

國立臺灣大學公共衛生學院流行病學與預防醫學研究所

博士論文

Institute of Epidemiology and Preventive Medicine

College of Public Health

National Taiwan University

Doctoral Dissertation



疾病率時空分析方法：台灣砷相關癌症之應用

Methods for Spatiotemporal Analysis of Disease Rates:

Application to Arsenic-related Cancers in Taiwan

莊景榮

Jing-Rong Jhuang

指導教授：李文宗 博士

Advisor: Wen-Chung Lee, M.D., Ph.D.

中華民國 112 年 8 月

August 2023

國立臺灣大學博士學位論文
口試委員會審定書

論文中文題目 疾病率時空分析方法：台灣砷相關
癌症之應用

論文英文題目 Methods for Spatiotemporal
Analysis of Disease Rates: Application to
Arsenic-related Cancers in Taiwan

本論文係莊景榮君（學號 F05H41001）在國立臺灣大學
流行病學與預防醫學研究所完成之博士學位論文，於民國
112 年 7 月 26 日承下列考試委員審查通過及口試及格，特此
證明。

口試委員：

李文宗

（簽名）

（指導教授）

杜祿康

林苑俞

溫石弘

廖勇柏

致 謝



回顧七年的碩博士班，非常感謝在求學路上幫助過我的師長們。感謝指導教授李文宗老師，所謂「師父領進門，修行在個人」，老師重視邏輯思考能力，讓我有雄厚的內力基礎，學習新技能會特別快。感謝口試委員杜裕康教授、林菟俞教授、陳忠信副教授、溫在弘教授、與廖勇柏教授惠賜寶貴意見，使得論文內容更加完整充實。感謝中研院陳建仁院士與高醫大李志宏教授，對實證應用提供寶貴意見，使得論文內容更臻完善。感謝溫敏杰主任、杜裕康所長、張淑惠教授、程蘊菁教授、與盧子彬教授，熱心分享學習與求職的經驗。特別感恩詹長權院長，與詹院長的合作使我逐漸習得跨部門溝通的能力。

我也感謝碩博士班期間幫助過我的朋友們。台灣癌症登記中心給予我課外學習的機會，培養我從懵懂的新鮮人成長為熟練的社會人。感謝濬如與雅雯教導我正確的處事態度，也謝謝瑋玲熱心給予關心，因為有妳們，讓我在工作與求學間得心應手。感謝霽修與宜蓉適時幫我加油打氣，也感謝學習時的夥伴馨尹、姿麟、明潔、正慈、美婷、芝潤、怡誼、與驊珊，在我感到疲累時對我伸出溫暖的手。

幸福家庭是我在求學路上的避風港，每當面臨徬徨無助時，家人們都是我的心靈導師，給予溫暖、鼓勵、與支持。最後我要感謝我的信仰，感恩天地神明在我求學期間保佑最愛的家人身體健康，特別感恩關聖帝君、保生大帝、文昌帝君、與朱衣神君賜予我健康身體與聰穎智慧，協助我保持正向的態度，得以順利完成學位。

莊景榮 謹識

中華民國一百一十二年八月八日

目 錄



口試委員審定書.....	#
誌謝.....	i
表目錄.....	iii
圖目錄.....	iv
中文摘要.....	v
英文摘要.....	vi
第一章 導論.....	1
第二章 研究方法.....	5
第三章 電腦模擬.....	10
第四章 實證資料應用.....	12
第五章 結論與展望.....	18
參考文獻.....	19
附錄.....	36

表目錄

表一、單一發生率熱點隨出生世代縮小的情境下，各種方法的對稱平均絕對百分 誤差(%).....	25
表二、單一發生率熱點不隨出生世代改變的情境下，各種方法的對稱平均絕對百 分誤差(%).....	26
表三、雙發生率熱點都隨出生世代縮小的情境下，各種方法的對稱平均絕對百分 誤差(%).....	27
表四、雙發生率熱點分別隨出生世代快速與緩慢縮小的情境下，各種方法的對稱 平均絕對百分誤差(%).....	28
表五、無發生率熱點的情境下，各種方法的對稱平均絕對百分誤差 (%).....	29

圖目錄

圖一、1995-2019 年砷相關癌症年齡標準化發生率長期趨勢.....	31
圖二、1913-1977 出生世代 60-64 歲族群的砷相關癌症發生率趨勢.....	32
圖三、1913-1977 出生世代 60-64 歲族群的肝細胞癌及小細胞與鱗狀細胞肺癌發生率地理分布.....	33
圖四、1913-1977 出生世代 60-64 歲族群的鮑恩氏病及基底細胞與鱗狀細胞皮膚癌發生率地理分布.....	34
圖五、1913-1977 出生世代 60-64 歲族群的膀胱泌尿上皮癌及上泌尿道泌尿上皮癌發生率地理分布.....	35

中文摘要



烏腳病是 1950 至 1990 年台灣西南沿海地區的流行病。本文欲利用台灣癌症登記資料比較砷相關癌症在烏腳病流行地區與非流行地區的發生率。然而 1995 年起癌症登記資料的品質才趨於穩定，一般的發生率長期趨勢難以看出烏腳病防治成效。再者烏腳病是一種地方性流行病，發生率的地理分布也是重要的議題。本文提出一種時空地圖的繪製方法。我先使用隨機效應年齡-年代-世代模型得到各個行政區域的時空參數估計，再對時空參數進行內插，求得精細至經緯度的時空參數。歷經前述步驟可繪製出疾病率在年齡、年代、與世代的等高線地圖，還能製作影片播放疾病率的時空變化。我也進行電腦模擬比較經驗貝氏法、核密度估計法、克里金法、以及穩定克里金法等內插方法，結果顯示穩定克里金法的表現為最佳。最後，我使用本文的方法分析台灣癌症登記中的砷相關癌症資料，發現烏腳病流行地區與非流行地區的癌症發生率差距在 1958-1962 出生世代後逐漸消失，這個轉折點恰好是烏腳病防治的起始時間，時空地圖也顯示砷相關癌症發生率群聚在西南沿海地區，在 1960 出生世代後逐漸消失。本文的方法不僅在電腦模擬有優於傳統方法的表現，在實例應用中也展示具有公共衛生價值的發現。

關鍵字：隨機效應年齡-年代-世代模型；內插；時空地圖；砷相關癌症；發生率

英文摘要



Blackfoot disease was endemic to southwestern Taiwan during the 20th century. This study aims to use the Taiwan Cancer Registry dataset to compare incidence rates of arsenic-related cancers between Blackfoot disease-endemic areas and the remaining areas of Taiwan. However, not until 1995 that the Taiwan Cancer Registry could provide high-quality data. Therefore, the long-term trend in the incidence rates cannot be used to evaluate the effectiveness of Blackfoot disease prevention. In addition, the geographical distribution of arsenic-related incidence rates is also an important issue. This study proposed a spatiotemporal mapping method. First, the random-effects age-period-cohort model is fitted to obtain spatiotemporal parameter estimations in each administration area. Then, these estimations are further interpolated to make predictions on each longitude and latitude. With the procedures, we can draw contour maps of disease rates according to age, period, and cohort and create a movie to show spatiotemporal dynamics vividly. The Monte-Carlo simulation compared interpolation methods, including empirical Bayes, kernel density estimation, kriging, and stabilized kriging. The stabilized kriging performed better among all. Finally, we analyzed arsenic-related cancers from the Taiwan Cancer Registry dataset. The incidence gap between Blackfoot disease-endemic areas and the remaining areas of Taiwan shrunk after the 1958-1962 birth cohort, coinciding with the initial prevention of Blackfoot disease. In addition, spatiotemporal clusters of high incidence rates were identified in southwestern Taiwan, and the clusters also started to dissipate after the 1960 birth cohort. Our method demonstrated appropriate ability and contributed to public health.

Keywords: random-effects age-period-cohort model; interpolation; spatiotemporal mapping; arsenic-related cancers; incidence rate

第一章 導論



1.1 研究動機

描述性流行病學是癌症監測的利器。國際癌症研究機構(International Agency for Research on Cancer)定期發表發生率統計。最近的結果顯示全球癌症負擔仍居高不下¹，不過高度開發國家因有效推動防治計畫而逐漸減輕癌症負擔^{2,3}。台灣癌症登記(Taiwan Cancer Registry)能提供可靠數據作為台灣癌症監測的重要依據⁴⁻⁶。我曾利用台灣癌症登記資料探討發生率的長期趨勢，也發現特定癌症的發生率趨勢在相關的防治計畫實施後逐漸轉為下降⁷⁻⁹。

烏腳病(Blackfoot disease)是 1950 至 1990 年台灣西南沿海地區的流行病。本文利用台灣癌症登記資料比較烏腳病流行地區與非流行地區的癌症發生率。然而，1995 年起台灣癌症登記資料的品質才趨於穩定，一般的發生率長期趨勢難以看出烏腳病防治的影響。再者烏腳病是一種地方性流行病，癌症發生率的地理分布也是重要的議題。本文回顧描述性流行病學的時間與空間的分析方法，包含年齡-年代-世代分析(age-period-cohort analysis)及癌症地圖的繪製方法(cancer mapping)，也列舉國內曾進行過的研究。接著提出一種時空地圖繪製方法(spatiotemporal mapping)，可以繪製疾病率在三種時間維度上等高線地圖。最後分析台灣癌症登記資料中的砷相關癌症(arsenic-related cancers)，探討發生率在各個出生世代的地理分布之變化。

1.2 研究背景

流行病學關心年齡(age)、年代(period)、及世代(cohort)共三種時間維度。流行病學家經常繪製疾病率的年齡曲線(age curve)或長期趨勢(long-term trend)，例如當代年齡曲線(cross-sectional age curve)或世代年齡曲線(longitudinal age curve)可呈現疾病率在年代或世代的年齡走勢；年齡標準化率(age-standardized rate)長期趨勢可表現年齡調整的疾病率隨著年代的波動；年齡別率(age-specific rate)曲線可顯現各年齡層的疾病率隨著年代或世代的轉變。流行病學家也建構年齡-年代-世代模型(age-period-cohort model)^{10,11}，將三種時間維度同時放入一個統計模型中，期待獲得這三種時間分別對於疾病率的獨立效應(independent effect)。三種時間效應在流行病學的意義有所不同，年齡效應(age effects)描述疾病率於不同階段生命歷程(life

course)的走勢，年代效應(period effects)顯示疾病率於各日曆年(calendar year)的波動，世代效應(cohort effects)表現疾病率於各出生世代的變化。然而由於三種時間維度為完全線性相關：出生世代 + 診斷年齡 = 發生年代，這使得年齡-年代-世代模型面臨一個不可甄別問題(non-identifiability problem)¹²：有無窮多組的模型參數能夠配適資料達到相同的適合度(goodness of fit)，因而無法從中甄別出唯一的一組參數解。

歷來各種解決不可甄別問題的辦法被提出^{13,14}，至今仍未有共識^{15,16}。Su and Lee 最近提出在年齡-年代-世代模型中加入一種公平離差率限制式(constant-relative-variation constraint, CRV constraint)¹⁷，這種做法視年代與世代效應為隨機(stochastic)而設定年齡效應為命定(deterministic)，適用年齡效應為主年代與世代效應為輔的情形。Su and Lee 的方法已被應用在台灣癌症登記資料¹⁸⁻²⁰，從中發現到許多癌症的發生率趨勢具有年齡效應為主而世代效應為輔的特性。另一種年齡-漂移模型(age-drift model)¹⁰⁻¹²便適用於針對年齡與世代效應較強的疾病，年齡-漂移模型雖無法拆解年齡、年代、及世代各自的獨立效應，但可以正確估計各個年齡與世代的疾病率。研究者曾以年齡-漂移模型發現台灣女性乳癌發生率在早期出生世代具有明顯的城鄉差距，且城鄉差距會隨著出生世代逐漸縮小²¹。

流行病學家也經常繪製疾病地圖(disease mapping)²²。疾病地圖可呈現疾病率的空間分佈的特徵，流行病學家可從中提出可能的成因，可以協助擬定假說，也可以作為衛生機構制定政策的參考。疾病地圖時而表現極大高低落差的疾病率，這種空間變異(spatial variability)的成因可能只是隨機誤差(random error)，也可能是有跡可循。托布勒地理學第一定律(Tobler's first law of geography)指出鄰近地區的事物比遙遠地區更相關，這種空間相依(spatial dependency)的現象能使疾病率在地圖上表現群聚(cluster)的樣貌。為確定群聚現象是隨機抑或命定，統計學者發明空間自相關係數(spatial autocorrelation)以量測空間相依的強度並提出假說檢定及熱點偵測(hotspot detection)的方法^{23,24}。統計學者也提出數種描述空間相依樣貌的數理模型^{25,26}，使得迴歸模型中的相關係數矩陣能夠考慮空間相依的特性。另一種空間變異來自地理學第二定律(the second law of geography)^{27,28}的空間異質性(spatial heterogeneity)，或稱空間不平穩(spatial nonstationary)，意指區域內固有的地理因素(如人口、交通、地形、氣候、與水文等)可產生區域間特徵的差異。疾病地圖常以行政區域為單位呈現疾病率的地理分布，各行政區域的人口數有高有低，小地區疾

病率的估計有較大的變異數，使得行政區域圖表現出明顯的空間異質性。為降低人口分佈的影響，流行病學家會使用經驗貝氏方法(empirical Bayes methods)^{29,30} 將人口數較少的行政區域之疾病率往族群平均做較大幅度的收縮(shrinkage)，而人口數較大的行政區域收縮幅度則較小，使得各行政區域的疾病率較不易顯現極大的高低落差。

疾病地圖會因為可調整面積單元問題(modifiable areal unit problem)³¹ 而有不同的判讀。例如當疾病率群聚在縣市交界處時，縣市地圖不易偵測出熱點(hot spot)，即便以鄉鎮市區地圖找到熱點，群聚也可能被數個鄉鎮市區分割而削弱。若以各個鄉鎮市區的中心點作為疾病率的觀測點，再內插(interpolation)推估任一未知點的疾病率，便能夠繪製出疾病率的等高線地圖，期待削弱可調整面積單元問題造成的判讀誤差。以下簡單介紹幾種常見的內插方法。距離反比加權法(inverse distance weighting)³² 以觀測點與未知點之間距離的倒數作為權重，計算加權平均數當成未知點的疾病率，計算簡便但易受離群值的干擾。趨勢面分析法(trend surface analysis)³³ 使用多項式迴歸(polynomial regression)建構一個平滑的空間曲面，模擬空間變化的趨勢，不過須留意模型設定對結果判讀的影響。核密度估計(kernel density estimation)^{31,34} 與克里金法(kriging)^{35,36} 都是現今常用的內插方法，分別以核函數(kernel)與半變異函數(semivariogram)推估任一未知點的疾病率，不僅充分利用資料空間相依的特性，也改進距離反比加權法的缺點。兩種內插方法的差別在於核函數的帶寬(bandwidth)是一個主觀認定的常數而半變異函數的參數是由資料估計而得。此外，核密度估計與克里金法也都能考慮空間異質性，衍生成為地理加權迴歸(geographically weighted regression)^{31,37} 與迴歸克里金法(regression kriging)³⁸。近來研究者發現克里金法繪製的疾病地圖在人口眾多之行政區域會過度平滑(oversmoothing)而在人口稀少之行政區域則又不夠平滑(undersmoothing)。穩定克里金法(stabilized kriging)³⁹ 在推估時多考慮行政區域疾病率的變異數，變異數較小者有較大的權重，變異數較大者則有較小的權重，加權平均後得到的推估疾病率會相對穩定，改善傳統克里金法的弱點。

1.3 研究目的

前述時間與空間的方法皆有蓬勃的發展，時空整合的方法則相對薄弱。本文提出一種疾病率的時空地圖繪製方法。先以隨機效應年齡-年代-世代模型估計出各個行政區域在三種時間維度的發生率。接著對模型中的時空參數進行內插，推估各個經緯座標點的發生率，得以繪製疾病率在三種時間維度的等高線地圖。我也進行電腦模擬，評估新方法在各種時空群聚變化的情境中的表現。最後，我分析台灣癌症登記中的砷相關癌症資料，比較烏腳病流行地區與非流行地區在各個出生世代的發生率趨勢，也繪製各個出生世代的發生率等高線地圖，評估烏腳病防治的成效。文章最後也對未來的研究方向提出建議。

第二章 研究方法



2.1 基於行政區域的時空分析方法

年齡-年代-世代模型與不可甄別性問題

假定疾病個案數服從一個卜瓦松分布(Poisson distribution)，則疾病率的年齡-年代-世代模型是一個對數線性模型，表示如下：

$$\log r_{ij} = \mu + \alpha_i + \beta_j + \gamma_k ,$$

其中 r_{ij} 是第 i 年齡層第 j 年代的發生率， μ 為截距項， α_i 、 β_j 、 γ_k 分別為年齡、年代、與世代效應的參數(分別有 I 、 J 、 K 個參數)。由於不可甄別性問題，參數估計會有無窮多組解。這三個時間參數可被拆解為線性效應(linear effect)和曲率效應(curvature effect)¹⁴，不可甄別性問題可被表示如下：

$$\begin{cases} \alpha_i^{(u)} = (\alpha_L + u) \times \left(i - \frac{I+1}{2}\right) + \alpha_{C,i} \\ \beta_j^{(u)} = (\beta_L + u) \times \left(j - \frac{J+1}{2}\right) + \beta_{C,j} \\ \gamma_k^{(u)} = (\gamma_L - u) \times \left(k - \frac{K+1}{2}\right) + \gamma_{C,k} \end{cases} ,$$

α_L 、 β_L 、及 γ_L 分別為年齡、年代、世代的線性效應的估計，而 α_C 、 β_C 、及 γ_C 分別為年齡、年代、世代的曲率效應的估計， u 是一個任意數(arbitrary number)。由上式可知，任兩組參數解的線性效應會不相等，不過曲率效應會相等。

隨機效應年齡-年代-世代模型

隨機效應年齡-年代-世代模型⁴⁰使用年齡-漂移的結構以規避不可甄別性問題，雖無法拆解三種時間維度的效應，但可以正確估計疾病率，表示如下：

$$\log r_{ij}^{(s)} = \mu^{(s)} + \alpha_i^{(s)} + \gamma_k^{(s)} + \alpha_{C,i} + \beta_{C,j} + \gamma_{C,k},$$

其中

$$\begin{cases} \mu^{(s)} = \mu + \tau_0^{(s)} \\ \alpha_i^{(s)} = \left(i - \frac{I+1}{2}\right) \times (\alpha_L + \beta_L + \tau_\alpha^{(s)}) \\ \gamma_k^{(s)} = \left(k - \frac{K+1}{2}\right) \times (\gamma_L + \beta_L + \tau_\gamma^{(s)}) \end{cases},$$

其中 $\tau_0^{(s)}$ 、 $\tau_\alpha^{(s)}$ 、 $\tau_\gamma^{(s)}$ 分別為第 s 個行政區域(有 S 個參數)的截距、年齡、及世代的線性隨機效應，並且假定隨機效應服從一個多變量常態分配：

$$\boldsymbol{\tau} = \begin{bmatrix} \tau_0^{(s)} \\ \tau_\alpha^{(s)} \\ \tau_\gamma^{(s)} \end{bmatrix} \sim \text{MVN} \left(\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_0^2 & \sigma_{0\alpha} & \sigma_{0\gamma} \\ \sigma_{0\alpha} & \sigma_\alpha^2 & \sigma_{\alpha\gamma} \\ \sigma_{0\gamma} & \sigma_{\alpha\gamma} & \sigma_\gamma^2 \end{bmatrix} \right),$$

其中 σ_0^2 、 σ_α^2 、及 σ_γ^2 為各隨機效應的變異數， $\sigma_{0\alpha}$ 、 $\sigma_{0\gamma}$ 、及 $\sigma_{\alpha\gamma}$ 為隨機效應間的共變異數。上述模型可由 R 軟體的 lme4 套件或 SAS 軟體的 PROC GLIMMIX 建構。另一方面，為了考慮行政區域間空間相依的特性，後續有研究者提出在上述模型中加入一種廣義多變量條件自相關結構 (generalized multivariate conditional autoregressive)⁴¹。這種自相關結構可由 R 軟體的 brms 與 rstan 套件建立，然而參數估計需要使用貝氏統計，經測試發現會耗費大量的運算時間，因此本文使用前述未考慮空間相依的隨機效應年齡-年代-世代模型。



2.2 基於經緯度的時空分析方法

時空參數的內插

隨機效應年齡-年代-世代模型建構完成後，每個行政區域都有自己的截距 $(\mu + \tau_0^{(s)}; \text{記作}\hat{\theta}_0^{(s)})$ 、年齡效應 $(\alpha_L + \beta_L + \tau_\alpha^{(s)}; \text{記作}\hat{\theta}_\alpha^{(s)})$ 、及世代效應 $(\gamma_L + \beta_L + \tau_\gamma^{(s)}; \text{記作}\hat{\theta}_\gamma^{(s)})$ 。接著再分別對 $\hat{\theta}_0^{(s)}$ 、 $\hat{\theta}_\alpha^{(s)}$ 、及 $\hat{\theta}_\gamma^{(s)}$ 進行內插。此處可使用各種內插方法，本文採用核密度估計、克里金法、及穩定克里金法。核密度估計是一種無母數(nonparametric)的地理統計方法，推估模型如下：

$$\hat{\theta}_l^{(\mu)} = \frac{1}{Sh_l} \sum_{s=1}^S K\left(\frac{d(\mu, s)}{h_l}\right) \times \hat{\theta}_l^{(s)}, l \in \{0, \alpha, \gamma\},$$

$K(\cdot)$ 是非負的核函數，列舉幾種常用的核函數如下：

$$K\left(\frac{d(\mu, s)}{h_l}\right) = \begin{cases} \exp\left\{-\frac{1}{2}\left(\frac{d(\mu, s)}{h_l}\right)^2\right\}, \text{高斯(Gaussian)} \\ \exp\left\{-\frac{d(\mu, s)}{h_l}\right\}, \text{指數(exponential)} \\ \left(1 - \left(\frac{d(\mu, s)}{h_l}\right)^2\right)^2, \text{雙平方(bisquare)} \end{cases},$$

$d(\mu, s)$ 表示位置 μ 與位置 s 之間的距離，可為歐式距離(Euclidean distance)或其他種距離定義。例如臺灣本島的東西向受中央山脈的阻隔，此時以交通距離定義較為適當。帶寬 h_l 是一個正數，可調控進入核內的觀測點數量。帶寬又可分為固定帶寬(fixed bandwidth)與適應性帶寬(adaptive bandwidth)，其中適應性帶寬已被驗證有較穩定的表現，因此本文採用適應性帶寬。此外，為避免主觀認定的帶寬影響分析結果，本文以留一交叉驗證(leave-one-out cross-validation)來決定帶寬的數值。

基於區域化變數理論(The Theory of Regionalized Variables)，假定資料服從等向性(isotropy)的固有平穩隨機場(intrinsically stationary random field)，再由高斯-馬爾可夫定理(Gauss-Markov theorem)得出最佳線性不偏(best linear unbiased estimator)的克里金推估模型：

$$\hat{\theta}_l^{(\mu)} = \sum_{s=1}^S w_{l,s} \hat{\theta}_l^{(s)}, l \in \{0, \alpha, \gamma\},$$

其中權重 w_s 可由下式求得

$$\begin{bmatrix} w_{l,1} \\ w_{l,2} \\ \vdots \\ w_{l,S} \\ \lambda \end{bmatrix} = \begin{bmatrix} \tau_l^2 & C_{l,12} & \cdots & C_{l,1S} & 1 \\ C_{l,21} & \tau_l^2 & \cdots & C_{l,2S} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ C_{l,S1} & C_{l,S2} & \cdots & \tau_l^2 & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} C_{l,1\mu} \\ C_{l,2\mu} \\ \vdots \\ C_{l,S\mu} \\ 1 \end{bmatrix},$$



