

報告大綱

- ◆ 資料蒐集面對的挑戰
- ◆ 健保資料庫與資料蒐集
- ◆ 門診就醫與常住地
- ◆ 健保資料庫的應用
- ◆ 結論與討論

什麼是大數據？



大數據：大、快、雜、疑

- 大數據（Big Data；或譯為「巨量資料」）
在2010年由IBM所提出，涵蓋四個V：
 - 大量化(Volume)：至少TB及PB以上
 - 快速化(Velocity)：即時處理
 - 多樣化(Variety)：視頻、GIS等資料
 - 真實性(Veracity)：資料品質(2014年加入！)
- 許多人認為資料蒐集方法（如：抽樣）在大數據時代沒有存在的必要。
 - 迷思之一：「 $n = all$ 」！



大數據等於一切？

■ 無限母體 vs. 有限母體

→ 再多的資料還是樣本！（如消費者行為）

■ 資料品質仍是主要關鍵！

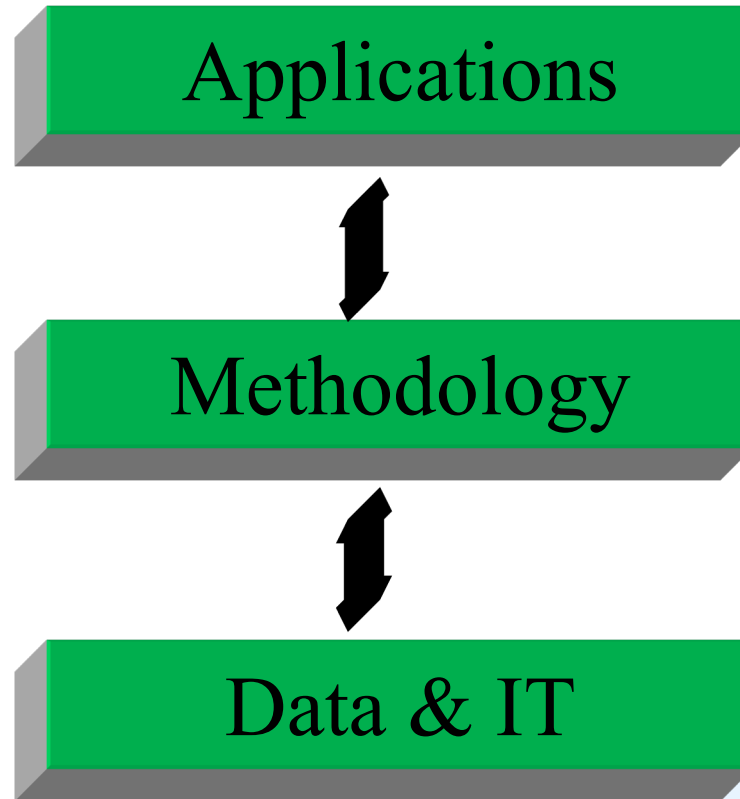
→ 人工智慧（AI）仰賴好的訓練資料，尋找背後的規則及趨勢（如 alphago）。

例如：Google Flu Trends透過搜尋引擎的「關鍵詞」，比美國疾管局（Center of Disease Control; CDC）更早偵測出流感盛行。

→ 族群及關鍵詞、從眾效應(Bandwagon effect)

大數據分析為跨領域合作

- 透過數量化分析，篩選出應用領域所需的重要訊息及知識。（三者配合！）



統計分析與資料蒐集

■ 資料筆數與變數個數

- 資料多可增加因應空間（資料縮減）；
- 尚無方法可系統性增加變數，變數設定與目標及問題有關（結構、非結構資料）。

■ 資料的多樣性（深度、廣度）

- 關鍵變數取決於問題需求。（如：內政部、衛福部；文字分析—文言文、白話文。）
- ◆ 問題：敏感性議題？（如：攤販、收入）

結構資料vs.非結構資料

- 結構資料(Hard Data)為容易量化的資料，以系統化方式蒐集；現行分析多針對結構資料。
- 非結構資料(Soft Data)無系統化的蒐集方法，定義方式取決於問題需要（如：質性資料）；但分析非結構資料尚無定論。

→ 文字分析(Text Mining)近年受到大眾青睞，但即便是中文「斷詞」也不容易。

例如：文言文、白話文
(中研院漢語平衡語料庫)



願付價值(Willingness-to-pay, WTP)

■ 條件評估法(Contingent Valuation, CV)用於探討WTP，範例之一支付卡(Payment Card)。

→例：若有種治療可使您的健康恢復至受傷前的狀態，您最多願意支付多少金額來接受治療？

☐ 0 ~ 10,000

☐ 10,001 ~ 20,000

☐ 20,001 ~ 30,000

☐ 30,001 ~ 40,000

☐ 40,001 ~ 50,000

☐ 50,001 ~ 60,000

☐ 60,001 以上

資料蒐集與大數據



http://www.abc.net.au/reslib/201204/r926508_9670000.jpg

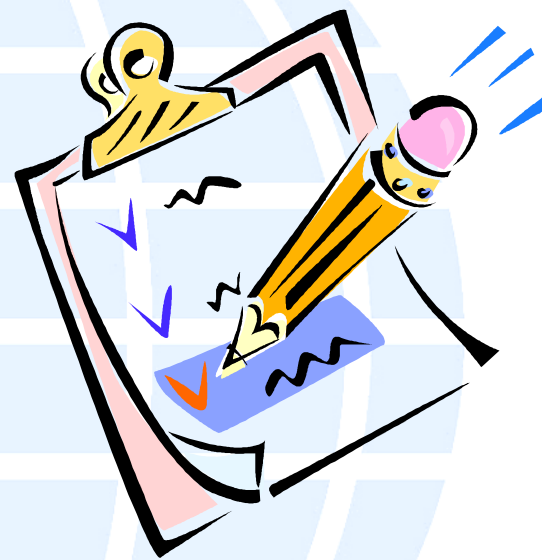
人口統計資料的來源

■ 臺灣的人口資料大多來自於：

1. 戶口普查(Census)
2. 抽樣調查(Surveys)
3. 公務統計(Official Statistics)

→ 生命統計(Vital Statistics ;
或譯為人口動態統計)

→ 戶籍登記(Population Register)





https://commons.wikimedia.org/wiki/File:1920_Empire_of_Japan_Census.jpg

<http://www.cpreview.org/blog/2018/1/political-consequences-from-a-void-in-census-leadership>

104年

農林漁牧業普查



聽頭家的心聲 顧土地的未來 105年4月1日至5月31日

主辦單位：行政院主計總處 協辦單位：縣市政府及鄉鎮市區公所



廣告



[https://www.stat.gov.tw/public/data/dgbas04/bc1/public/2015census/432%E8%BE%B2E6%9E%97%E6%BC%81%E7%89%A7%E6%A5%AD%E6%99%AE%E6%9F%A5%0%E5%AE%A3%E5%82%B3%E5%B8%83%E5%B9%94\(240x60cm\).jpg](https://www.stat.gov.tw/public/data/dgbas04/bc1/public/2015census/432%E8%BE%B2E6%9E%97%E6%BC%81%E7%89%A7%E6%A5%AD%E6%99%AE%E6%9F%A5%0%E5%AE%A3%E5%82%B3%E5%B8%83%E5%B9%94(240x60cm).jpg)

普查(Census)

- 普查指特定地區及特定時刻，對全部人口的自然與法律身份事項，及其有關的經濟與社會特定資料加以調查、蒐集、整理、分析等全部作業。
- 最早普查是西元前3800年巴比倫，以瓦片記錄維持人民的所需食物總數。
→ 埃及、中國、波斯、希臘、印度古代有類似普查的制度，紀錄稅收、徵兵等資訊。



普查（續）

- 古羅馬重視普查，普查官員的地位很高，每隔五年普查一次；羅馬城的居民必須向官方申報自己的相關資訊，否則可能被當作奴隸販賣。
- 中國戶口普查紀錄最早可追溯至西漢平帝元始二年（西元二年），當時約有1,236萬戶、5,767萬人，以現代標準算是非常精確。（註：宋徽宗時約12,000萬人）

中國古代的人口普查

- 《周禮》記：「小司徒之職，掌建邦之教法，以稽國中及四鄰都鄙之夫家九比之數，以辨其貴賤、老幼、廢疾

及頒比法於六鄰之大夫，使登其鄉之眾寡、六畜、車輦，辨其物，以歲時入其數，以施政教，行征令。及三年則大比，大比則受邦國之比要。」



中國古代的戶籍制度

- 在什伍組織的基礎上，進行戶口的登記與統計，記載於文書之上，便成為戶籍。
- 春秋戰國時的戶籍有兩種：版籍、計簿。
 - 版籍類似出生死亡記錄，由「司民」負責管全國人口的總數。
 - 計簿統計內容較為廣泛，是徵斂兵役、賦役的依據。
- 註：計簿調查項目包括性別、年齡、健康狀況、身份、財產等。

普查調查的對象

■ 兩種不同普查的方法：

1. de facto（實際）「現在人口」：紀錄普查當時的人口結構，聯合國推薦的方法。
2. de jure（法理或權利）「常住人口」：紀錄實際居住當地人口，我國、美國、加拿大等國使用的方法。

■ 過去我國普查之後辦理抽樣複查。

註：1990年採5%系統抽樣、2000年略過。

臺灣的戶口普查

- 中央政府於遷台後，於1956年9月16日舉辦第一次臺閩地區戶口普查，兼查「常住人口」及「現在人口」。
- 普查標準日在1965年延至12月16日。
- 臺閩地區每十年舉辦一次普查，最近一次全人口普查在2000年12月進行，由主計總處負責。（1980、1990年內政部負責。）
- 2010年12月以抽樣調查（16%）代替全面性的戶口普查。

美國的人口普查

- 美國的戶口普查每十年進行一次，在西元為「0」結尾的年份之四月一日進行。美國普查問卷有長、短問卷之分，約有 1/6 及 5/6 的比例填寫長、短問卷。

註：長卷問項的推估有賴抽樣設計。

→2010年改以美國社區調查(ACS；
American Community Survey)代替長卷，
在年內取得15%的樣本資料（每年3%）。

普查面臨的挑戰(台灣，7+6+2次)

1. 民眾配合度的降低

→都市化的程度愈高，完訪率愈低。

註：2000年普查的未回覆率約1/3。

2. 調查成本的上升

→2000年的預算原為二十七億元，因九二一地震而刪為不足十億。

3. 資料的精確性

→常住人口與戶籍人口何者為主？

2010年臺灣的戶口普查

- 因為經費及人力等諸多考量，雖然國內學術單位反對，2010年以公務統計、配合抽樣調查（16%），取代傳統的戶口普查。
- **問題**：公務統計是戶籍人口、抽樣調查是常住人口，兩者如何整併？（另外，缺乏常住人口的母體清冊，如何抽樣？）
- 個人戶籍地與居住地存在著一定程度的落差，差異性似乎逐年擴大（陳肇男與劉克智，2002；顏貝珊與余清祥，2011）。

臺灣普查的潛在問題：戶籍及常住差異

戶籍人口與居住人口差異狀況：以高雄縣為例

單位：千人

居住人口	狀況	戶籍登記		非戶籍登記		總計
	是	籍在人在		人在籍不在		1,186
		1,109	89%	77	6%	
	否	籍在人不在				
		135	11%			
	總計	1,244	100%			

資料來源：行政院主計處國內遷徙調查

戶籍人口與居住人口差異狀況：以八八風災小林村災情為例

單位：人

常住人口	狀況	戶籍登記		非戶籍登記		總計
	死亡	籍在人在		人在籍不在		434
		386	52%	48	11%	
	存活	籍在人不在				
		359	48%			
	總計	745	100%			

