

華人家庭動態資料庫簡介

訪問架構與資料串連方式



黃奕嘉
2015.9.16

大綱

- × 樣本結構與訪問狀態
- × 跨期資料串連
- × 未來工作概況
- × 東南中國調查

台灣調查概況-樣本結構簡介

×四群主樣本

- ×1999：35-46歲（1953-1964年出生）
- ×2000：46-65歲（1935-1954年出生）
- ×2003：27-39歲（1964-1976年出生）
- ×2009：26-32歲（1977-1983年出生）

×子女（主）樣本

- ×子女主樣本：1976年以後滿25歲的主樣本子女。
- ×子女樣本：滿16歲未滿25歲的主樣本子女。
- ×子女主樣本的子女（第三代）不進行追蹤。

主樣本抽樣—1999及2000年樣本

× 抽樣範圍：台灣地區

× 1999年以35歲至46歲為對象，預計完成1,000案，實際完成999案。

× 2000年增加46歲至55歲及56歲至65歲二個年齡層，預計分別完成1,000案，最終共完成1,959案。

× 抽樣方法：

× 使用「分層三段等機率」抽樣原則抽取合格樣本，並另外抽出1.5倍~2倍之備取樣本。

× 以羅啟宏先生所著之「台灣省鄉鎮發展類型之研究」的分層原則進行分層，以1997年底內政部人口統計資料進行抽樣。

主樣本抽樣—2003年

- × 抽樣範圍：台灣地區
- × 預計訪問對象及樣本數：27-39歲國民，預計完成樣本數為1,300案，實際完成1,152案。
- × 抽樣方法：
 - × 依據1999年的抽樣村里抽取合格樣本
 - × 依照每一分層以往之完成率，採用膨脹比例方法抽取樣本個案，膨脹比例介於1.6倍至2.1倍之間，每一抽取樣本個案皆為必訪對象。

主樣本抽樣—2009年

- × 訪問範圍：台灣地區
- × 預計訪問對象及樣本數：26-32歲國民，預計完訪2,000案，實際完成2,092案。
- × 抽樣方法：分層等機率三階段抽樣法（probability proportional to size, PPS），依據侯佩君等人「台灣鄉鎮市區類型之研究」進行地區分層，並依據2008年6月人口統計檔計算每一層級的人口比例，抽出鄉鎮、村里，最後採用膨脹比例方法抽取樣本個案，膨脹比例介於1.5倍至2.5倍之間。

樣本流失：主樣本，臺灣 (%)

1953-1964 Cohort

調查次	訪問年月	成功樣本	流失樣本	流失率	累積流失率
1	1999.2	999	—	—	—
2	2000.2	802	197	19.72	19.72
3	2001.2	726(651)	76	7.65	26.97
4	2002.2	782(607)	-56	-5.63	21.33
5	2003.2	751(481)	31	3.12	24.45
6	2004.2	706(440)	45	4.53	28.98
7	2005.2	677(419)	29	2.92	31.89
8	2006.2	663(396)	14	1.41	33.30
9	2007.2	637(372)	26	2.62	35.92
10	2008.2	616(353)	21	2.11	38.03
11	2009.2	594(340)	22	2.21	40.24
12	2010.2	564(324)	30	3.02	43.26
13	2011.2	554(310)	10	1.01	44.27
14	2012.2	545(300)	9	1.62	45.17
15	2014.2	518(288)	27	4.95	47.89

流失率：主樣本，臺灣 (%)

1935-1954 Cohort

調查次	訪問年月	成功樣本	流失樣本	流失率	累積流失率
1	2000.2	1,959	—	—	—
2	2001.2	1,730	229	11.69	11.69
3	2002.2	1,642(1,548)	88	4.49	16.18
4	2003.2	1,566(1,193)	76	3.88	20.06
5	2004.2	1,489(1,074)	77	3.93	23.99
6	2005.2	1,396(967)	93	4.75	28.74
7	2006.2	1,341(896)	55	2.81	31.55
8	2007.2	1,288(841)	53	2.70	34.25
9	2008.2	1,259(795)	29	1.48	35.73
10	2009.2	1,202(755)	57	2.91	38.64
11	2010.2	1,158(717)	44	2.25	40.89
12	2011.2	1,144(672)	14	0.71	41.60
13	2012.2	1,066(642)	26	2.38	45.58
14	2014.2	977(593)	89	8.35	50.13