穩定克里金法在上述的矩陣公式中加入了各個行政區域參數估計的變異數：

$$\begin{bmatrix} w_{l,1} \\ w_{l,2} \\ \vdots \\ w_{l,S} \\ \lambda \end{bmatrix} = \begin{bmatrix} \tau_l^2 + v_{l,1} & C_{l,12} & \cdots & C_{l,1S} & 1 \\ C_{l,21} & \tau_l^2 + v_{l,2} & \cdots & C_{l,2S} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ C_{l,S1} & C_{l,S2} & \cdots & \tau_l^2 + v_{l,S} & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} C_{l,1\mu} \\ C_{l,2\mu} \\ \vdots \\ C_{l,S\mu} \\ 1 \end{bmatrix},$$

λ 為 Lagrange 乘數(使 $\sum_{s=1}^S w_s=1$)， v_s 是 $\hat{\theta}_l^{(s)}$ 的變異數是已知常數，未知參數有碎塊效應(τ^2)、與共變異函數 C_{ij} ($C_{ij} = C_{ji}$, $i = 1, \dots, S$; $j = i + 1, \dots, S$)、及 $C_{i\mu}$ ($i = 1, \dots, S$; μ 為欲推估的位置點)，可利用半變異圖(semivariogram)估計。我們從 S 筆資料任取兩筆，可以計算出 $\frac{1}{2} \times S \times (S - 1)$ 個成對樣本半變異數： $\frac{1}{2} \left(\hat{\theta}_l^{(k)} - \hat{\theta}_l^{(j)} \right)$ ($k \neq j$)。再將這些成對樣本半變異數針對兩筆資料間的空間距離(d)作圖即為半變異圖。接著，我們對其配適一個隨著 d 單調遞增的半變異函數 $\gamma_l(d)$ ，列舉幾種常見的半變異函數如下：

$$\gamma_l(d) = \begin{cases} \tau_l^2 + \text{sill}_l \times \left[1 - \exp \left\{ - \left(\frac{d}{h_l} \right)^2 \right\} \right], & \text{高斯(Gaussian)} \\ \tau_l^2 + \text{sill}_l \times \left[1 - \exp \left\{ - \frac{d}{h_l} \right\} \right], & \text{指數(exponential)} \\ \tau_l^2 + \text{sill}_l \times \left[\frac{3}{2} \frac{d}{h_l} - \frac{1}{2} \left(\frac{d}{h_l} \right)^3 \right], & \text{球形(spherical)} \end{cases}$$

h 類似帶寬的功能，在克里金法的理論中被稱為影響範圍(range)，當 d 極大時可達半變異函數的上限(此上限又稱門檻)，將門檻(sill_l)減去 $\gamma_l(d)$ 即可得到共變異函數(covariance function; 記作 $C(d)$)，矩陣公式中參數之估計即為： $\hat{\tau}^2 = C(0)$ ， $\hat{C}_{ij} = C(h_{ij})$ ， $\hat{C}_{i\mu} = C(h_{i\mu})$ 。

2.3 地圖繪製與影片製作

我產生固定間距的方形網狀座標系統(經度 120.00° 至 122.01° ，緯度 21.90° 至 25.30° ，共 129,524 個座標點)覆蓋整個台灣本島(詳見附錄一)。時間座標系統以 1 年為間距取得時間點。我使用前述的方法分別求得三種時間維度在任一地理座標上的發生率，舉例來說，當年齡與年代的範圍分別是 40-84 與 1995-2019，世代的範圍便是 1913-1977，則年齡與世代分別可求得 5,699,056 ($44 \times 129,524$) 及 8,289,536 ($64 \times 129,524$) 個時空座標點的發生率。最後使用 R 軟體的 ggplot2 與 av 套件，分別繪製三種時間維度的等高線地圖與製作時空動態地圖影片。

第三章 電腦模擬



3.1 模擬設定

我設定固定間距的方形網狀座標系統(共 2,075 個座標點)覆蓋嘉義與台南共 57 個鄉鎮市區(詳見附錄二)。時間座標系統以 1 年為間距取得時間點。我設定年齡為 9 個 5 齡分組，年代為 5 個 5 年分組，總共會有 2,565(9×5×57)筆模擬資料。我以對數常態分布(log-normal distribution)產生人口數，考慮平均人口數較大(10 萬人)與較小(5 萬人)的情況，及人口數變異較大(變異係數為 1)與較小(變異係數為 0.5)的情況。得到 57 鄉鎮市區的模擬人口數後，再乘以期望值便得到模擬個案數。

方法章節中的年齡-年代-世代模型生成發生率的期望值。假定模型中的截距、年齡、與世代效應服從等向性的固有平穩隨機場，年代效應則被視為命定。我設定截距與世代線性效应有群聚現象，年齡效應沒有群聚現象。我隨機指定 2,075 個座標點中的其中一點為群聚的中心點(x)，並使用高斯函數將 μ 座標點的參數真值設為：

$$\begin{cases} -7.725 + \exp\left\{-\frac{(d(\mu, x))^2}{h}\right\}, \text{截距} \\ -0.10 \times \exp\left\{-\frac{(d(\mu, x))^2}{h}\right\}, \text{世代線性效應} \end{cases},$$

我根據上述的函數生成單一發生率熱點與雙發生率熱點的情境，藉由調整 h 的大小，熱點可隨世代快速縮小、緩慢縮小、以及不隨世代改變的情況。我設定無年代效應，年齡效應與世代效應都設為凹函數(concave function)。

我比較原始發生率、經驗貝氏法、卜瓦松混合模型(即隨機效應年齡-年代-世代模型)、核密度估計、克里金法、以及穩定克里金法。原始發生率即以個案數除以人口數的發生率。經驗貝氏法是按年齡與年代配適資料，配適時又分為考慮空間相依及未考慮空間相依的情境，總共建構 90(5×9×2)個負二項模型(negative binomial model)。由於原始發生率、經驗貝氏法、及卜瓦松混合模型只能得到行政區域的發生率，因此內插時，位於同一行政區域內的經緯點被指定有相同的發生率推估值。此外克里金法又分為未調整碎塊效應與調整碎塊效應的兩種情況。

接著以對稱平均絕對百分誤差(symmetric mean absolute percentage error)評估各種方法的表現： $\frac{100\%}{2,075} \sum_{i=1}^{2,075} \frac{|\hat{r}_i - r_i|}{\hat{r}_i + r_i}$ ，其中 r_i 為第 i 個經緯座標點的發生率的期望值， $\hat{r}_i$ 為第 i 個經緯座標點的發生率的觀測值。對稱平均絕對百分誤差的數值愈小表示預測表現愈好，反之則愈差。另一方面，年齡-年代-世代資料有未觀測資料的問題。舉例來說，附錄四的交叉分類表格中，完整的出生世代應為1913-1977，但實際在各年齡層可觀測到的世代卻有侷限，如60-64歲族群只能觀測到1933-1957出生世代的資料，無法觀測的1913-1932與1958-1977世代就必須使用年齡-年代-世代模型推估。因此我也區別出觀測資料與未觀測資料，分別計算對稱平均絕對百分誤差。

3.2 模擬結果與討論

表一至表五呈現模擬結果。綜觀所有的情境，經緯度層級的方法的表現都優於行政區域層級的方法。行政區域層級的方法中以原始發生率的表現最差，經驗貝氏法的表現最好。經緯度層級的方法中以未調整碎塊效應的傳統克里金法表現最差，穩定克里金法的表現最好。此外穩定克里金法與核密度估計在各種人口分布的情境中都相對穩定，未調整碎塊效應的傳統克里金法表現則容易受人口分布干擾。另一方面，經緯度層級的方法在推估未觀測資料的表現都不如觀測資料，但仍以穩定克里金法表現稍好。

上述模擬得到穩定克里金法表現最穩定的結果，但仍需考量方法的限制。例如當資料違反各向同性的假設時，克里金法便會發生偏差。此外參數估計的標準誤有偏差時，穩定克里金法的表現也會被打折扣。另一方面，在行政區域層級的方法中，卜瓦松混合模型的表現與原始發生率相差不多，由於卜瓦松混合模型是所有的資料一次配適完成，可能會受限於模型結構的限制，才導致表現不佳。

第四章 實證資料應用



4.1 台灣的烏腳病防治

台灣西南沿海的居民曾因長期飲用砷污染的井水而導致慢性砷中毒⁴²。不僅引起皮膚病變(Skin lesions)或烏腳病(Blackfoot disease)，還提高多種癌症的發生與死亡風險。1960年起政府開始在烏腳病流行地區(Blackfoot disease-endemic areas)鋪設自來水管線以供給乾淨的飲用水，直到1970年大部分居民已停止飲用含砷的井水，流行病學研究也發現當地居民的砷相關癌症(arsenic-related cancer)的死亡風險也逐年降低⁴³⁻⁴⁵。然而台灣癌症登記資料品質直到1995年後才趨於成熟穩定，因此流行病學家不易透過癌症發生率的年代趨勢看出飲用水減砷(arsenic mitigation)的成效。本論文分析砷相關癌症發生率的出生世代趨勢，現今的台灣癌症登記可提供1913至1977出生世代的資料，在此世代的範圍內便可以看到癌症發生率趨勢於飲用水減砷前後的變化。

4.2 資料整理與統計分析

根據過去文獻及國際惡性腫瘤分類(ICD-O-3)，我考慮六種砷相關癌症：肝細胞癌(hepatocellular carcinoma)、小細胞與鱗狀細胞肺癌(small and squamous cell lung cancer)、鮑恩氏病(Bowen's disease)、基底細胞與鱗狀細胞皮膚癌(basal and squamous cell skin cancer)、膀胱泌尿上皮癌(urothelial bladder cancer)、及上泌尿道泌尿上皮癌(upper tract urothelial carcinoma)，詳細的定義見附錄三。1995至2019年確診癌症且診斷年紀介於40至84歲的病人被納入分析，這些病人被歸為9個5齡分組(40-44、45-49、...、75-79、80-84)及5個5年分組(1995-1999、2000-2004、2005-2009、2010-2014、2015-2019)，總計有45個年齡-年代交叉分類(詳見附錄四)，衍生13個出生世代分組(1913-1917、1918-1922、...、1973-1977)。這些病人也根據戶籍代碼被歸為349個在台灣本島的鄉鎮市區分組，其中有21個被陳拱北教授定義為烏腳病流行地區，有51個被衛生福利部定義為C型肝炎流行地區(詳見附錄五)。依據前述時間與空間的分組，總計有15,705個年齡-年代-地區交叉分類(9×5×349)，每一分類都對應六種砷相關癌症的個案數，再由內政部人口統計網取得每一分類的人口數。

基於世界衛生組織的 2000 年世界標準人口(詳見附錄六)，我計算 1995 至 2019 年的年齡標準化發生率並繪製長期趨勢。我也使用本論文提出的時空地圖繪製方法。首先建構隨機效應年齡-年代-世代模型，模型中允許 349 個鄉鎮市區可以有不同的截距(intercept)、年齡與世代的線性效應(linear effects)係數、以及年齡、年代、與世代的二次效應(quadratic effects)係數。為了比較烏腳病流行區與非流行區的癌症發生率出生世代趨勢，每一個出生世代都針對烏腳病流行地區與非流行地區分別計算配適發生率(fitted incidence)的中位數。接著使用穩定克里金法，針對隨機效應年齡-年代-世代模型中的各個時空參數進行內插，得到任一經緯度的配適發生率，如此能夠繪製等高線地圖並以影片播放發生率隨出生世代的地理分佈的變化。上述的資料整理與統計分析分別使用 SAS 9.4 與 R 4.2.2 進行。

4.3 結果

1995 至 2019 年之砷相關癌症年齡標準化發生率長期趨勢

圖一顯示 1995 至 2019 年期間所有癌症的年齡標準化發生率在烏腳病流行地區都明顯高於非流行地區。兩個地區肝細胞癌的年齡標準化發生率皆於 1995 年起上升，約 2000 年趨於平穩趨勢，2010 年後轉為下降。兩個地區小細胞與鱗狀細胞肺癌的年齡標準化發生率都呈現下降趨勢，不過烏腳病流行地區下降的速度較為緩慢。鮑恩氏病的年齡標準化發生率在烏腳病流行地區呈現波動趨勢，在非烏腳病流行地區卻呈現上升趨勢。兩個地區基底細胞與鱗狀細胞皮膚癌的年齡標準化發生率起初都快速上升並於 2005 年翻倍，不過之後逐漸轉為緩慢增加。兩個地區膀胱泌尿上皮癌的年齡標準化發生率起初都快速上升，2000 年後趨於平穩，更於 2007 年後轉為下降，而且烏腳病流行地區下降的速度更快。最後，兩個地區上泌尿道泌尿上皮癌的年齡標準化發生率起初都快速增加，不過 2000 年後都轉為緩慢的上升趨勢。

1913 至 1977 出生世代之砷相關癌症發生率趨勢

圖二顯示 1913-1977 出生世代 60-64 歲族群的砷相關癌症發生率趨勢在烏腳病流行地區與非流行地區都呈現下凹形狀(concave-downward shape)，其中烏腳病流行地區的癌症發生率也較高。肝細胞癌發生率起初呈上升趨勢，在 1943-1947 出生世代時發生率增加約 10 倍，之後轉為快速下降。兩個地區發生率的差距從



1923-1927 出生世代開始增加，1943-1947 出生世代到達每 10 萬人口 80 例，之後發生率差距逐漸縮小並於 1968-1972 出生世代趨於 0。小細胞與鱗狀細胞肺癌發生率起初遞增，在 1928-1932 出生世代時發生率增加約 4 倍，之後開始緩慢下降並於 1963-1967 出生世代後更迅速地下降。兩個地區的發生率差距從 1928-1932 出生世代開始增加，1943-1947 出生世代到達每 10 萬人口 9 例，之後發生率差距縮小並於 1963-1967 出生世代趨於 0。兩個地區鮑恩氏病的發生率都顯現上升趨勢，其中烏腳病流行區的發生率在所有出生世代都較高。基底細胞與鱗狀細胞皮膚癌的發生率起初遞增，在 1943-1947 出生世代時發生率增加約 4 倍，之後維持平穩趨勢。兩個地區的發生率差距從 1913-1917 出生世代開始增加，1953-1957 出生世代到達每 10 萬人口 18.4 例，之後發生率差距緩慢地遞減。膀胱泌尿上皮癌的發生率起初遞增，在 1928-1932 出生世代時發生率增加約 2 倍，之後維持平穩，再於 1953-1957 出生世代後轉為下降。兩個地區的發生率差距在 1913-1917 出生世代時為每 10 萬人口 4.6 例，1953-1957 出生世代增為每 10 萬人口 15.9 例，之後發生率差距快速縮小並於 1963-1967 出生世代趨於 0。最後，上泌尿道泌尿上皮癌的發生率起初快速增加，在 1933-1937 出生世代時發生率增加約 2 倍，之後便維持平穩趨勢。兩個地區的發生率差距在 1913-1917 出生世代時為每 10 萬人口 3.3 例，1943-1947 出生世代增為每 10 萬人口 14.3 例，之後發生率差距開始緩慢縮小。

附錄七呈現 1913-1977 出生世代 45-49 歲及 75-79 歲族群的砷相關癌症發生率趨勢，整體上結果與 60-64 歲族群相似，些微的差異在於 45-49 族群中兩個地區鮑恩氏病發生率的差距逐漸消失。此外我也進行性別分層分析並得出與圖二相似的結果(附錄八)，些微的差異在於男女基底細胞與鱗狀細胞皮膚癌的發生率差距保持存在，男性鮑恩氏病與上泌尿道泌尿上皮癌的發生率差距逐漸減少，女性小細胞與鱗狀細胞肺癌以及上泌尿道泌尿上皮癌的發生率差距都保持存在。

我也繪製 1913-1977 出生世代 60-64 歲族群在烏腳病或 C 型肝炎流行地區的砷相關癌症發生率趨勢(詳見附錄九)。肝細胞癌發生率在 C 型肝炎流行地區更高於烏腳病流行地區，然而兩個地區的發生率差距在 1958-1962 出生世代後漸趨於 0。我也按部位(頭頸部: C440-C444; 其他部位: C445-C449)與年齡繪製 1913-1977 出生世代的基底細胞與鱗狀細胞皮膚癌發生率趨勢(詳見附錄十)，烏腳病流行地區與非流行地區的發生率差距在頭頸部會逐漸縮小但在其他部位則保持存在。

1913 至 1977 出生世代之砷相關癌症發生率地理分布

圖三、圖四、與圖五呈現 1913-1977 出生世代 60-64 歲族群的砷相關癌症發生率地理分布，每一張圖都附有一個時空地圖的影片網址。肝細胞癌發生率雖於烏腳病流行地區較高，但地理分佈更近似 C 型肝炎流行地區，無論烏腳病或 C 型肝炎流行地區，肝細胞癌發生率都於 1960 出生世代開始下降，並於 1975 出生世代後接近於非流行地區的發生率。對於肝細胞癌以外的其他癌症，都在北門、學甲、布袋、義竹等地發現了高發生率的時空群聚，而這些鄉鎮正是烏腳病流行的重災區。對於小細胞與鱗狀細胞肺癌以及膀胱泌尿上皮癌，群聚於 1960 出生世代開始下降並於 1975 出生世代後消失，不過對於鮑恩氏病、基底細胞與鱗狀細胞肺癌、及上泌尿道泌尿上皮癌，群聚於 1960 出生世代開始緩慢消散，到了 1977 出生世代仍然存在。除此之外，由附錄十一的 Z 值地理分布可知，砷相關癌症發生率在西南沿海的烏腳病流行地區有顯著的群聚。由附錄十二的半變異圖可知，群聚的影響範圍由大至小依序為肝細胞癌、基底細胞與鱗狀細胞皮膚癌、泌尿上皮癌、小細胞與鱗狀細胞肺癌、鮑恩氏病。

時空群聚也被發現在非烏腳病流行地區或非 C 型肝炎流行地區，包含位於北部且逐漸消退的肝細胞癌群聚；位於北部、中部、及東部且迅速消失的小細胞與鱗狀細胞肺癌群聚；位於東部且持續存在的基底細胞與鱗狀細胞皮膚癌群聚；位於北部、中部、及東部的泌尿上皮癌群聚，其中膀胱癌群聚隨出生世代逐漸消散，但上泌尿道癌群聚則保持存在。另一方面，附錄十三呈現 1913-1977 出生世代 60-64 歲族群的基底細胞與鱗狀細胞皮膚癌發生率地理分布，頭頸部與其他部位的發生率有相似的地理分佈，些微的不同在於頭頸部發生率群聚在花蓮較宜蘭明顯，其他部位發生率群聚則只有位於宜蘭。

45-49 歲及 75-79 族群砷相關癌症的發生率時空分布呈現於附錄十四至十九，結果與 60-64 歲族群相近。此外性別分層後的發生率時空地圖呈現於附錄二十至二十五，男性與女性的地理分布也相似，些微不同在於男性小細胞與鱗狀細胞肺癌有明顯的群聚位於宜蘭，女性則是宜花東三地都有群聚；女性鮑恩氏病與上泌尿道泌尿上皮癌在烏腳病流行地區的群聚都逐漸消散，男性的群聚卻維持存在。另一方面，我以肺腺癌做為小細胞與鱗狀細胞肺癌的陰性對照，並以泌尿上皮癌作為膀胱泌尿上皮癌與上泌尿道泌尿上皮癌的陽性對照。結果呈現於附錄二十六與二十七，肺腺癌在烏腳病流行區與非流行區的發生率差距為 0，泌尿上皮癌的

結果也與膀胱泌尿上皮癌非常相似。



砷相關癌症發生率之世代效應分析

最後我使用加入公平離差率限制式的年齡-年代-世代模型比較各個癌症在烏腳病流行地區與非流行地區的世代效應。結果呈現於附錄二十八至三十，烏腳病地區肝細胞癌與上泌尿道泌尿上皮癌的發生率都於 1963-1967 出生世代後下降，烏腳病地區小細胞與鱗狀細胞肺癌及膀胱泌尿上皮癌的發生率都於 1943-1947 出生世代後下降，鮑恩氏病發生率也於 1938-1942 出生世代後下降。

4.4 討論

本論文探索台灣烏腳病流行地區的飲用水減砷後對於砷相關癌症發生率的時空變化。1995 年至 2019 年所有癌症的年齡標準化發生率在台灣烏腳病流行地區都保持較高，看不出飲用水減砷後的影響。然而從出生世代的觀點來看，所有癌症的烏腳病流行地區與非流行地區的發生率差距都於 1958-1962 出生世代後逐漸縮小，此時間點恰好吻合自來水管線開始鋪設的年代。由時空地圖也發現到台灣西南沿海有明顯的高發生率時空群聚，群聚中心恰好位於烏腳病流行區的重鎮，且群聚也於 1960 出生世代後開始消散。

本研究的結果說明飲用水減砷有助於減少相關癌症的風險。過去有研究者使用 1980 至 2000 年的癌症登記數據，也發現飲用水減砷後對於癌症發生率有正面的效果^{46,47}。然而，這些研究僅探討皮膚癌、膀胱癌、與肺癌，且只分析 1953 以前出生的病人。反觀本研究考慮多種砷相關癌症，並根據文獻選出特定的組織型態，還能完整展示每一個出生世代發生率的地理分布。另一方面，相同的結果也被發現在美國、比利時、智利、及匈牙利⁴⁸⁻⁵²。值得注意的是智利的研究指出至少需要 40 年才能根除砷暴露對於癌症的危害^{49,50}，本研究觀察到肝細胞癌、小細胞與鱗狀細胞肺癌、以及膀胱泌尿上皮癌在兩個地區的發生率差距也都需要 40 年才能消除，鮑恩氏病、基底細胞與鱗狀細胞皮膚癌、與上泌尿道泌尿上皮癌需要至少 50 年才可能消除。砷暴露容易引起皮膚病變或腎功能障礙，若不立即治療很可能會導致癌症，況且當時台灣的醫療資源不充足，我推測這些因素造成皮膚癌與腎癌發生率下降較緩慢。

1990 年台灣東北沿海也發現砷污染的飲用水⁴²。政府記取過去的經驗利用 2

年就在當地安裝好自來水供給系統。儘管及時防治，後續的流行病學調查發現當地的癌症發生率明顯提高⁵³⁻⁵⁶。本研究的時空地圖也發現東北沿海有小細胞與鱗狀細胞肺癌、鮑恩氏病、基底細胞與鱗狀細胞皮膚癌、膀胱泌尿上皮癌、及上泌尿道泌尿上皮癌的發生率群聚。

紫外線是鮑恩氏病以及基底細胞與鱗狀細胞皮膚癌很重要的危險因子。這兩個癌症的發生率都在 1943-1947 出生世代轉變為緩慢上升趨勢，1950 年代台灣開始工業化，從事農業的比例大幅減少，這可能會造成皮膚癌發生率趨勢的轉變⁵⁷。此外，頭頸部相較其他部位更容易暴露於紫外線，本研究發現兩個地區的頭頸部皮膚癌發生率差距逐漸縮小，而其他部位皮膚癌的發生率差距較沒有減少。另外，其他部位皮膚癌的發生率群聚在東北沿海相較於頭頸部來得更明顯，而東北沿海恰好也曾是砷污染地區。上述這些結果說明砷暴露對於皮膚癌的影響主要在於頭頸部以外的其他部位。

慢性病毒性肝炎是肝細胞癌最重要的危險因子。台灣也流行慢性肝炎，不過施行疫苗介入與治療計畫後，台灣肝細胞癌發生率與死亡率都顯著下降⁵⁸⁻⁶⁰。肝細胞癌的時空地圖顯示發生率的地理分布與 C 型肝炎流行地區非常接近，有趣的是，深入分析居住在烏腳病流行地區但不在 C 型肝炎流行地區的居民，發現這些居民的發生率仍比其他非流行地區的居民來得高。所以肝細胞癌發生率的下降可能同時與飲用水減砷及肝炎防治有關。