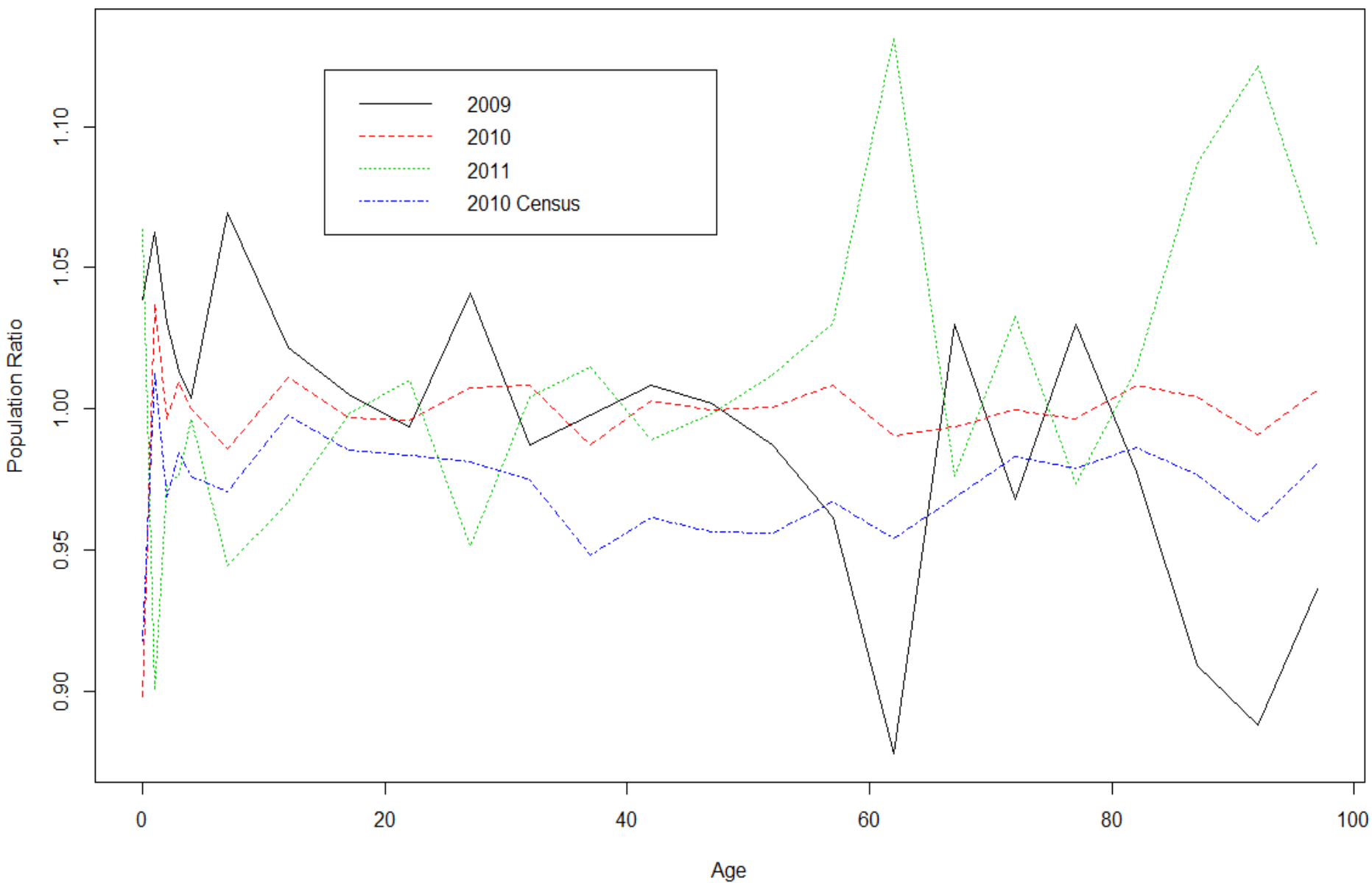
註：17鄰有1位外籍配偶死亡，無法列入戶籍人口。

資料來源：高雄縣甲仙鄉戶政事務所

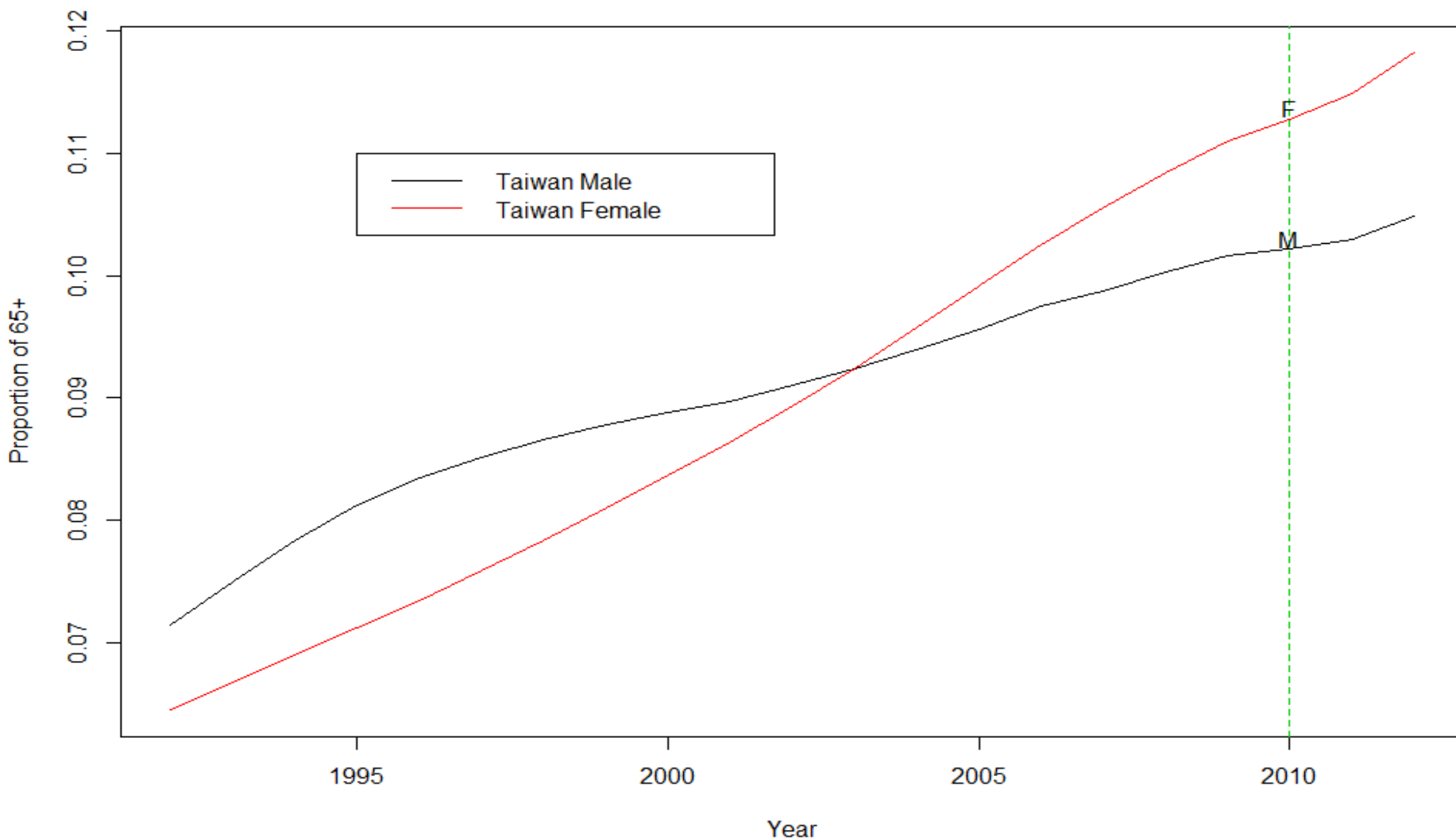
普查與戶籍人口的比較

- 檢視2009-2011年戶籍及普查人口，發現人數、年齡結構等都有不少差異。
 - 2010年普查僅涵蓋臺閩地區約16%人口，再以調查結果推估總人口數及其結構。
- 2010年戶籍資料與主計處公布的2010年普查結果最為接近，而2009-2011年三年平均的戶籍資料也算接近。
 - 戶籍人口數高出普查人口數 2%~3%；

普查與戶籍人口比較（男性2009-2011戶籍為分母）



歷年65歲以上人口比例（戶籍人口）



註：2010年普查65+比例均高於2010年年底戶籍人口

2003年老年人口普查（品質？）

■ 為了確定老年人口資料的正確性，臺灣政府在2003年8月至2004年2月對89歲以上的老年人進行普查，人數約有5萬人。有別於2000年戶口普查，本次普查的老人必須親自回答問題。（若無法回答，則由其照護者代答。）

→ 雖然耗費不少經費及人力，仍有不少老年人無法找到。

2003 臺灣老年人口普查結果

	89歲	90~94歲	95~99歲	100歲以上	加總
總人數	12,597	23,898	11,190	2,484	50,169
完訪數	92.17%	91.39%	88.38%	66.39%	89.67%
未完訪	7.83%	8.61%	11.62%	33.61%	10.33%

註：5,169 位未完訪

普查面臨的挑戰(美國，23次)

■ 民眾配合度的降低

→ 美國1990年普查郵寄問卷中約有 25.9% 的住戶未寄回。

註：2000年普查的未回覆率大約1/3。

■ 調查成本的上升

→ 1990年普查總預算26億美元。

註：2000/2010年普查花費45/130億美元。

■ 資料的精確性

→ Undercount 的比率升高。

普查替代方案？

- 除了上述隱憂外，無法取得即時資料、資料項目有限等也是普查的問題。
 - 美國與法國的滾動式普查(Rolling Census)，可解決部分即時及有限的需求。
- 我國的抽樣代替普查，並非分成五年、每年調查3%全國樣本，至多節省成本。
 - 問題：即時的常住人口可能嗎？
 - 臺灣學術界提議以健保推估常住人口。

健保資料庫、 資料蒐集



<http://h30507.www3.hp.com/t5/Journey-through-Enterprise-IT/Analyze-This-Big-Data-is-insurance-against-losing-a-competitive/ba-p/143577#.UgZmpLQVEqQ>

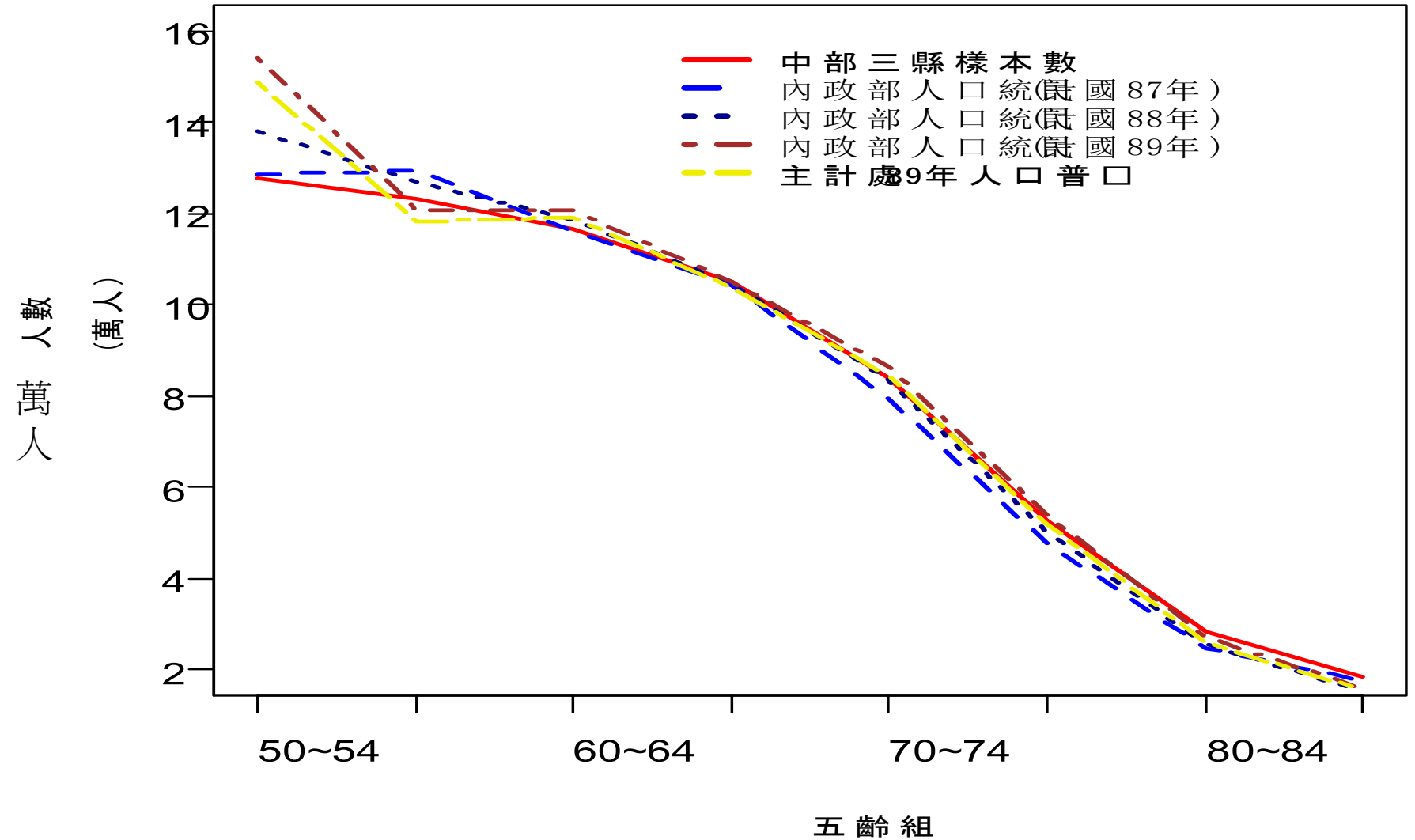
全民健保與人口資料

- 一代健保達成了醫療可近性(Accessibility)，學術界一般認為小病就醫在經常活動地區。例如：蔡文正等人(2003)及吳依凡(2004)。
 - 林民浩(2011)以上呼吸道感染之門診記錄，推論出70.1%被保險人居住的鄉鎮市區。
 - 林敬昇(2016)利用感冒記錄判定常住地，探討人口移動和醫療資源。
- 農民、地區人口投保地和居住地一致比例高，如：林民浩等人(2011)、連賢明(2011)。

評估健保就醫的人口資料

- 除了感冒外，也有學者提議以牙科、皮膚科作為常住地的參考依據。
- 問題：如何評估這些人口資料是否合宜？
- 以《921震災對中老年人死亡風險的影響》為例，參考陳芝嘉、余清祥、蔡偉德(2015)。
- 高齡人口在感冒就醫、戶籍登記、戶口普查三者的吻合度高。
- 問題：如何由健保資料中獲取全國人口的抽樣方法？（註：涵蓋各年齡層）

中部三縣、內政部與主計處



臺中縣、南投縣、彰化縣三種人口資料的比較

健保就醫與抽樣調查

- **研究目標**：「母體」為健保資料庫2005年百萬人抽樣檔(LHID2005)，尋求樣本代表性較佳的「抽樣」方式。
 - 根據健保就醫資料，找出就醫習性作為常住地推估的基礎，進而估計常住人口的特性。
- 問題：不同標準所定義之母體，本身未必擁有相同的人口特質。
 - 戶籍、常住、就醫、投保四種人口的異同？

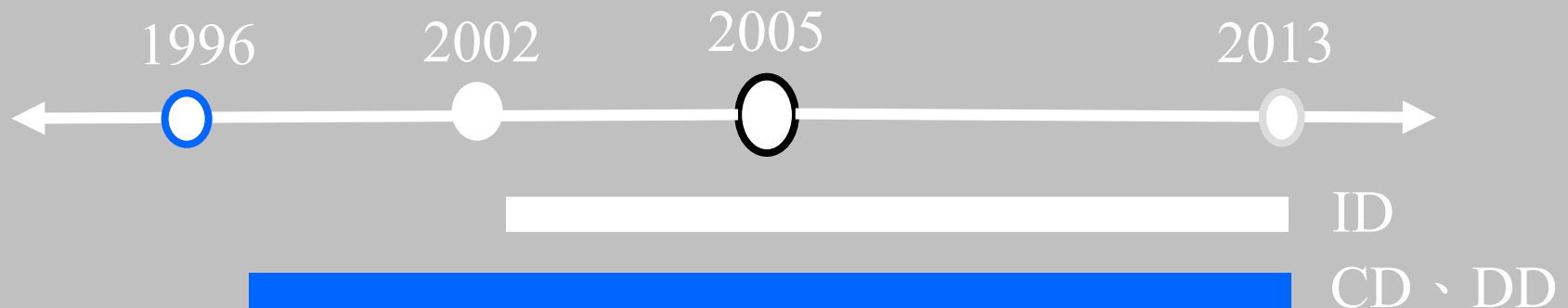
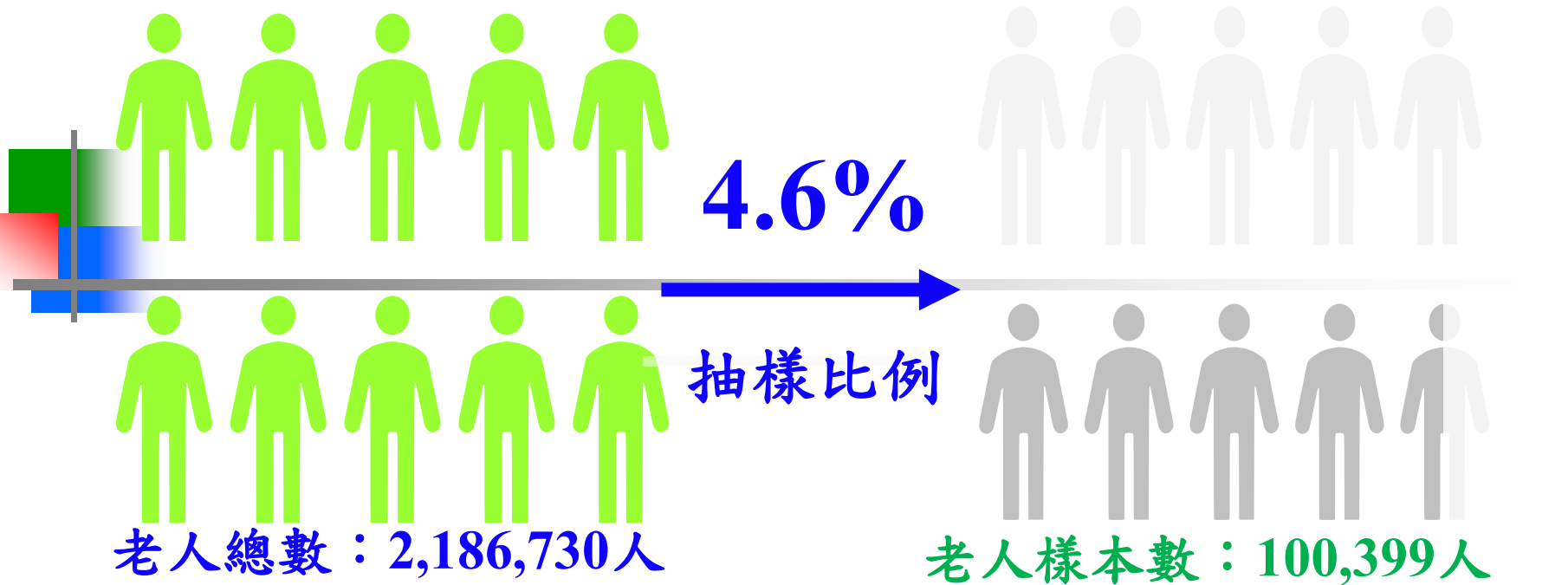
不同標準之人口資料特性

