

修勻學(Graduation) — Small Area Estimation

授課教師：余清祥教授

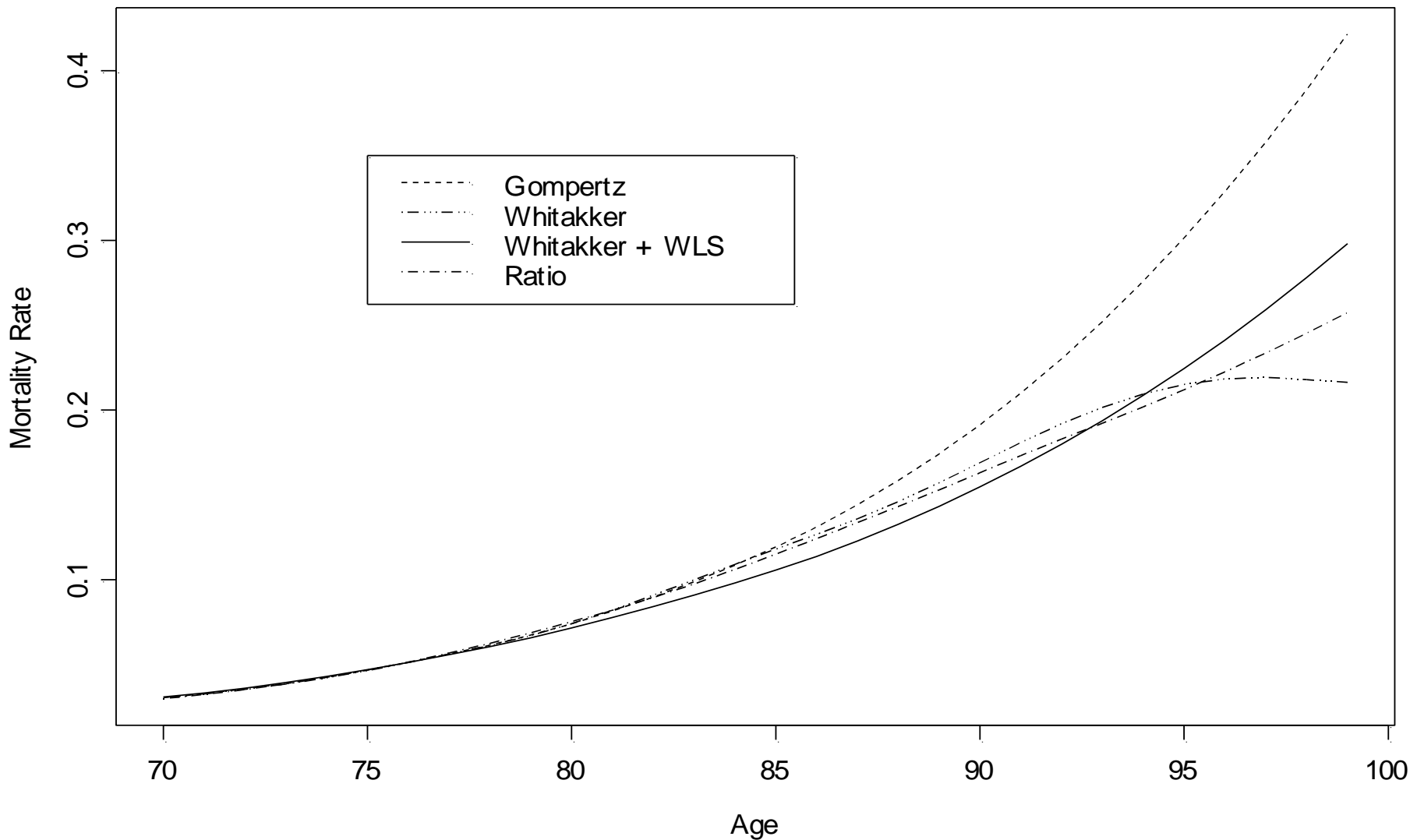
課程日期：2024年12月11日

資料下載：

<http://csyue.nccu.edu.tw>



Taiwan Male Mortality Rates (1999-2001)

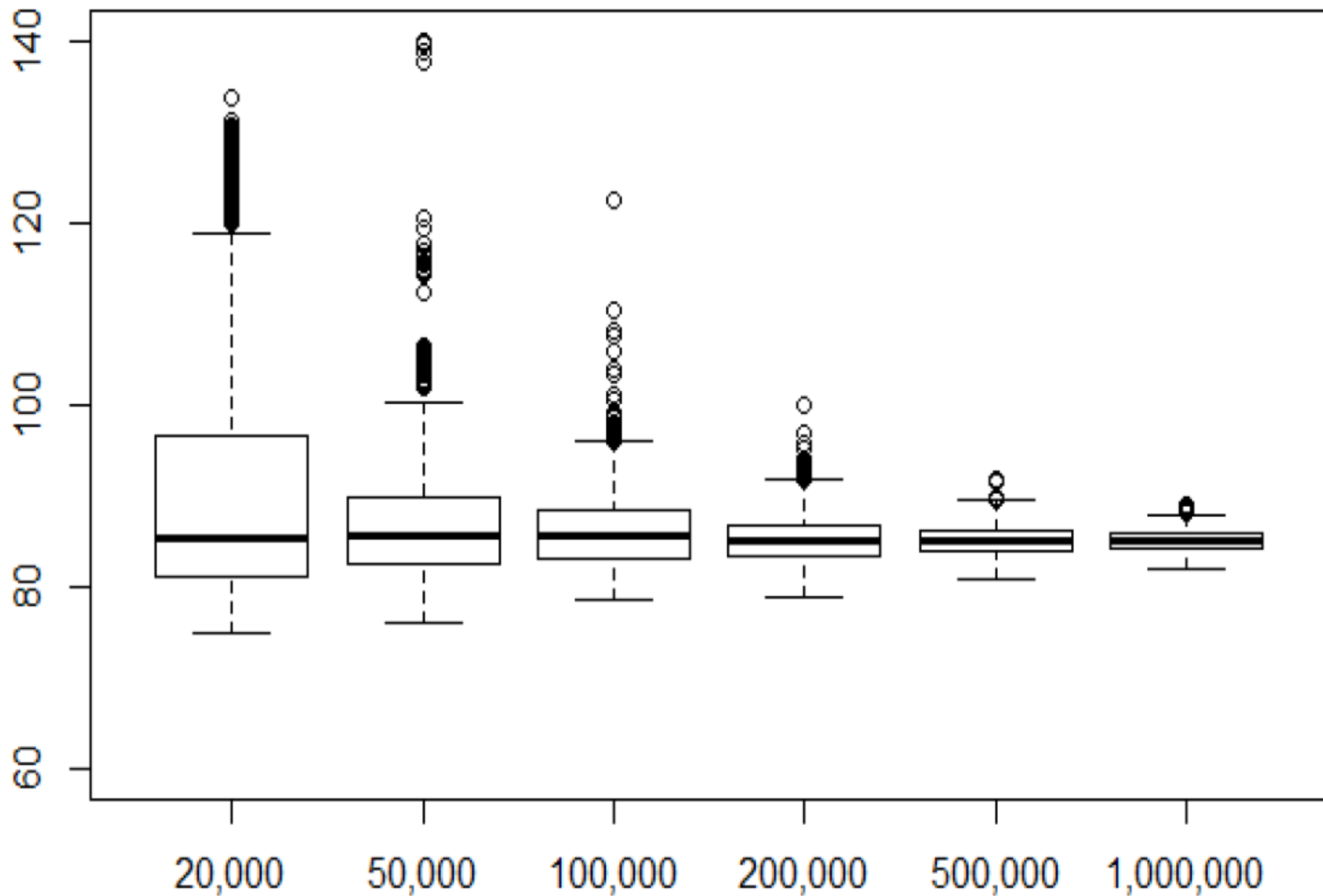


不同修勻方法與高齡死亡率

小區域人口的修勻

- 修勻方法的選擇在人數較少死亡率會有較大影響，人數多時不應有重大差異。
 - 變異數與樣本數為反比關係！
- 修勻可視為增加樣本數，累積數年資料(同地異時)、參考類似母體(異地同時)較為常見。
 - 累積三年(國民生命表)、五年(經驗生命表)、九年資料較為常見。
 - 貝氏修勻是較常見的異地同時方法，但選擇「類似」參考母體不容易。