肝細胞癌、小細胞與鱗狀細胞肺癌、膀胱泌尿上皮癌、及上泌尿道泌尿上皮癌的發生率都顯現由遞增轉為遞減的出生世代趨勢，這種下凹狀趨勢也可能與菸害防制及馬兜鈴酸禁用有關^{7,9,60-62}。此外上述癌症也在北部及中部發現發生率的時空群聚，這可能也對應到當地有較高的吸菸率及馬兜鈴酸食用率，有待進一步研究。

本研究有一些限制。首先，這是一項生態型研究，研究未考量其他已知的致癌因素(如吸菸與馬兜鈴酸)，也未能納入個體層級的資訊以調整干擾因子，例如家族史、累積砷暴露劑量、及遺傳因素等。再者，砷不一定只由飲用水進入人體，非流行地區的居民也可能吸入或食用到含砷的物質。另一方面，在較早與較晚的出生世代僅能觀測到較少的癌症個案，這可能導致不穩定的參數估計。最後，本研究僅能取得病人確診癌症時的居住地，未能考慮遷移戶口的影響。

第五章 結論與展望



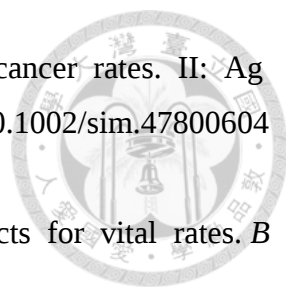
本文提出一種疾病率時空地圖的繪製方法。新方法可繪製出疾病率分別在年齡、年代、與世代的等高線地圖。電腦模擬也展示新方法的優勢與能力。在實證資料應用中，我發現烏腳病流行地區與非流行地區的癌症發生率差距在 1960 出生世代後逐漸消失，這個轉折點恰好是烏腳病防治的起始時間。新方法不僅展示烏腳病防治的成果，也看到肝病防治在肝細胞癌發生率的成效。

綜合本文的研究結果，我提出幾點未來研究方向。在方法的部分，時間也可拆解為存活年、年代、世代共三種維度，建議可發展一種隨機效應存活年-年代-世代模型(random-effects survivorship-period-cohort model)，繪製出存活率在三種時間維度的等高線地圖，期待成為健康不平等研究的利器。另一方面，由電腦模擬的結果發現到核密度估計的表現略輸穩定克里金法，然而核密度估計還未考慮疾病率的變異數，因此建議發展穩定核密度估計(stabilized kernel density estimation)，預期能有更好的表現。在實證資料應用的部分，研究者曾以穩定時空克里金法(stabilized spatiotemporal kriging)⁶³發現台灣口腔癌發生率的時空群聚與檳榔種植分布相似，然而該研究僅分析年代趨勢的變化，我建議也以本文的方法分析出生世代趨勢的變化。另一方面，空氣汙染是當今台灣重要的健康議題，我將利用本文的方法探討台灣肺腺癌的時空群聚分布，期待結果可以作為衛生機構制定政策的參考。

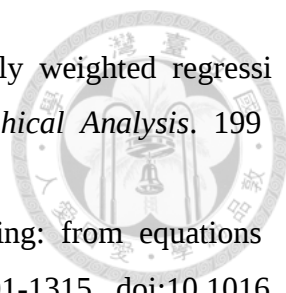
參考文獻

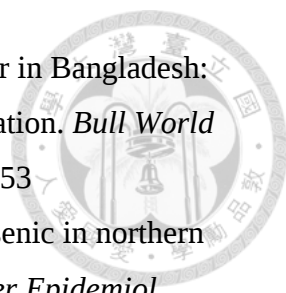


1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin*. 2021;71(3):209-249. doi:10.3322/caac.21660
2. Thun MJ, DeLancey JO, Center MM, Jemal A, Ward EM. The global burden of cancer: priorities for prevention. *Carcinogenesis*. 2010;31(1):100-110. doi:10.1093/carcin/bgp263
3. Jemal A, Center MM, DeSantis C, Ward EM. Global patterns of cancer incidence and mortality rates and trends. *Cancer Epidemiol Biomarkers Prev*. 2010;19(8):1893-1907. doi:10.1158/1055-9965.EPI-10-0437
4. Chiang CJ, You SL, Chen CJ, Yang YW, Lo WC, Lai MS. Quality assessment and improvement of nationwide cancer registration system in Taiwan: a review. *Jpn J Clin Oncol*. 2015;45(3):291-296. doi:10.1093/jjco/hyu211
5. Chiang CJ, Wang YW, Lee WC. Taiwan's Nationwide Cancer Registry System of 40 years: Past, present, and future. *J Formos Med Assoc*. 2019;118(5):856-858. doi:10.1016/j.jfma.2019.01.012
6. Kao CW, Chiang CJ, Lin LJ, et al. Accuracy of long-form data in the Taiwan cancer registry. *J Formos Med Assoc*. 2021;120(11):2037-2041. doi:10.1016/j.jfma.2021.04.022
7. Jhuang JR, Chiang CJ, Su SY, Yang YW, Lee WC. Reduction in the Incidence of Urological Cancers after the Ban on Chinese Herbal Products Containing Aristolochic Acid: An Interrupted Time-Series Analysis. *Sci Rep*. 2019;9(1):19860. Published 2019 Dec 27. doi:10.1038/s41598-019-56394-y
8. Jhuang JR, Su SY, Chiang CJ, et al. Forecast of peak attainment and imminent decline after 2017 of oral cancer incidence in men in Taiwan. *Sci Rep*. 2022;12(1):5726. Published 2022 Apr 6. doi:10.1038/s41598-022-09736-2
9. Jhuang JR, Chiu PC, Hsieh TC, Chen CH, Pu YS, Lee WC. Latency period of aristolochic acid-induced upper urinary tract urothelial carcinoma. *Front Public Health*. 2023;11:1072864. Published 2023 Mar 9. doi:10.3389/fpubh.2023.1072864
10. Clayton D, Schifflers E. Models for temporal variation in cancer rates. I: Age-period and age-cohort models. *Stat Med*. 1987;6(4):449-467. doi:10.1002/sim.4780060405

- 
11. Clayton D, Schifflers E. Models for temporal variation in cancer rates. II: Age-period-cohort models. *Stat Med*. 1987;6(4):469-481. doi:10.1002/sim.4780060406
 12. Holford TR. The estimation of age, period and cohort effects for vital rates. *Biometrics*. 1983;39(2):311-324.
 13. Yang Y, Land KC. *Age-Period-Cohort Analysis: New Models, Methods, and Empirical Applications*. 1st ed. Chapman and Hall/CRC; 2013. doi:10.1201/b13902
 14. Bell A. *Age, Period and Cohort Effects: Statistical Analysis and the Identification Problem*. 1st ed. Routledge; 2020. doi:10.4324/9780429056819
 15. Fosse E, Winship C. Analyzing Age-Period-Cohort Data: A Review and Critique. *Annu Rev Sociol*. 2019;45:467-492. doi:10.1146/annurev-soc-073018-022616
 16. Bell A. Age period cohort analysis: a review of what we should and shouldn't do. *Ann Hum Biol*. 2020;47(2):208-217. doi:10.1080/03014460.2019.1707872
 17. Su SY, Lee WC. Age-period-cohort analysis with a constant-relative-variation constraint for an apportionment of period and cohort slopes. *PLoS One*. 2019;14(12):e0226678. Published 2019 Dec 19. doi:10.1371/journal.pone.0226678
 18. Su SY, Chen WT, Chiang CJ, Yang YW, Lee WC. Oral cancer incidence rates from 1997 to 2016 among men in Taiwan: Association between birth cohort trends and betel nut consumption. *Oral Oncol*. 2020;107:104798. doi:10.1016/j.oraloncology.2020.104798
 19. Chiang CJ, Jhuang JR, Yang YW, et al. Association of Nationwide Hepatitis B Vaccination and Antiviral Therapy Programs With End-Stage Liver Disease Burden in Taiwan. *JAMA Netw Open*. 2022;5(7):e2222367. Published 2022 Jul 1. doi:10.1001/jamanetworkopen.2022.22367
 20. Liu HI, Chiang CJ, Su SY, et al. Incidence trends and spatial distributions of lung adenocarcinoma and squamous cell carcinoma in Taiwan. *Sci Rep*. 2023;13(1):1655. Published 2023 Jan 30. doi:10.1038/s41598-023-28253-4
 21. Lee PJ, Jhuang JR, Chen YC, et al. Urban-Rural Disparity in Birth Cohort Effects on Breast Cancer Incidence. *J Urban Health*. 2023;100(2):341-354. doi:10.1007/s11524-023-00718-x
 22. Lawson AB, Biggeri A, Bohning D, Lesaffre E, Viel JF, Bertollini R. *Disease Mapping and Risk Assessment for Public Health*. Wiley; 1999.
 23. MORAN PA. Notes on continuous stochastic phenomena. *Biometrika*. 1950;37

- (1-2):17-23.
24. Anselin L. Local Indicators of Spatial Association—LISA. *Geographical Analysis*. 1995;27:93-115. doi:10.1111/j.1538-4632.1995.tb00338.x
 25. De Oliveira V, Song JJ. Bayesian analysis of simultaneous autoregressive models. *Sankhyā: The Indian Journal of Statistics, Series B (2008-)*. 2008;70(2):323-350.
 26. De Oliveira V. Bayesian analysis of conditional autoregressive models. *Ann Inst Stat Math*. 2012;64:107-133. doi:10.1007/s10463-010-0298-1
 27. Goodchild MF. The Validity and Usefulness of Laws in Geographic Information Science and Geography. *Annals of the Association of American Geographers*. 2004;94(2):300-303. doi:10.1111/j.1467-8306.2004.09402008.x
 28. Zhu AX, Guonian L, Liu J, Qin CZ, Zhou C. Spatial prediction based on Third Law of Geography. *Annals of GIS*. 2018;24(4):225-240. doi:10.1080/19475683.2018.1534890
 29. Clayton D, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*. 1987;43(3):671-681.
 30. Marshall RJ. Mapping disease and mortality rates using empirical Bayes estimators. *J R Stat Soc Ser C Appl Stat*. 1991;40(2):283-294.
 31. Wen TH. *Spatial Analysis: Methods and Applications*. Yeh Yeh Book Gallery, Taipei; 2015. [In Chinese]
 32. Lu GY, Wong DW. An adaptive inverse-distance weighting spatial interpolation technique. *Computers & Geosciences*. 2008;34(9):1044-1055. doi:10.1016/j.cageo.2007.07.010
 33. Liaw YP, Lee WC, Chen CJ. Trend surface analysis for cancer mapping: application to breast cancer mortality in Taiwan. *Taiwan Journal of Public Health*. 1998;17(6):474-484.
 34. Parzen E. On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics*. 1962;33(3):1065-1076.
 35. Carrat F, Valleron AJ. Epidemiologic mapping using the "kriging" method: application to an influenza-like illness epidemic in France. *Am J Epidemiol*. 1992;135(11):1293-1300. doi:10.1093/oxfordjournals.aje.a116236
 36. Berke O. Exploratory disease mapping: kriging the spatial risk function from regional count data. *Int J Health Geogr*. 2004;3(1):18. Published 2004 Aug 26. doi:10.1186/1476-072X-3-18

- 
37. Brunson C, Fotheringham AS, Charlton ME. Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical Analysis*. 1996;28:281-98.
 38. Hengl T, Heuvelink G, Rossiter DG. About regression-kriging: from equations to case studies. *Computers & Geosciences*. 2007;33(10):1301-1315. doi:10.1016/j.cageo.2007.05.001
 39. Hsu CC, Tsai DR, Su SY, et al. A Stabilized Kriging Method for Mapping Disease Rates. *J Epidemiol*. 2023;33(4):201-208. doi:10.2188/jea.JE20210276
 40. Chernyavskiy P, Little MP, Rosenberg PS. A unified approach for assessing heterogeneity in age-period-cohort model parameters using random effects. *Stat Meth Med Res*. 2019;28:20-34. doi: 10.1177/0962280217713033
 41. Chernyavskiy P, Little MP, Rosenberg PS. Spatially varying age-period-cohort analysis with application to US mortality, 2002-2016. *Biostatistics*. 2020;21:845-859. doi: 10.1093/biostatistics/kxz009
 42. Jean JS, Bundschuh J, Chen CJ, Liu CW, Guo HR, Lin TF, Chen YH. *The Taiwan crisis: a showcase of the global arsenic problem*. CRC Press, Taylor & Francis Group, Leiden; 2010.
 43. Yang CY, Chiu HF, Wu TN, Chuang HY, Ho SC. Reduction in kidney cancer mortality following installation of a tap water supply system in an arsenic-endemic area of Taiwan. *Arch Environ Health*. 2004;59(9):484-488. doi:10.1080/00039890409603430
 44. Chiu HF, Ho SC, Yang CY. Lung cancer mortality reduction after installation of tap-water supply system in an arseniasis-endemic area in Southwestern Taiwan. *Lung Cancer*. 2004;46(3):265-270. doi:10.1016/j.lungcan.2004.05.012
 45. Chiu HF, Ho SC, Wang LY, Wu TN, Yang CY. Does arsenic exposure increase the risk for liver cancer?. *J Toxicol Environ Health A*. 2004;67(19):1491-1500. doi:10.1080/15287390490486806
 46. Su CC, Lu JL, Tsai KY, Lian IeB. Reduction in arsenic intake from water has different impacts on lung cancer and bladder cancer in an arseniasis endemic area in Taiwan. *Cancer Causes Control*. 2011;22(1):101-108. doi:10.1007/s10552-010-9679-2
 47. Cheng PS, Weng SF, Chiang CH, Lai FJ. Relationship between arsenic-containing drinking water and skin cancers in the arseniasis endemic areas in Taiwan. *J Dermatol*. 2016;43(2):181-186. doi:10.1111/1346-8138.13058

- 
48. Flanagan SV, Johnston RB, Zheng Y. Arsenic in tube well water in Bangladesh: health and economic impacts and implications for arsenic mitigation. *Bull World Health Organ.* 2012;90(11):839-846. doi:10.2471/BLT.11.101253
 49. Steinmaus CM, Ferreccio C, Romo JA, et al. Drinking water arsenic in northern Chile: high cancer risks 40 years after exposure cessation. *Cancer Epidemiol Biomarkers Prev.* 2013;22(4):623-630. doi:10.1158/1055-9965.EPI-12-1190
 50. Smith AH, Marshall G, Roh T, Ferreccio C, Liaw J, Steinmaus C. Lung, Bladder, and Kidney Cancer Mortality 40 Years After Arsenic Exposure Reduction. *J Natl Cancer Inst.* 2018;110(3):241-249. doi:10.1093/jnci/djx201
 51. Yang Q, Flanagan SV, Chillrud S, et al. Reduction in drinking water arsenic exposure and health risk through arsenic treatment among private well households in Maine and New Jersey, USA. *Sci Total Environ.* 2020;738:139683. doi:10.1016/j.scitotenv.2020.139683
 52. Pál L, Jenei T, McKee M, et al. Health and economic gain attributable to the introduction of the World Health Organization's drinking water standard on arsenic level in Hungary: A nationwide retrospective study on cancer occurrence and ischemic heart disease mortality. *Sci Total Environ.* 2022;851(Pt 2):158305. doi:10.1016/j.scitotenv.2022.158305
 53. Chiou HY, Chiou ST, Hsu YH, et al. Incidence of transitional cell carcinoma and arsenic in drinking water: a follow-up study of 8,102 residents in an arseniasis-endemic area in northeastern Taiwan. *Am J Epidemiol.* 2001;153(5):411-418. doi:10.1093/aje/153.5.411
 54. Tsai TL, Kuo CC, Hsu LI, et al. Association between arsenic exposure, DNA damage, and urological cancers incidence: A long-term follow-up study of residents in an arseniasis endemic area of northeastern Taiwan. *Chemosphere.* 2021;266:129094. doi:10.1016/j.chemosphere.2020.129094
 55. Chen CL, Chiou HY, Hsu LI, Hsueh YM, Wu MM, Chen CJ. Ingested arsenic, characteristics of well water consumption and risk of different histological types of lung cancer in northeastern Taiwan. *Environ Res.* 2010;110(5):455-462. doi:10.1016/j.envres.2009.08.010
 56. Chen CL, Chiou HY, Hsu LI, et al. Arsenic in drinking water and risk of urinary tract cancer: a follow-up study from northeastern Taiwan. *Cancer Epidemiol Biomarkers Prev.* 2010;19(1):101-110. doi:10.1158/1055-9965.EPI-09-0333
 57. Mao YK, Chi S. *Agricultural and industrial development in Taiwan*. Published for

the International Food Policy Research Institute by Johns Hopkins University Press;1995: Chapter 2. Accessed April 1, 2023.

<https://www.ifpri.org/cdmref/p15738coll2/id/129337/filename/129548.pdf>

58. Chiang CJ, Jhuang JR, Yang YW, et al. Association of Nationwide Hepatitis B Vaccination and Antiviral Therapy Programs With End-Stage Liver Disease Burden in Taiwan. *JAMA Netw Open*. 2022;5(7):e2222367. Published 2022 Jul 1. doi:10.1001/jamanetworkopen.2022.22367
59. Kee KM, Chen CH, Hu JT, et al. Secular Trends of Clinical Characteristics and Survival of Hepatocellular Carcinoma in Taiwan from 2011 to 2019. *Viruses*. 2022;15(1):126. Published 2022 Dec 31. doi:10.3390/v15010126
60. Su SY, Chiang CJ, Yang YW, Lee WC. Secular trends in liver cancer incidence from 1997 to 2014 in Taiwan and projection to 2035: An age-period-cohort analysis. *J Formos Med Assoc*. 2019;118(1 Pt 3):444-449. doi:10.1016/j.jfma.2018.07.001
61. Liao CI, Fang HC, Lee PT, et al. Trends in the incidence of urothelial carcinoma in Taiwan after the ban on aristolochic acid-containing Chinese herbal preparations, 2001-2018: a national population-based cohort study [published online ahead of print, 2023 Apr 15]. *J Cancer Res Clin Oncol*. 2023;10.1007/s00432-023-04771-6. doi:10.1007/s00432-023-04771-6
62. Liu HI, Chiang CJ, Su SY, et al. Incidence trends and spatial distributions of lung adenocarcinoma and squamous cell carcinoma in Taiwan. *Sci Rep*. 2023;13(1):1655. Published 2023 Jan 30. doi:10.1038/s41598-023-28253-4
63. Tsai DR, Jhuang JR, Su SY, Chiang CJ, Yang YW, Lee WC. A stabilized spatiotemporal kriging method for disease mapping and application to male oral cancer and female breast cancer in Taiwan. *BMC Med Res Methodol*. 2022;22(1):270. Published 2022 Oct 13. doi:10.1186/s12874-022-01749-9

表一、單一發生率熱點隨出生世代縮小的情境下，各種方法的對稱平均絕對百分誤差(%)

鄉鎮市區人口數		資料型態	行政區域層級的方法				經緯度層級的方法			
平均數	變異係數		原始發生率	經驗貝氏法 ¹	經驗貝氏法 ²	卜瓦松混合模型	核密度估計	傳統克里金法 ³	傳統克里金法 ⁴	穩定克里金法
50,000	0.5	觀測資料	19.73	18.30	18.31	19.66	11.89	18.53	14.05	10.88
		未觀測資料	-	-	-	-	19.36	25.41	20.73	17.98
50,000	1.0	觀測資料	20.03	19.44	19.42	20.02	11.95	17.65	14.14	11.26
		未觀測資料	-	-	-	-	19.87	24.72	21.42	18.83
10,000	0.5	觀測資料	19.61	18.65	18.67	19.56	11.81	17.47	13.25	11.31
		未觀測資料	-	-	-	-	19.80	24.78	20.70	18.94
10,000	1.0	觀測資料	20.26	19.56	19.53	20.24	11.75	19.36	12.87	11.13
		未觀測資料	-	-	-	-	19.42	26.50	19.98	18.46

¹ 負二項模型；未考慮空間自相關結構：單一地區的疾病率由所有地區共同估計之。

² 負二項模型；考慮空間自相關結構：單一地區的疾病率由本身及其鄰近地區共同估計之。

³ 未調整碎塊效應(nugget effect)。

⁴ 有調整碎塊效應。

表二、單一發生率熱點不隨出生世代改變的情境下，各種方法的對稱平均絕對百分誤差(%)

鄉鎮市區人口數		資料型態	行政區域層級的方法				經緯度層級的方法			
平均數	變異係數		原始發生率	經驗貝氏法 ¹	經驗貝氏法 ²	卜瓦松混合模型	核密度估計	傳統克里金法 ³	傳統克里金法 ⁴	穩定克里金法
50,000	0.5	觀測資料	17.67	16.02	16.11	17.47	11.50	17.86	13.69	10.65
		未觀測資料	-	-	-	-	20.20	25.98	21.60	19.00
50,000	1.0	觀測資料	17.45	16.75	16.74	17.42	11.41	15.82	12.79	10.71
		未觀測資料	-	-	-	-	20.12	24.18	20.87	19.03
10,000	0.5	觀測資料	16.97	15.79	15.88	16.90	11.36	17.02	11.68	10.70
		未觀測資料	-	-	-	-	19.83	24.98	19.76	18.78
10,000	1.0	觀測資料	17.59	15.28	15.52	17.20	11.27	17.01	13.11	10.69
		未觀測資料	-	-	-	-	19.95	24.90	21.06	18.95

¹ 負二項模型；未考慮空間自相關結構：單一地區的疾病率由所有地區共同估計之。

² 負二項模型；考慮空間自相關結構：單一地區的疾病率由本身及其鄰近地區共同估計之。

³ 未調整碎塊效應(nugget effect)。

⁴ 有調整碎塊效應。

表三、雙發生率熱點都隨出生世代縮小的情境下，各種方法的對稱平均絕對百分誤差(%)

鄉鎮市區人口數		資料型態	行政區域層級的方法				經緯度層級的方法			
平均數	變異係數		原始發生率	經驗貝氏法 ¹	經驗貝氏法 ²	卜瓦松混合模型	核密度估計	傳統克里金法 ³	傳統克里金法 ⁴	穩定克里金法
50,000	0.5	觀測資料	17.53	15.98	15.98	17.41	12.08	12.69	11.69	11.19
		未觀測資料	-	-	-	-	20.12	20.30	19.55	18.89
50,000	1.0	觀測資料	17.78	17.11	17.10	17.75	11.77	17.00	11.94	11.09
		未觀測資料	-	-	-	-	19.75	24.26	19.54	18.71
10,000	0.5	觀測資料	17.55	15.46	15.70	17.30	11.95	18.06	12.12	11.21
		未觀測資料	-	-	-	-	20.08	25.48	19.97	19.01
10,000	1.0	觀測資料	17.42	16.33	16.41	17.34	11.98	19.77	12.04	11.35
		未觀測資料	-	-	-	-	20.25	27.33	20.03	19.33

¹ 負二項模型；未考慮空間自相關結構：單一地區的疾病率由所有地區共同估計之。

² 負二項模型；考慮空間自相關結構：單一地區的疾病率由本身及其鄰近地區共同估計之。

³ 未調整碎塊效應(nugget effect)。

⁴ 有調整碎塊效應。

表四、雙發生率熱點分別隨出生世代快速與緩慢縮小的情境下，各種方法的對稱平均絕對百分誤差(%)