	戶籍人口	常住人口	投保人口	就醫人口
特色	紀錄完整、 有關權利義務	反映實際 生活狀況	健保資料	健保資料、 類似常住
限制	籍在人不在、 人在籍不在	十年一次 項目較少	集體投保 (大公司)	就醫相關

LHID2005

2005年所有年齡在保者

2005年百萬人抽樣檔



研究目標

■ 評估不同就醫方式得出之資料特性：

→感冒就醫；（一年65~70%、三年80~85%）

→感冒＋其他疾病就醫；

→健保門診金額較少之就醫；

→基層醫療院所之就醫；

註：涵蓋率、樣本代表性等為判斷標準。

■ 假設條件：「小病」就近看診、各地區都有能應付小病的醫療院所。

「小病」的特性

■ 何謂小病？

→ 罹患及就診的機會高、醫療院所的服務差異性小、選擇常住地附近的基層院所就醫。

註：差異性小或許可由醫療費用取代。

→ 就醫率、就醫金額、基層院所門診比例

■ 除了感冒（上呼吸道感染）外，消化系統（含牙醫、胃腸）、循環系統（含心臟、血管）可作為輔助之用。

→ 整體就醫率排名：感冒、消化、循環。

研究步驟

■ 2005年百萬人抽樣檔的樣本代表性

→ 全臺灣(母體)、抽樣檔的人口結構比較；

註：可查詢衛福部網站，在性別、年齡、平均投保金額等因子，簡單隨機抽樣的抽樣檔與母體沒有顯著差異。

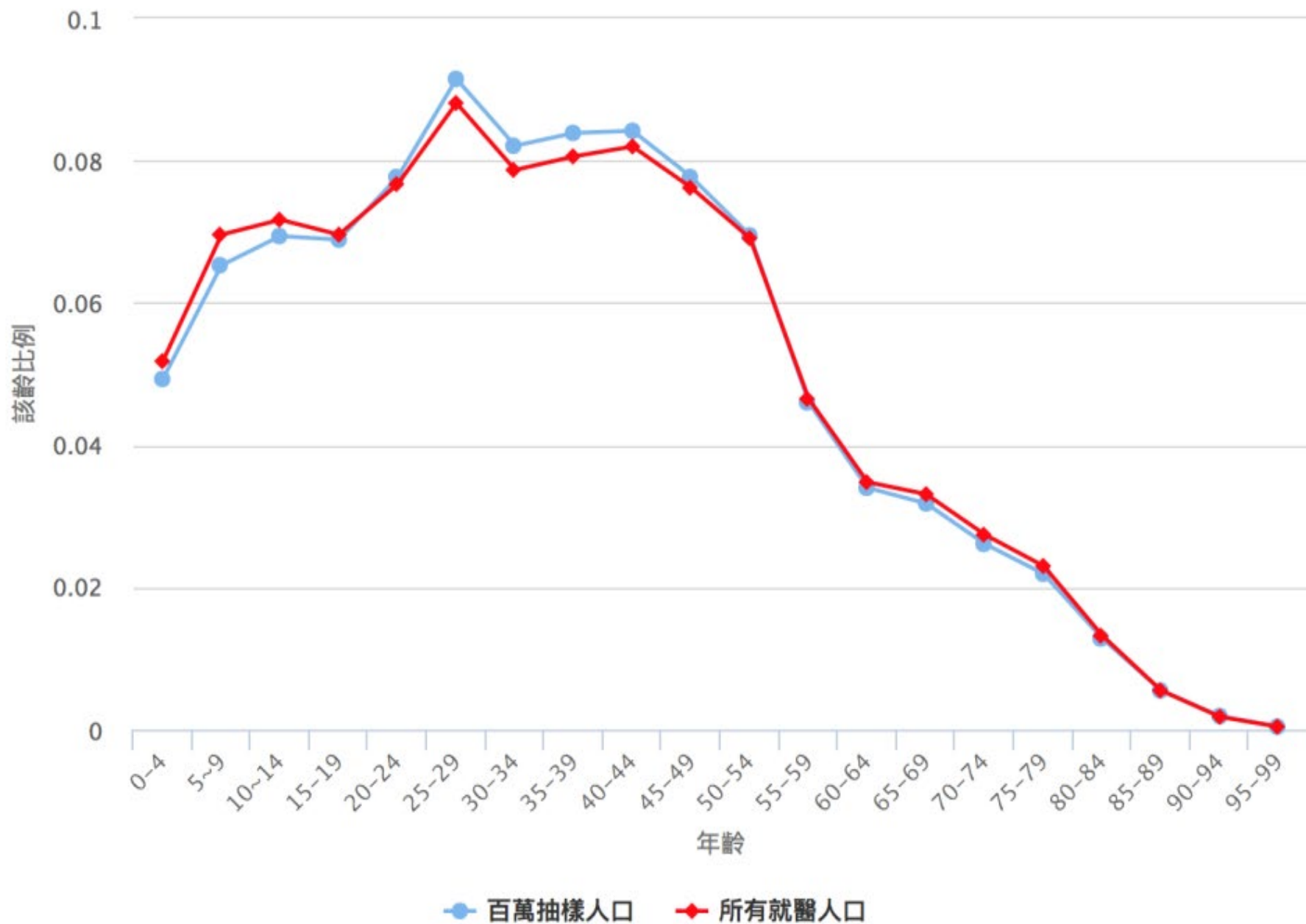
→ 全臺灣健保就醫、抽樣檔就醫比較。

■ 感冒就醫是否適合用於估計常住人口？

→ 加入其他疾病能否改善；

→ 考量低門診金額、基層醫療院所就醫。

百萬抽樣人口 與 所有就醫人口 年齡結構



全臺灣及健保抽樣檔的門診就醫比較

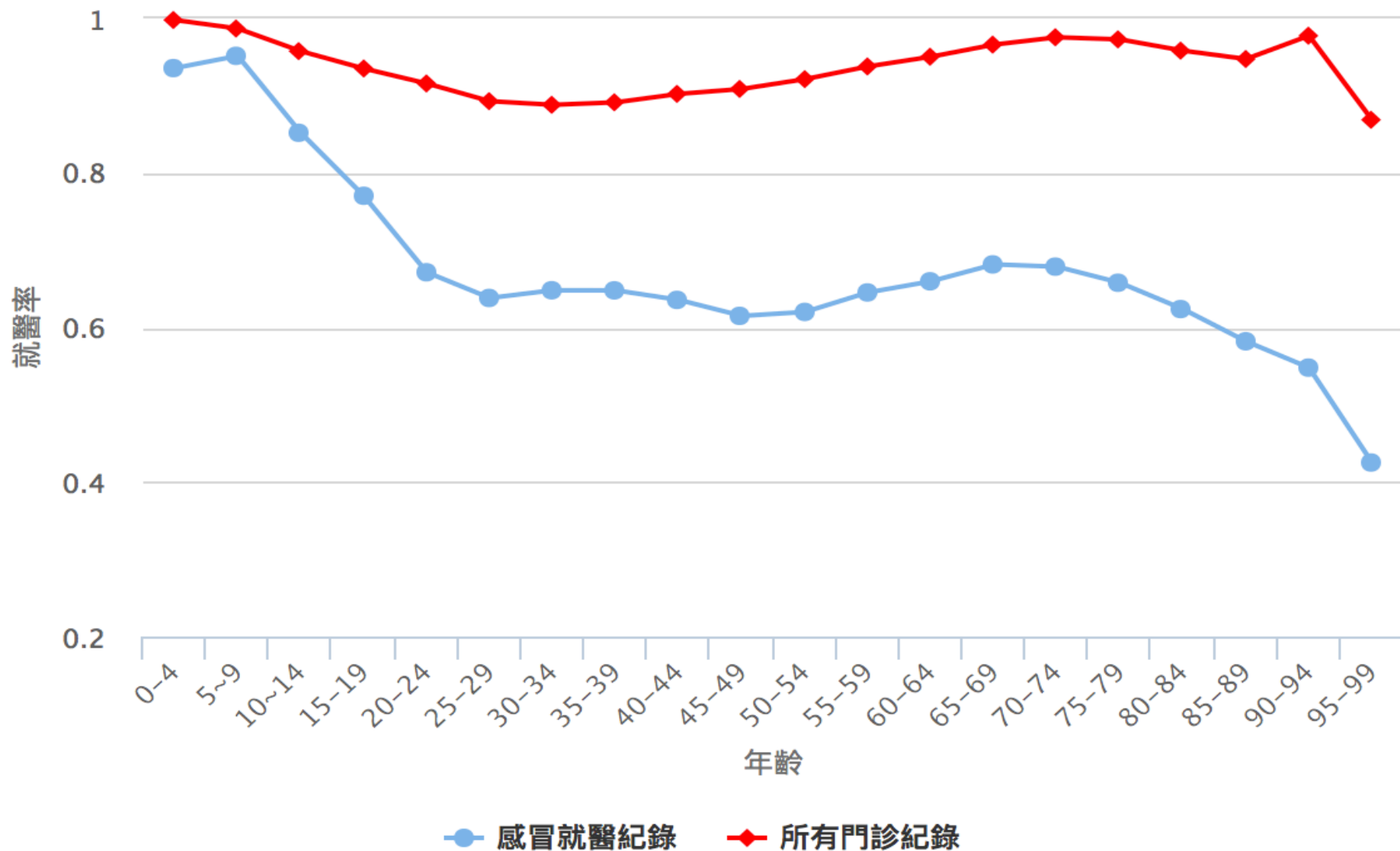
年度	人均就醫件數		人均就醫次數	
	官方	抽樣檔	官方	抽樣檔
2005	16.210	16.249	24.231	24.212
2008	16.047	16.094	25.352	25.437
2011	16.469	17.132	26.532	28.352

註：次數—門診紀錄可紀錄一至三個ICD診斷碼，

感冒就醫

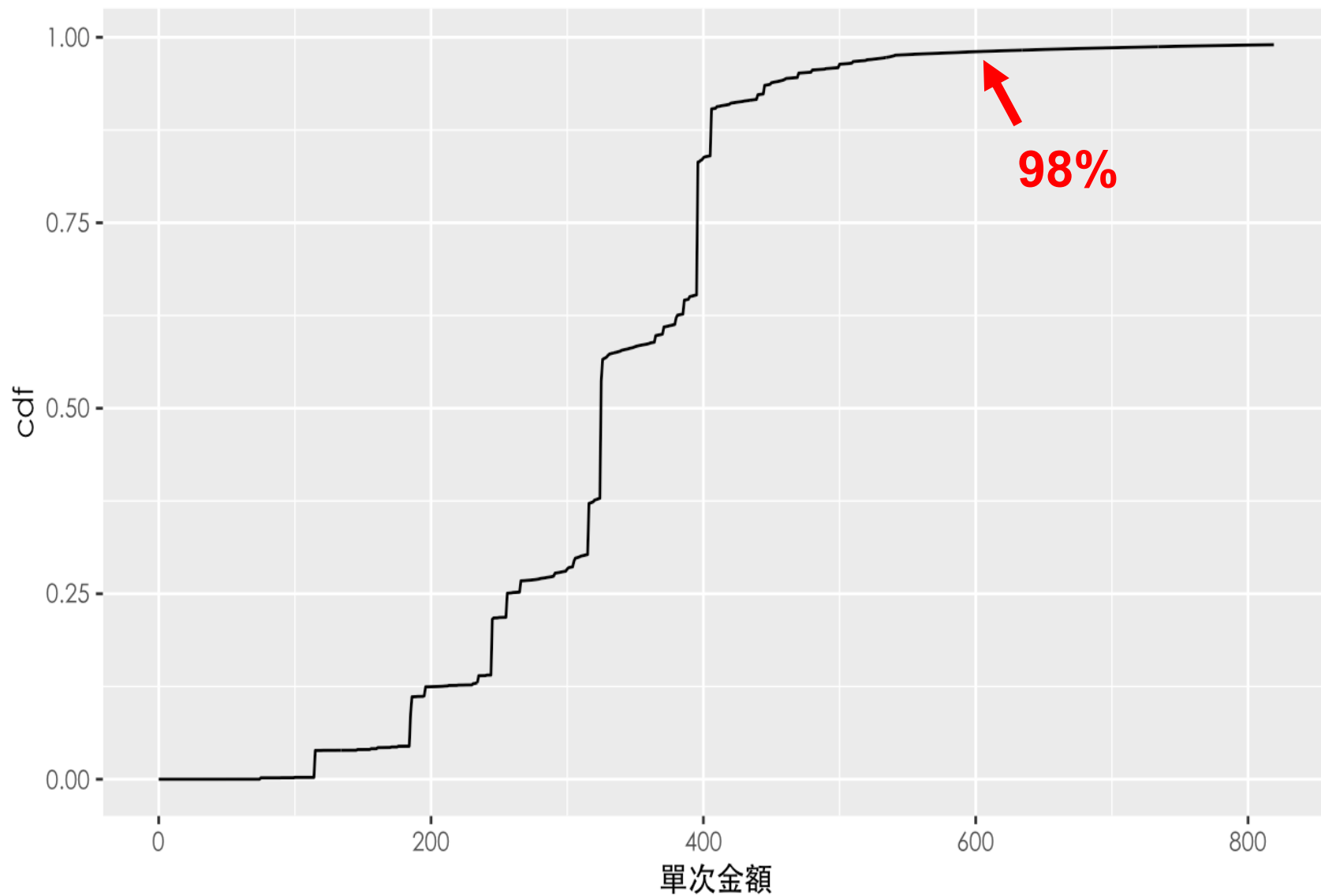
- 幼齡人口的感冒就醫率較高，65歲以上高齡人口的感冒就醫率平均僅約五成。
- 國人使用健保相當普遍，每年平均有95%人口就醫，最低的青壯年也有90%左右。
- 感冒門診金額普遍不高，而且基層醫療院所的感冒門診量可達九成。
- 感冒是不錯的小病判斷標準，或許可嘗試「感冒＋其他疾病」提升整體就醫率。