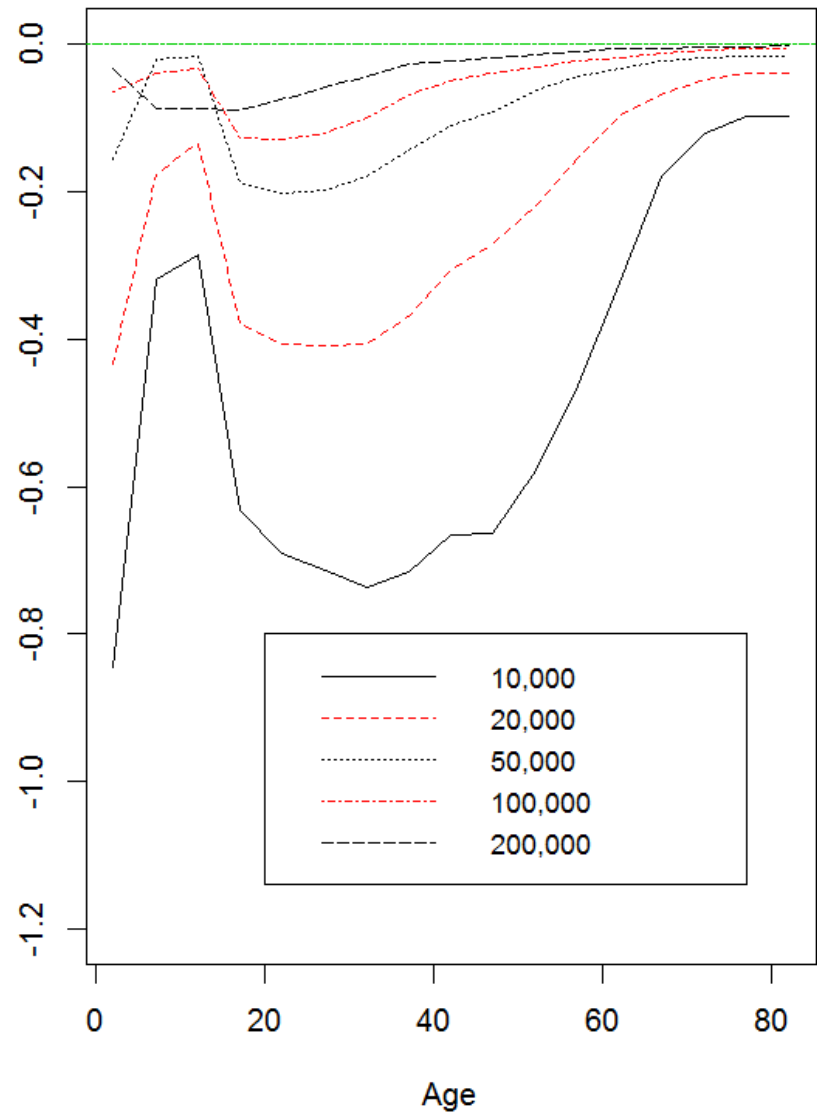
零歲平均餘命與樣本數的關係(台灣女性、模擬)



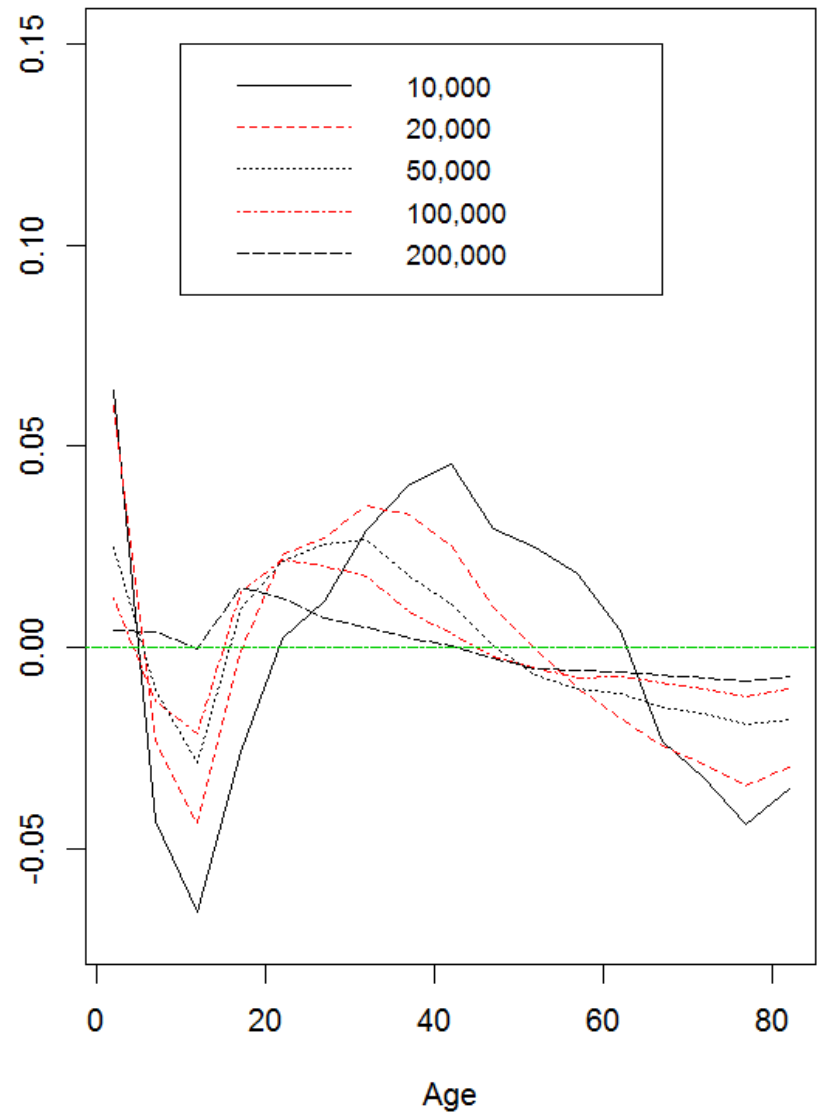
修勻與死亡率模型

- 死亡率模型可視為一種修勻方法，多半可算是「同地異時」形式的增加樣本。
 - 死亡率模型在樣本數少時也會遭遇問題。
- 死亡率模型的參數估計在人數少時有偏誤。
 - Lee-Carter模型的 α_x 及 β_x 估計值，在人數少於200,000會有明顯偏誤。
 - Cairns-Blake-Dowd模型(Chen et al., 2017)及羅吉斯模型的參數估計值有較大變異數。

$$\alpha_x - \hat{\alpha}_x$$



$$\beta_x - \hat{\beta}_x$$



Lee-Carter模型的參數估計值之偏誤

如何增加樣本數？

- 增加樣本數是解決死亡率模型參數估計最直接的方法，但如何增加？
 - MWA之類的修勻方法可用嗎？
- 加入死亡率特性類似的樣本是較為可行的作法，但如何定義「類似」？
 - 死亡率指標？(Li and Lee, 2004 & 2005)
 - 仿造「同地同時」的傳統修勻方法，建立特性類似的參考母體。

Partial Standard Mortality Ratio

- Lee (2003) 提出參考大區域，在死亡率比值與 SMR 間求取加權幾何平均的修勻方法。

$$v_x = u_x^* \times \exp \left(\frac{d_x \times \hat{h}^2 \times \log(d_x / e_x) + (1 - d_x / \sum d_x) \times \log(\text{SMR})}{d_x \times \hat{h}^2 + (1 - d_x / \sum d_x)} \right)$$

$$\text{SMR} = \frac{\sum_x d_x}{\sum_x n_x \cdot u_x^*}, \quad \hat{h}^2 = \max \left(\frac{\sum (d_x - e_x \times \text{SMR})^2 - \sum d_x}{\text{SMR}^2 \times \sum e_x^2}, 0 \right)$$

→ n_x 及 d_x 為小區域 x 歲人數及死亡數

→ u_x^* 為大區域 x 歲死亡率



Whittaker Mortality Ratio

- Whittaker比值法

- 以SMR調整死亡率比值異常的年齡。

- 令 x 歲死亡率比值 $r_x = u_x / u_x^*$

- 當 $r_x = 0$ 或 $r_x > 2 \times \text{SMR}$ ，則該年齡死亡率比值以SMR取代。

- 以Whittaker修勻死亡率比值。

- 相較於Partial SMR，參考SMR訊息較低，且多考慮年齡間的變化。



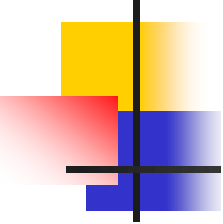
死亡率的調整值

- 因為人口結構不同，若直接以粗死亡率的數值當作死亡率高低的判斷標準，則老人比例高的地區，死亡率通常比較高。
- 根據比較地區的死亡資料是否完整，選用兩種調整方式：
 1. 直接調整法(Direct Method of Adjustment)
 2. 間接調整法(Indirect Method of Adjustment)

- 
- 如果要比較的地區之年齡別死亡率可完整獲得，可使用直接調整法：

$$ADR_D = \frac{\sum_x {}_n P_x^s \cdot {}_n m_x^j}{\sum_x {}_n P_x^s} = \frac{\sum_x {}_n P_x^s \cdot {}_n m_x^j}{P^s}$$