鄉鎮市區人口數		資料型態	行政區域層級的方法				經緯度層級的方法			
平均數	變異係數		原始發生率	經驗貝氏法 ¹	經驗貝氏法 ²	卜瓦松混合模型	核密度估計	傳統克里金法 ³	傳統克里金法 ⁴	穩定克里金法
50,000	0.5	觀測資料	18.95	16.71	17.03	18.72	11.81	12.11	12.03	11.29
		未觀測資料	-	-	-	-	19.84	20.26	19.93	19.10
50,000	1.0	觀測資料	19.14	17.85	18.06	19.05	12.05	12.09	11.83	11.23
		未觀測資料	-	-	-	-	20.15	19.99	19.84	19.11
10,000	0.5	觀測資料	19.57	18.77	18.82	19.54	12.07	12.24	11.81	11.26
		未觀測資料	-	-	-	-	20.22	20.14	19.81	19.09
10,000	1.0	觀測資料	20.29	18.61	18.74	20.14	12.14	13.31	11.80	11.25
		未觀測資料	-	-	-	-	20.28	20.88	19.73	19.12

¹ 負二項模型；未考慮空間自相關結構：單一地區的疾病率由所有地區共同估計之。

² 負二項模型；考慮空間自相關結構：單一地區的疾病率由本身及其鄰近地區共同估計之。

³ 未調整碎塊效應(nugget effect)。

⁴ 有調整碎塊效應。

表五、無發生率熱點的情境下，各種方法的對稱平均絕對百分誤差(%)

鄉鎮市區人口數		資料型態	行政區域層級的方法				經緯度層級的方法			
平均數	變異係數		原始發生率	經驗貝氏法 ¹	經驗貝氏法 ²	卜瓦松混合模型	核密度估計	傳統克里金法 ³	傳統克里金法 ⁴	穩定克里金法
50,000	0.5	觀測資料	11.26	10.45	10.49	11.20	6.89	7.78	7.00	6.81
		未觀測資料	-	-	-	-	14.05	14.66	14.05	14.02
50,000	1.0	觀測資料	11.93	10.21	10.46	11.68	7.01	8.05	7.29	6.99
		未觀測資料	-	-	-	-	14.46	15.34	14.83	14.66
10,000	0.5	觀測資料	12.06	9.87	10.36	11.69	6.91	7.83	7.14	6.90
		未觀測資料	-	-	-	-	14.35	15.05	14.71	14.42
10,000	1.0	觀測資料	11.91	10.61	10.77	11.79	7.08	7.53	7.25	7.06
		未觀測資料	-	-	-	-	14.79	14.94	14.95	14.54

¹負二項模型；未考慮空間自相關結構：單一地區的疾病率由所有地區共同估計之。

²負二項模型；考慮空間自相關結構：單一地區的疾病率由本身及其鄰近地區共同估計之。

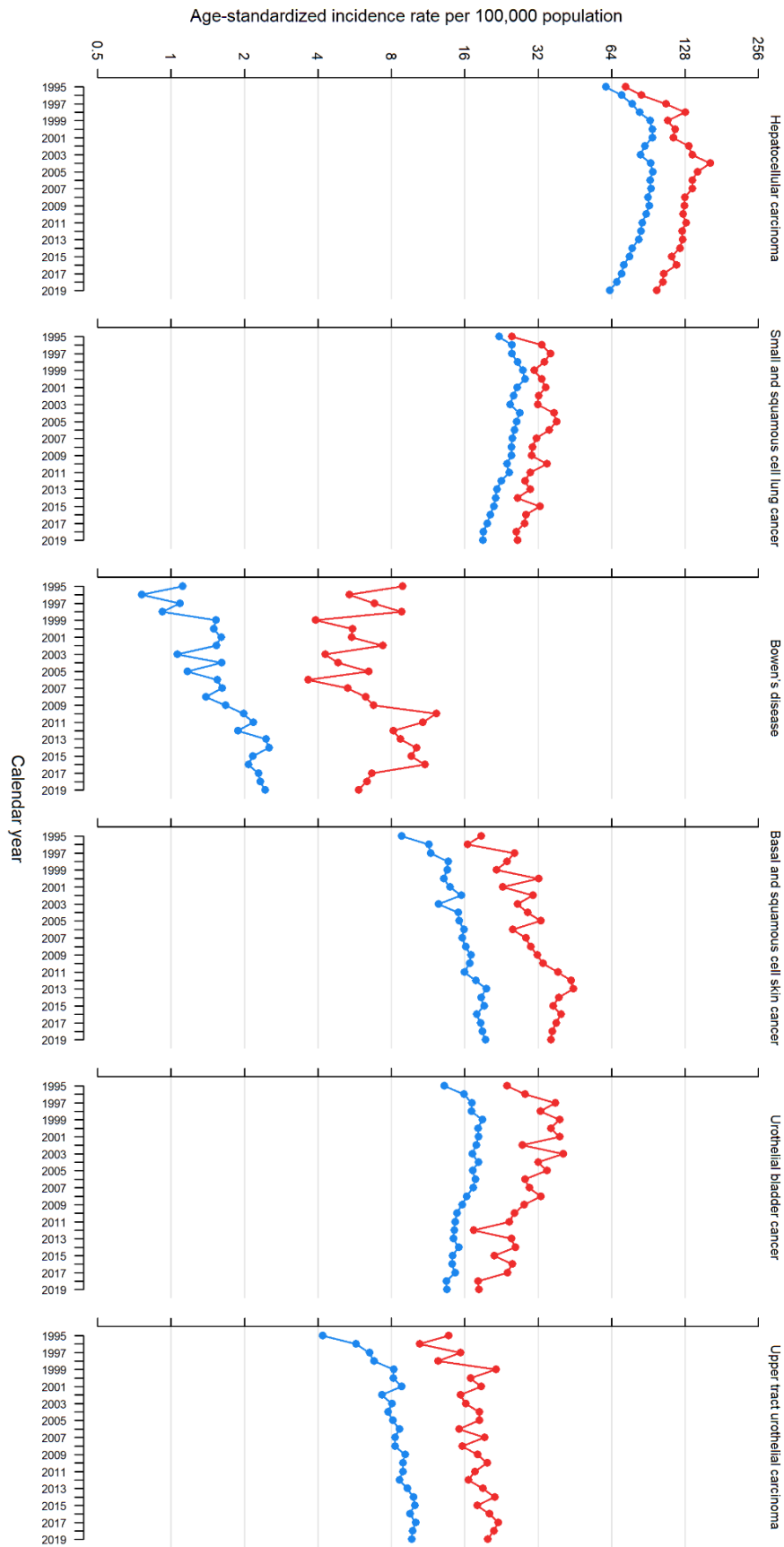
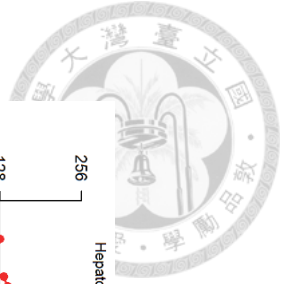
³未調整碎塊效應(nugget effect)。

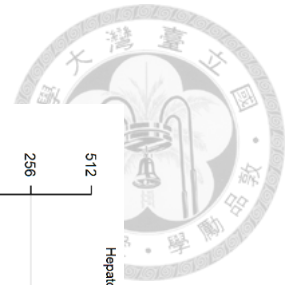
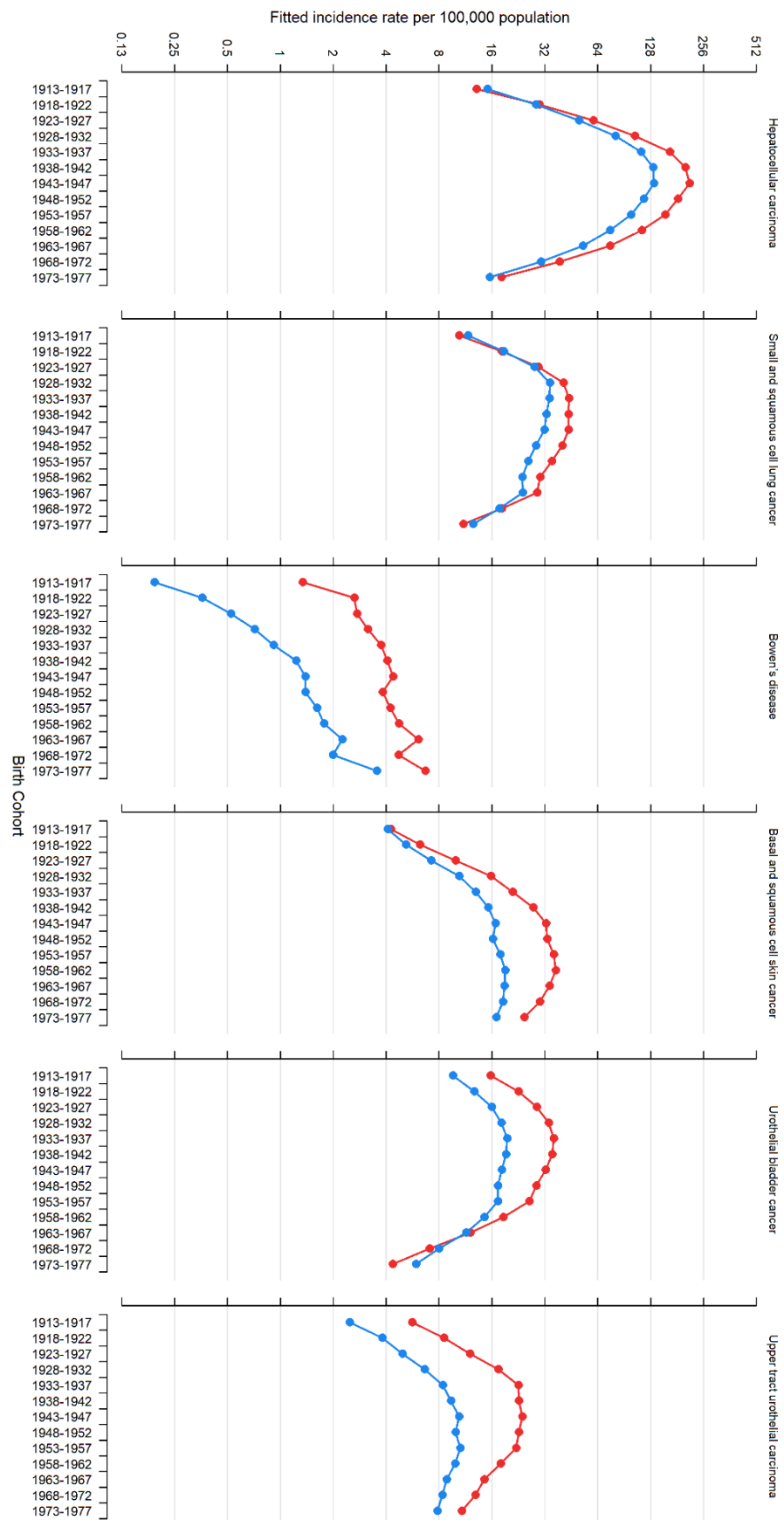
⁴有調整碎塊效應。

圖註



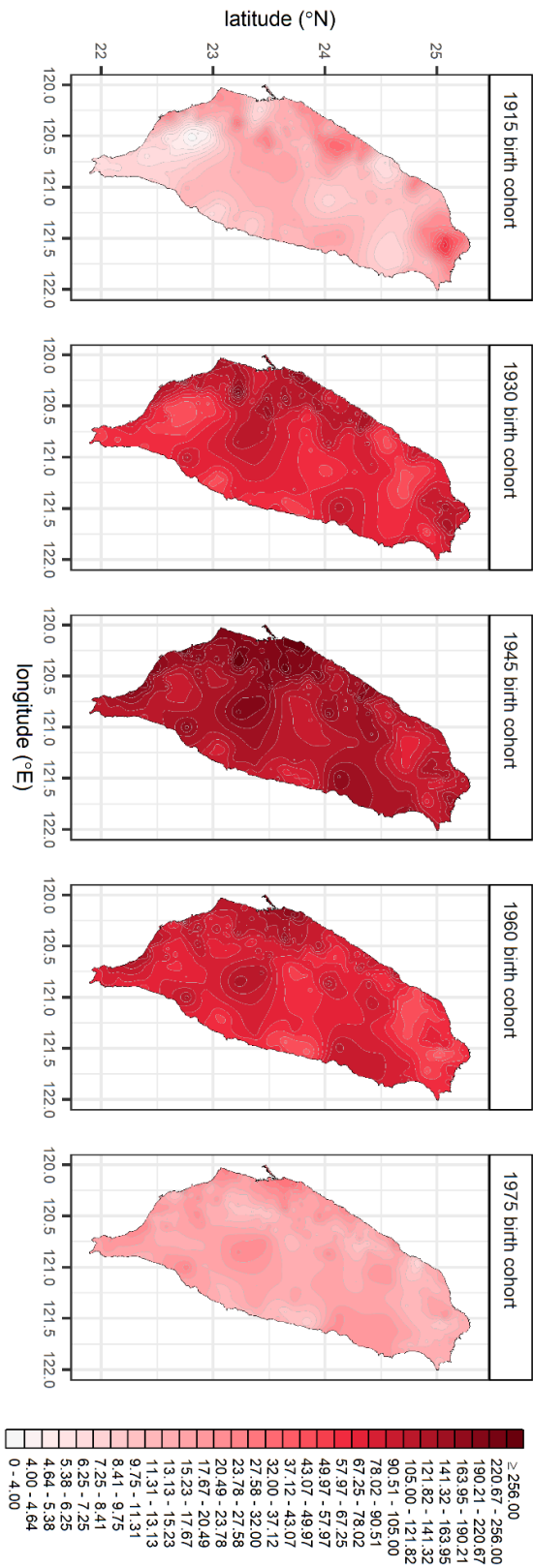
- 圖一、 1995-2019 年砷相關癌症年齡標準化發生率長期趨勢。紅色線為烏腳病流行地區，藍色線為非烏腳病流行地區。
- 圖二、 1913-1977 出生世代 60-64 歲族群的砷相關癌症發生率趨勢。紅色線為烏腳病流行地區，藍色線為非烏腳病流行地區。
- 圖三、 1913-1977 出生世代 60-64 歲族群的肝細胞癌及小細胞與鱗狀細胞肺癌發生率地理分布。影片網址：https://drive.google.com/file/d/1jTqpStpaSVczlBI_y8oiygPLHeDnLwfK/view?usp=share_link
- 圖四、 1913-1977 出生世代 60-64 歲族群的鮑恩氏病及基底細胞與鱗狀細胞皮膚癌發生率地理分布。影片網址：https://drive.google.com/file/d/1wwJYMT6sZ9t2K3wq8UqAkou3F87uicxl/view?usp=share_link
- 圖五、 1913-1977 出生世代 60-64 歲族群的膀胱泌尿上皮癌及上泌尿道泌尿上皮癌發生率地理分布。影片網址：https://drive.google.com/file/d/1rhVZHwr1L1z6jvG5NUnh52I2wVUXJ29P/view?usp=share_link