感冒就醫紀錄 與 所有門診紀錄 就醫率

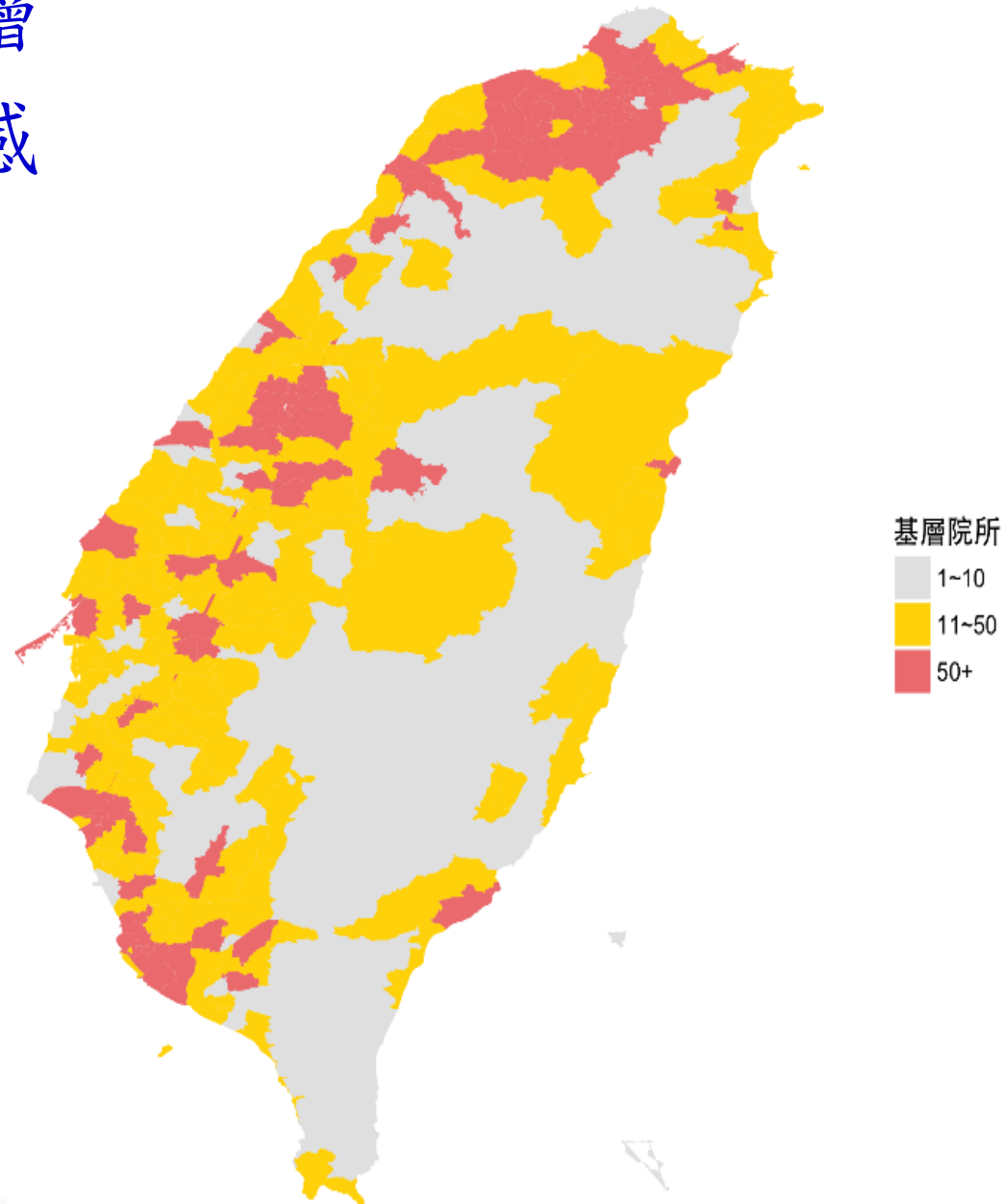
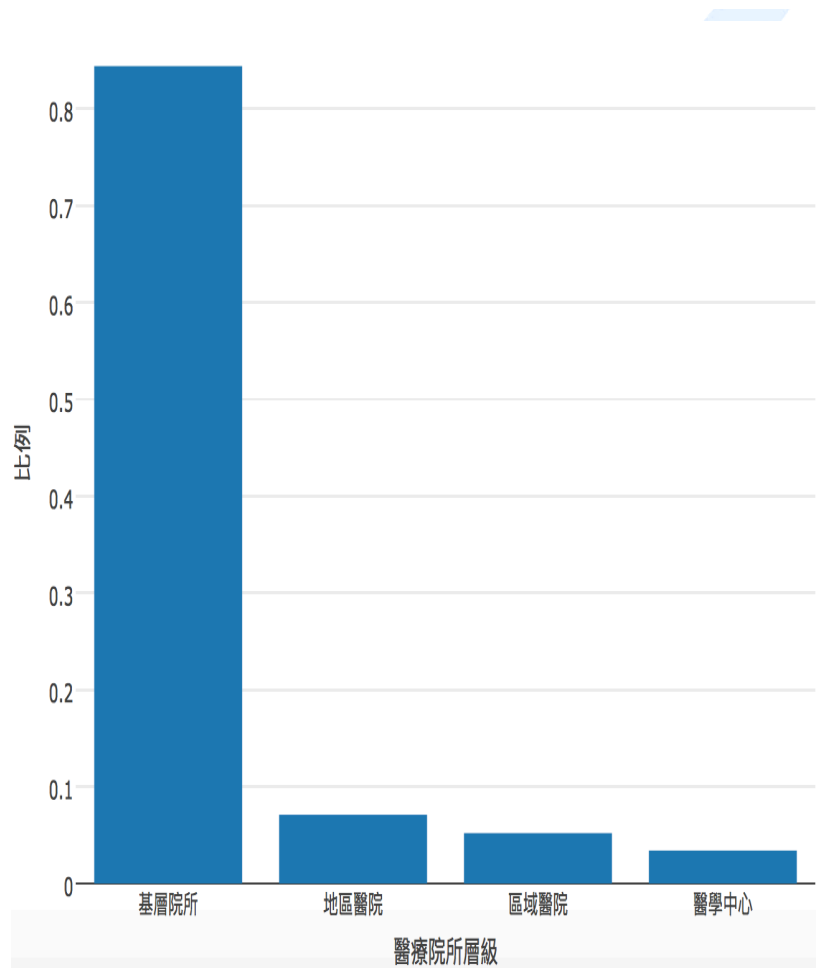