其中 ${}_n P_x^s$ 為標準母體 x 到 $x+n$ 歲的人數
 ${}_n m_x^j$ 為第 j 地區 x 到 $x+n$ 歲的死亡率

- 
- 如果無法獲得年齡別死亡率，只能獲得各年齡層的人數，可使用間接調整法：

$$ADR_I = \frac{D^j}{\sum_x {}_n m_x^s \cdot {}_n P_x^j} \cdot \frac{D^s}{P^s}$$

也就是說每一地區藉由與標準母體的標準死亡比(Standard Mortality Ratio ; SMR)來比較：

$$SMR = \frac{D^j}{\sum_x {}_n m_x^s \cdot {}_n P_x^j}$$

範例(Brown)：

密西根州與佛羅里達州

MICHIGAN

Age Group	Population on July 1, 1985 (thousands)	Percent	Deaths in 1985	$n m_x \cdot 10^3$
0-5	662	7.3	1,889	2.85
5-15	1,366	15.0	385	0.28
15-25	1,568	17.3	1,543	0.98
25-35	1,600	17.6	2,049	1.28
35-45	1,186	13.1	2,592	2.19
45-55	842	9.3	4,512	5.36
55-65	844	9.3	11,460	13.58
65-75	618	6.8	18,264	29.55
75-85	306	3.4	20,637	67.44
85+	95	1.0	15,381	161.91
Total	9,087		78,712	8.66

FLORIDA

Age Group	Population on July 1, 1985 (thousands)	Percent	Deaths in 1985	$n m_x \cdot 10^3$
0-5	750	6.6	2,241	2.99
5-15	1,348	11.9	419	0.31
15-25	1,677	14.8	1,847	1.10
25-35	1,775	15.6	2,713	1.53
35-45	1,402	12.3	3,270	2.33
45-55	1,105	9.7	5,986	5.42
55-65	1,308	11.5	15,301	11.70
65-75	1,201	10.6	29,875	24.88
75-85	641	5.6	36,292	56.62
85+	<u>161</u>	1.4	<u>23,131</u>	143.67
Total	11,368		121,075	10.65

標準母體：全美國人口

UNITED STATES

Age Group	Population on July 1, 1985 (thousands)	Percent	Deaths in 1985	$n m_x \cdot 10^3$
0-5	18,004	7.54	47,369	2.63
5-15	33,923	14.21	8,933	0.26
15-25	39,551	16.57	37,935	0.96
25-35	42,027	17.60	51,852	1.23
35-45	31,764	13.31	65,815	2.07
45-55	22,589	9.46	116,634	5.16
55-65	22,337	9.36	286,480	12.83
65-75	17,009	7.12	482,646	28.38
75-85	8,836	3.70	568,848	64.38
85+	<u>2,695</u>	1.13	<u>419,051</u>	155.49
Total	238,735		2,086,440*	8.74

- 若依直接調整法計算，可得

$$ADR_D(\text{Michigan}) = 9.14224$$

$$ADR_D(\text{Florida}) = 8.12808$$

→ Michigan的死亡率較高！

- 若依間接調整法計算，可得

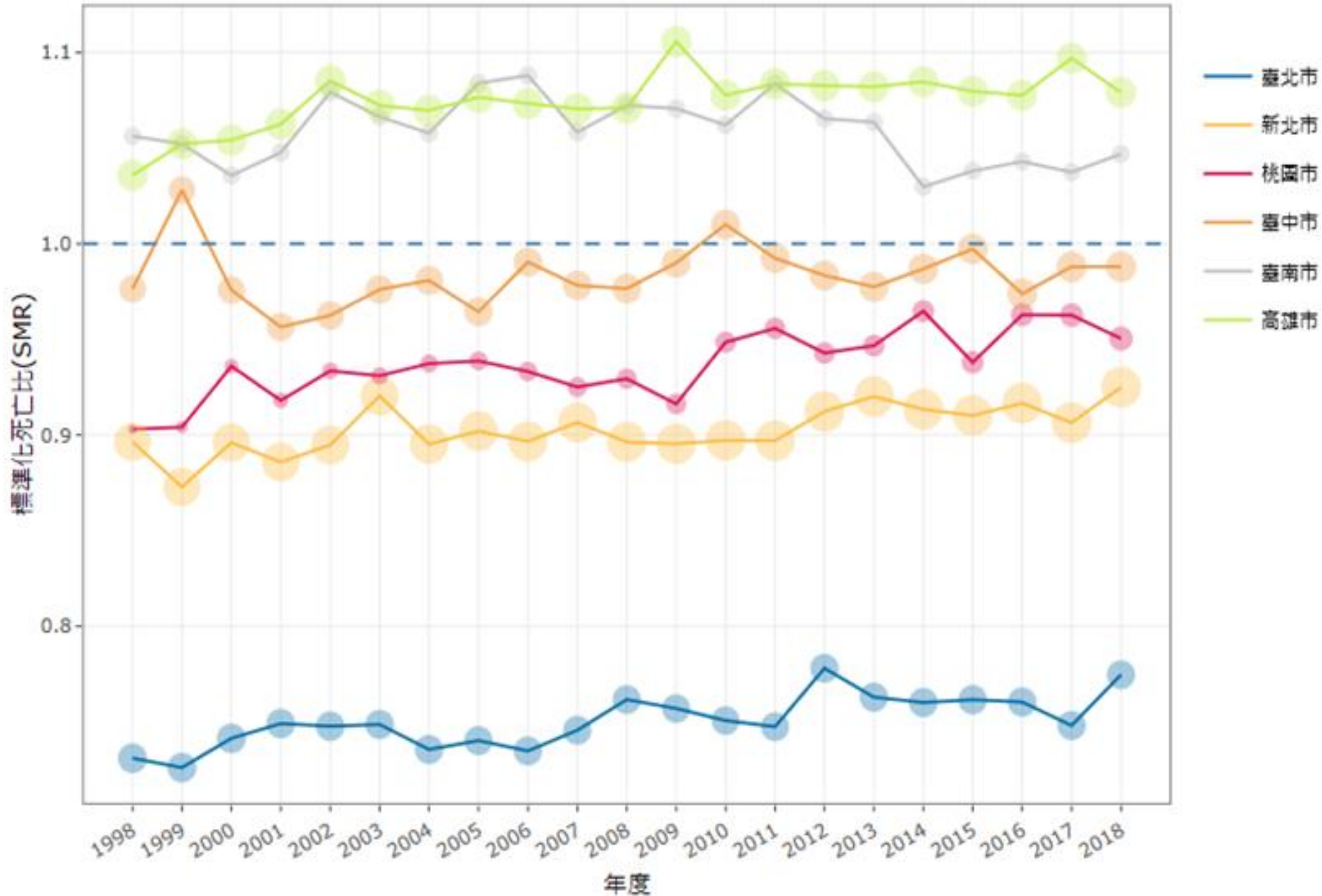
$$ADR_I(\text{Michigan}) = 9.14670$$

$$ADR_I(\text{Florida}) = 8.02307$$

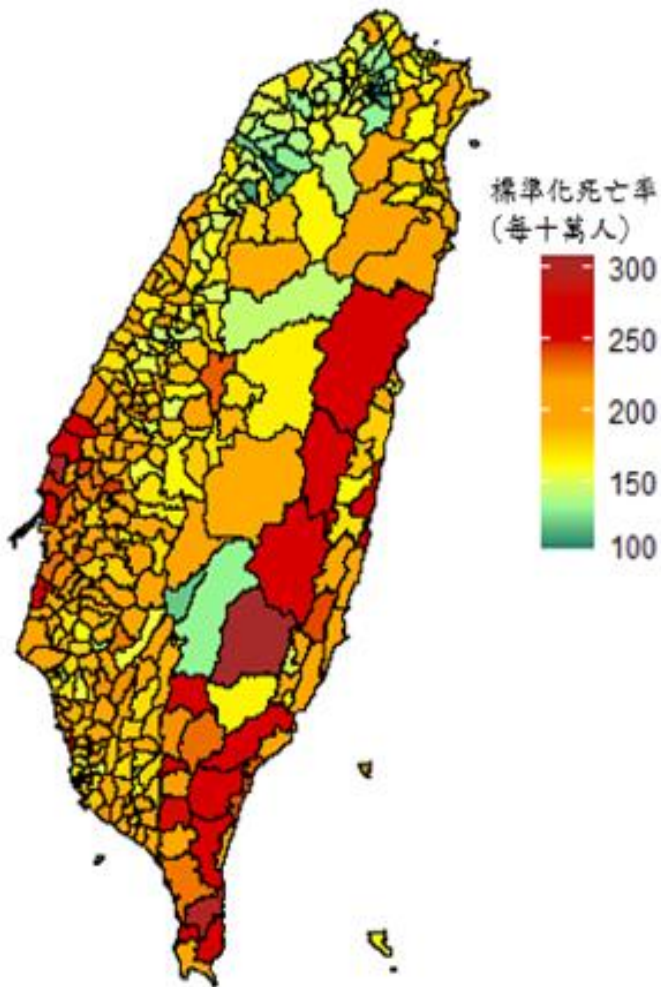
→ 其中 $SMR(\text{Michigan}) = 1.04658$

$$SMR(\text{Florida}) = 0.918017$$

六個院轄市SMR的趨勢 (1998~2018年男性)