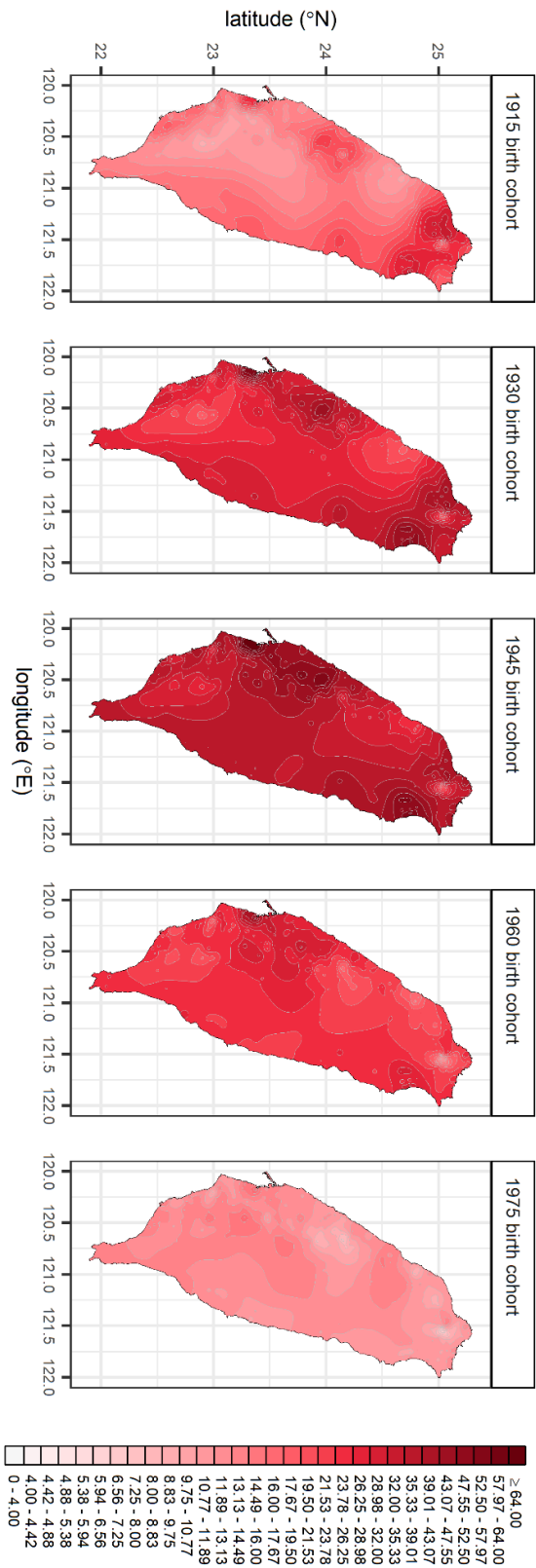




(A) Hepatocellular carcinoma

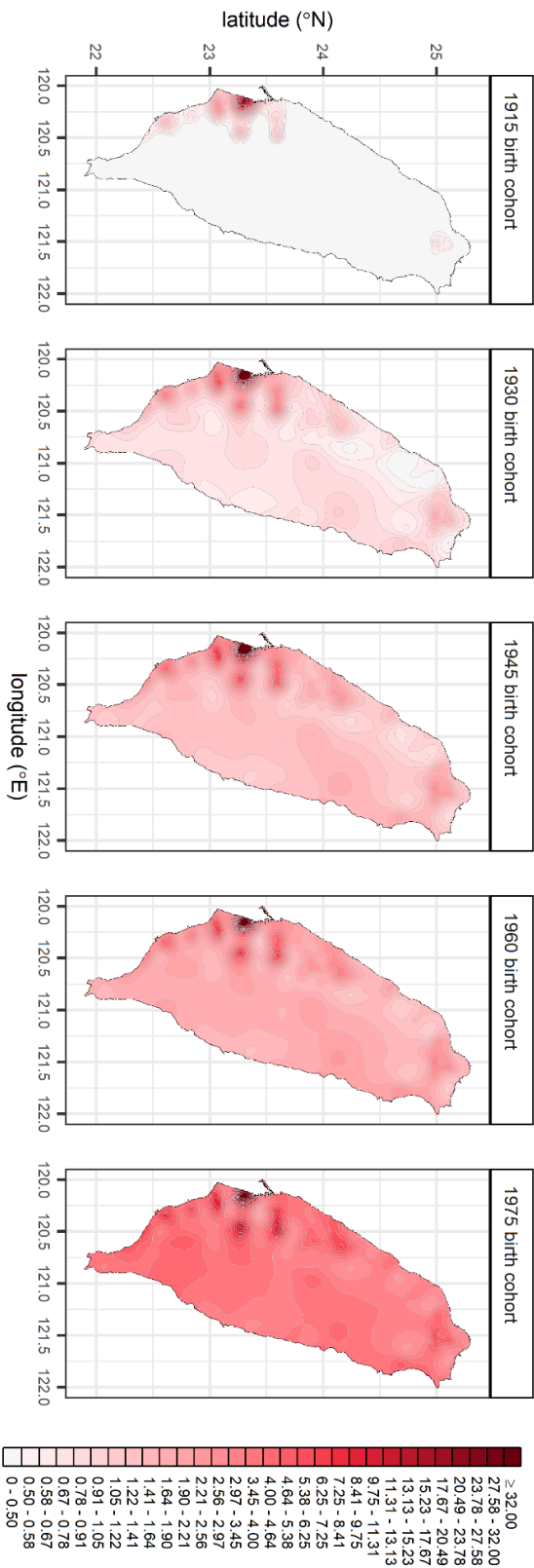


(B) Small and squamous cell lung cancer

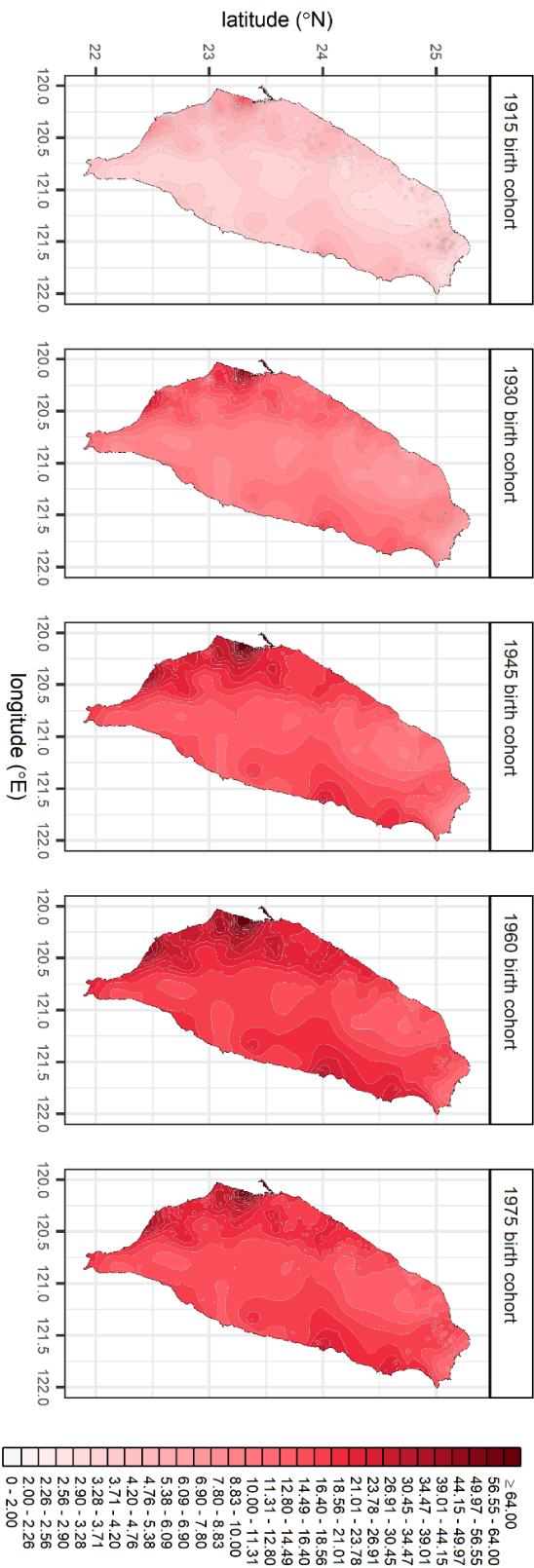




(A) Bowen's disease

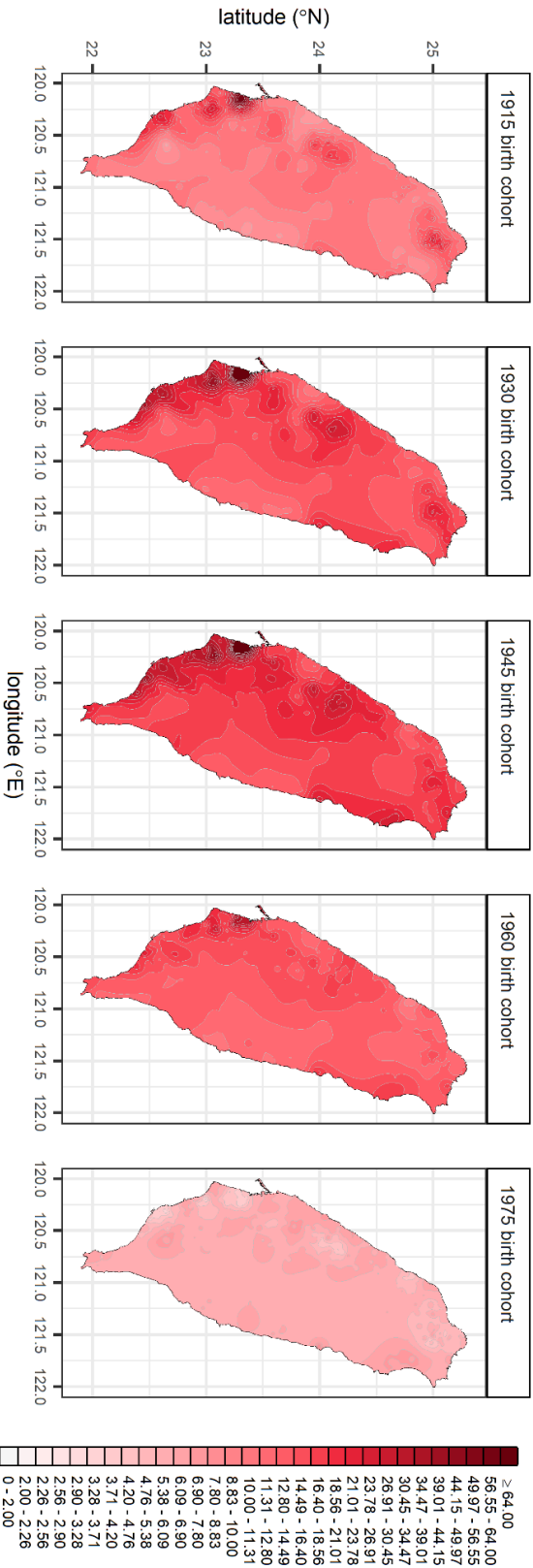


(B) Basal and squamous cell skin cancer

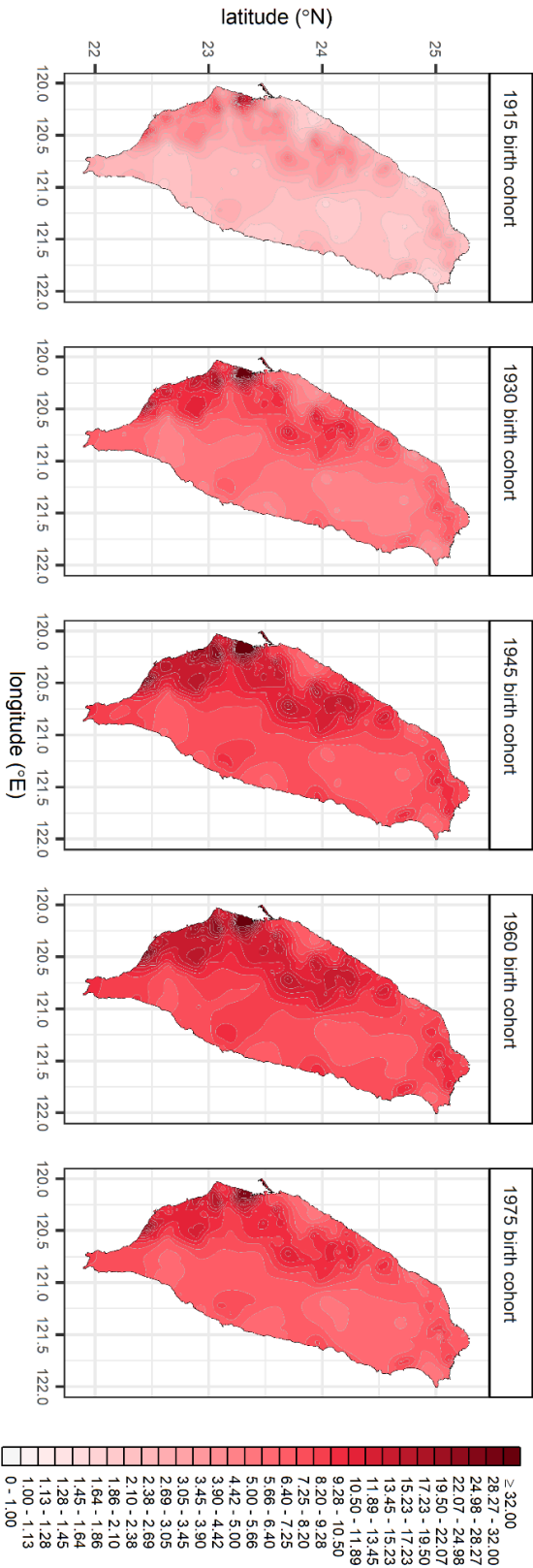




(A) Urothelial bladder cancer



(B) Upper tract urothelial carcinoma



圖五

附錄目錄



- 附錄一、覆蓋台灣本島的 129,524 個經緯座標點
- 附錄二、覆蓋嘉義與台南的 2,075 個經緯座標點及 57 個鄉鎮市區中心座標點
- 附錄三、砷相關癌症的定義，依據國際腫瘤分類標準(ICD-O-3)
- 附錄四、年齡、年代、與世代交叉分類
- 附錄五、台灣本島的 349 個鄉鎮市區地圖
- 附錄六、世界衛生組織的 2000 年世界標準人口
- 附錄七、1913-1977 出生世代 45-49 歲及 75-79 歲族群的砷相關癌症發生率趨勢
- 附錄八、1913-1977 出生世代 60-64 歲男性與女性的砷相關癌症發生率趨勢
- 附錄九、1913-1977 出生世代 60-64 歲族群在烏腳病或 C 肝流行地區的砷相關癌症發生率趨勢
- 附錄十、1913-1977 出生世代的基底細胞與鱗狀細胞皮膚癌發生率趨勢，依年齡與部位呈現
- 附錄十一、60-64 歲族群砷相關癌症發生率 Z 值的地理分布
- 附錄十二、砷相關癌症在各隨機效應的半變異圖
- 附錄十三、1913-1977 出生世代 60-64 歲族群的基底細胞與鱗狀細胞皮膚癌發生率地理分布，依部位呈現，影片網址：https://drive.google.com/file/d/1BdwaadhFK5ALK8-FcHFTMtkDmTYjUYuc/view?usp=drive_link
- 附錄十四、1913-1977 出生世代 45-49 歲族群的肝細胞癌及小細胞與鱗狀細胞肺癌發生率地理分布，影片網址：https://drive.google.com/file/d/1dfVOR0DNTHQ5LRd_dNUQMBMxucJh1LcA/view?usp=share_link
- 附錄十五、1913-1977 出生世代 45-49 歲族群的鮑恩氏病及基底細胞與鱗狀細胞皮膚癌發生率地理分布，影片網址：https://drive.google.com/file/d/1S4Lut3MaOS-VtqZ2k4He_e2lS_xV3lqK/view?usp=share_link
- 附錄十六、1913-1977 出生世代 45-49 歲族群的膀胱泌尿上皮癌及上泌尿道泌尿上皮癌發生率地理分布，影片網址：https://drive.google.com/file/d/1wg4jw4AZfYITFQqj-jGp0aLt2P3sdKUm/view?usp=share_link
- 附錄十七、1913-1977 出生世代 75-79 歲族群的肝細胞癌及小細胞與鱗狀細胞肺癌發生率地理分布，影片網址：https://drive.google.com/file/d/1HwUxUBylHVcKGt7mBIAe0Sq-X4C_5XNM/view?usp=share_link
- 附錄十八、1913-1977 出生世代 75-79 歲族群的鮑恩氏病及基底細胞與鱗狀細胞皮膚癌發生率地理分布，影片網址：<https://drive.google.com/file/d/1BCm>



G9VDxivQOijuMPnfYq9bz3vCBITh/view?usp=share_link

附錄十九、1913-1977 出生世代 75-79 歲族群的膀胱泌尿上皮癌及上泌尿道泌尿上皮癌發生率地理分布，影片網址：https://drive.google.com/file/d/1gYRga4rLjm-1ePxHQC7gYSPlnYHRFEm/view?usp=share_link

附錄二十、1913-1977 出生世代 60-64 歲男性的肝細胞癌及小細胞與鱗狀細胞肺癌發生率地理分布，影片網址：https://drive.google.com/file/d/1NaK11s3VcAAk7go1i0WufpQP2h7bXuzX/view?usp=share_link

附錄二十一、1913-1977 出生世代 60-64 歲男性的鮑恩氏病及基底細胞與鱗狀細胞皮膚癌發生率地理分布，影片網址：https://drive.google.com/file/d/1C47C5fpkxPkRKQssysHcP6cS8i8Fc8V7/view?usp=share_link

附錄二十二、1913-1977 出生世代 60-64 歲男性的膀胱泌尿上皮癌及上泌尿道泌尿上皮癌發生率地理分布，影片網址：https://drive.google.com/file/d/13yB_OpXKYMgmnJaIxM8QhO-fR3QpKCOc/view?usp=share_link

附錄二十三、1913-1977 出生世代 60-64 歲女性的肝細胞癌及小細胞與鱗狀細胞肺癌發生率地理分布，影片網址：https://drive.google.com/file/d/1uTLQf-cqUQWWwr6vq08GwPMcLvCn-JxS/view?usp=share_link

附錄二十四、1913-1977 出生世代 60-64 歲女性的鮑恩氏病及基底細胞與鱗狀細胞皮膚癌發生率地理分布，影片網址：https://drive.google.com/file/d/1As61FTJBqqRyQTKynxTohc3JMdQwWGsC/view?usp=share_link

附錄二十五、1913-1977 出生世代 60-64 歲女性的膀胱泌尿上皮癌及上泌尿道泌尿上皮癌發生率地理分布，影片網址：https://drive.google.com/file/d/1DUH1gKG7dqXTEmuNVouWtTJe7IdDl7S/view?usp=share_link

附錄二十六、1913-1977 出生世代 60-64 歲族群的肺腺癌與泌尿上皮癌發生率趨勢

附錄二十七、1913-1977 出生世代 60-64 歲肺腺癌與泌尿上皮癌發生率地理分布

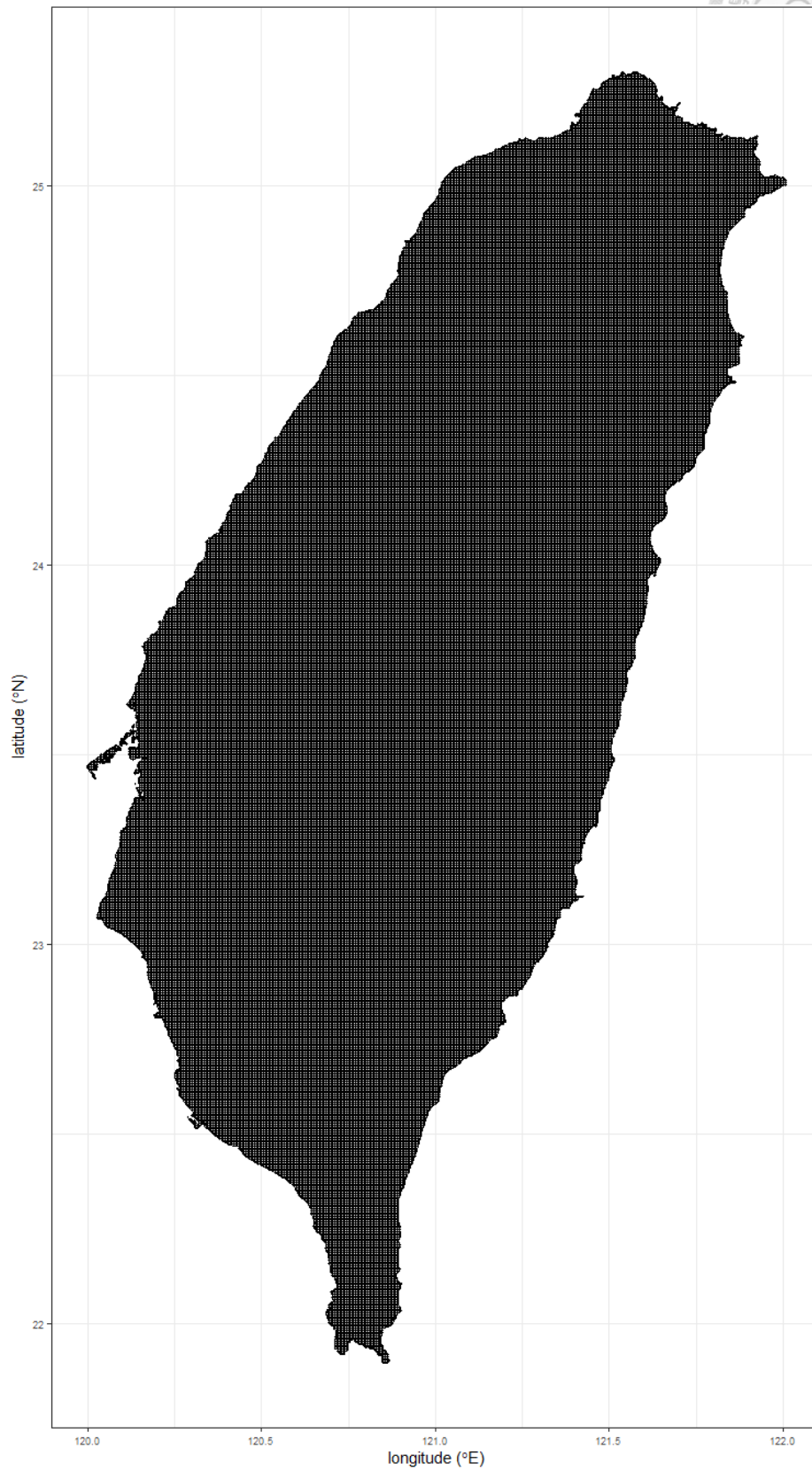
附錄二十八、肝細胞癌及小細胞與鱗狀細胞肺癌在烏腳病流行區與非流行區的年齡別發生率及世代效應趨勢

附錄二十九、鮑恩氏病及基底細胞與鱗狀細胞皮膚癌在烏腳病流行區與非流行區的年齡別發生率及世代效應趨勢

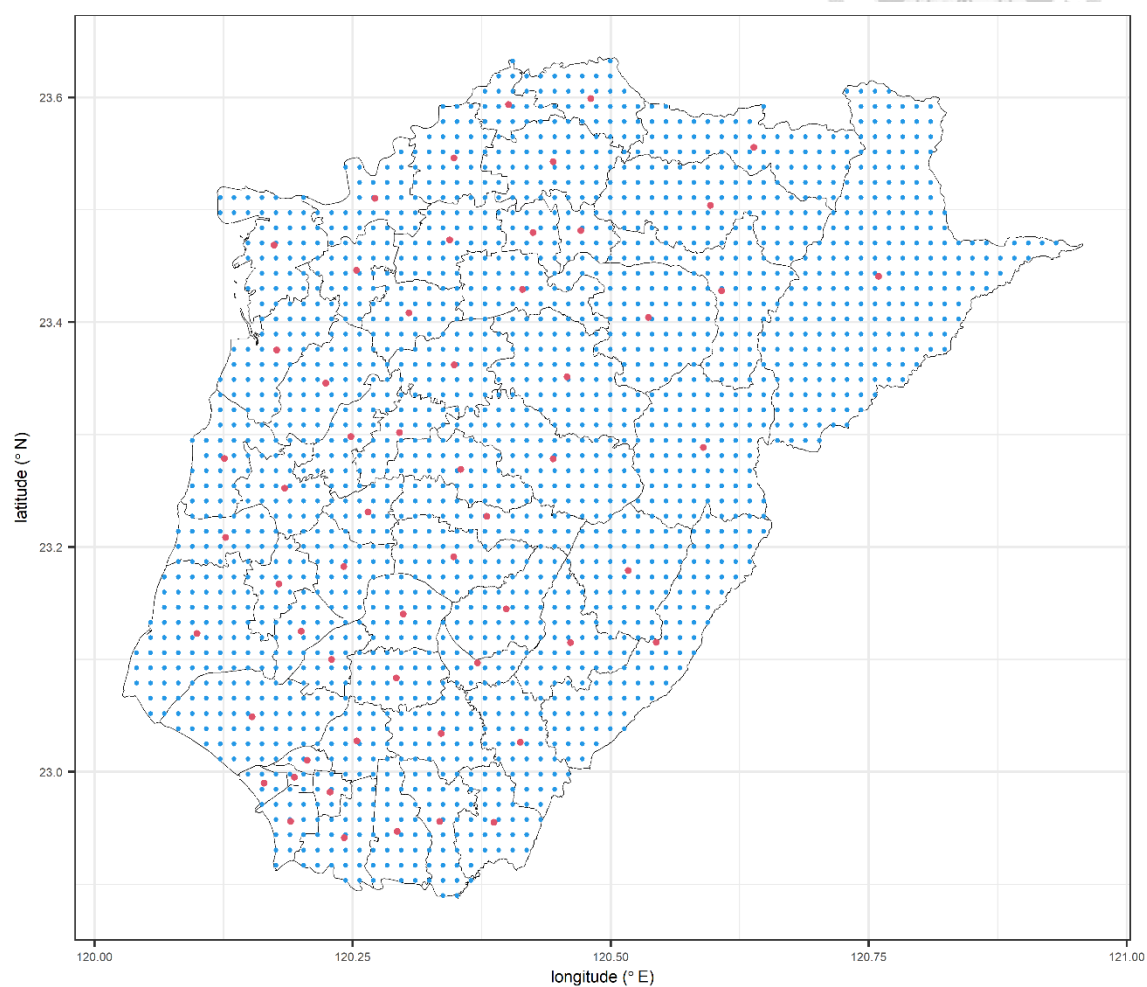
附錄三十、膀胱泌尿上皮癌及上泌尿道泌尿上皮癌在烏腳病流行區與非流行區的年齡別發生率及世代效應趨勢

附錄三十一、電腦模擬之 R 程式碼

附錄一、覆蓋台灣本島的 129,524 個經緯座標點



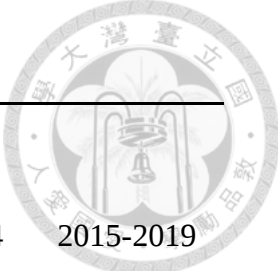
附錄二、覆蓋嘉義與台南的 2,075 個經緯座標點(藍色點)及 57 個鄉鎮市區中心座標點(紅色點)



附錄三、砵相關癌症的定義，依據國際腫瘤分類標準(ICD-O-3)

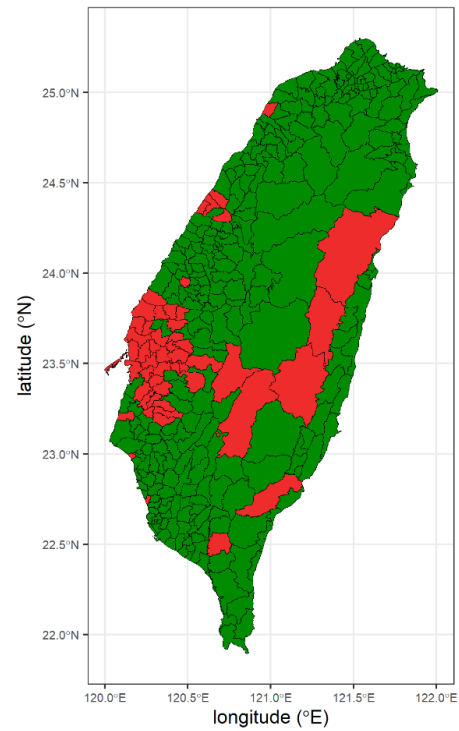
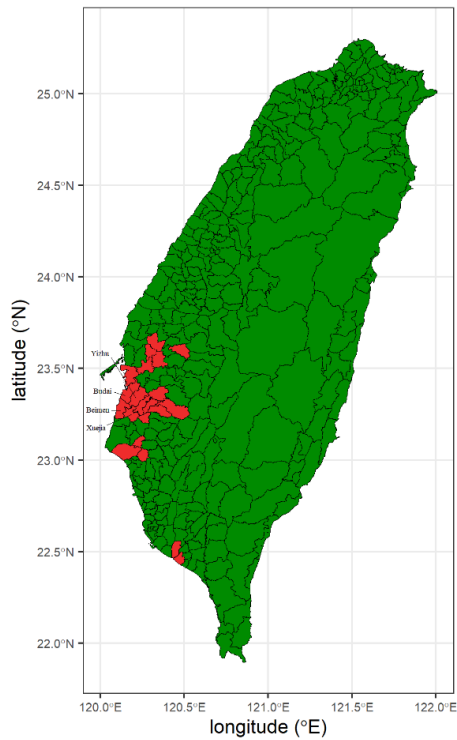
Cancer types	ICD-O-3 topography codes	ICD-O-3 morphology codes	ICD-O-3 behavior code
肝細胞癌 Hepatocellular carcinoma	C220	8170-8175, 8180, 9990	/3 (malignant tumors)
小細胞與鱗狀細胞肺癌 Small and squamous cell lung cancer	C34	8002, 8041-8045, 8051-8052, 8070-8076, 8083-8084	/3 (malignant tumors)
基底細胞與鱗狀細胞皮膚癌 Basal and squamous cell skin cancer	C44	8090-8095, 8097, 8051-8052, 8070-8076, 8083-8084	/3 (malignant tumors)
鮑恩氏病 Bowen's disease	C44	8010, 8052, 8072, 8076, 8070-8071, 8075, 8077, 8081, 8120	/2 (in situ lesions)
膀胱泌尿上皮癌 Urothelial bladder cancer	C67	8120, 8122, 8130-8131	/3 (malignant tumors)
上泌尿道泌尿上皮癌 Upper tract urothelial carcinoma	C659, C669	8120, 8122, 8130-8131	/3 (malignant tumors)

附錄四、年齡、年代、與世代交叉分類



年 代 年 齡	1995-1999	2000-2004	2005-2009	2010-2014	2015-2019
40-44	1953-1957	1958-1962	1962-1967	1968-1972	1973-1977
45-49	1948-1952	1953-1957	1958-1962	1963-1967	1968-1972
50-54	1943-1947	1948-1952	1953-1957	1958-1962	1963-1967
55-59	1938-1942	1943-1947	1948-1952	1953-1957	1958-1962
60-64	1933-1937	1938-1942	1943-1947	1948-1952	1953-1957
65-69	1928-1932	1933-1937	1938-1942	1943-1947	1948-1952
70-74	1923-1927	1928-1932	1933-1937	1938-1942	1943-1947
75-79	1918-1922	1923-1927	1928-1932	1933-1937	1938-1942
80-84	1913-1917	1918-1922	1923-1927	1928-1932	1933-1937

附錄五、台灣本島的 349 個鄉鎮市區地圖。左圖中紅色部分的 21 個鄉鎮市區是烏腳病流行地區(水中砷濃度達 350 ng/ml)；右圖中紅色部分的 51 個鄉鎮市區是 C 型肝炎流行地區(anti-HCV 陽性率達 10%)