感冒單次門診就醫點數

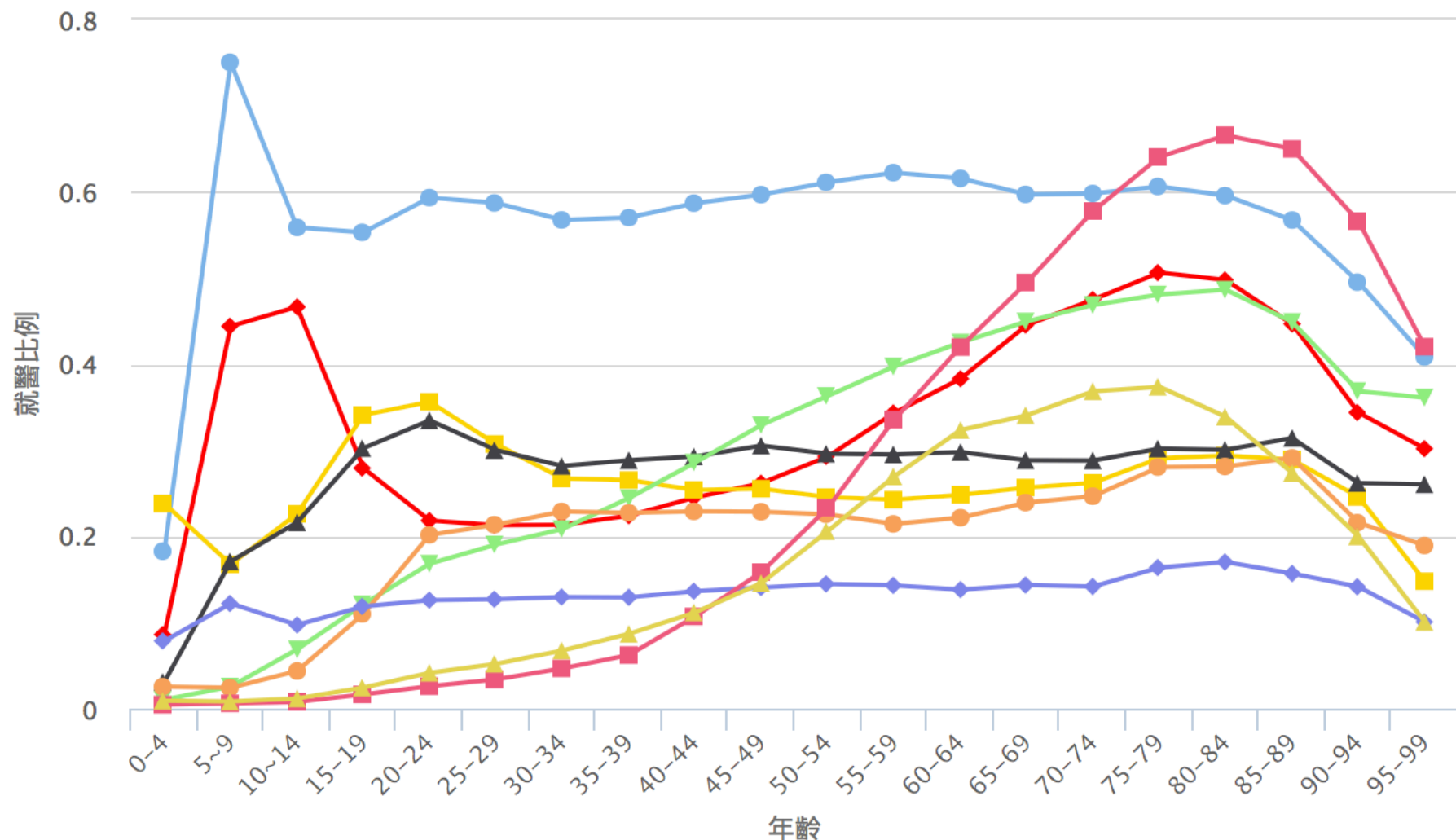
上呼吸道感染



- 各鄉鎮市區皆有基層院所，基層院所佔感冒就醫90%。



未就診上呼吸道感染者於其他疾病的就醫情形

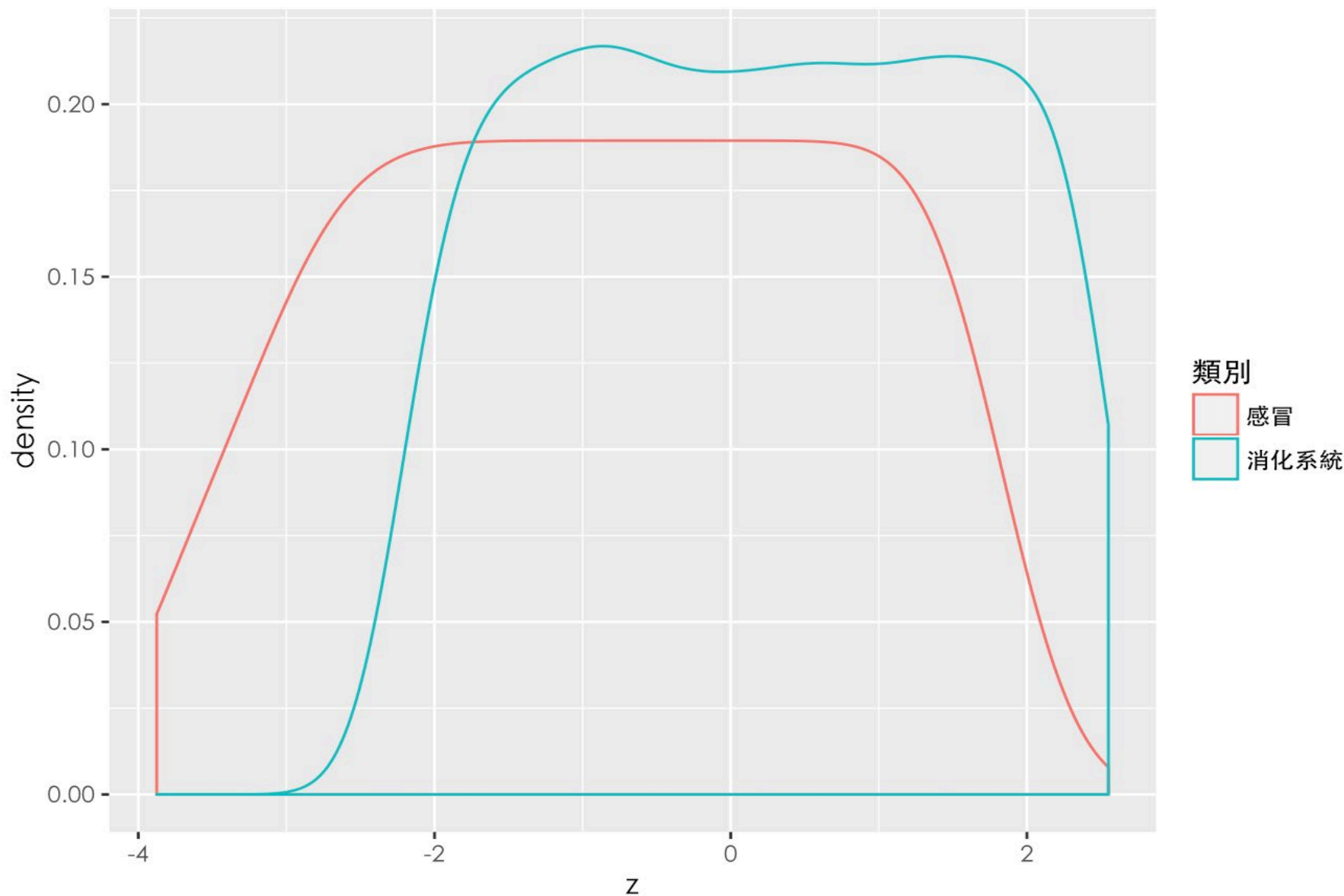


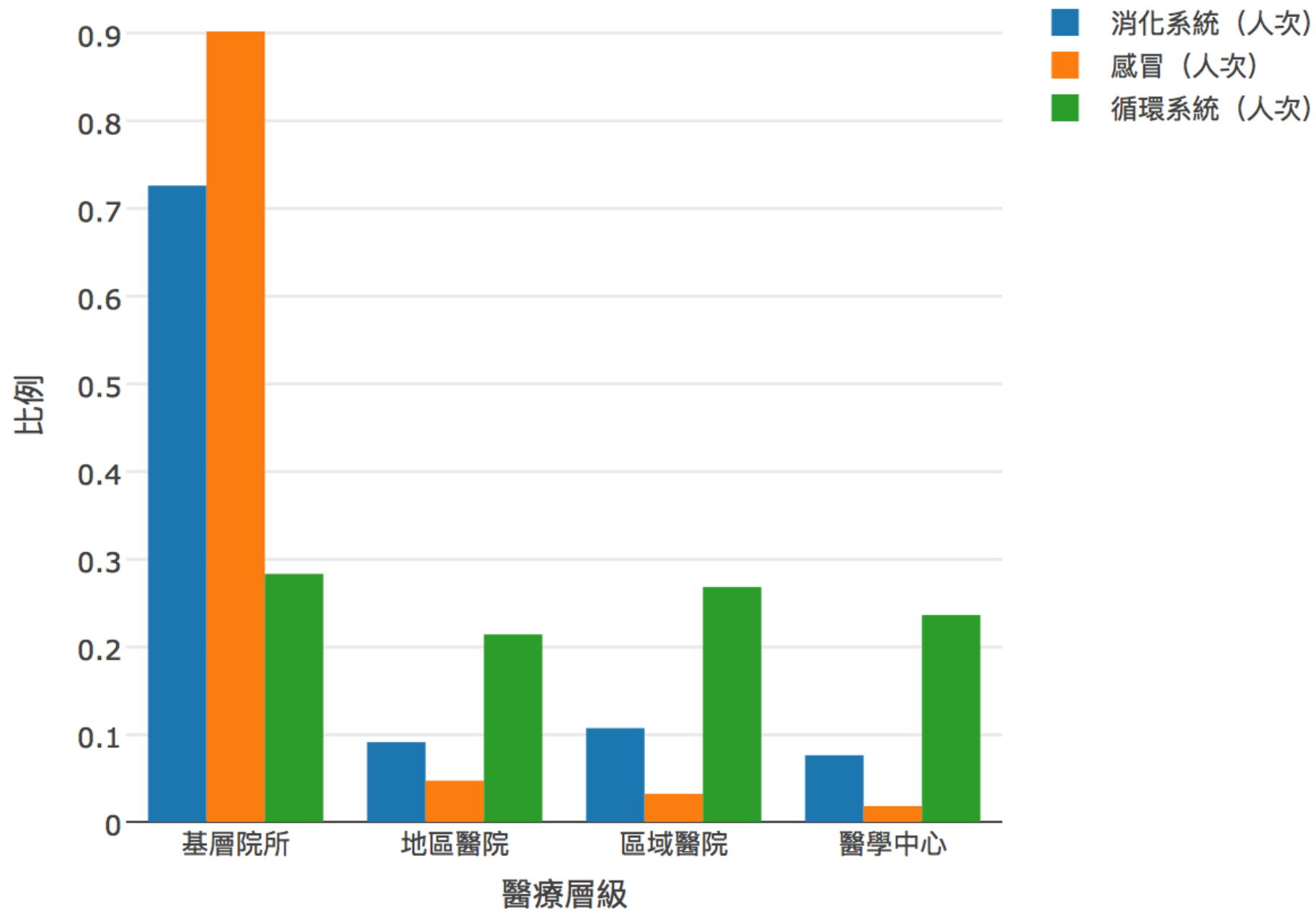
- 消化系統疾病比例
- 神經系統及感覺器官之疾病比例
- 皮膚及皮下組織疾病比例
- ▲ 損傷及中毒比例
- ▼ 骨骼肌肉系統及結締組織疾病比例
- 泌尿生殖系統疾病比例
- ◆ 感染病及寄生蟲疾病比例
- 循環系統疾病比例
- ▲ 內分泌營養及新陳代謝疾病與免疫性疾患比例

感冒＋其他疾病

- 本文考慮加入消化系統、循環系統疾病，分別可提升整體、高齡人口就醫率。
 - 以門診金額為比較標的，消化系統較為接近感冒，但循環系統的差異很大。
 - 比較不同疾病的醫療院所層級，循環系統接近50%在大型醫院就醫。
- 選擇感冒＋消化系統疾病，再以門診金額篩選兩種疾病。
 - 整體就醫率提升將近10%。

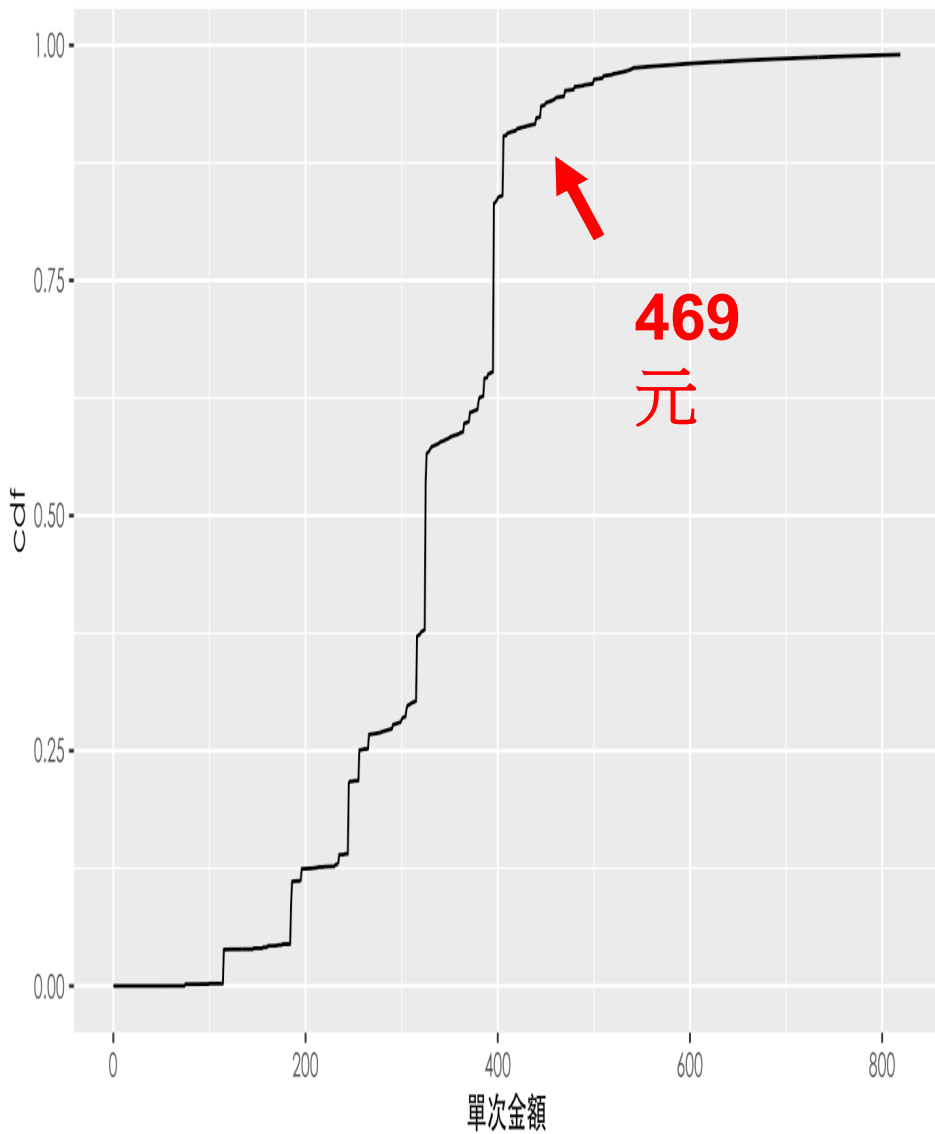
感冒、消化疾病就醫金額的標準化分佈



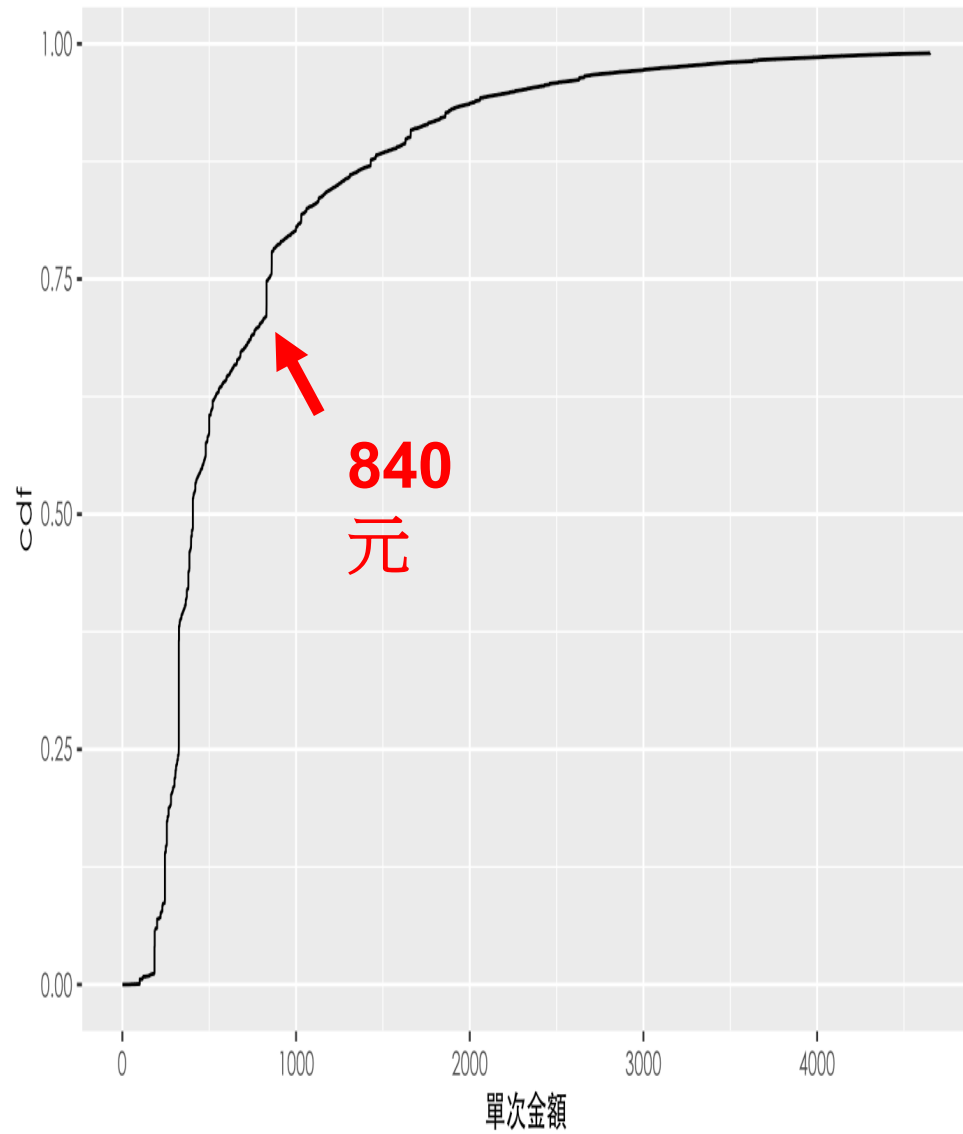


感冒、消化單次就醫金額篩選

上呼吸道感染

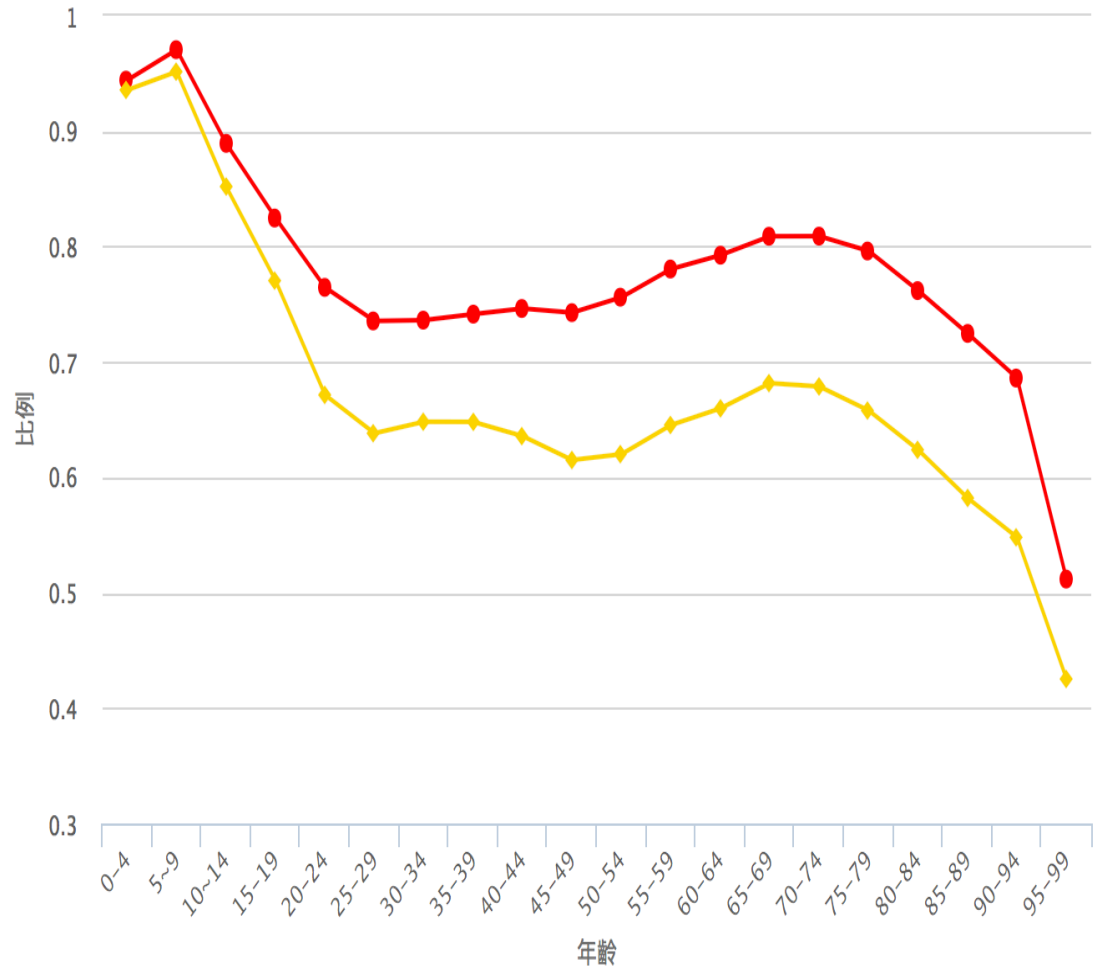


消化系統疾病



感冒+消化 vs.感冒的就醫率

各年齡層就醫比例（上呼吸道感染與消化系統）



● 上呼吸道感染+消化（已篩選金額） ● 上呼吸道感染（填補前）

估計方法

總就醫率

上呼吸道感染
(未篩選金額)

0.701

上呼吸道感染+
消化系統疾病
(皆有篩選金額)