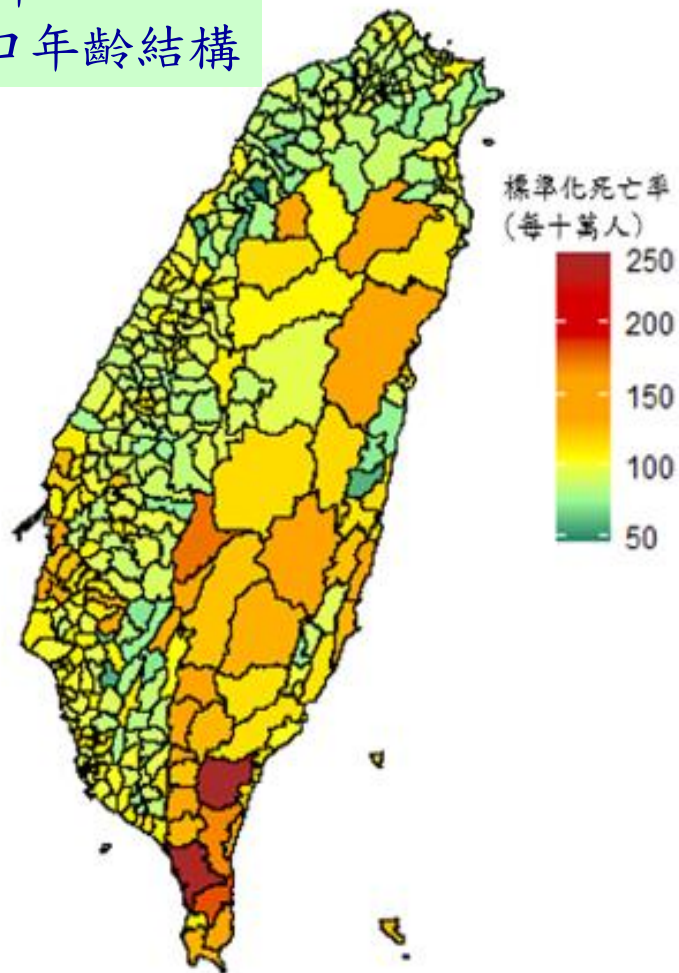


男



女

2000年WHO
標準人口年齡結構



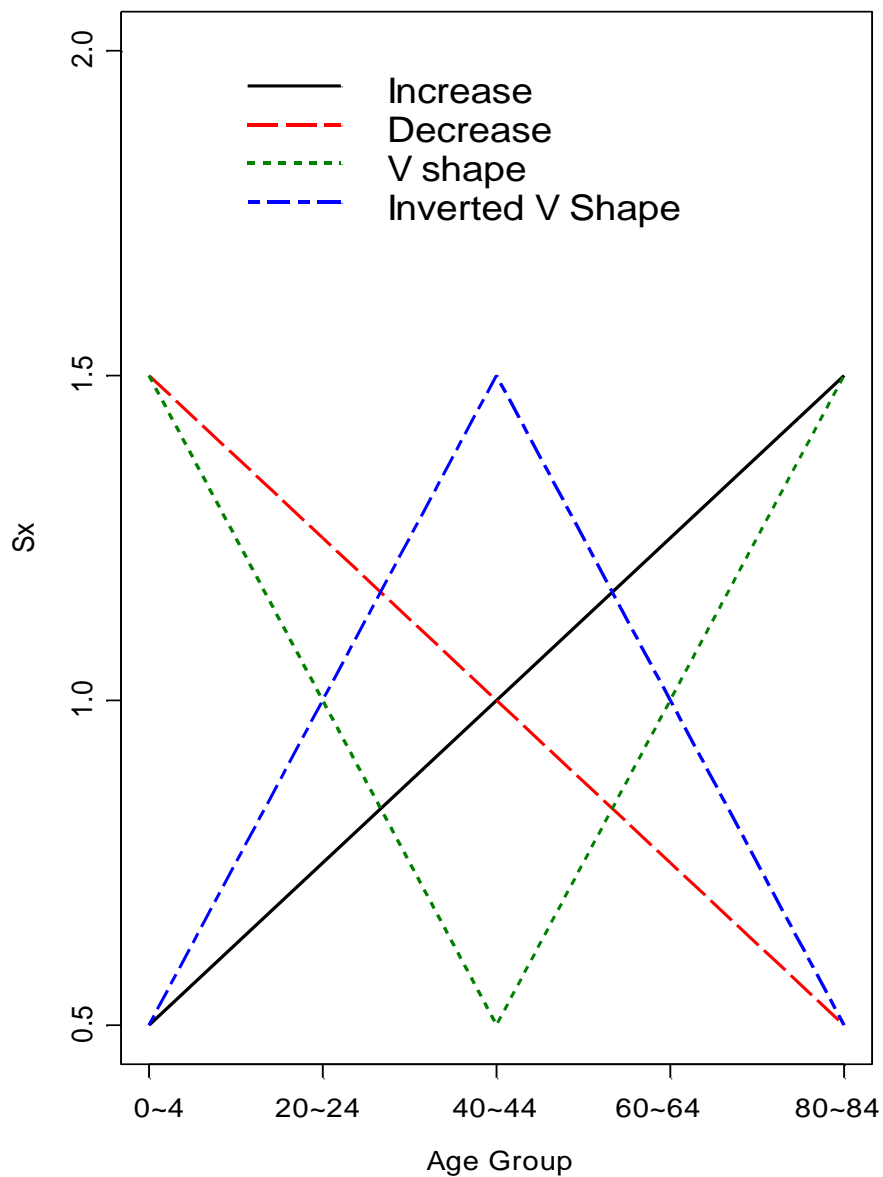
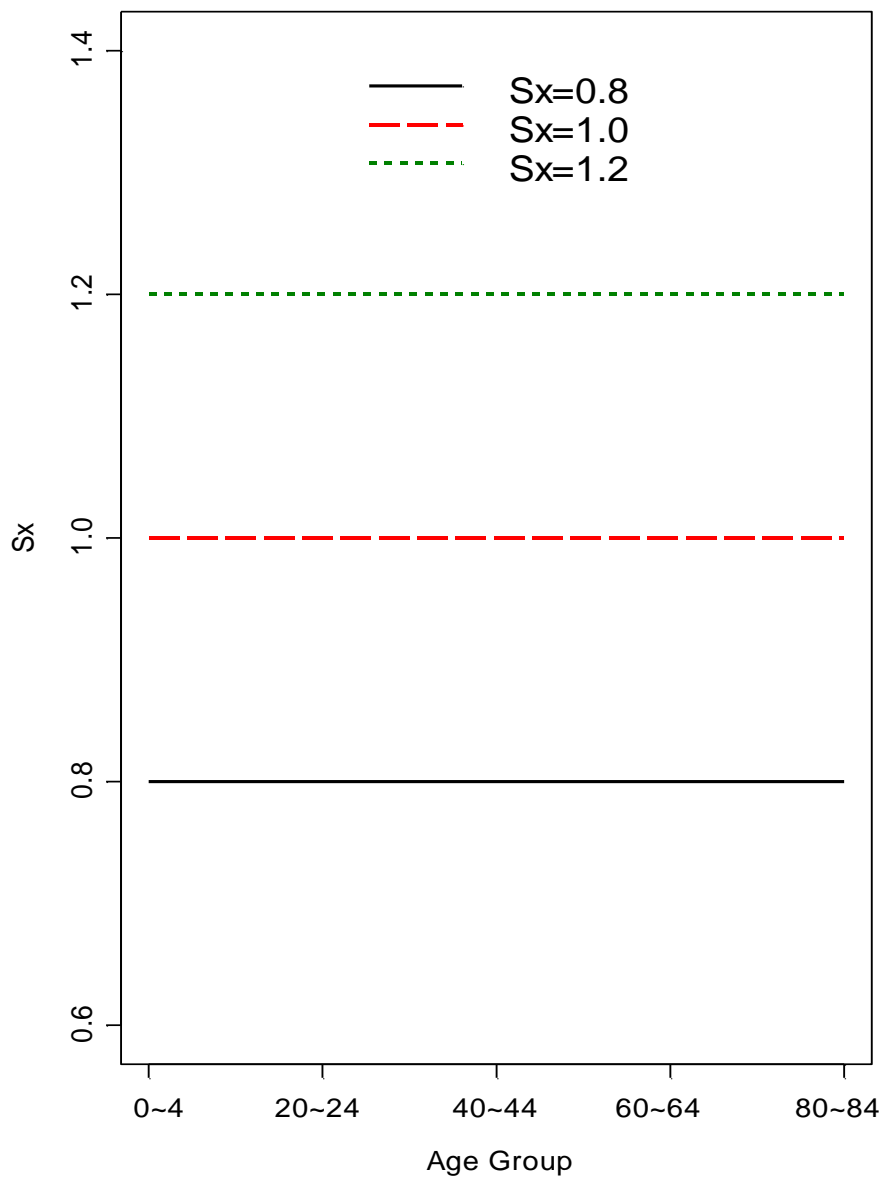
鄉鎮市區癌症死因的標準化死亡率(2014~2018年)



