run;
```

日期由後至前排，去重複後留下最晚日期的那一筆

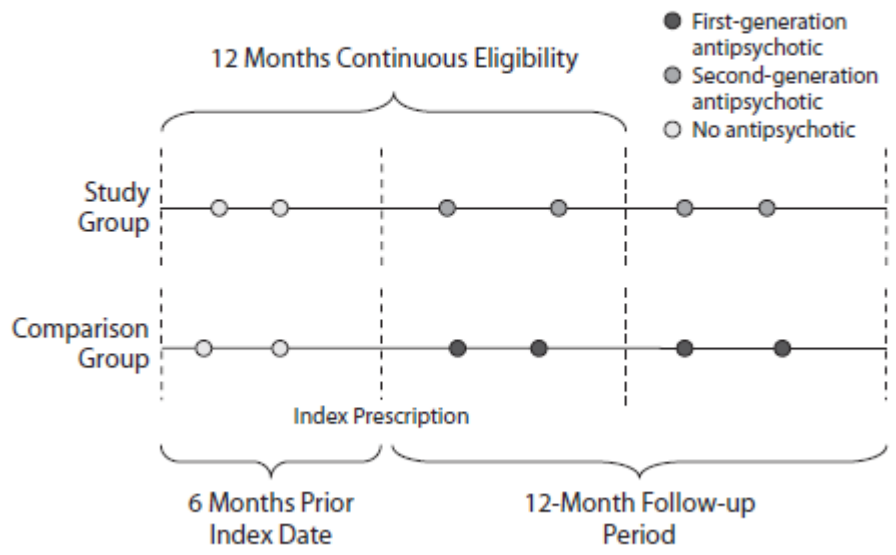
```
proc sort data=n05_index1;
   by id index_date;run;
data test1;
   set n05_index1;
   by id;
   if first.id;
run;
```

If last.id; 取最後一筆

Inclusion criteria

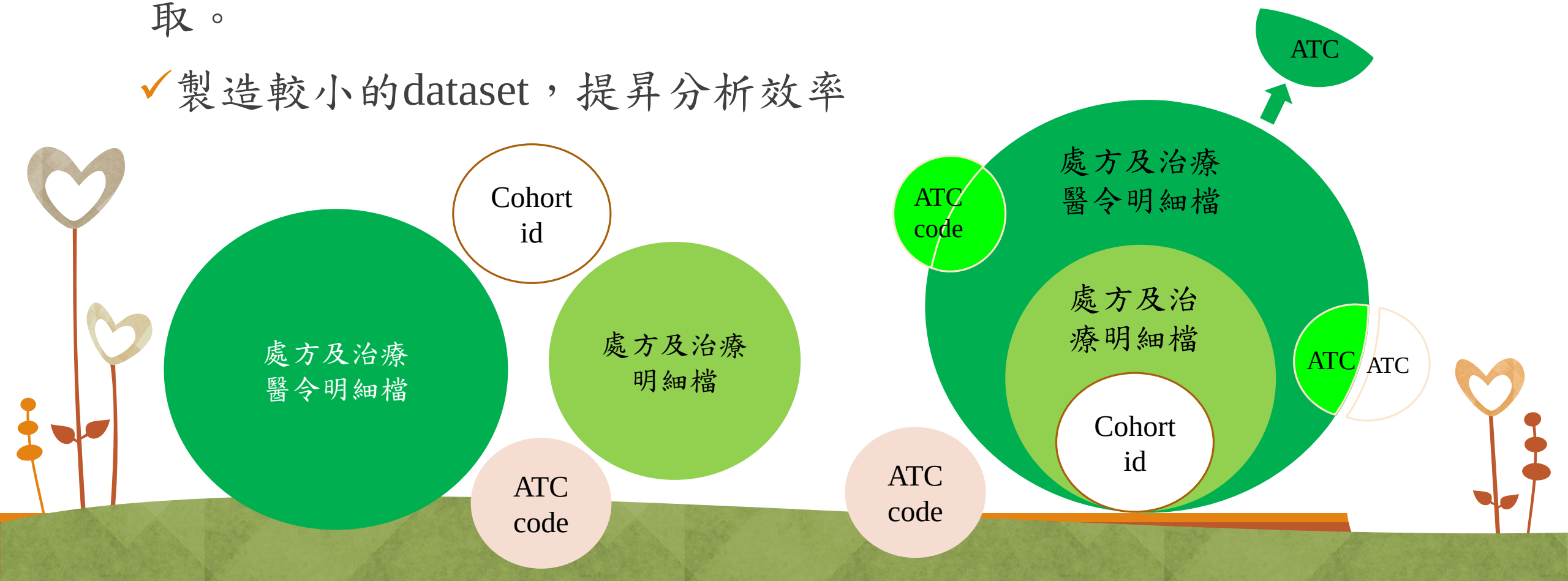
1. All adults aged 50 years and above
2. On antipsychotic medication from July 1, 2000, to December 31, 2007
3. New user
4. Eligibility

Figure 1. Cohort Identification for Study Group and Comparison Group



2nd step
extract all medical record of target population

- ✓利用目標研究對象之ID串聯其所有就醫紀錄，以利後續變數擷取。
- ✓製造較小的dataset，提昇分析效率



Outcome

- Cerebrovascular events (stroke)
- Discontinuation
- Switching
- Death
- End of continuous eligibility
- End of maximum follow-up