樣本流失：主樣本，(%)

1964-1976 Cohort

調查次	訪問年月	成功樣本	流失樣本	流失率	累積流失率
1	2003.2	1,152	—	—	—
2	2004.2	832	320	27.78	27.78
3	2005.2	808(681)	24	2.08	29.86
4	2006.2	768(602)	40	3.47	33.34
5	2007.2	745(556)	23	1.99	35.33
6	2008.2	716(511)	29	2.52	37.85
7	2009.2	704(486)	12	1.04	38.89
8	2010.2	685(462)	19	1.65	40.54
9	2011.2	654(430)	31	2.69	43.23
10	2012.2	652(419)	2	0.31	43.40
11	2014.2	602(396)	50	7.69	47.74

樣本流失:主樣本，臺灣 (%)

1977-1984 Cohort

調查次	訪問年月	成功樣本	流失樣本	流失率	累積流失率
1	2009.2	2,182	—	—	—
2	2010.2	1,721	461	21.13	21.13
3	2011.2	1,685(1,550)	36	1.65	22.78
4	2012.2	1,644(1,459)	41	2.43	24.66
5	2014.2	1,539(1,343)	105	6.39	29.47

台灣調查概況-樣本結構簡介

× 子女樣本

- × 樣本來源：抽樣主樣本所回答的子女資訊。
- × 滿 16 歲未滿 25 歲採用子女問卷，隔年進行訪問（問卷代號以 C 開頭）。每次訪問樣本數約為 800~1,000 筆
- × 主樣本子女於滿 25 歲時，採用主樣本第一次受訪的 RCI 問卷訪問。每次訪問樣本數約為 200 筆左右。之後與抽樣主樣本一同追蹤。

× 父母樣本

2000 年時，針對 1999 年抽樣主樣本的父親（或母親）進行訪問，但之後未持續追蹤。該次共訪問 260 位主樣本父母

× 兄弟姊妹樣本

2000 年時，針對 1999 年抽樣主樣本的兄弟姊妹（擇一）進行訪問，但之後未持續追蹤。該次共訪問 271 位受訪者

台灣調查概況-樣本結構簡介

× 問卷類別與訪問對象

- × RI問卷：第一次訪問主問卷。1999、2000、2003抽樣主樣本，首波訪問使用。
- × RCI問卷：滿25歲子女及2009抽樣主樣本採用RCI問卷。問卷內容以RI2003為藍本進行修訂。主要的目的為受訪者進入主樣本追蹤的第一年蒐集受訪者基本資訊。
- × RR問卷：主問卷。2000~2004以R開頭後以羅馬字標明訪問次數及訪問年（例如RII2000）。2005起統一以RR開頭後接訪問年（RR2006）。
- × C問卷：子女問卷。以C開頭，後接羅馬字標示訪問次數，後接訪問年。針對滿16歲子女訪問。主要問題在於就學的歷程。

台灣調查概況-樣本結構簡介

× 受訪者受訪歷程

- × 抽樣主樣本：第一年接受RI（RCI）問卷訪問，其後每年以RR問卷進行追蹤（2012年以後改為隔年追蹤）
- × 抽樣主樣本的子女：民國65年以後出生的子女，當其滿16歲時，以C問卷進行訪問，採隔年追蹤。當其滿25歲時以RCI問卷進行訪問，訪問成功後，比照抽樣主樣本進行追蹤。
 - × 2005年後RCI問卷與C問卷隔年交互進行訪問。RCI問卷每次訪問兩個年齡層，因此，部分年齡層樣本C與RCI間間隔2年。
 - × 如果受訪者於應訪問RCI年度未完訪，將持續以RCI問卷嘗試接觸，直至訪問過RCI問卷後，才比照主樣本進行追蹤。
 - × 例外：2009年前，部分子女於滿25歲時未訪問RCI問卷，之後變以RR問卷進行追蹤。2011年時，嘗試補回相關資訊。(107/142)