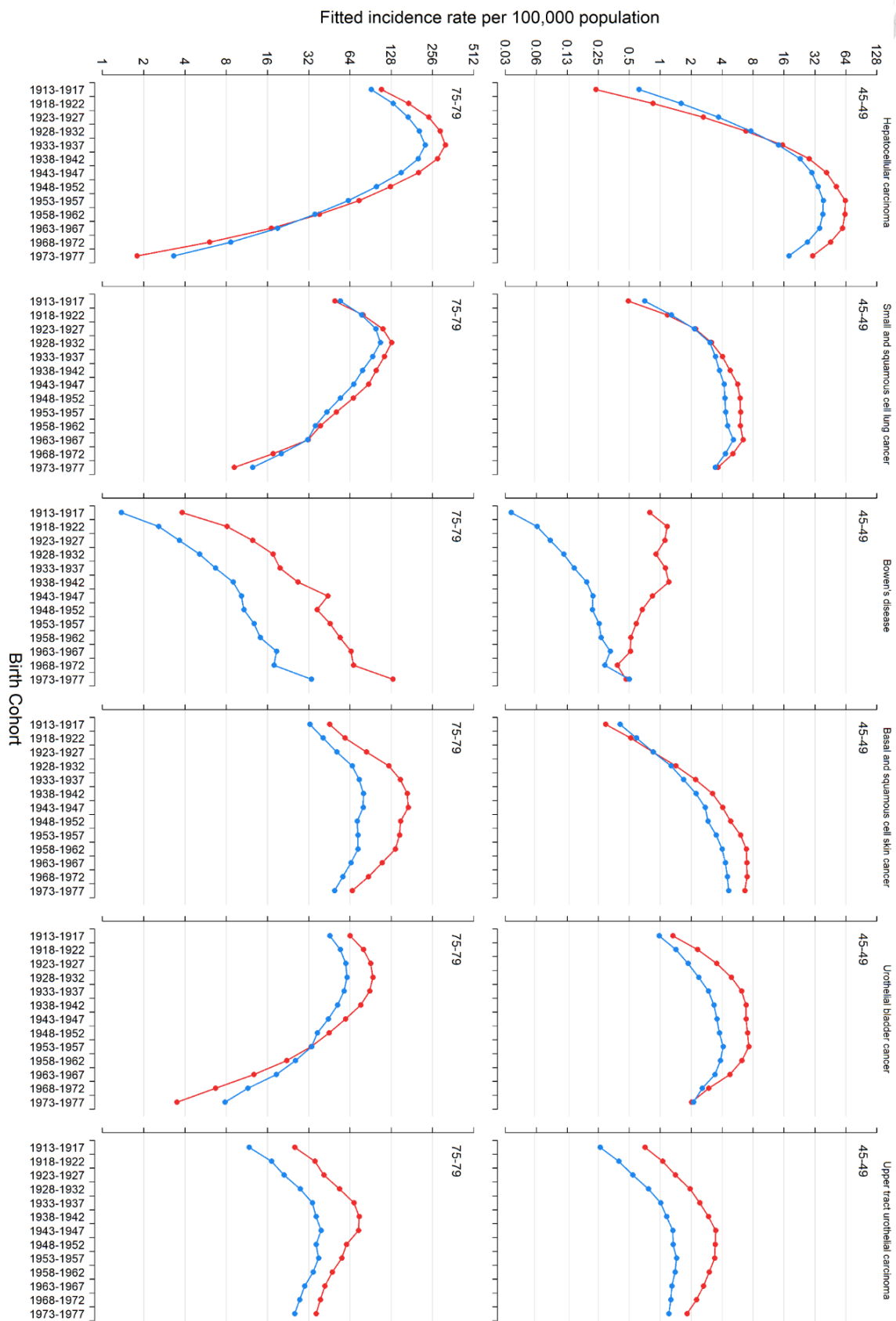


附錄六、世界衛生組織的 2000 年世界標準人口

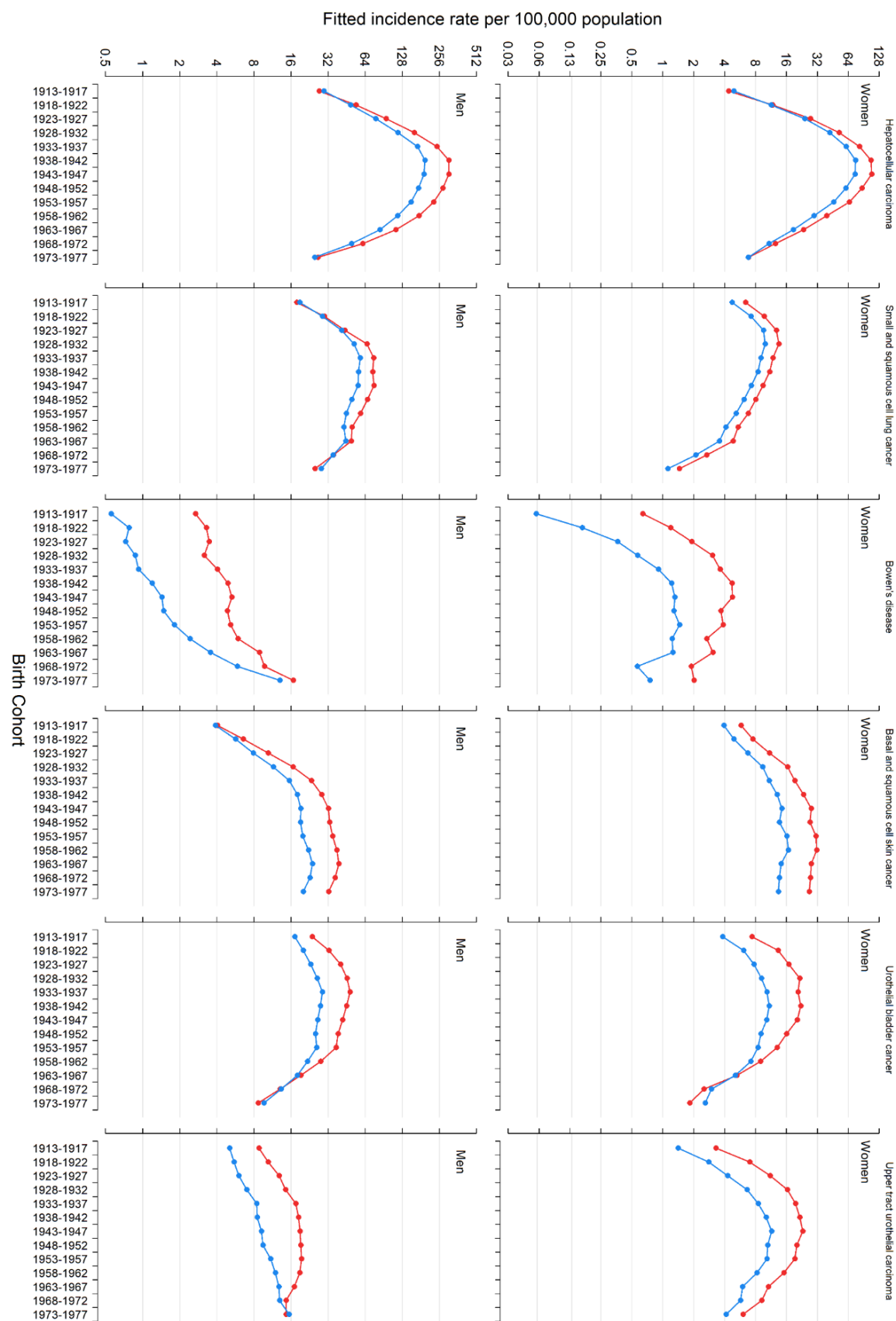
年齡區間	全年齡標準人口數	40-84 歲標準人口數
0-4	8,800	0
5-9	8,700	0
10-14	8,600	0
15-19	8,500	0
20-24	8,200	0
25-29	7,900	0
30-34	7,600	0
35-39	7,200	0
40-44	6,600	19,469
45-49	6,000	17,699
50-54	5,400	15,929
55-59	4,600	13,569
60-64	3,700	10,914
65-69	3,000	8,850
70-74	2,200	6,490
75-79	1,500	4,425
80-84	900	2,655
85+	600	0



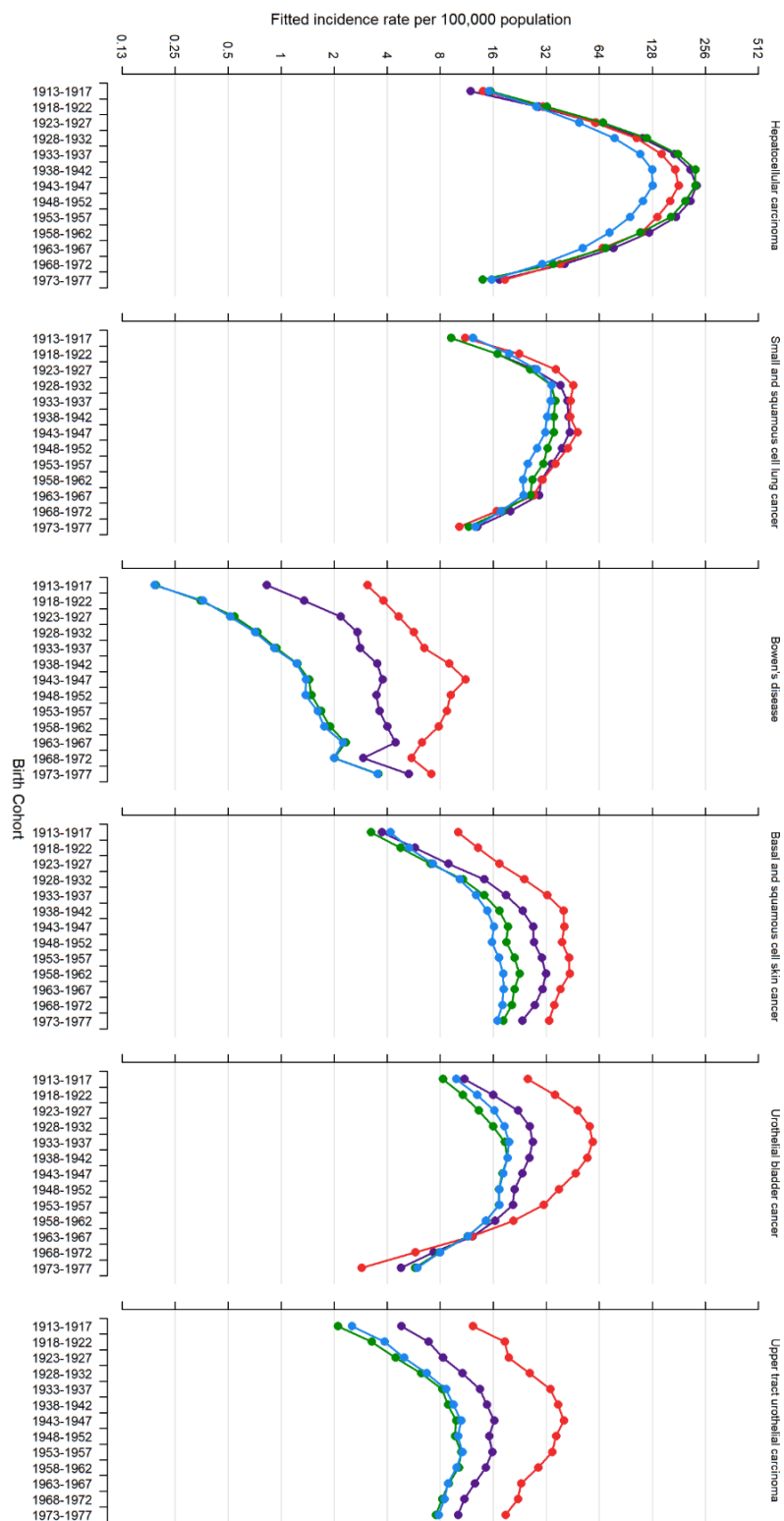
附錄七、1913-1977 出生世代 45-49 歲及 75-79 歲族群的砷相關癌症發生率趨勢，
紅色線為烏腳病流行地區，藍色線為非烏腳病流行地區



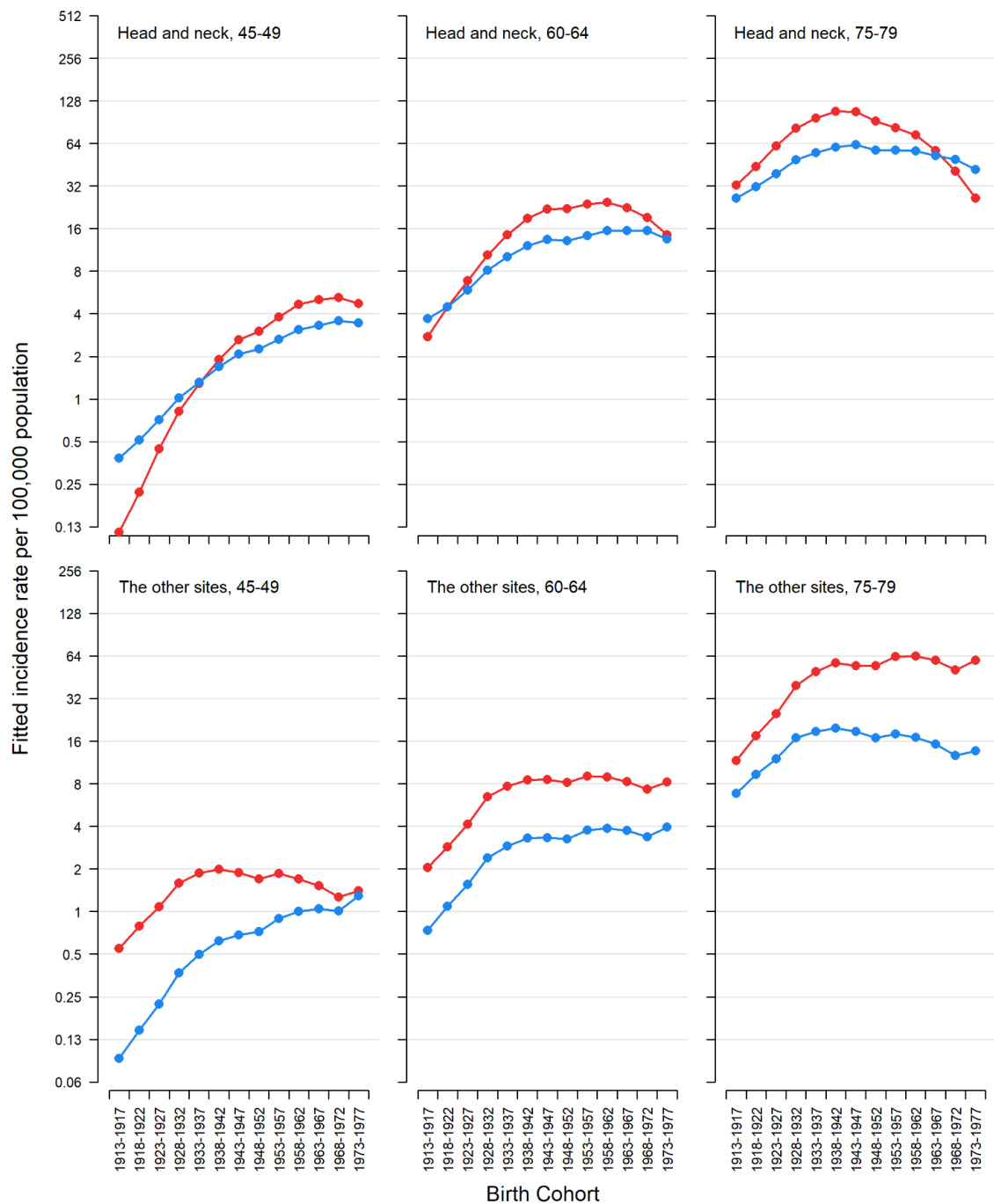
附錄八、1913-1977 出生世代 60-64 歲男性與女性的砷相關癌症發生率趨勢，紅色線為烏腳病流行地區，藍色線為非烏腳病流行地區



附錄九、1913-1977 出生世代 60-64 歲族群在烏腳病或 C 型肝炎流行地區的砷相關癌症發生率趨勢，紫色線是烏腳病與 C 型肝炎流行地區，紅色線是僅為烏腳病流行地區，綠色線是僅為 C 型肝炎流行地區，藍色線為非烏腳病或 C 型肝炎流行地區

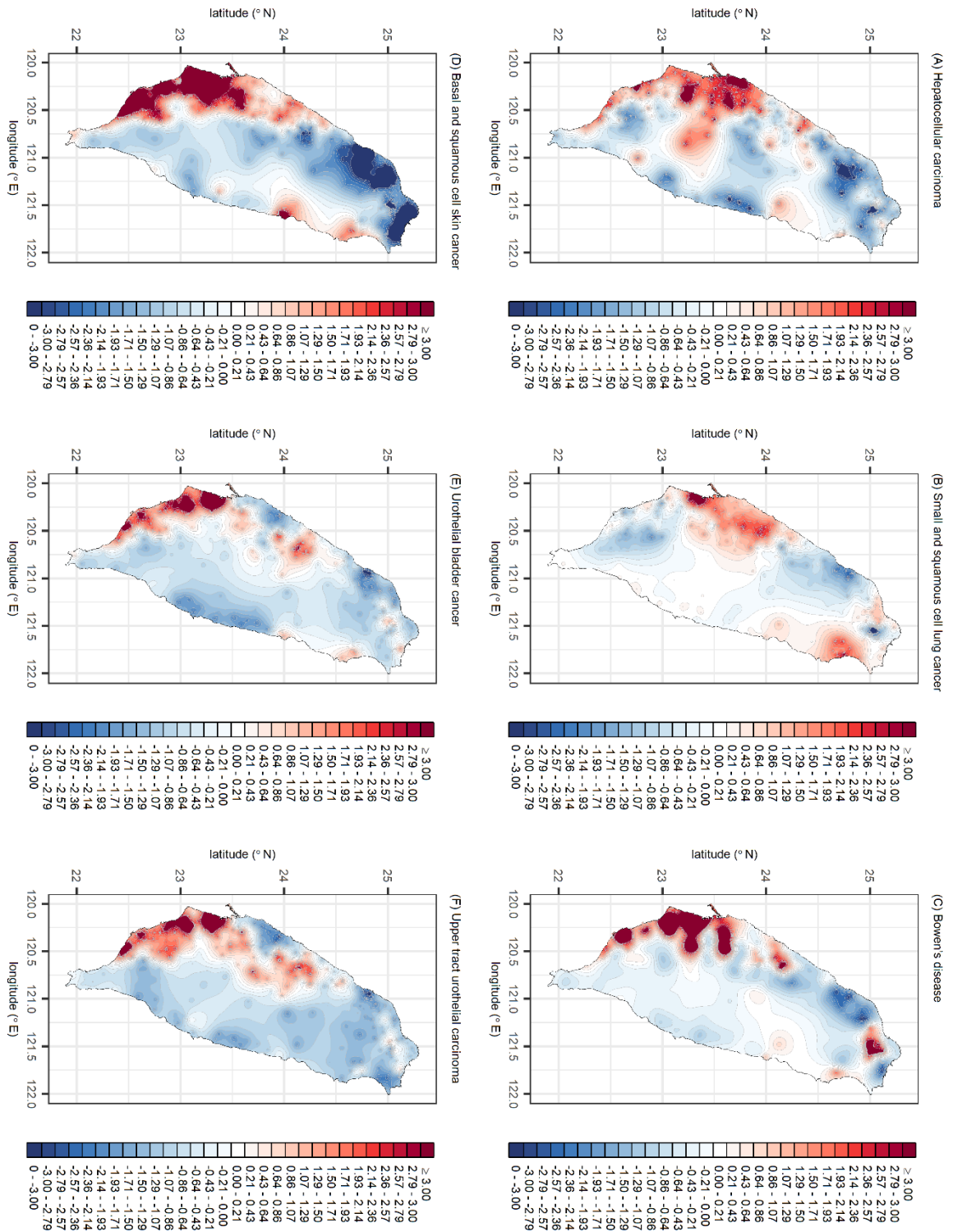


附錄十、1913-1977 出生世代的基底細胞與鱗狀細胞皮膚癌發生率趨勢，依年齡與部位呈現，紅色線為烏腳病流行地區，藍色線為非烏腳病流行地區

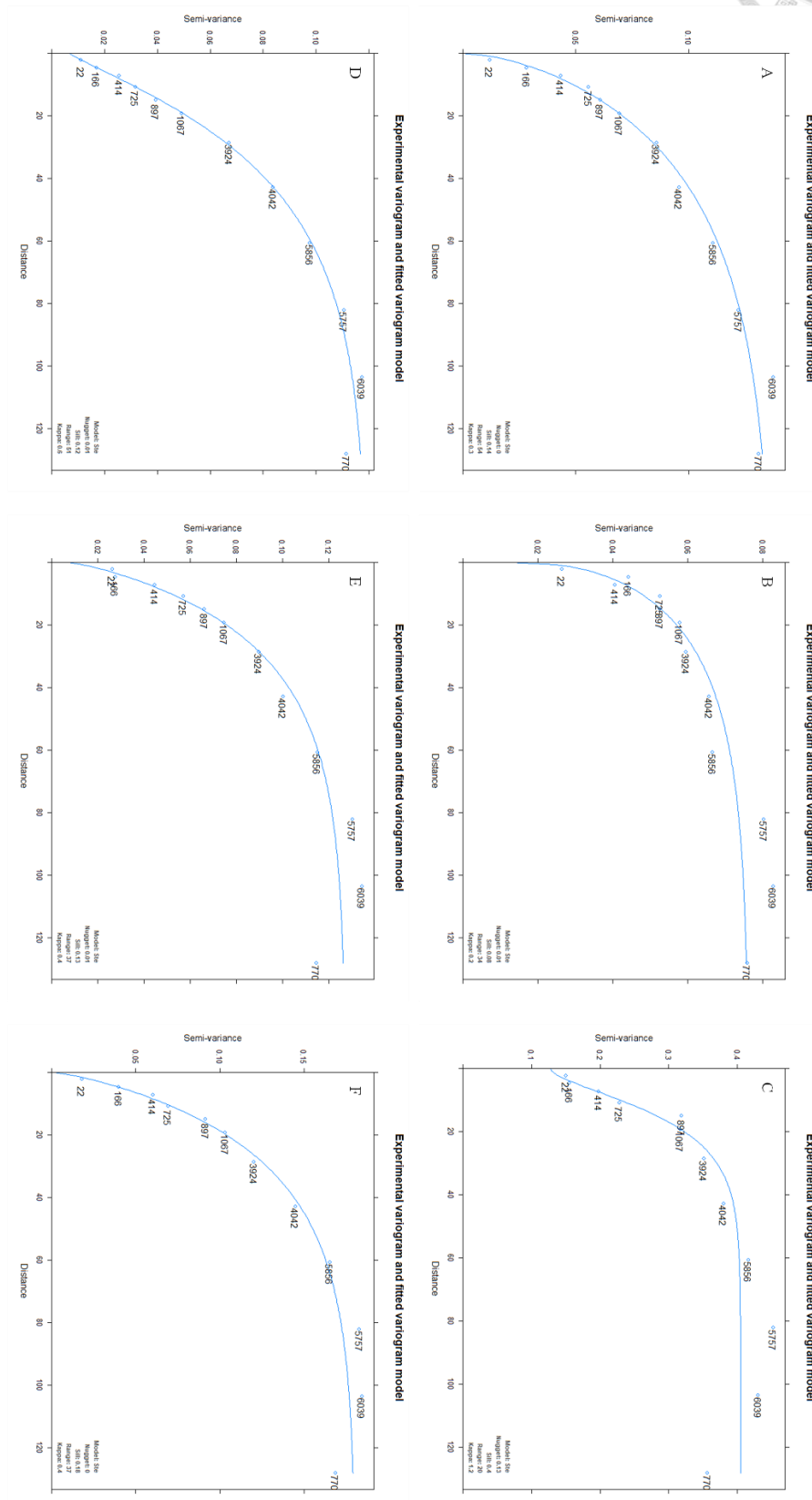




附錄十一、60-64 歲族群砷相關癌症發生率 Z 值的地理分布



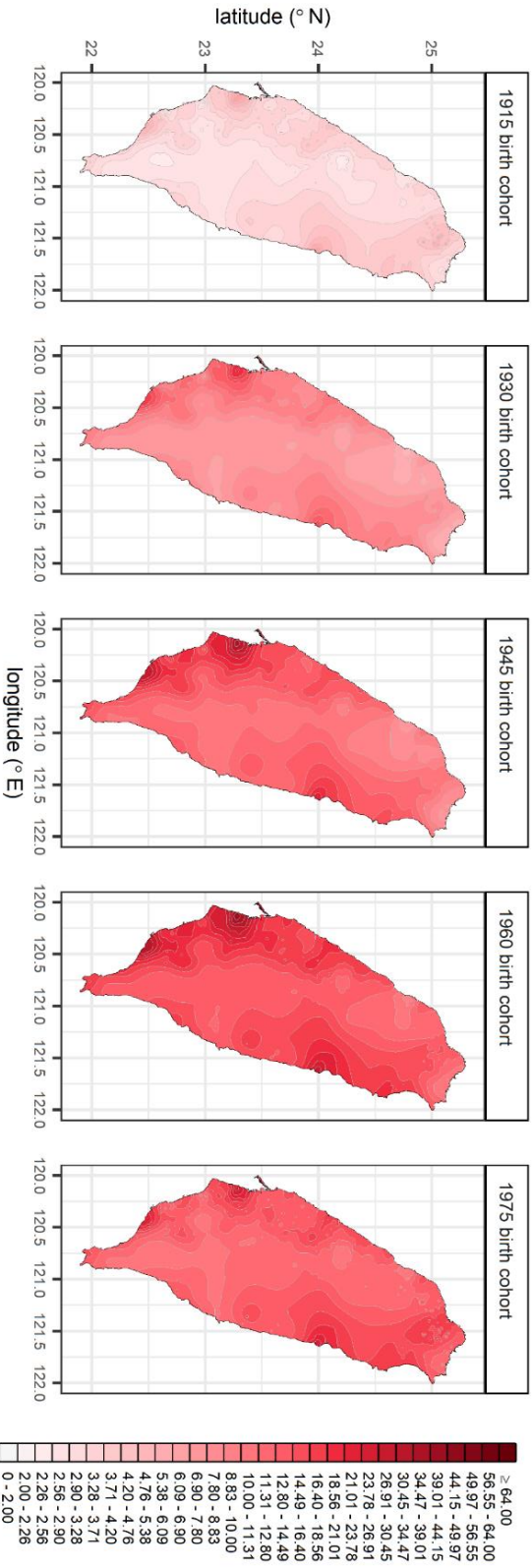
附錄十二、砷相關癌症在各隨機效應的半變異圖，圖 A 至圖 F 依序分別是肝細胞癌、小細胞與鱗狀細胞肺癌、鮑恩氏病、基底細胞與鱗狀細胞皮膚癌、膀胱泌尿上皮癌、及上泌尿道泌尿上皮癌