0.793

各醫療院所的疾病就醫比例（感冒+消化）

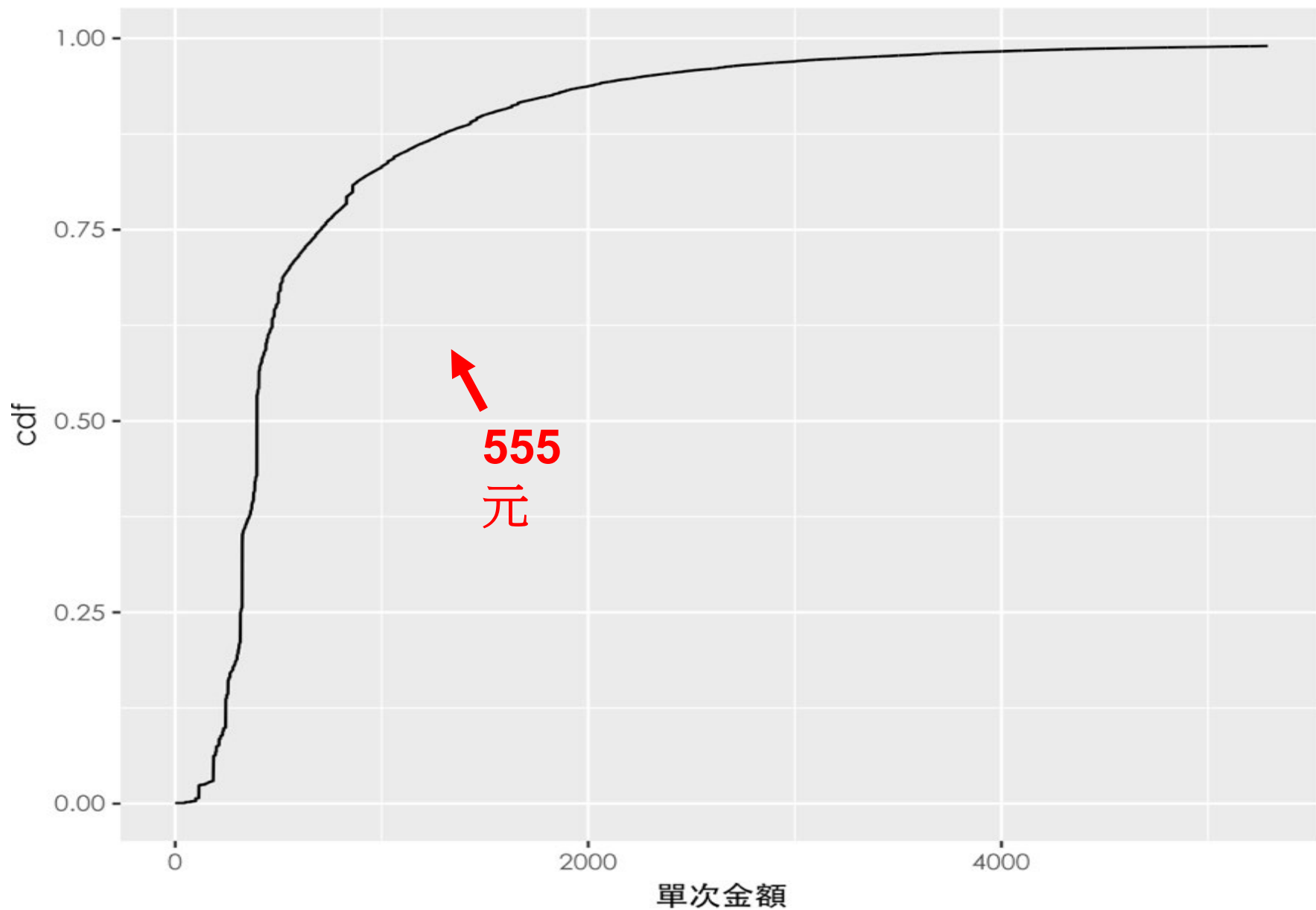
	基層院所	地區醫院	區域醫院	醫學中心
上呼吸道感染	0.902	0.047	0.032	0.018

	基層院所	地區醫院	區域醫院	醫學中心
消化系統疾病 (篩選金額前)	0.726	0.091	0.107	0.076
消化系統疾病 (篩選金額後)	0.838	0.051	0.061	0.050

就醫金額、基層醫療院所

- 除了就醫病因外，也可透過就醫金額（健保補助）、就醫院所層級判斷「小病」。
- 仿造感冒、消化系統就醫金額，計算所有病因就醫金額的分佈，再選取某個門檻金額，低於555元列為「小病」。
- 感冒及消化疾病大多在基層醫療院所，因此在這一层級就醫者視為小病。
- 註：因為有些人只在醫學中心或教學醫院就醫，只考慮基層院所仍會出現遺漏值。

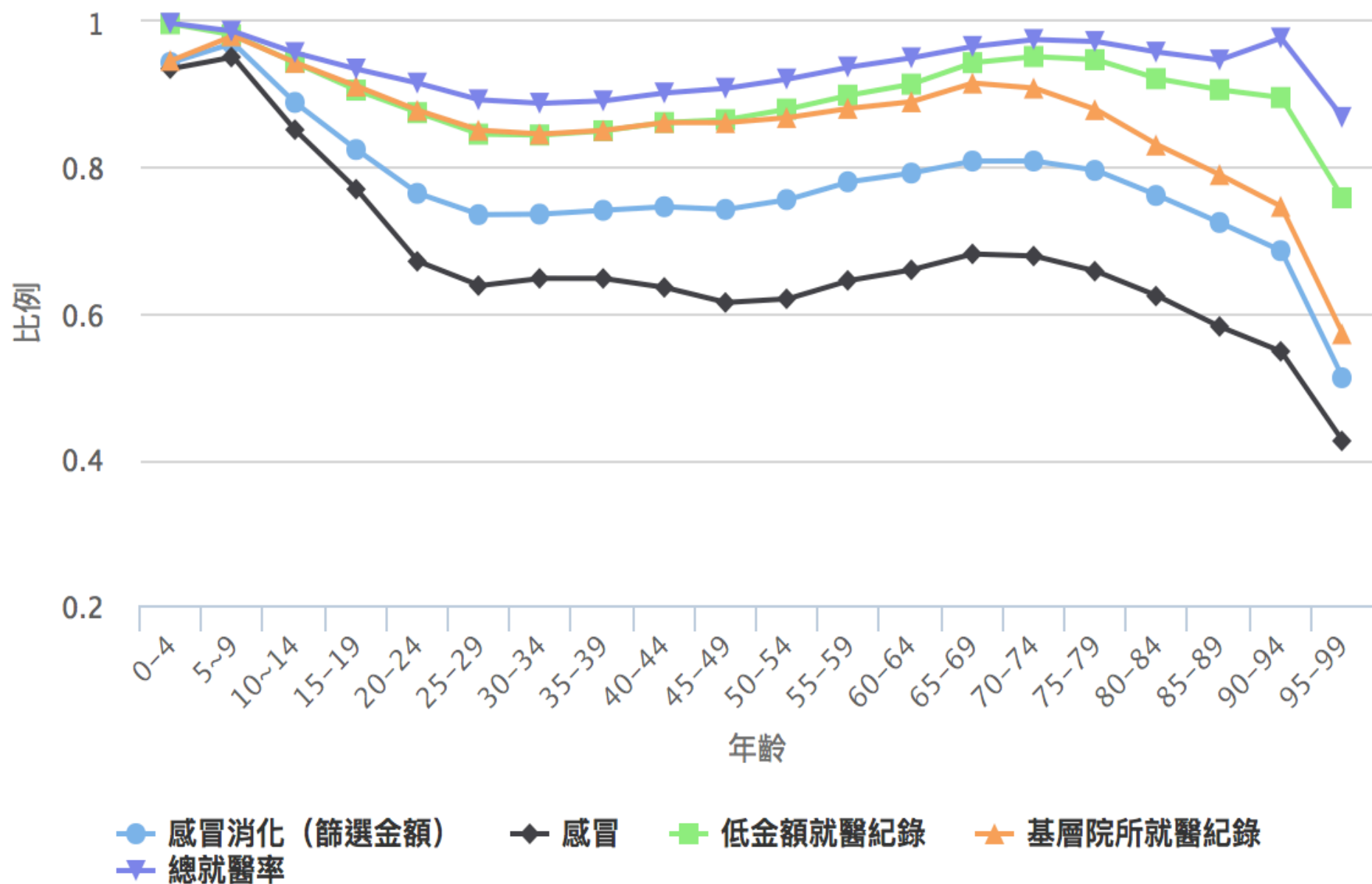
單次就醫金額CDF與篩選



各方法的估計結果比較

判斷準則	就醫率	各層級醫療院所比例			
		基層院所	地區醫院	區域醫院	醫學中心
感冒	0.701	0.902	0.047	0.032	0.018
感冒＋消化 (篩選金額)	0.798	0.838	0.051	0.061	0.050
金額低於 555點	0.964	0.844	0.071	0.052	0.034
基層院所	0.958	1.000	---	---	---

各推估準則的母體涵蓋率



不同準則與樣本代表性

- 透過上述準則決定就醫人口，以年齡及性別比例（整體就醫率為輔）作為判斷依據，比較抽樣檔、四種準則的相似性。
- 四種準則的年齡結構卡方檢定（樣本代表性）p-value都小於0.01。
- 年齡結構的MAPE誤差以單次就醫金額小於555元最佳、基層醫療院所次之。（年齡別就醫率的結果也類似。）

各方法的整體估計誤差

準則	感冒	感冒+消化	低金額	基層院所
年齡 MAPE	0.1456	0.0794	0.0117	0.0480

各方法的年齡別估計誤差

年齡	整體門診	感冒	感冒補入消化	低金額	基層院所
5~19歲	0.1998	0.2463	0.2232	0.2031	0.2058
20~44歲	0.3898	0.3843	0.3843	0.3860	0.3922
45~64歲	0.2848	0.2750	0.2752	0.2842	0.2821
65歲以上	0.1256	0.1084	0.1173	0.1268	0.1199

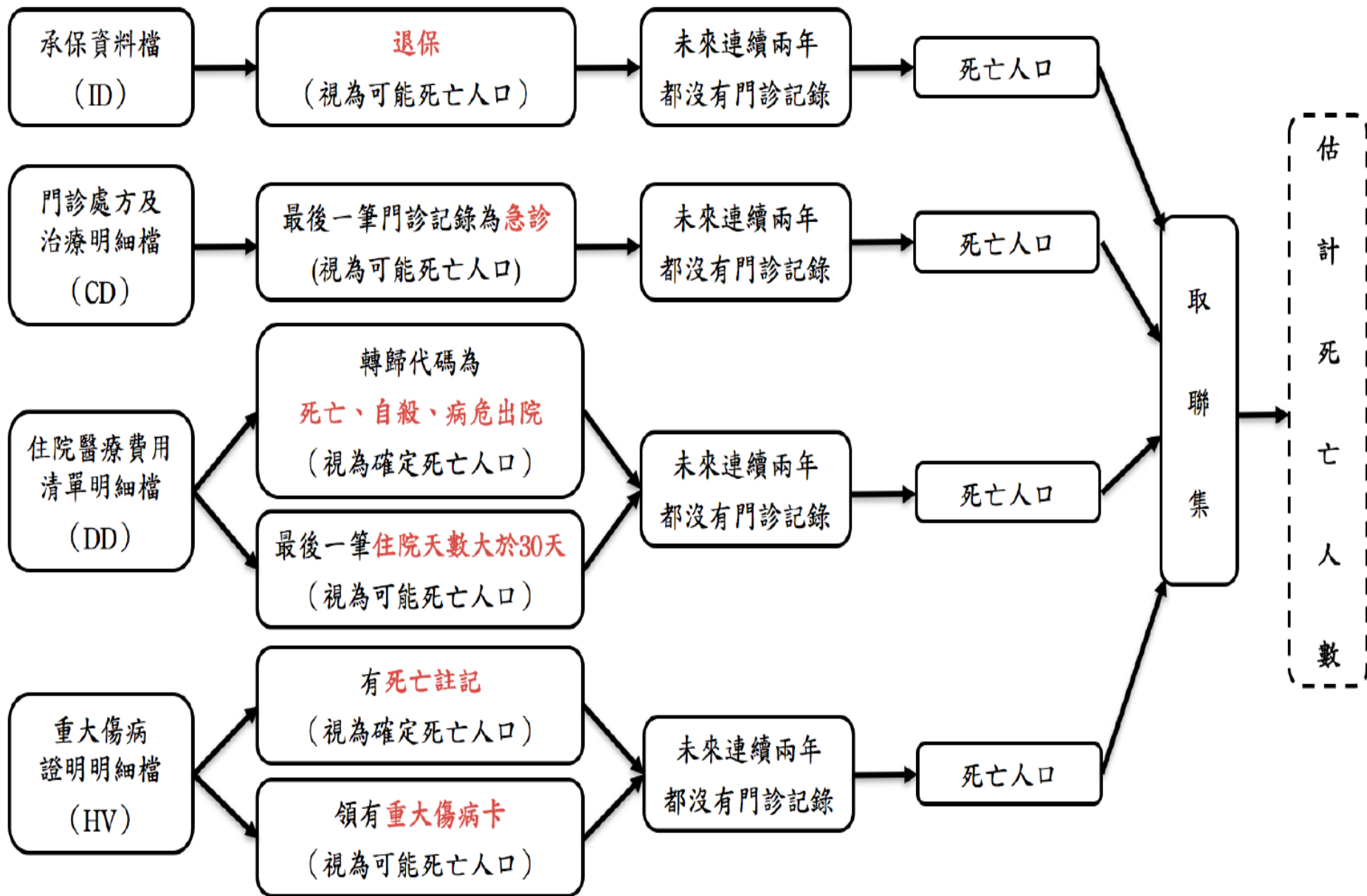
健保資料庫的應用



<http://www.tnooz.com/2012/01/04/how-to/big-data-and-the-infinite-possibilities-for-the-travel-industry/>