關於死亡率的調整值

- 直接調整法與間接調整法可能不同嗎？
- 直接調整法類似「分配函數的大小」，即若對任意 t , $P(X \leq t) \geq P(Y \leq t)$ ，則 $X \geq^D Y$ 。
→ 更嚴格的標準為類似歷次國民生命表，後面年度的年齡別死亡率都較低。
- 間接調整法以小地區人口結構為加權對象，例如：若小地區高齡死亡率較低、且高齡人口比例也高，可能會得到SMR小於一。

大小區域的七種死亡情境設定



小區域修勻方法的模擬結果

MAPE (%)

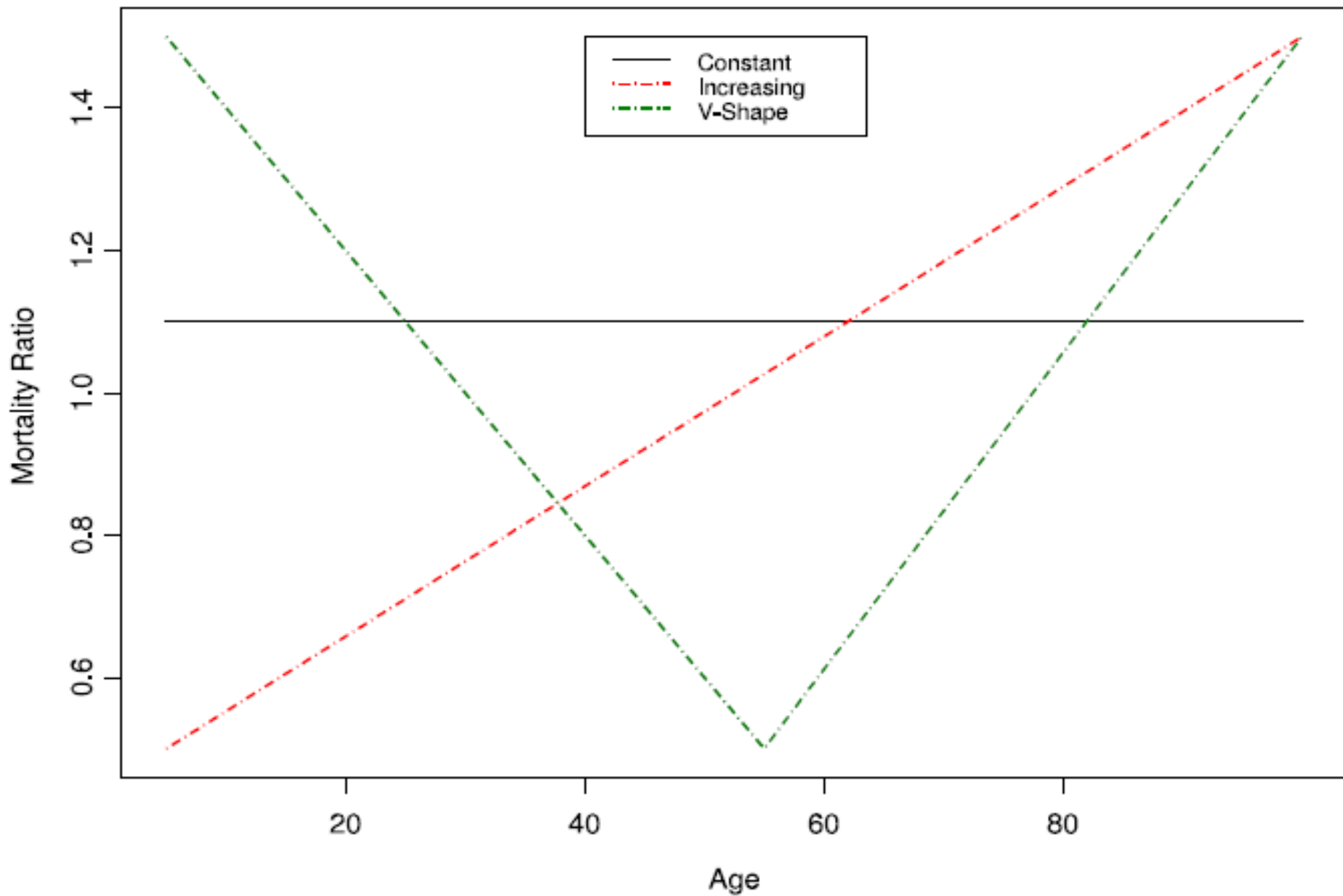
	80%	100%	120%	Incr.	Decr.	V	Λ
Raw	40.8	36.7	33.1	42.5	33.4	35.2	39.6
Whittaker	31.7	29.6	27.9	32.9	28.4	31.7	29.8
<input checked="" type="checkbox"/> Partial SMR	4.3	3.8	3.6	46.7	25.6	25.5	18.5
<input checked="" type="checkbox"/> Whittaker Ratio	12.7	12.7	12.0	32.8	22.3	13.5	16.0

註：小區域以信義區為基準，臺北市為參考母體。

異地同時的修勻建議

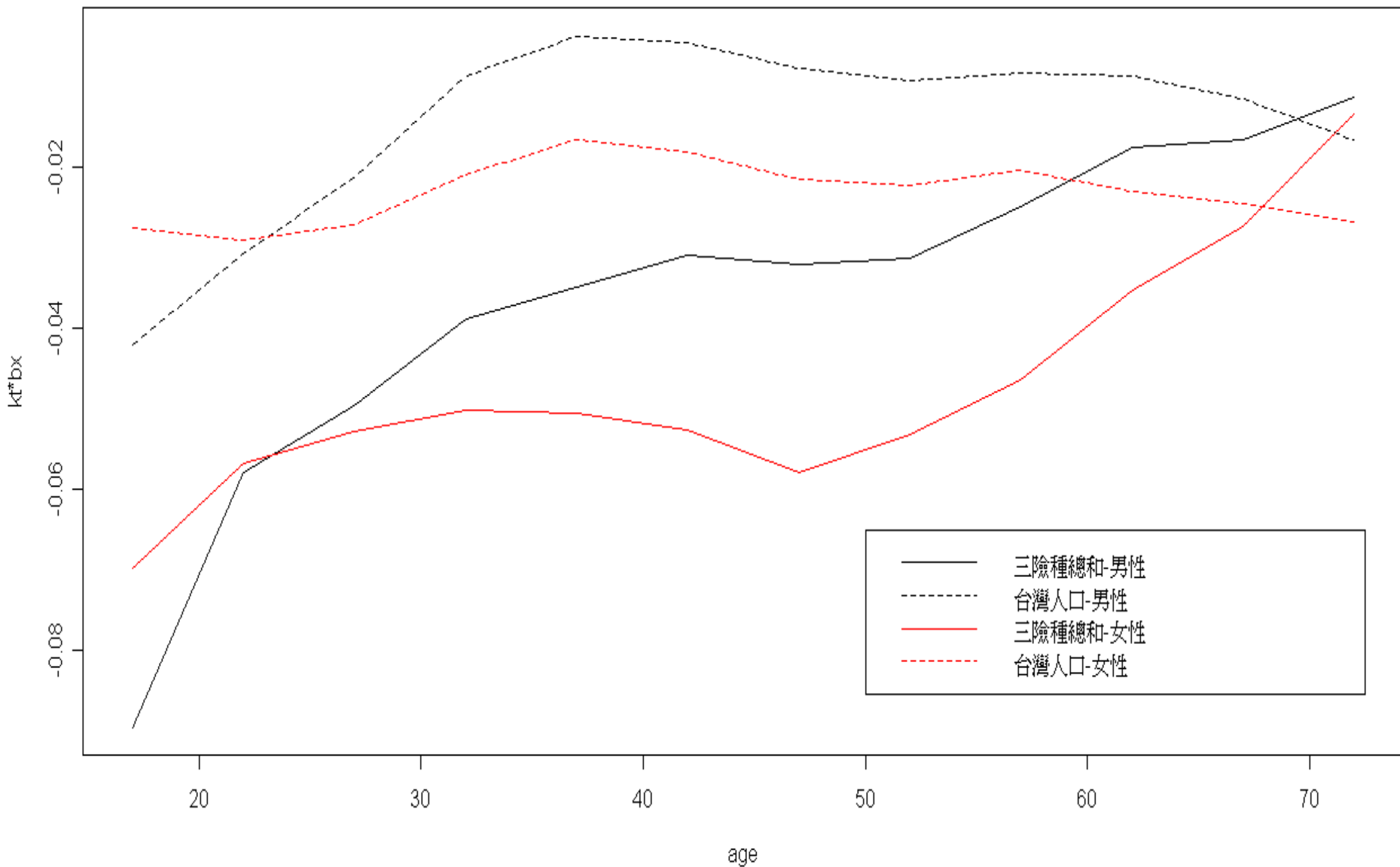
- Partial SMR及Whittaker死亡率比值都是不錯的修勻方法，只要小區域與大母體的死亡率趨勢類似。（例如： $a \leq 0.3$ ）
 - 不容易選擇與死亡率類似的母體，但若可蒐集小區域的歷史資料，例如：累積過去20年資料就像是參考20倍人數的大母體。
- 若死亡率變化很有規律，加總小區域人數的「大母體」可與Partial SMR及Whittaker死亡率比值結合。（異地同時與同地異時本質相通。）

簡化的三種死亡情境設定



死亡率改善程度比較(商業保險vs.全台灣人口)

死亡率下降幅度



Partial SMR與Whittaker Ratio的比較

MAPE (%) of 3 mortality scenarios (Population size = 50,000).