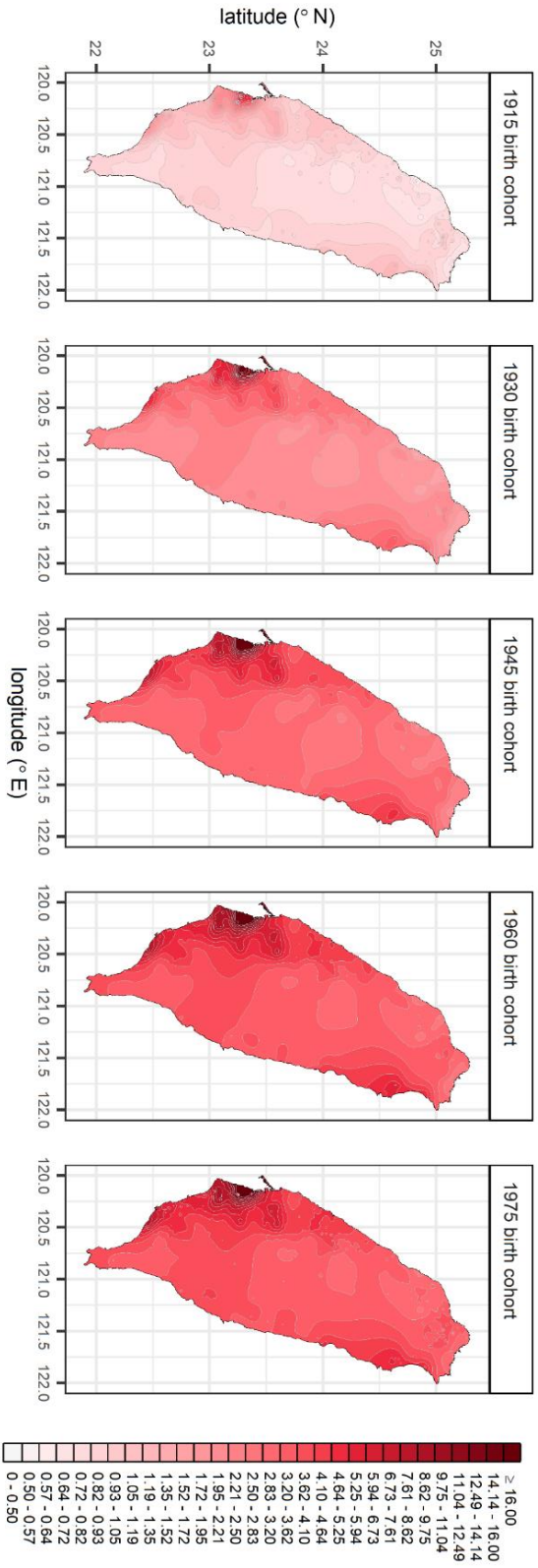
附錄十一、1913-1977 出生世代 60-64 歲族群的基底細胞與鱗狀細胞皮膚癌發生率
地理分布，依部位呈現



(A) Head and neck of basal and squamous cell skin cancer



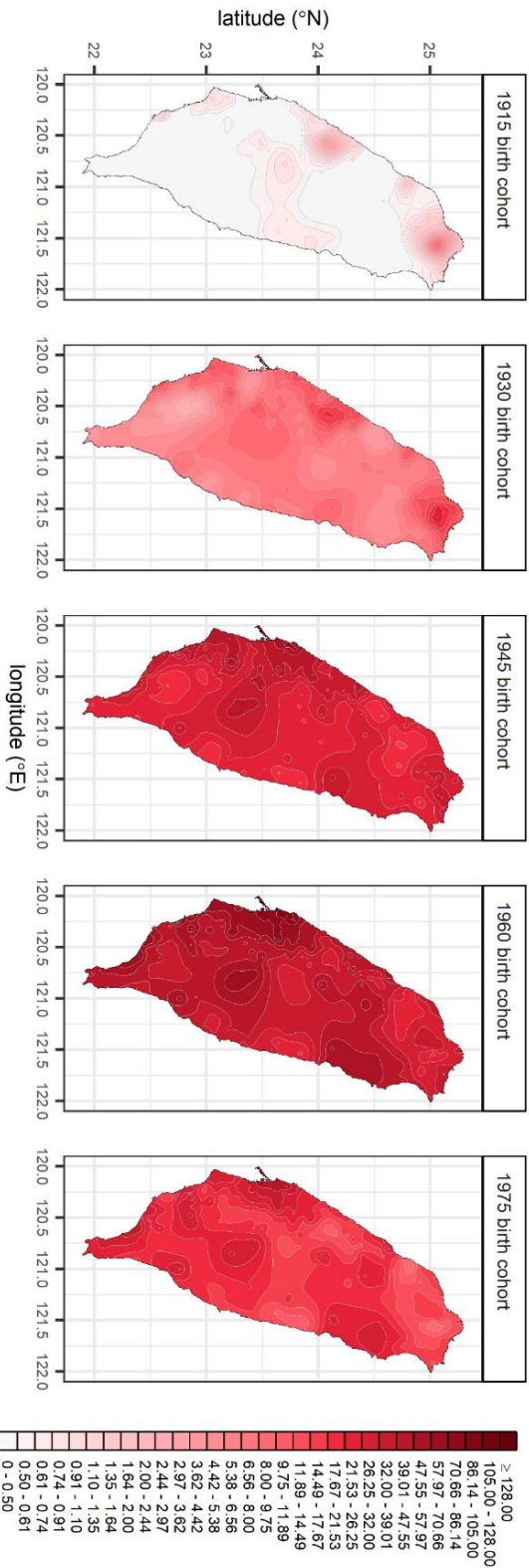
(B) The other sites of basal and squamous cell skin cancer



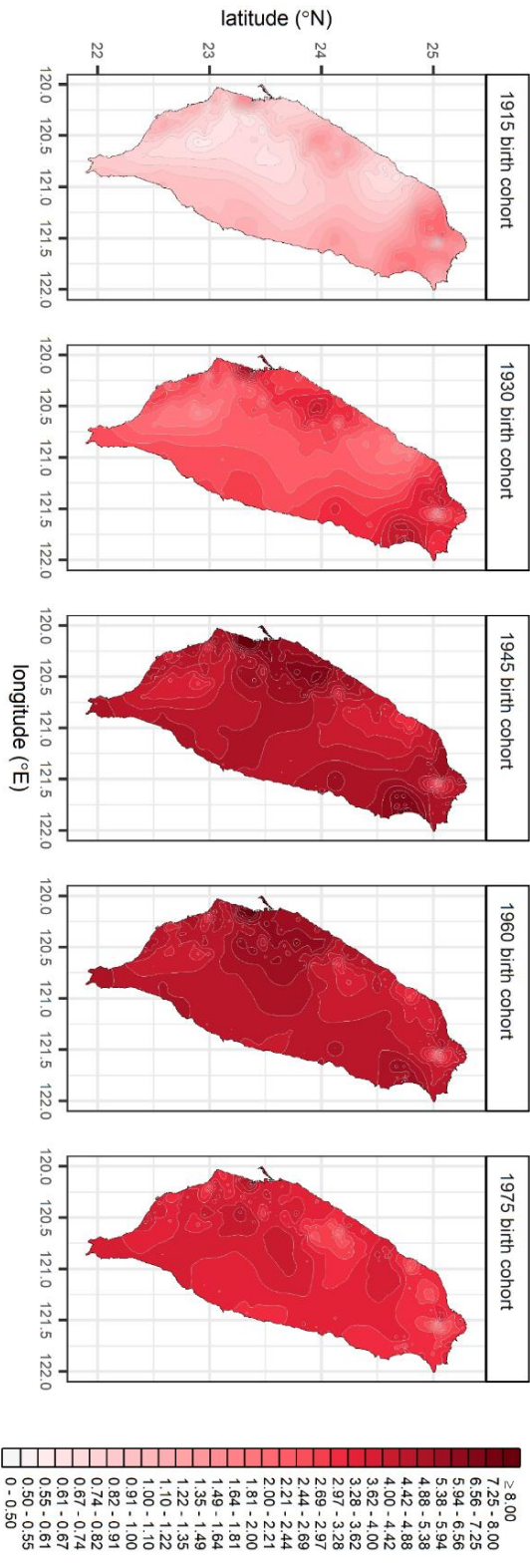
附錄十二、1913-1977 出生世代 45-49 歲族群的肝細胞癌及小細胞與鱗狀細胞肺癌
發生率地理分布



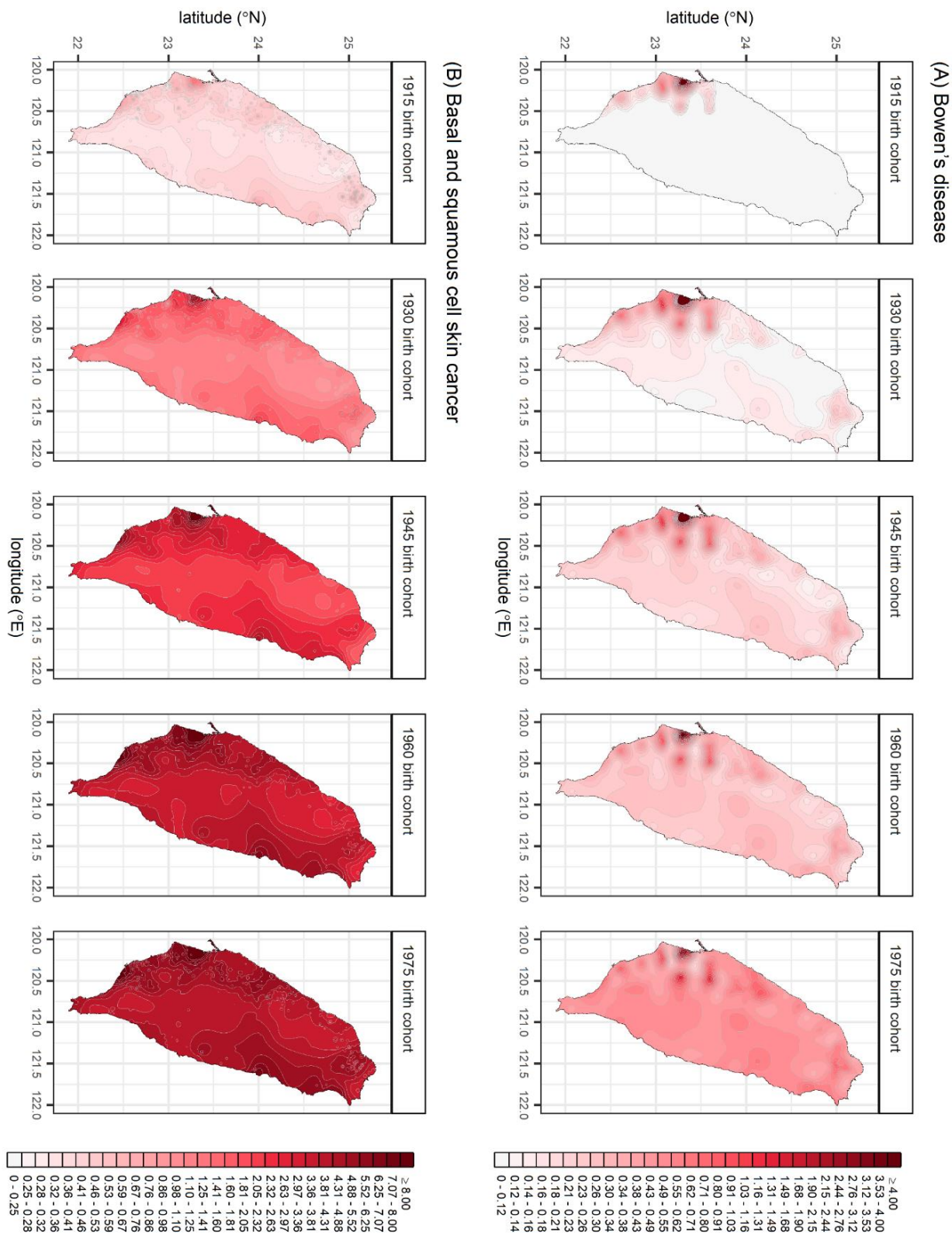
(A) Hepatocellular carcinoma



(B) Small and squamous cell lung cancer



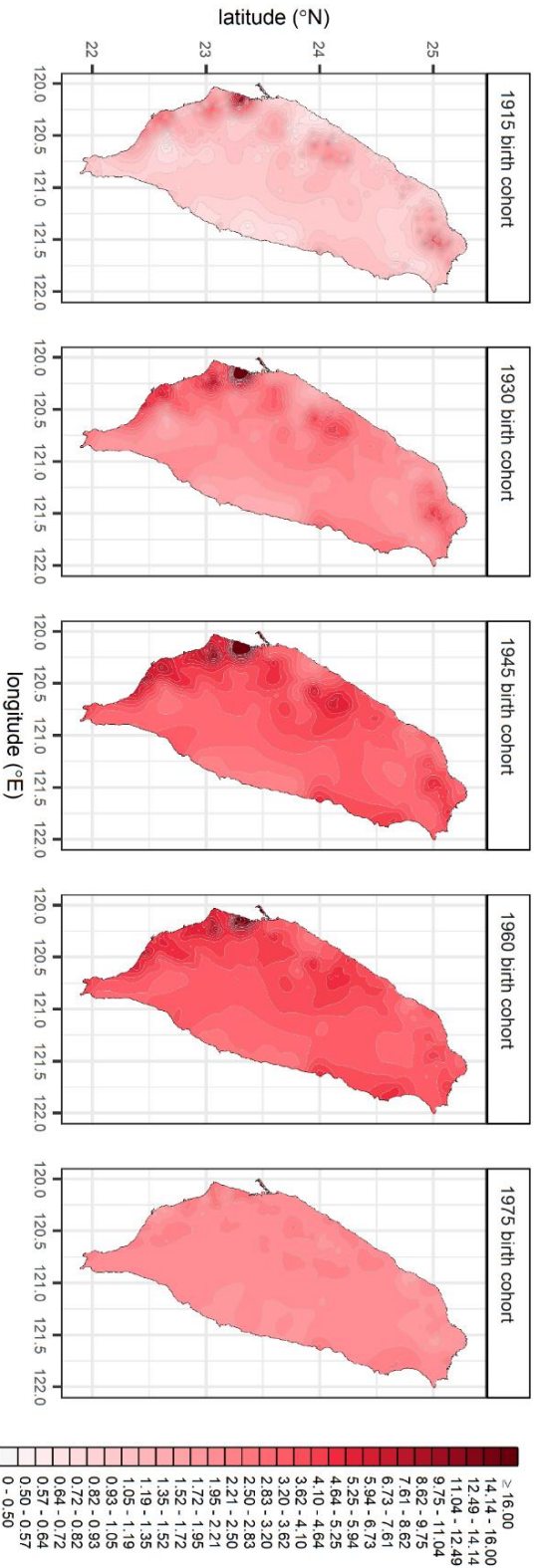
附錄十三、1913-1977 出生世代 45-49 歲族群的鮑恩氏病及基底細胞與鱗狀細胞皮膚癌發生率地理分布



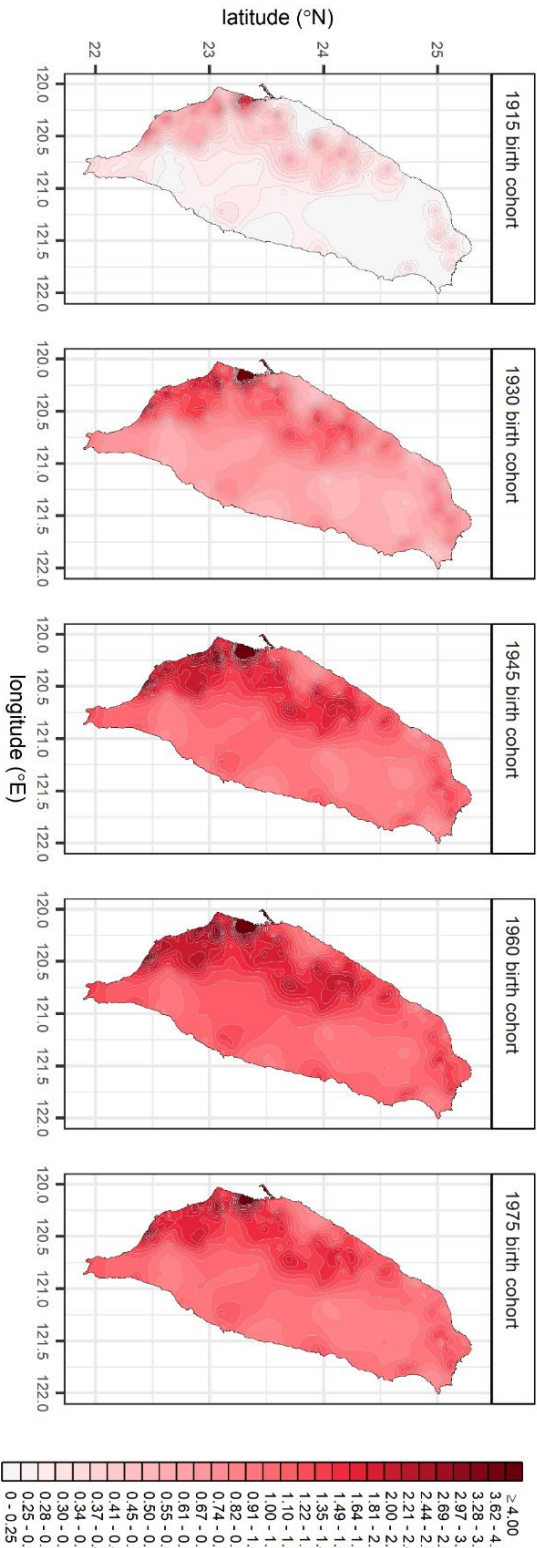
附錄十四、1913-1977 出生世代 45-49 歲族群的膀胱泌尿上皮癌及上泌尿道泌尿上皮癌發生率地理分布