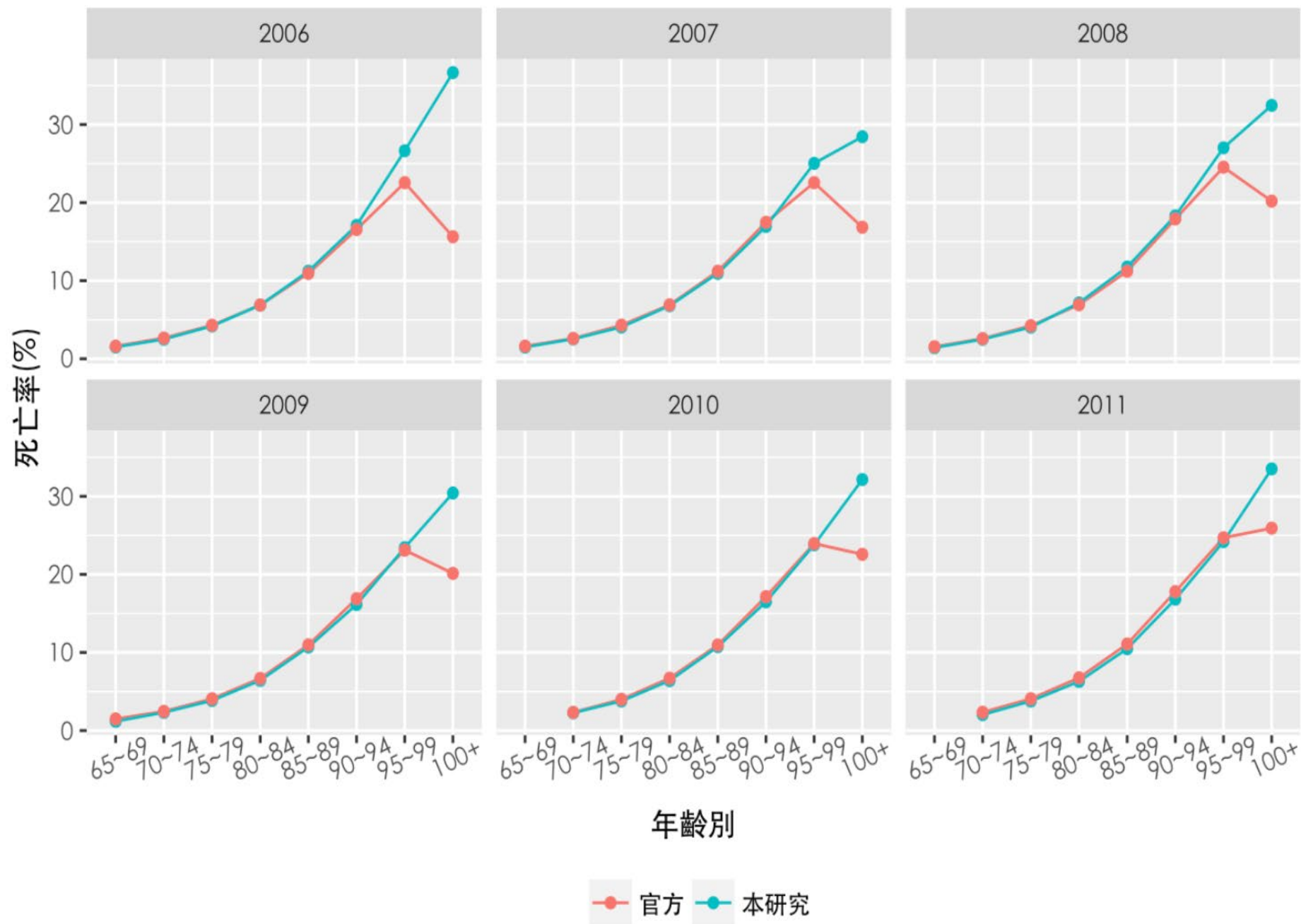
健保資料與人口研究

- 透過上述就醫人口的判斷，健保資料也可用於人口特質的推估，包括死亡率（平均餘命）、疾病歷程、人口遷移等。
 - 但健保資料庫缺乏完整的死亡紀錄，需先建立死亡判斷準則。
- 以高齡人口（65歲以上）的死亡率為例，以退保、門診紀錄等為判斷標準。
 - 全國、各縣市死亡率與官方結果接近。

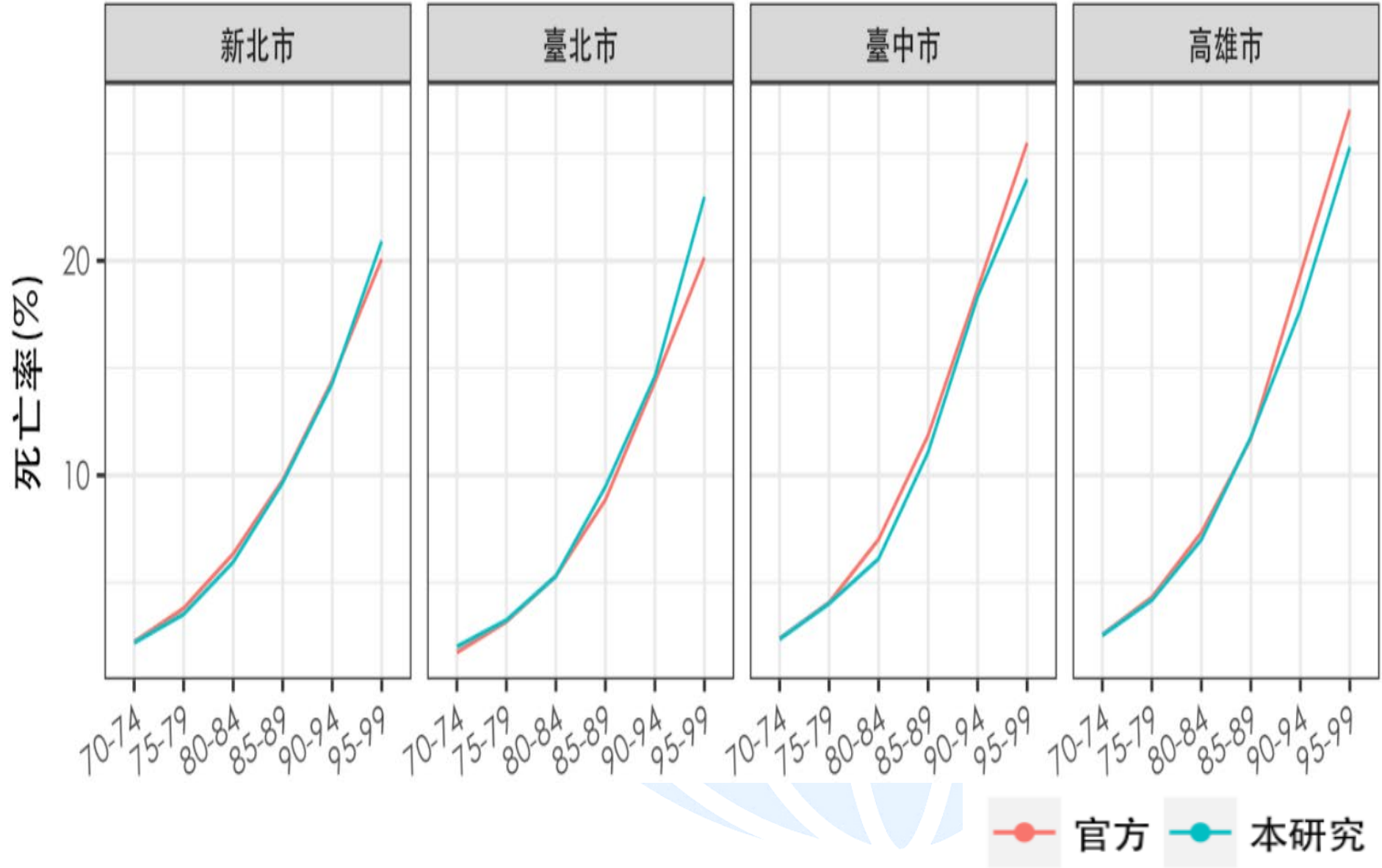


高齡人口的健保死亡判斷準則

各年度死亡率比較



2010年官方與健保資料推估死亡率的比較



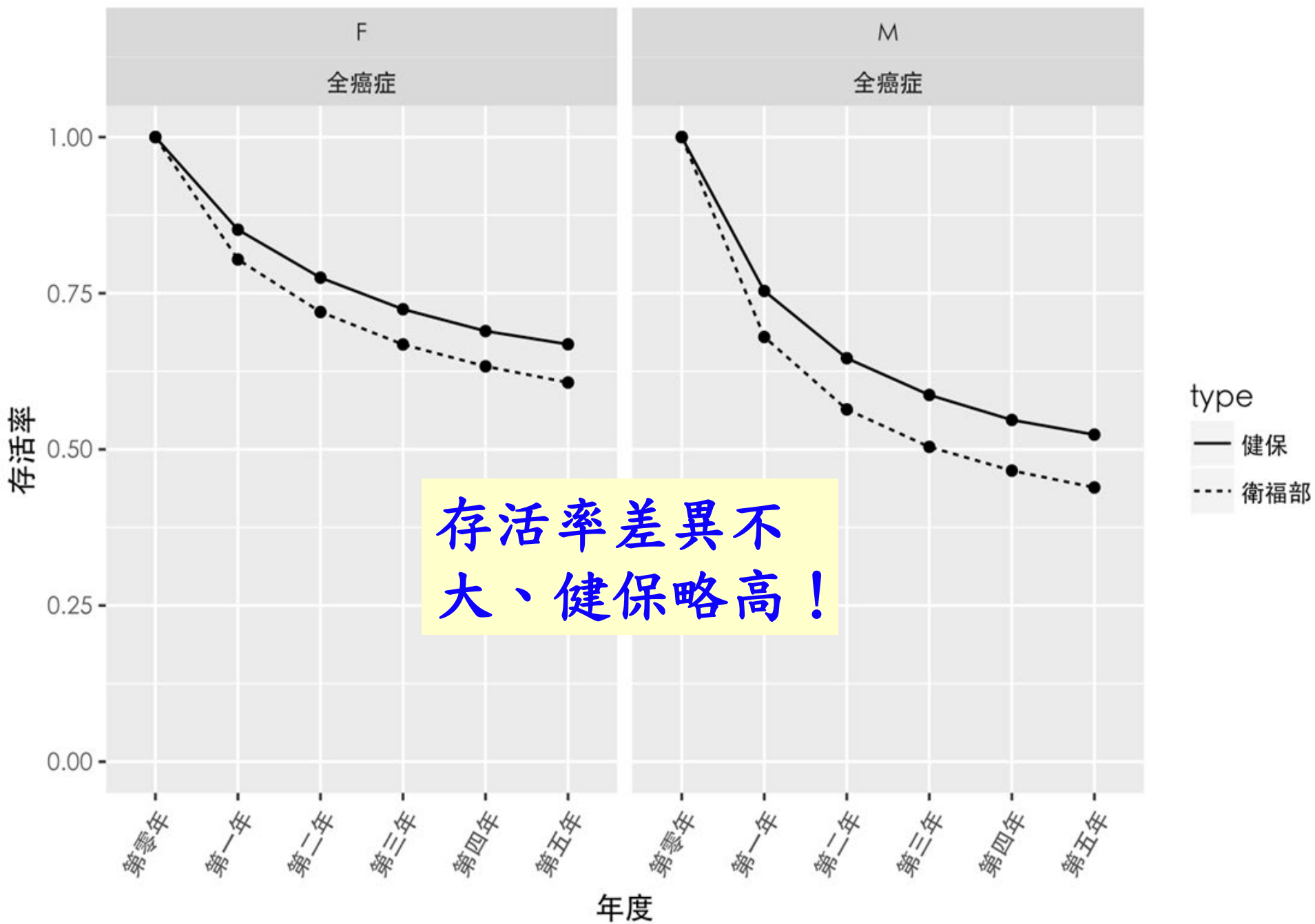
高齡人口的死亡率

- 高齡(65歲以上)老年人資料品質已難掌控，
高高齡(85歲以上)更難確認。(Wilmoth, 1994)
- 美國有不錯的社會保險制度，老年人可享有
Medicare的保障，可由美國社會安全機構獲得
老年人資料。

$$q_x = \frac{1}{11} \left[(95 - x) q_x^c + (x - 84) q_x^M \right], \text{ for } 85 \leq x \leq 94$$

Medicare data : Ages 84 and higher
Census data : Ages 95 and lower

10大癌症存活率_健保與衛福部比較



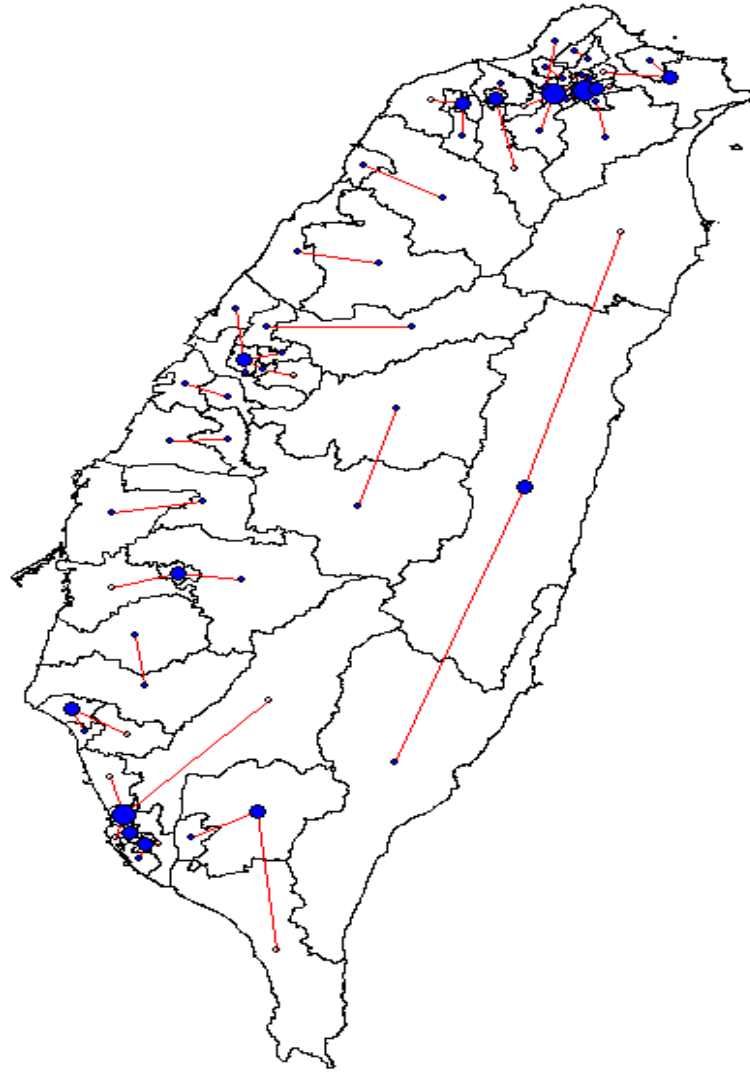
健保資料與人口遷移

- 戶政司僅有全省、院轄市見的遷移紀錄，且僅有遷移人數，未必標示年齡性別。
 - 以健保就醫判定「常住地」，比較各年度（或半年）間的地點差異。
- 以選舉區為例展示臺灣的人口遷移特性。
 - 各選舉區人數介於20～50萬，避免縣市、鄉鎮市區間人口差異大、人數不足的問題。
 - 資料分析顯示人口遷移有明顯的區域特性，其中花東為淨遷出，新竹、台中淨遷入。

	總計	國外	新北	臺北	桃園	臺中	臺南	高雄	臺灣	福建	同縣市
總計	1040247	45780	93445	73971	67867	46655	23304	34038	105366	10054	514650
新北	183112	9764	0	42851	24899	5245	2610	3684	35570	2811	72132
臺北	146691	14802	46410	0	9255	4067	1595	2562	17382	949	55708
桃園	80874	3660	9685	4934	0	2425	1150	1680	10883	1446	44966
臺中	125441	3988	4389	4008	4148	0	2123	2687	22490	1002	83306
臺南	74990	2143	2654	2165	2385	2594	0	6731	8287	573	49010
高雄	128685	4373	4021	3435	3394	3441	7444	0	16774	1142	86850
臺灣	347483	9395	31767	19398	22786	30088	8947	17384	0	3095	149940
福建	6557	88	913	442	1000	398	195	409	1107	0	2731