<i>a</i>		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Constant	Raw	30.1	28.7	27.6	26.5	25.3	24.7	23.3	23.0	22.4	21.8
	Whittaker	13.2	12.8	12.5	12.3	11.9	11.7	11.5	11.4	11.1	10.9
	PSMR	5.0	4.6	4.6	4.2	4.1	4.1	3.9	3.8	3.7	3.7
Increasing	Raw	30.2	32.4	36.0	41.3	47.6	56.2	66.2	79.8	99.2	143.3
	Whittaker	13.3	15.1	18.5	23.2	28.9	36.3	45.3	57.9	78.0	122.3
	PSMR	4.9	7.5	12.7	19.4	27.4	37.0	48.8	64.9	88.8	140.9
V-shape	Raw	30.2	31.0	33.6	38.0	43.7	51.4	60.1	72.7	90.9	130.6
	Whittaker	13.3	14.2	17.1	21.5	27.1	33.7	41.4	52.4	68.7	105.8
	PSMR	4.9	7.5	12.5	18.8	26.4	35.2	45.4	59.1	79.2	123.0

Population size = 200,000

<i>a</i>		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Constant	Raw	15.0	14.3	13.6	13.1	12.7	12.3	11.9	11.5	11.3	10.9
	Whittaker	8.5	8.3	8.0	7.7	7.4	7.3	7.0	6.9	6.8	6.6
	PSMR	2.5	2.4	2.2	2.2	2.1	2.0	2.0	1.9	1.9	1.8
Increasing	Raw	14.9	17.2	22.2	27.8	35.5	44.0	54.7	68.3	90.0	132.9
	Whittaker	8.4	10.4	14.6	19.4	25.8	32.8	42.0	53.5	72.5	112.8
	PSMR	2.4	6.2	12.5	19.9	28.5	38.2	50.1	65.6	89.4	138.6
V-shape	Raw	14.9	16.7	21.6	27.4	34.4	42.9	52.9	65.8	85.2	125.0
	Whittaker	8.5	10.5	14.9	20.5	26.9	34.5	43.4	54.9	72.2	109.5
	PSMR	2.4	6.2	12.4	19.5	27.5	36.5	47.0	60.5	80.3	121.0

註：臺灣地區(1990-2009)男性五齡組為基準。

同地同時、同地異時的修勻誤差比較

MAPE of simulation (Taiwanese Male, 1990-2009, 5-age) (Unit: %).

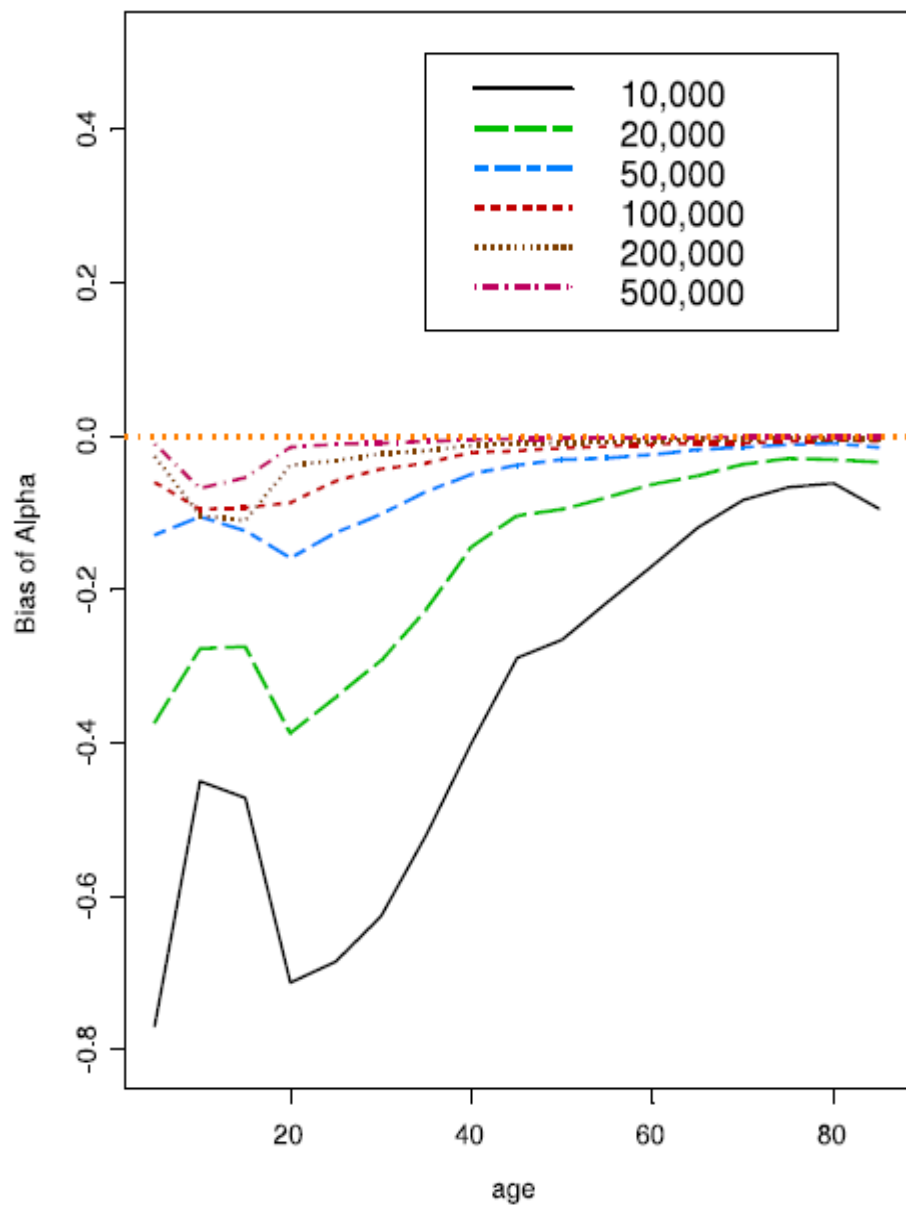
	10,000	20,000	50,000	100,000	200,000	500,000
Raw	68.23	50.59	32.90	22.88	16.28	10.27
Lee-Carter	33.57	23.67	15.53	10.97	8.66	6.05
Partial SMR	14.31	11.75	9.68	8.70	8.09	7.50

MAPE of simulation (Taiwanese Male, 1990-2009, single-age) (Unit: %).