台灣調查概況-樣本結構簡介

華人家庭計畫歷年調查樣本及問卷結構圖

受訪者年 次(民國)	調查年(西元)															
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2014	
24-43		RI-2000	RII-2001	RIII-2002,	RIV2003,	RVI2004,										
42-53	RI-1999	RII-2000	RIII-2001	RIV2002	RV2003	RV2004,										
53-65					RI-2003 (新抽樣)	RII2004	RR2005	RR2006	RR2007							
65																
66						RCI-2004				RR2008	RR2009					
67-68												RR2010	RR2011			
69							RCI-2005							RR2012	RR2014	
70		CI-2000														
71				CII-2002					RCI-2007							
72						CIII-2004										
73											RCI-2009					
74																
75								CIV2006								
76																
77										CV2008						
78																
79-80																
81-82																
83-84																
85-86																
其他	FI2000 (RI1999及RI2000的父母於2000年的調查) SI2000 (RI1999及RI2000的兄弟姊妹於2000年的調查)															

備註：藍色應為主機本；黃色應為子女機本；綠色應為新抽樣主機本或主機本屆25歲子女機本首次納入主機本處理；同欄框內代表使用同一份問卷訪問。

計畫首頁->調查架構及資料介紹->資料關係示意圖

<http://psfd.sinica.edu.tw/web/check.htm>

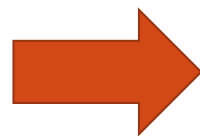
樣本編號的變革與擴充

- × 調查訪問之初，樣本編碼為5碼數字編碼（3碼地區編號+2碼流水號），2003年擴充樣本並整合子女樣本編碼，擴充為7碼數字編碼。
- × 利用樣本編號規則可以分辨樣本屬於哪一群樣本群
 - × 1999年抽樣樣本群：第2、4、7碼皆為0
 - × 2000年抽樣樣本群：第2碼大於0，第4、7碼為0
 - × 2003年抽樣樣本群：第4碼為1，第7碼為0
 - × 2009年抽樣樣本群：第4碼為2或3，第7碼為0
 - × 子女樣本、子女主樣本：第7碼大於0
- × 樣本編碼前6碼可視為戶號，相同者為同一家人。
- × SI2000及PI2000仍維持原始的5碼數字編碼。

樣本編號的變革與擴充

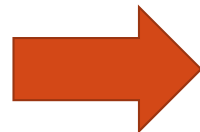
範例

× 0010100



1999樣本

× 1030030



2000樣本

× 1140110



2003樣本

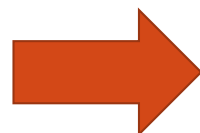
× 0131010



2009樣本

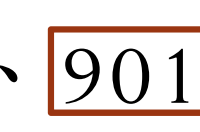
× 1041170

× 7012030



× 5243320

× 9010080



9010082

一家人

跨期資料串連-基本串連程序

× 篩選所需變數：

- × 若變數有無變動跳答設計，則需一併考慮先前各期資料。
- × 確認各期資料選項或過錄規則是否一致，若不一致需重新編碼。

× 整理受訪者編號

- × 確認受訪者編號皆為7碼版本
- × 確認受訪者編號的欄位屬性及長度一致

× 變數重新命名：各期變數名稱除受訪者編號外，不可重複。

- × 不同波次的同一變數給予相同名稱，後接年度區別(occ_2005)