註：資料來源為內政部戶政司（2013 年至 2015 年平均值）；桃園市在 2015 始有資料，因此僅以 2015 年為代表。

Taiwan Region

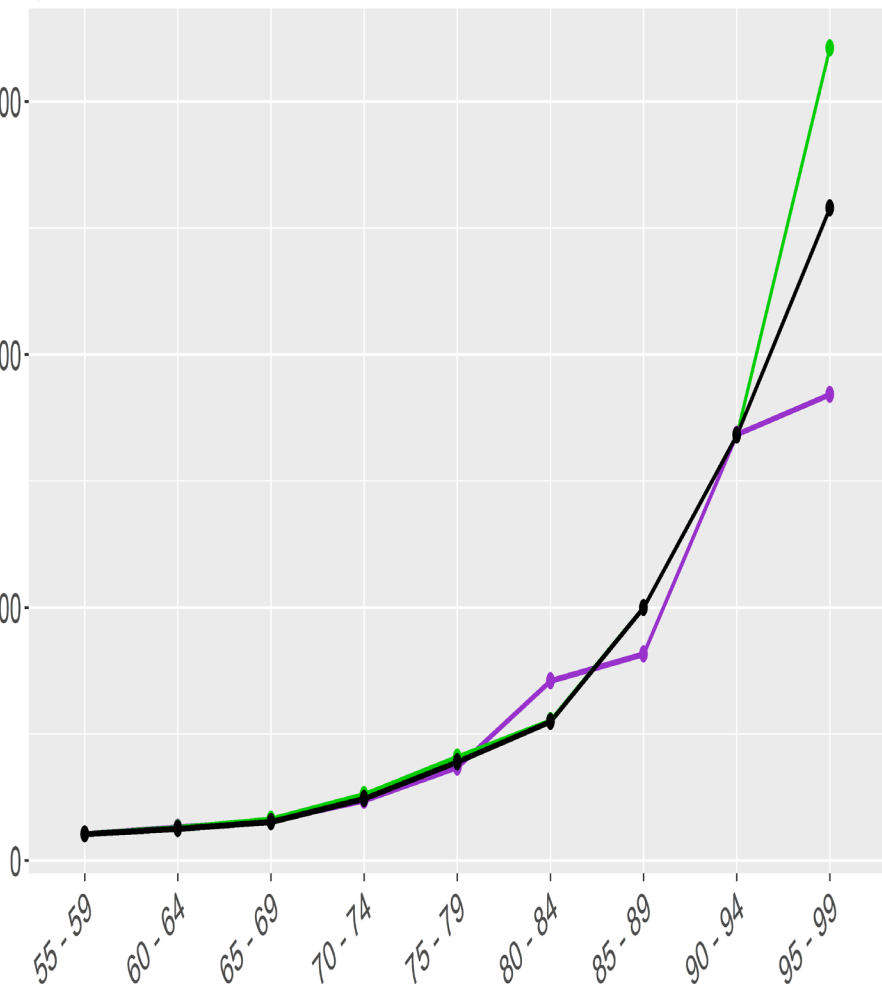


臺灣遷移較頻繁地區（健保、選舉區）

縣市間、縣市內移動者的死亡率比較（2006年）

中高齡不同類型的死亡率(女性)

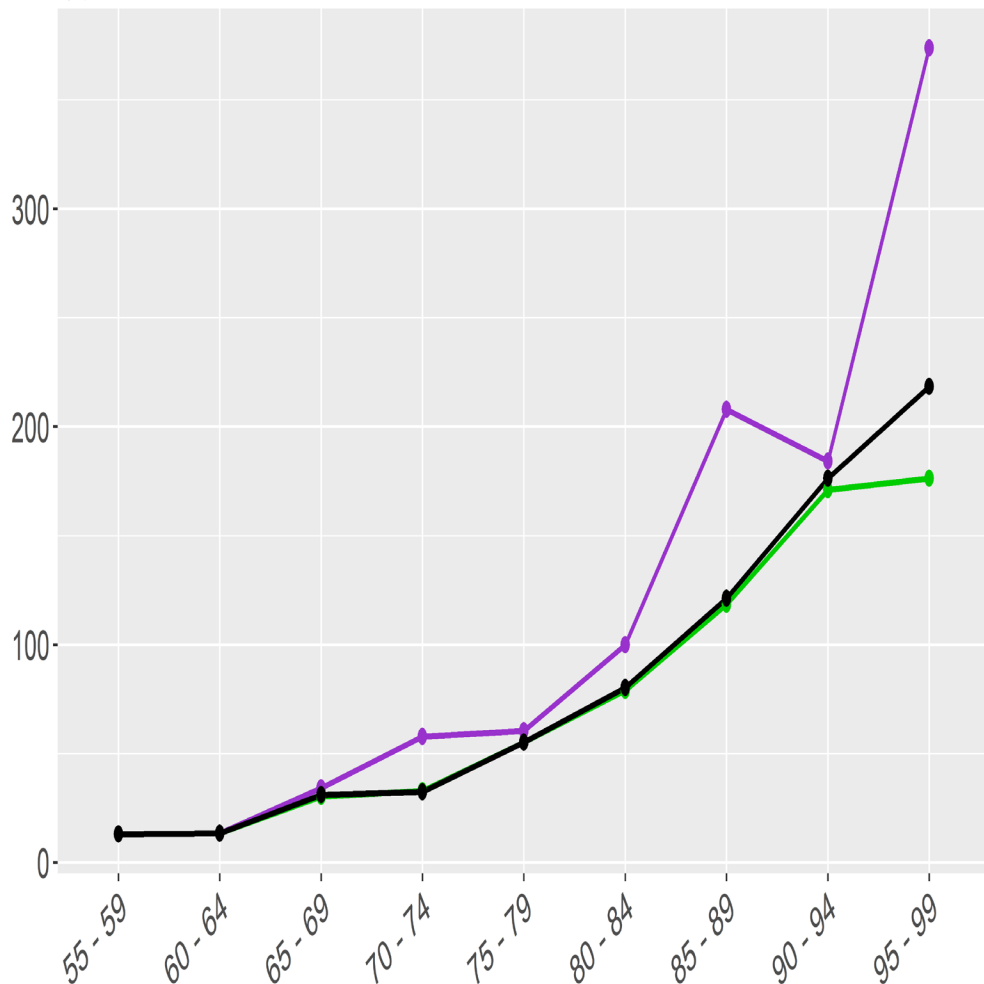
單位: ‰



● 縣市間 ● 縣市內 ● 內政部

中高齡不同類型的死亡率(男性)

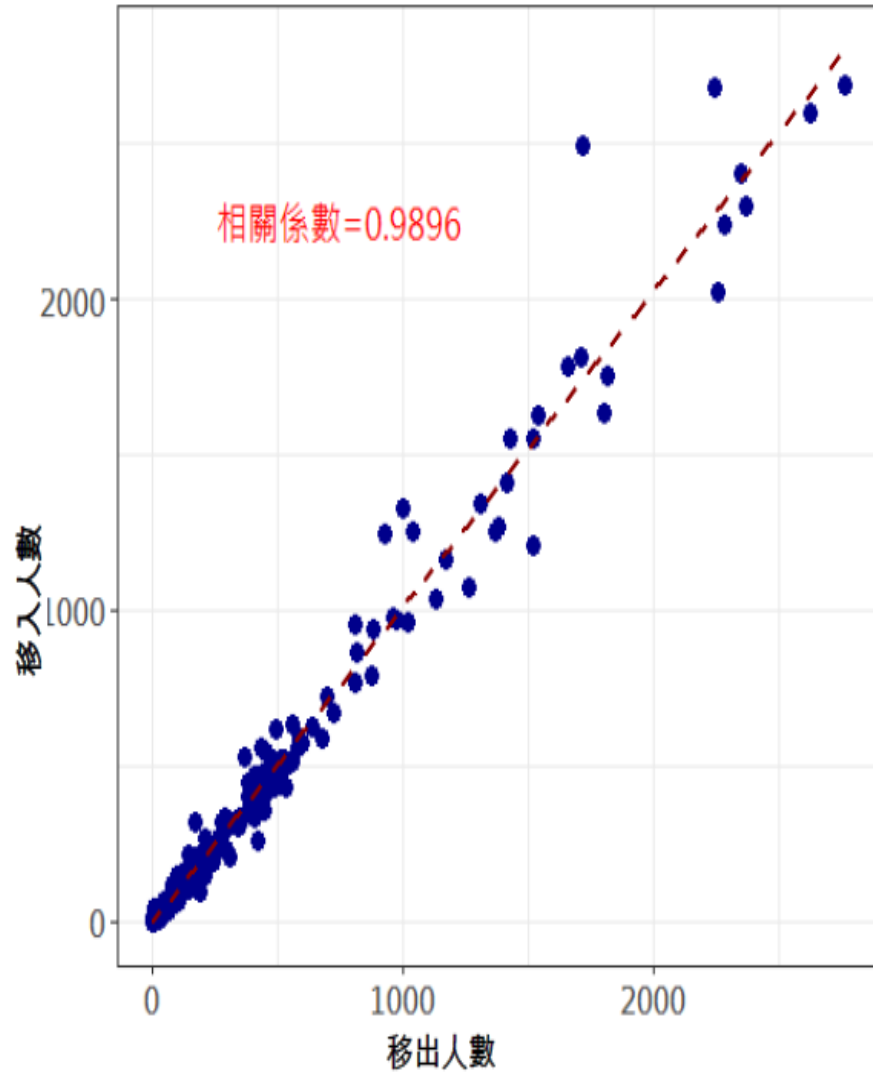
單位: ‰



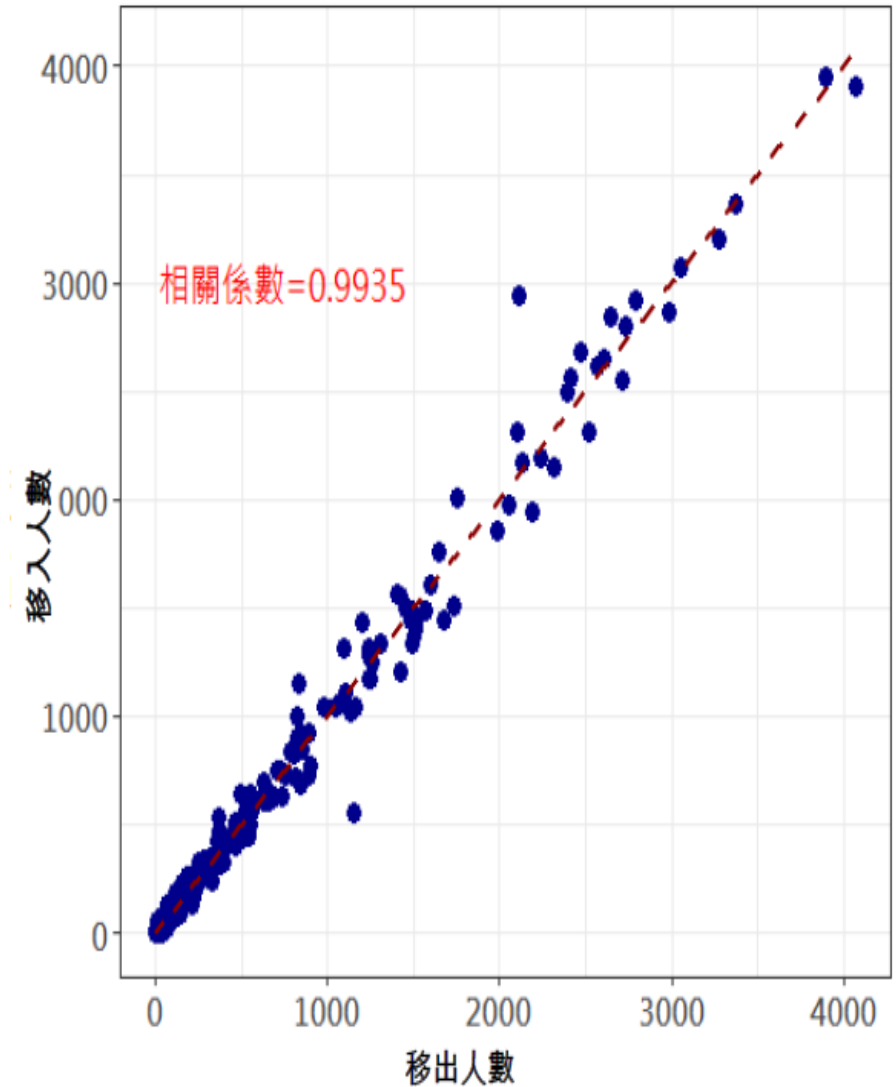
● 縣市間 ● 縣市內 ● 內政部

移入人數和移出人數散佈圖（鄉鎮市區層級）

縣市間



縣市內



全民健保職業類別及保費負擔比例

保險對象類別			負擔比例(%)		
			被保險人	投保單位	政府
第一類	公務人員 公職人員	本人及眷屬	30	70	0
	私校教職員	本人及眷屬	30	35	35
	公民營事業、機構等有一定雇主的受雇者	本人及眷屬	30	60	10
	雇主 自營業主 專門職業及技術人員自行執業者	本人及眷屬	100	0	0
第二類	職業工會會員 外僱船員	本人及眷屬	60	0	40
第三類	農民、漁民 水利會會員	本人及眷屬	30	0	70
第四類	義務役軍人、替代役役男、軍校軍費生、在卹遺眷、矯正機關之收容人	本人	0	0	100
第五類	低收入戶	本人	0	0	100
第六類	榮民、榮民遺眷家戶代表	本人	0	0	100
		眷屬	30	0	70
	其他地區人口	本人及眷屬	60	0	40

健保資料庫缺

結論

- 受到個資法等影響，受訪者回覆意願下降，影響調查的資料品質及成功率。
 - 不少國家在2010年使用戶口普查替代方案，大致可分為「公務資料＋調查」、「滾動式普查」兩類。
- 本文以全民健保資料庫為素材，探索可反映母體特性的資料攫取方法。
 - 適當的抽樣設計蒐集類似母體結構的樣本，可用於推估母體的其他特性。

討論與建議

- 大數據世代的來臨，資料蒐集、分析及附加價值等議題也愈受重視。
- 類似新型態普查方法，問卷調查是否也可透過串連資料庫、或是其他方法進行？
- 個資保護由「個人同意制」，轉成資料使用者的「使用責任制」？（例如：丹麥）
- 成立國家統計局，統合各類資料庫，再以混碼去除個資。（仿造全民健保資料庫）
- 跨領域專家的結合（數據科學家）？

參考文獻：

- ◆ 顏貝珊與余清祥(2010),「2010年各國人口普查制度之研究」,《人口學刊》
- ◆ 余清祥、王信忠、許筱翎(2019),「從全民健保評估高齡人口的醫療需求」,《人口學刊》
- ◆ Wang, H., Yue, C.J., & Wang, T. (2019), “Do Domestic Immigrants Live Longer? An Approach for Estimating Life Expectancy of Small Populations”, *Migration Letters*
- ◆ 余清祥、簡于閔、梁穎誼(2020),「健保資料與抽樣調查」,《調查研究—方法與應用》
- ◆ 余清祥、梁穎誼、林佩柔(2022),「健康、醫療利用與人口移動的關聯」,《地理學報》