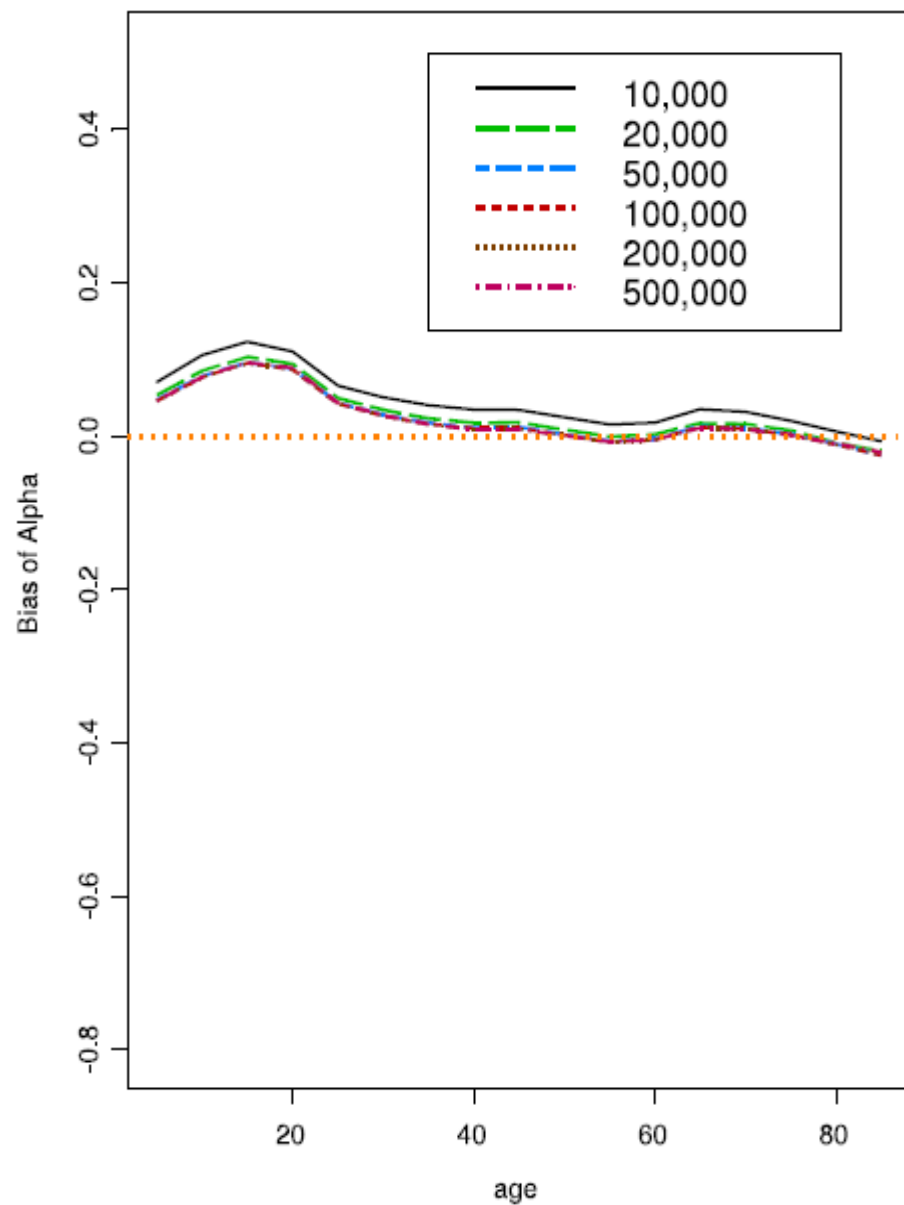
	10,000	20,000	50,000	100,000	200,000	500,000
Raw	125.56	101.45	73.01	54.89	39.40	24.60
Lee-Carter	69.77	52.70	31.11	19.93	13.30	8.32
Partial SMR	18.02	14.82	12.47	11.39	10.72	10.19

註：臺灣地區(1990-2009)男性五齡組為基準。

LC



PSMR+LC

**Fig. 7.** α_x estimates bias for different models.

MAPE (%) of various aggregation years (5-age).

	10,000	20,000	50,000	100,000	200,000	500,000
PSMR+20 years	14.31	11.75	9.68	8.70	8.09	7.50
PSMR+15 years	13.47	10.78	8.72	7.78	7.10	6.39
PSMR+10 years	13.80	11.21	9.19	8.25	7.51	6.57
PSMR+5 years	14.83	12.50	10.67	9.75	8.93	7.73