× 資料合併：

- × 各期變數以受訪者編號排序
- × 串連資料

× 跳答資訊補回

跨期資料串連-跳答資訊補回

× 哪些題目需跨期找答案？

- × 出生年次與性別：2001~2005年主問卷，從第三波訪問起未詢問受訪者的性別及出生年次。
- × 受訪者工作資訊：若受訪者工作沒有變動或僅部分變動，視情形略過「為誰工作」、「公司人數」、「行職業」等資訊。
- × 配偶基本資訊及工作資訊：若婚姻狀態無變動，略過結婚（離婚\分居\喪偶）時間、配偶出生年次、最高教育程度等資訊。配偶工作跳題情形與受訪者部分相同。

跨期資料串連-補回出生年次及性別

- × 由第一波(RI)或第二波(RII)問卷中補回。
- × 若歷年問卷出生年次不同該如何處理？
 - × 問卷設計希望受訪者回答「實際」的出生年次，但受訪者可能有時回答「身份證」上的出生年次，或國曆/農曆的年次，而產生差異。
 - × 從資料檢索及合併系統下載基本資訊。
- × 性別不一致該如何處理？
 - × 請聯絡計畫小組確認。

跨期資料串連-補回工作資訊

- × 一般補回原則：若受訪者工作無變動，則
 - × 本期工作資訊 = 前一期工作資訊（2003年除外）。
 - × RR2003工作資訊，在RR2004問卷裡。
 - × RR2003詢問年譜表，該年度的工作資訊於2004年補問。
 - × RR2004詢問所有人當時的工作狀況（未設計跳答），然後再詢問其2003年的工作狀況。

- × RI1999、RI2000、RII2000、RI2003、RR2004、RI2009無須回補

- × 其餘資料補回順序：2000 ← 2001、2001 ← 2002、
2003 → 2004、2004 ← 2005、2005 ← 2006.....

跨期資料串連-補回工作資訊

A6. 請問您目前的主要工作與 100 年 1 月的主要工作有何不同？（可多選）

- | | |
|----------------------------|-------------------------------------|
| ① 雇主改變 <u>(續問 A7a 題)</u> | ② 100 年 1 月後才開始工作 <u>(跳問 A8a 題)</u> |
| ③ 工作內容改變 <u>(跳問 A9a 題)</u> | ④ 職位改變 <u>(跳問 A9a 題)</u> |
| ⑤ 工作地點改變 <u>(跳問 A9b 題)</u> | ⑥ 沒有改變 <u>(跳問 A9b 題)</u> |
| ⑦ 其他 (請說明) _____ | <u>(跳問 A8a 題)</u> |

- × 雇主改變：所有工作資訊完整詢問
- × 去年訪問後才開始工作：不問離開前一份工作原因
- × 工作內容或職位改變：省略「為誰工作」、「公司人數」，但有行職業資訊
- × 沒有改變：省略「為誰工作」、「公司人數」、「行職業」