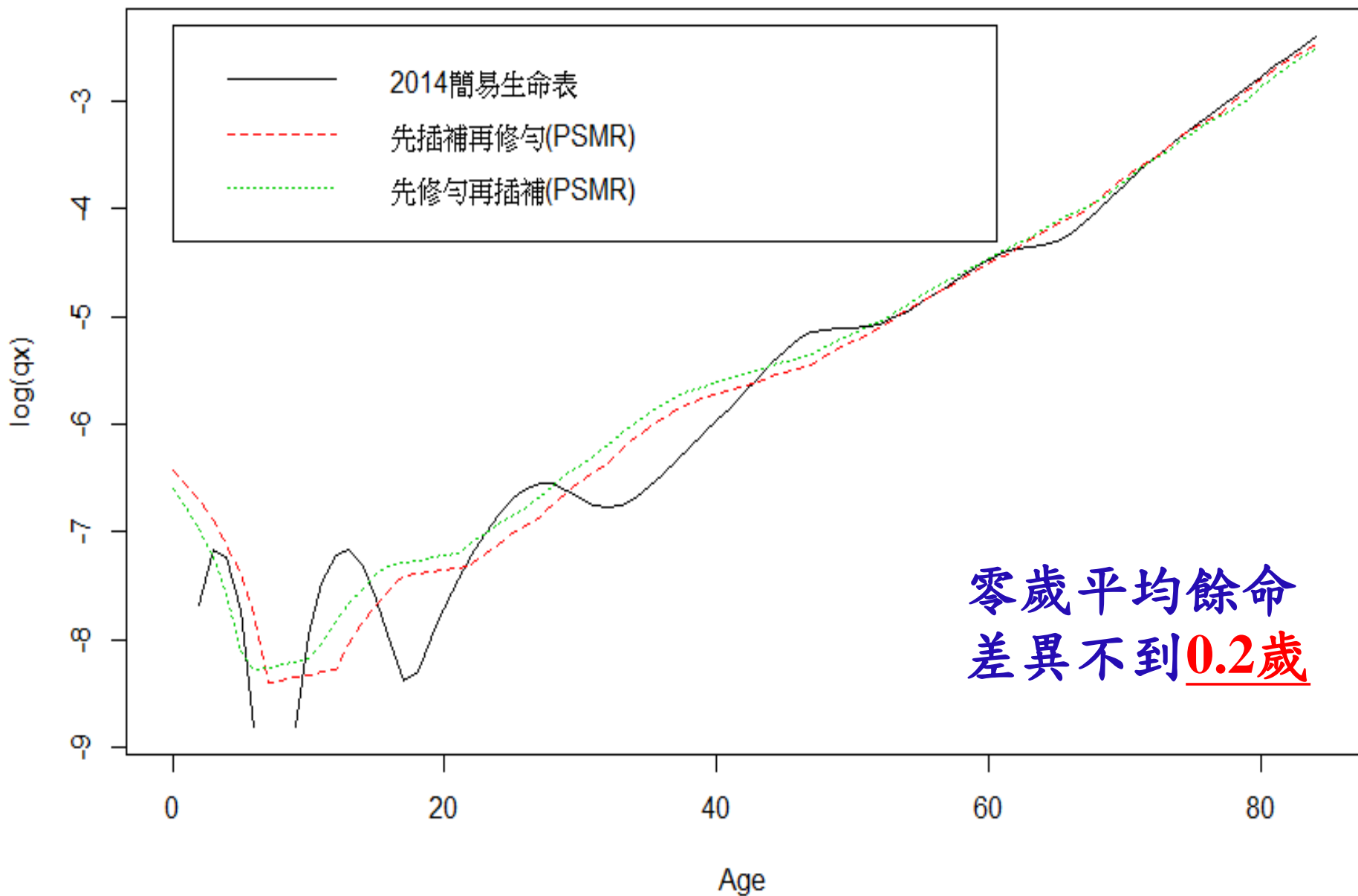
■ 合併年數

→ 為了取得足夠的樣本數，建議累積15~20年的人口資料；或可拆解成前十年、後十年分別估計模型參數，如果參數估計值接近，則累積樣本的作法可行。

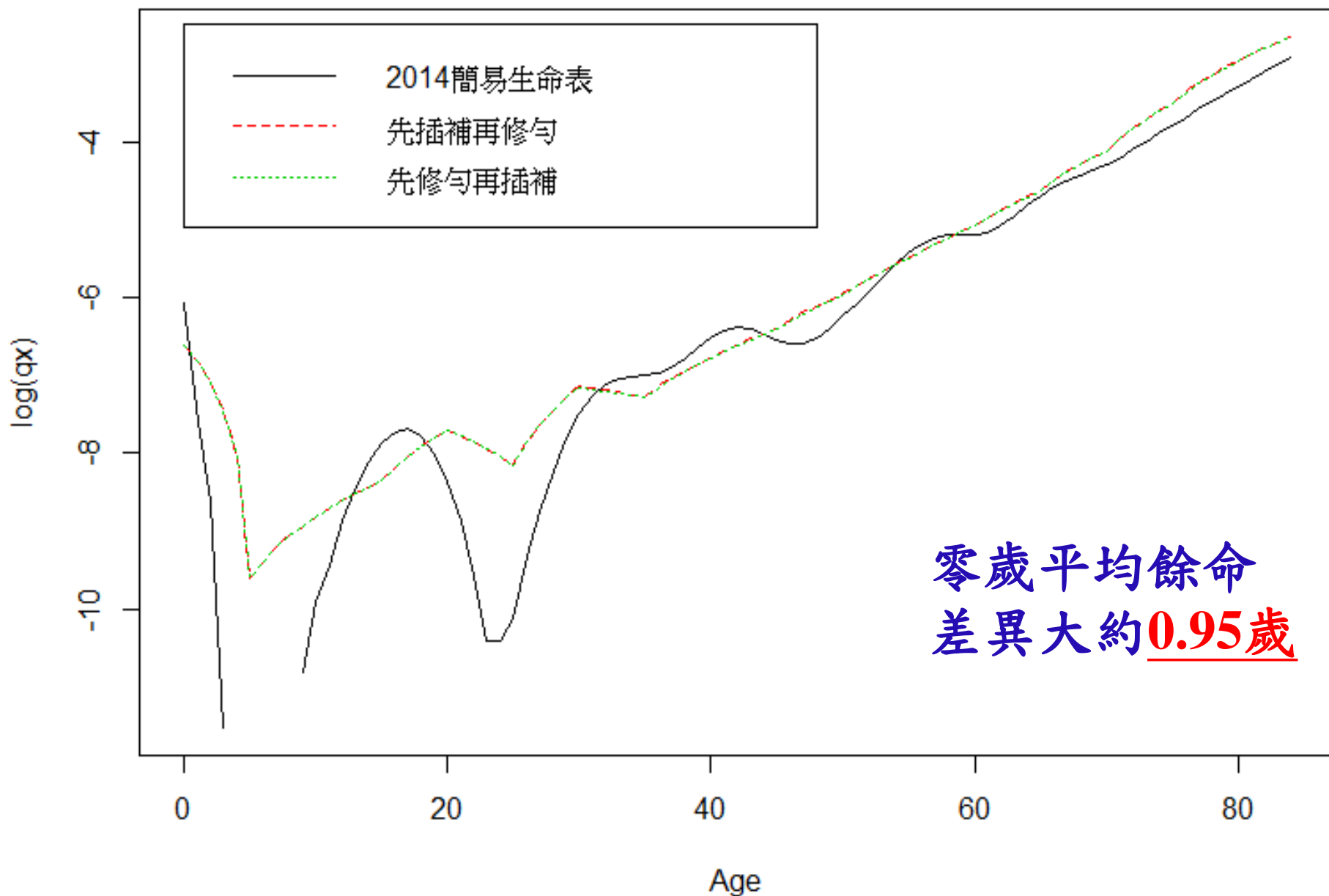
簡易生命表編算方法的瑕疵

- 現行簡易生命表修勻主要採行Greville三次九項公式，亦即以編算年齡（單齡）為中心，往左、往右各加入四個單齡的死亡率。
 - 編算地區人數少，五齡組死亡數可能為零，修勻後的單齡死亡率為0（澎湖縣）。
- 使用上述「同地異時」可避免傳統修勻在五齡組死亡數為零遇到的問題。
 - 換言之，簡易生命表會因為人數引起死亡率修勻瑕疵，是否應設定最小編算人數？

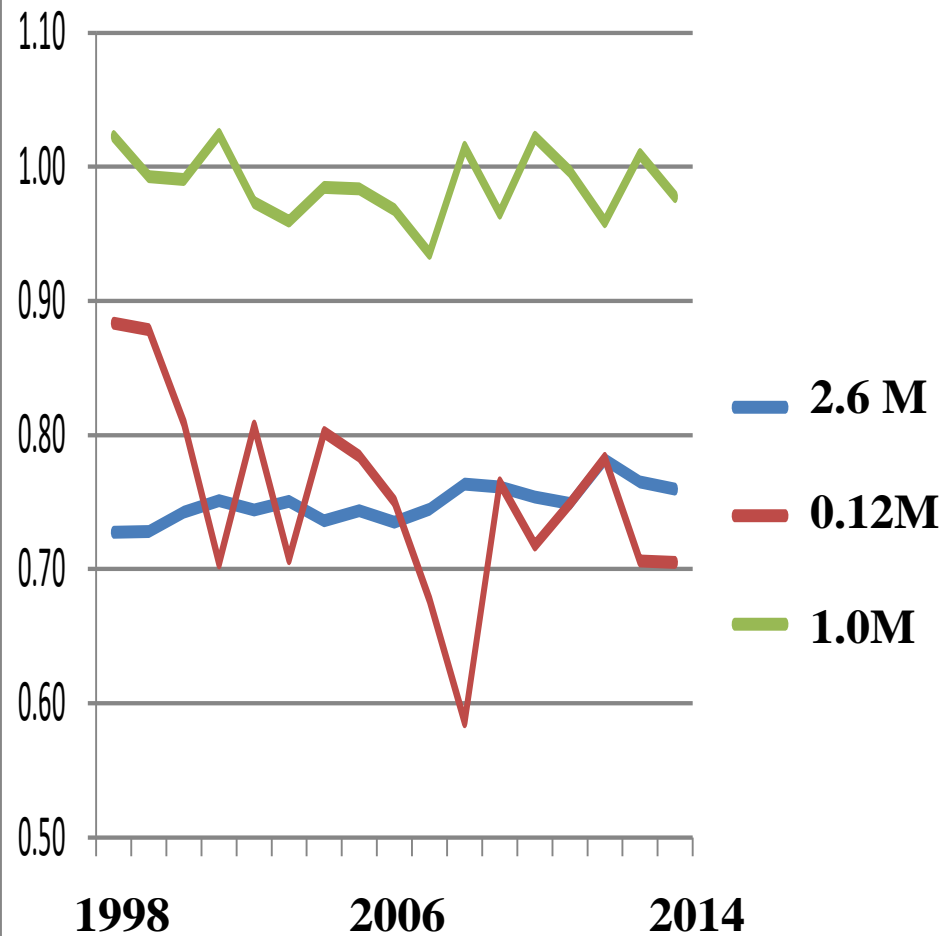
小區域編算方法實例（2014年澎湖縣男性）



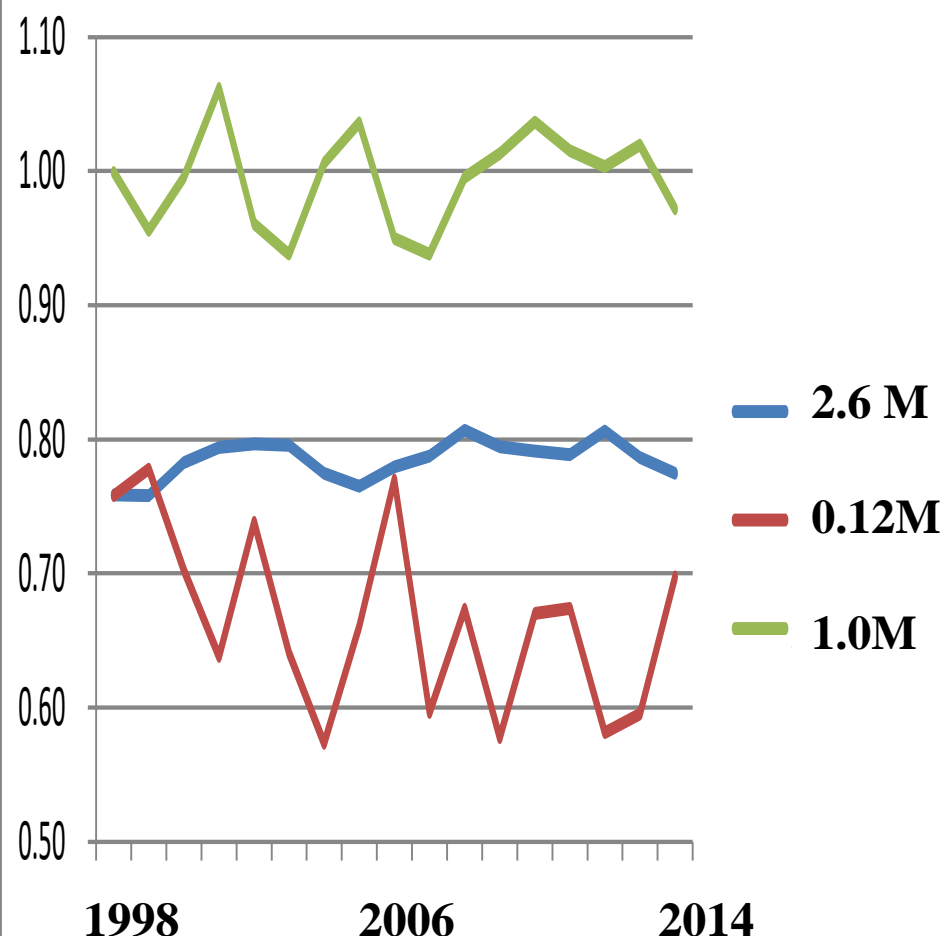
小區域編算方法實例（2014年澎湖縣女性）



男性



女性



人數少時SMR也會有比較大的震盪



小區域編算方法的建議

- 傳統修勻能降低年齡死亡率的震盪，單齡或五齡組在人數不少於20萬時可行。
 - 若能給定歷史資料，「同地異時」方法明顯優於傳統修勻方法。
 - 人數較少的地區（如澎湖縣）可選用「同地異時」，現行的修勻方法會有瑕疵。
- 人數太少的縣市可編算生命表嗎？