跨期資料串連-補回工作資訊

A12.請問您在去年（民國 92 年）的一月間是否有工作？

(1) 有（跳問第 A14 題）

(2) 沒有

A14.請問您去年（民國 92 年）一月所做的工作，與今年（民國 93 年）一月所從事的主要工作，是否為同一份工作？

(1) 是同一份工作（跳問第 A16a 題）

(2) 不是同一份工作

25

A16a.如果您繼續從事去年(民國 92 年)一月份的工作，請問您的職位、工作地點或工作內容有改變嗎？（可多選）

(1) 職位改變

(2) 工作地點改變

30 31

(3) 工作內容改變

(4) 沒有改變（跳問第 A19a 題）

32 33

✗ RI2003 受訪者未詢問此部分。

✗ 如工作完全相同，則省略行職業、公司人數、為誰工作、工作地、上下班交通時間。可以同份問卷A5a~A6b詢問目前的狀況補回。

跨期資料串連-補回工作資訊

- × 若受訪者訪問曾中斷，則工作資訊補回可加入工作年資條件後，嘗試利用更前期資料補回資訊。

A11b. 請問您這份工作做了多久了？ _____ 年 _____ 月

- × 如果受訪者的工作年資超過2年，則可考慮以再前一期的工作資訊替代回補。
- × 行職業碼變動：
 - × 1999-2008採用自訂行職業編碼（參考iso code 88及主計處代碼修訂）
 - × 2009年起採用主計處行職業代碼（職業碼第五次修訂版，行業碼第八次修訂版）
 - × 2011年起採用主計處行職業編碼（職業碼第六次修訂版，行業碼第八次修訂版）
 - × 2012起採用主計處行職業編碼（職業碼第六次修訂版，行業碼第九次修訂版）

跨期資料串連-補回婚姻及配偶資訊

× 部分婚姻狀態並未設計跳答，不需回補資訊。

A16. 請問您目前結婚了嗎？

- ① 未婚（跳問第 14 頁 B. 居住安排）
- ② 同居
- ③ 已婚（第一次結婚），民國 _____ 年結婚
- ④ 離婚再婚，民國 _____ 年再婚
- ⑤ 喪偶再婚，民國 _____ 年再婚
- ⑥ 分居（跳問第 14 頁 B. 居住安排）
- ⑦ 離婚（跳問第 14 頁 B. 居住安排）
- ⑧ 喪偶（跳問第 14 頁 B. 居住安排）
- ⑨ 其他（請說明）_____（跳問第 14 頁 B. 居住安排）

跨期資料串連-補回婚姻及配偶資訊

- ✗ RIII2001、2002、2005、2011年開始，採用類似工作的方式設計，一直是已婚的情形略過配偶出生年次、最高教育程度等資訊。
- ✗ 補回的方式大致與工作相同。

A16a. 請問從（民國 99 年 1 月以後／從上次訪問您以來），您個人的婚姻狀況有沒有改變（例如，從單身變成已婚、離婚後再婚）？情形怎樣的呢？

- ① 沒有變化；仍是已婚（跳問 A18 題）
- ② 沒有變化；仍是同居（跳問 A18 題） ⑨⑥ ⑨⑧
- ③ 沒有變化；仍然未婚單身（跳問第14頁 B. 居住安排）
- ④ 沒有變化；仍然離婚（跳問第14頁 B. 居住安排）
- ⑤ 沒有變化；仍然分居（跳問第14頁 B. 居住安排）
- ⑥ 沒有變化；仍然喪偶（跳問第14頁 B. 居住安排）
- ⑦ 有變化；初次結婚／再婚
- ⑧ 有變化；分居（跳問 A16c 題）
- ⑨ 有變化；離婚（跳問 A16c 題）
- ⑩ 有變化；喪偶（跳問 A16c 題）
- ⑨⑦ 其他（請說明）_____（跳問第14頁 B. 居住安排）

跨期資料串連-補回婚姻及配偶資訊

× 2003年婚姻狀態可利用2004年資訊替代轉換。

A24a.請問從民國 91 年 1 月以後，您個人的婚姻狀況有沒有改變？情形怎樣的呢？

- (01)沒有變化；仍是已婚（含同居）（跳問 A28 題）
- (02)沒有變化；仍然單身（跳問第 10 頁 B、進修教育情形）
- (03)有變化；在最近二年內結婚（續問 A25 題）
- (04)有變化；在最近二年內分居（跳問第 10 頁 B、進修教育情形）
- (05)有變化；在最近二年內離婚（跳問第 10 頁 B、進修教育情形）
- (06)有變化；在最近二年內喪偶（跳問第 10 頁 B、進修教育情形）
- (97)其他 _____（跳問第 10 頁 B、進修教育情形）

A24b.請問從民國 92 年 1 月以後，您個人的婚姻狀況有沒有改變？情形怎樣的呢？

- (01)沒有變化；仍是已婚（含同居）（跳問 A28 題）
- (02)沒有變化；仍然單身（跳問第 10 頁 B、進修教育情形）
- (03)有變化；在最近一年內結婚（續問 A25 題）
- (04)有變化；在最近一年內分居（跳問第 10 頁 B、進修教育情形）
- (05)有變化；在最近一年內離婚（跳問第 10 頁 B、進修教育情形）
- (06)有變化；在最近一年內喪偶（跳問第 10 頁 B、進修教育情形）
- (97)其他 _____（跳問第 10 頁 B、進修教育情形）

跨期資料串連-補回婚姻及配偶資訊

- × 訪問中斷情形，需參考結婚的時間，確認配偶是否為同一人。
 - × 對於沒有結婚時間的年度，對訪問中斷補回資訊較難判斷。
- × 再考慮配偶的工作年資，確認工作是否為同一個。

跨期資料串連-子女主樣本的串連

- × 子女主樣本的「源頭」在哪？
 - × 看年次
 - × 最早一次RR問卷訪問的前一次RCI訪問。
 - × 不管了！把所有的RCI都拿來串就對了！
 - × 如果還是串不到？！
 - × Sorry!!

合併家戶資料

- × 合併主樣本及其子女
 - × 以樣本編號前六碼進行合併
 - × 末碼為0者為主樣本。
 - × 末碼大於0為子女，其順序並非代表排行，請自行按出生年排序。
 - × 子女的出生年次與主樣本回答的子女資訊不合？
 - × 建議以子女自己回答的年次為準。
- × 合併主樣本及其父母(PI2000)、兄弟姊妹(SI2000)
 - × 父母及兄弟姊妹由於後續不再追蹤，並未另外編定樣本編號，僅需將編碼更新為7碼後，直接以樣本編號串連即可。

未來展望與規劃

- × 台灣調查：自 2012 年起，主樣本將與子女樣本同步採用隔年調查的方式追訪，預計於 2016 年進行下一波調查
- × 2011 年以前資料已完成釋出。進行 2012、2014 訪問資料清理及釋出。
- × 進行行職業碼重新編碼
- × 進行資料加權
 - × RI1999、RI2000 加權已釋出
 - × RI2003、RI2009 加權進行中，預計明年可完成。
 - × 研究追蹤資料的加權
- × 進行跨期資料檢核。
- × 建構官方版的合併資料。

東南中國調查

× 樣本結構：

- × 樣本出生年的範圍與台灣相同，出生於1935-1976間。
- × 每兩年面訪，2004 and 2006, 但第二次追蹤在 2011，2013正進行第三次追蹤。

× 調查地區：上海、浙江、福建

× 抽樣：四階段隨機分層抽樣

- × “縣”或區
- × 都市化程度分成四大類
- × 鄉鎮或市區
- × 村或街道
- × 家戶：以 Kish 方法抽出家戶中的受訪者。

東南中國調查

- ✘ 調查執行：由臺灣 PSFD 團隊與中國社科院人口與勞動經濟研究所合作
- ✘ 樣本規模：2004年首次訪問共完成福建地區1,925案，浙江地區1,856案，上海地區903案

調查次	訪問年	成功樣本	流失樣本	流失率	累積流失率
1	2004	4,684	—	—	—
2	2006	4,370	314	6.70	6.70
3	2011	3,871	499	11.42	17.36
4	2013	3,627	224	5.79	22.57

東南中國調查

× 資料釋出

× 2004、2006年調查資料已於SRDA公開釋出。

× 2011年資料已完成檢誤，如有使用需求請洽計畫小組。

× 2013年資料正在檢誤中，完成後將比照2011年資料供研究使用。

× 未來規劃：持續與中國社會科學院人口與勞動經濟研究所合作推展調查，並預計於今年進行第5波訪問，同時擴增樣本。

謝謝各位聆聽！敬請指教！！

計畫聯絡方式：

Email: psfd@gate.sinica.edu.tw, cakewalk@gate.sinica.edu.tw

網站： <http://psfd.sinica.edu.tw/>

電話：02-27821693轉261, 269