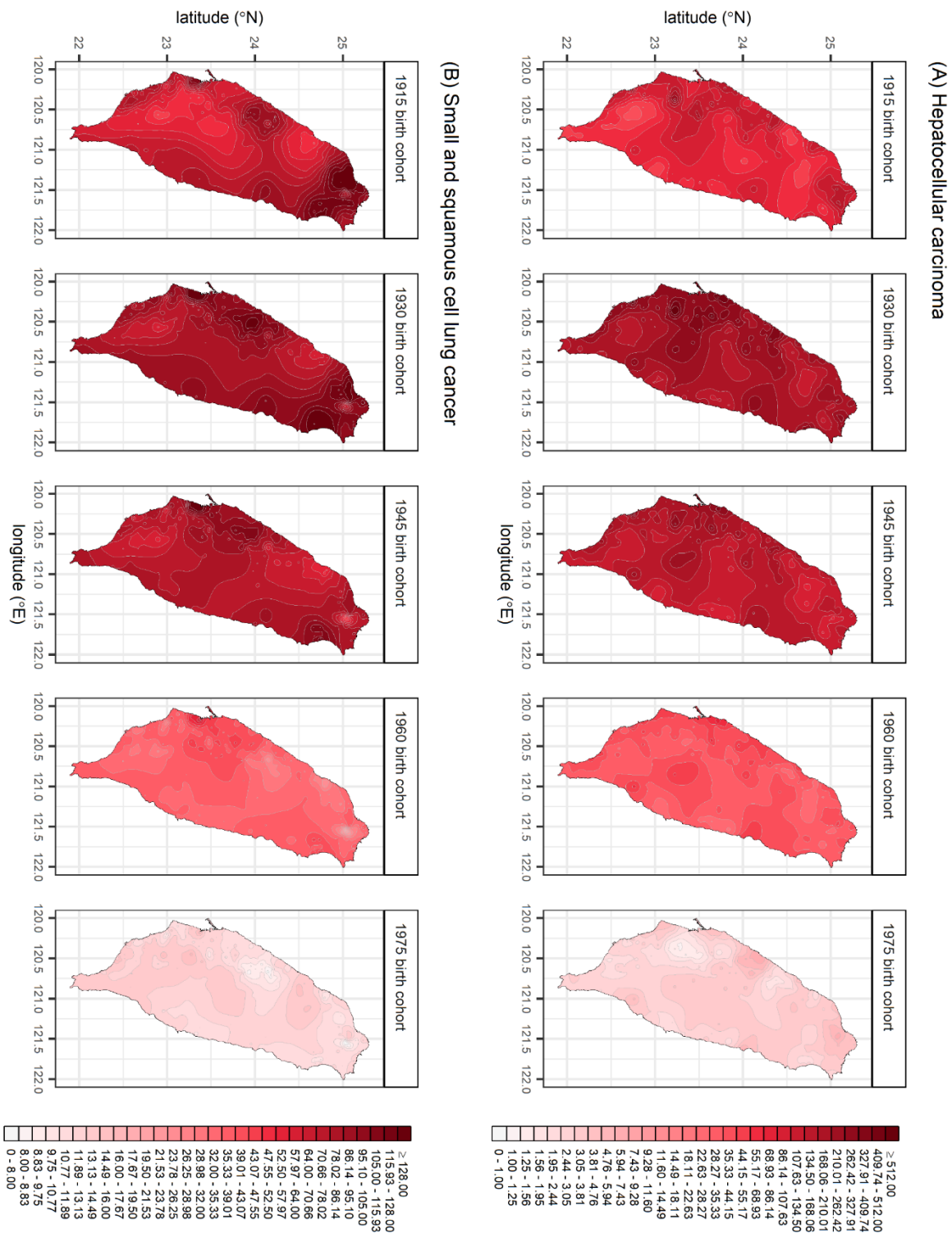
(A) Urothelial bladder cancer



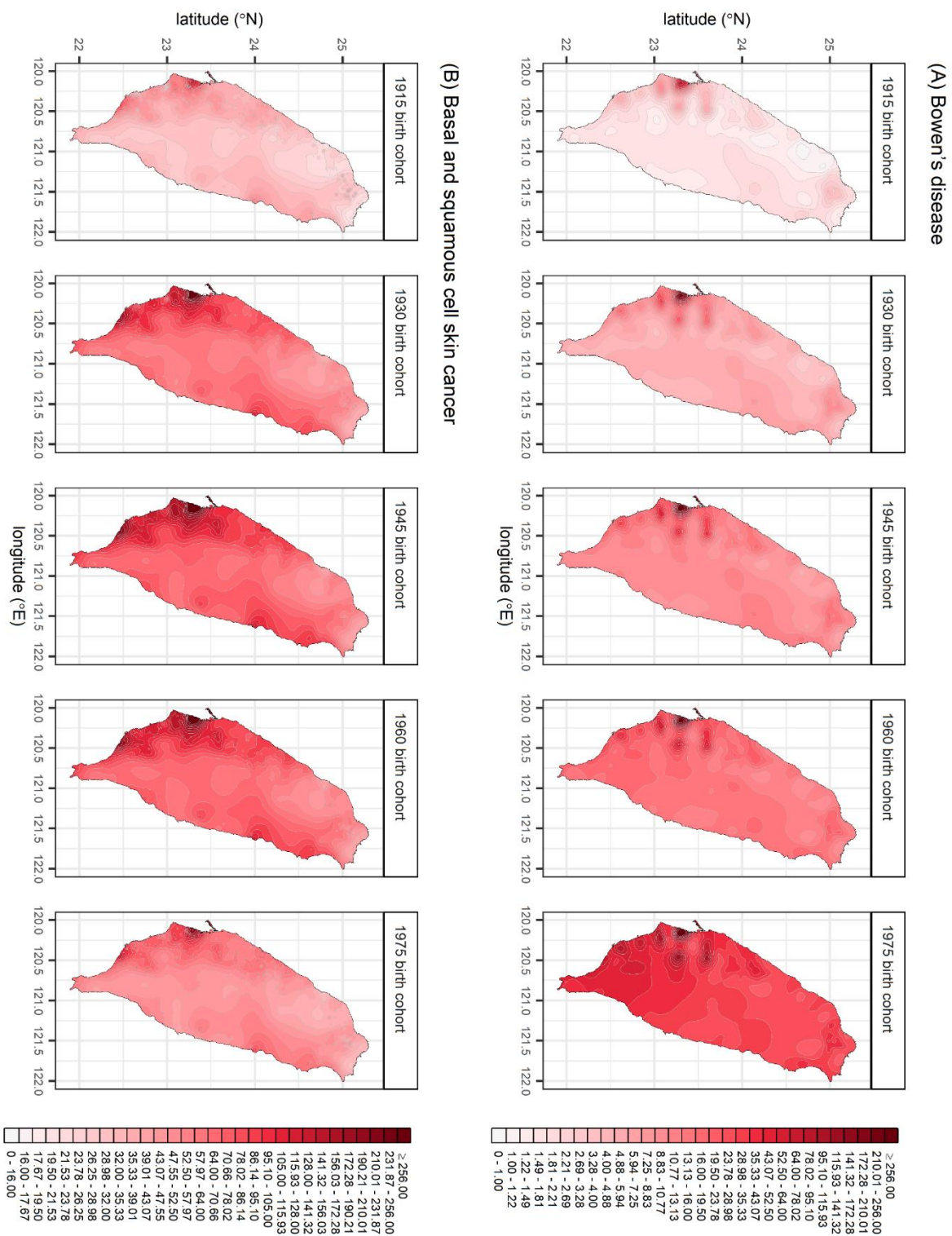
(B) Upper tract urothelial carcinoma



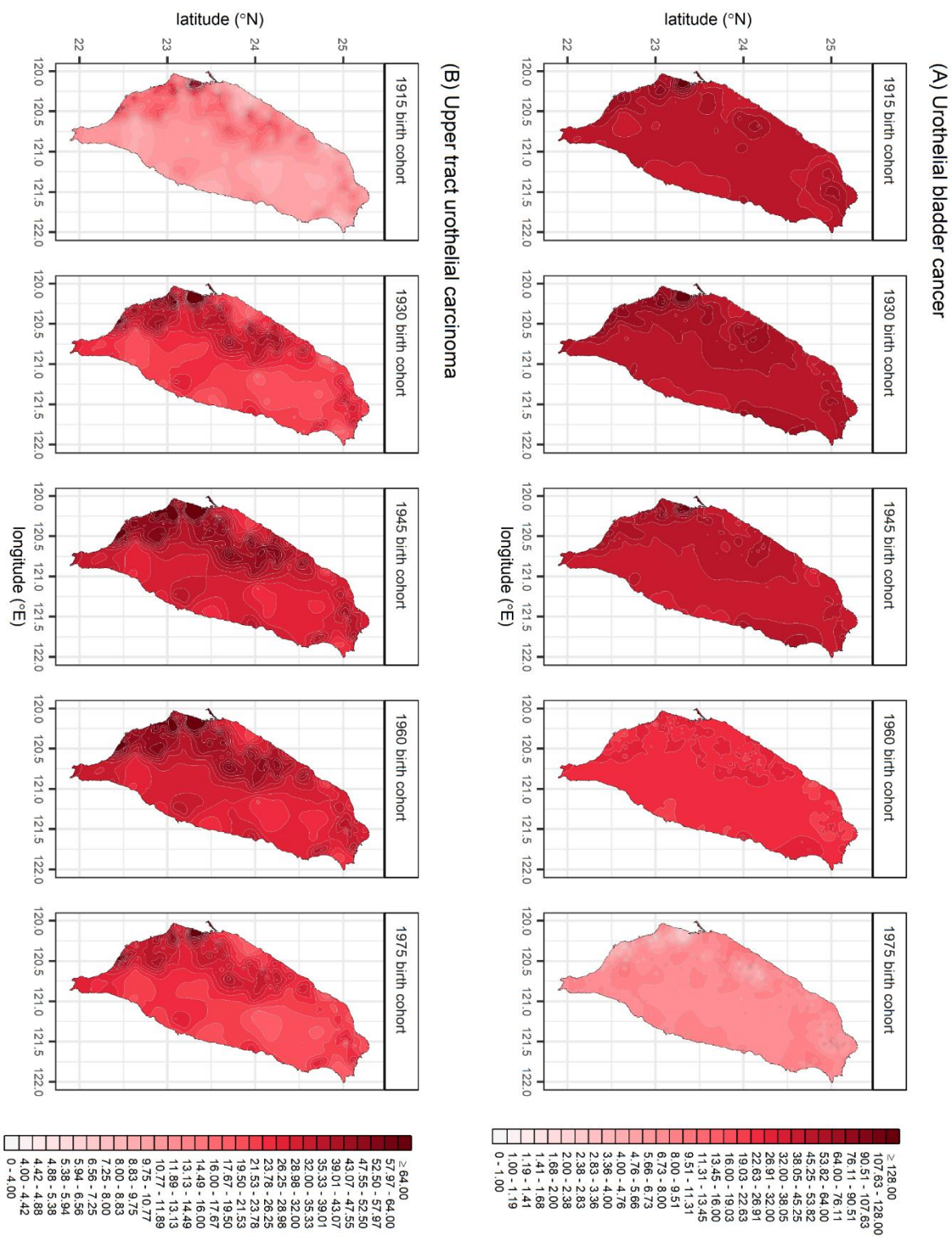
附錄十五、1913-1977 出生世代 75-79 歲族群的肝細胞癌及小細胞與鱗狀細胞肺癌發生率
地理分布



附錄十六、1913-1977 出生世代 75-79 歲族群的鮑恩氏病及基底細胞與鱗狀細胞皮膚癌發生率地理分布



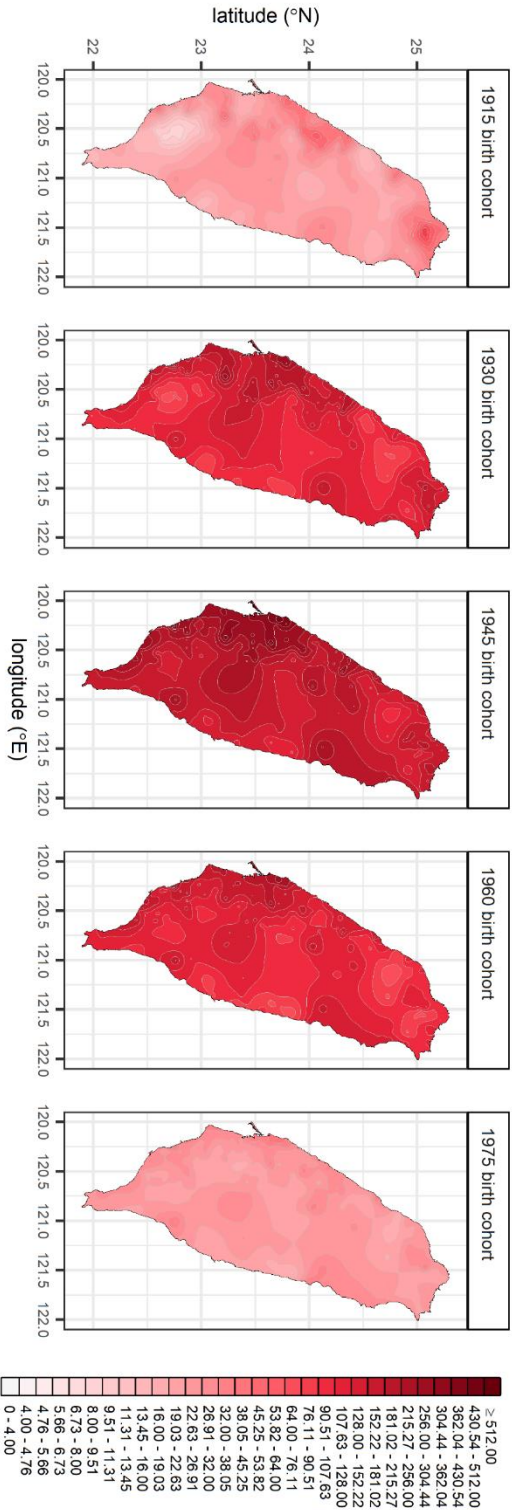
附錄十七、1913-1977 出生世代 75-79 歲族群的膀胱泌尿上皮癌及上泌尿道泌尿上皮癌發生率地理分布



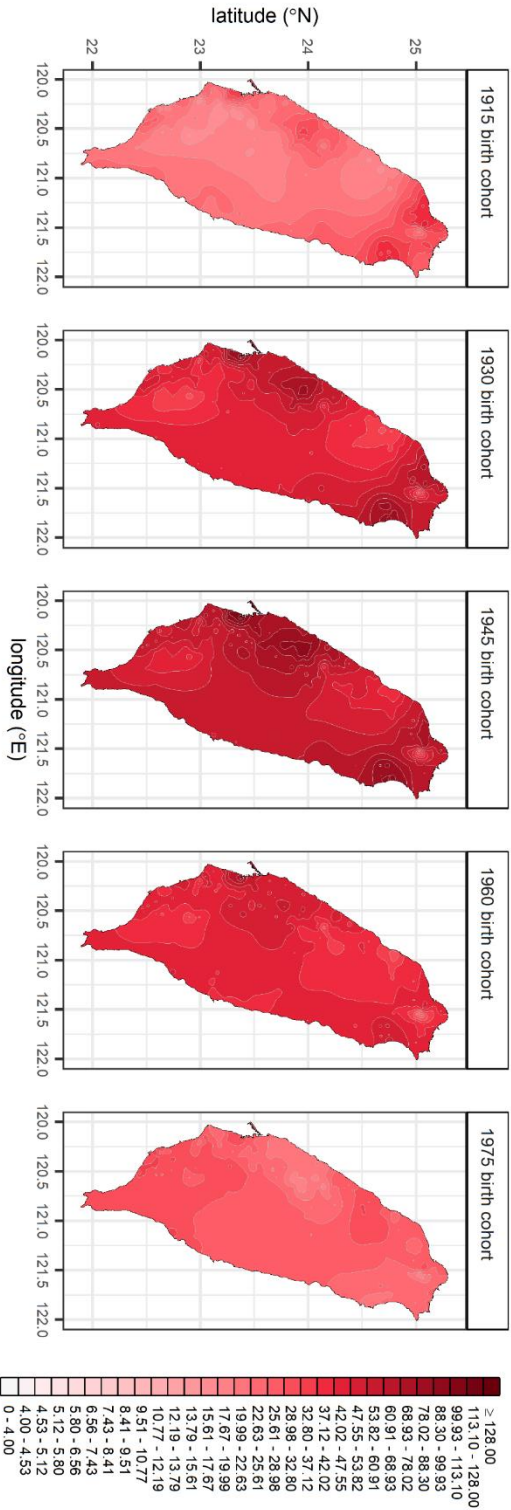
附錄十八、1913-1977 出生世代 60-64 歲男性的肝細胞癌及小細胞與鱗狀細胞肺癌
發生率地理分布



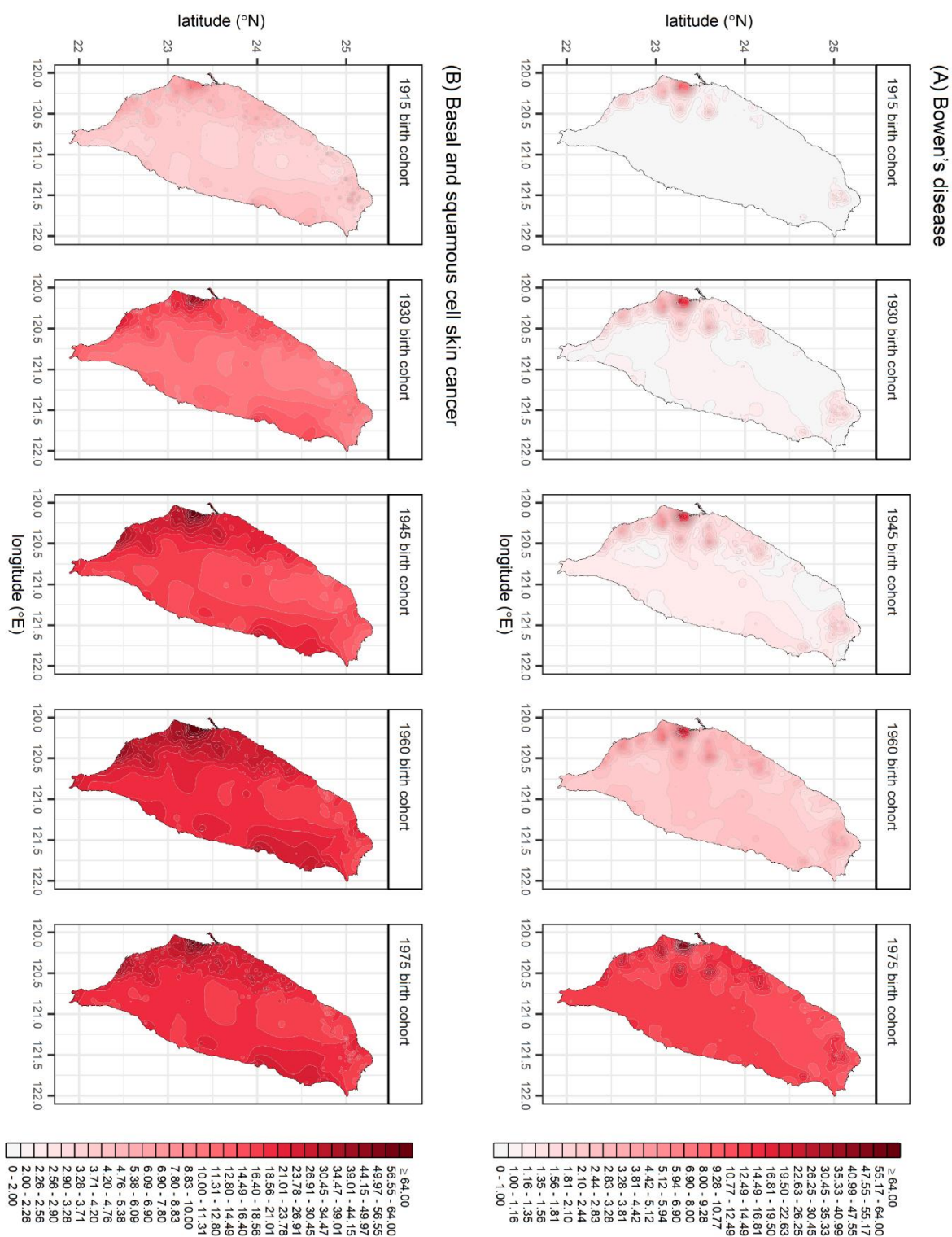
(A) Hepatocellular carcinoma



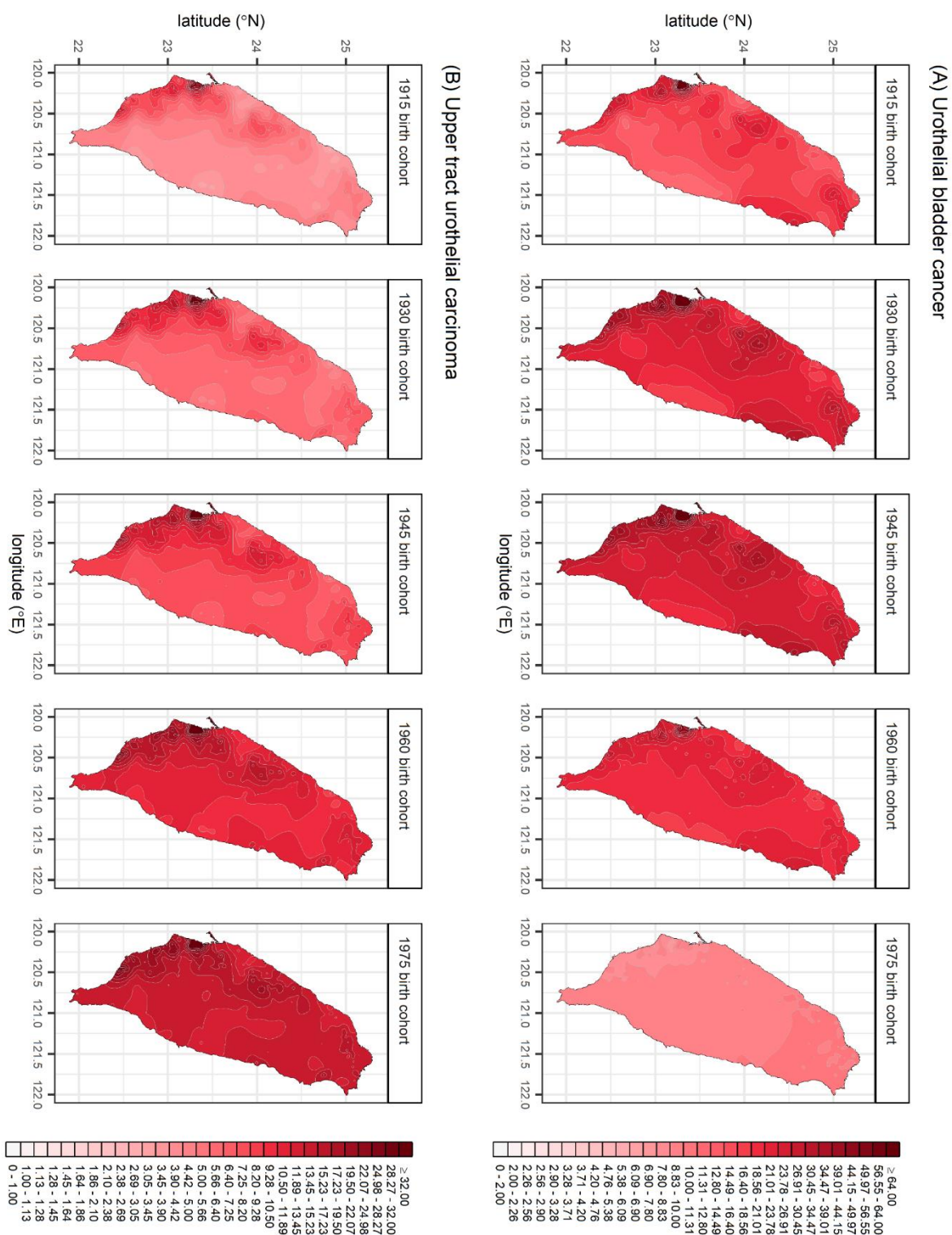
(B) Small and squamous cell lung cancer



附錄十九、1913-1977 出生世代 60-64 歲男性的鮑恩氏病及基底細胞與鱗狀細胞皮膚癌發生率地理分布



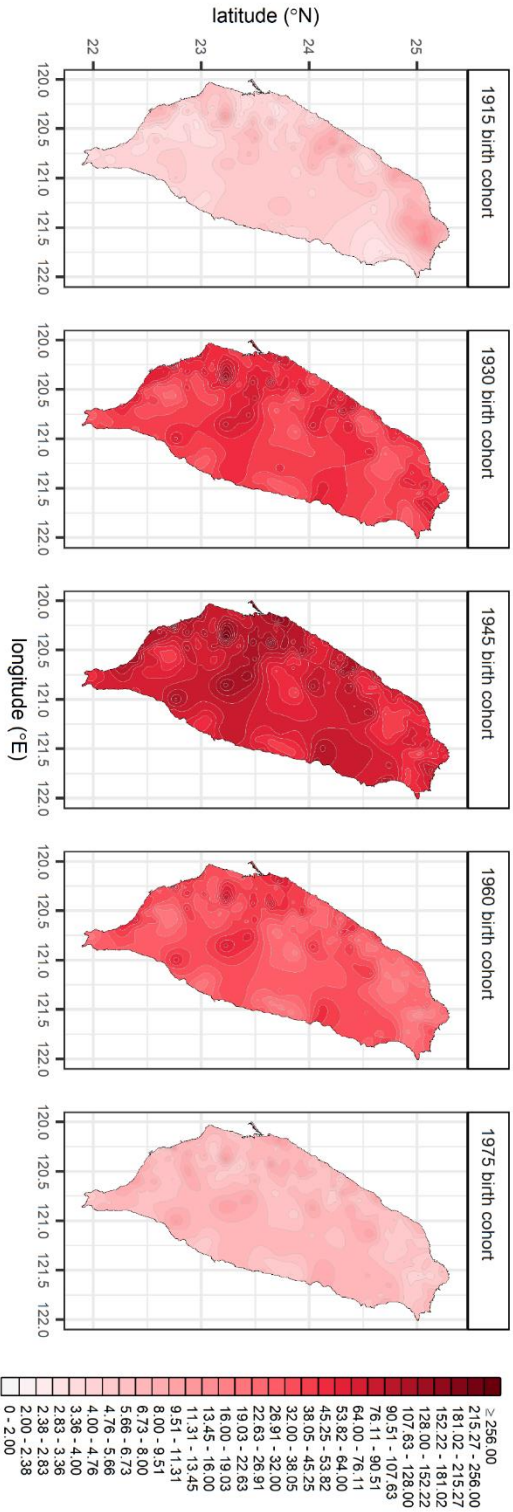
附錄二十、1913-1977 出生世代 60-64 歲男性的膀胱泌尿上皮癌及上泌尿道泌尿上皮癌發生率地理分布



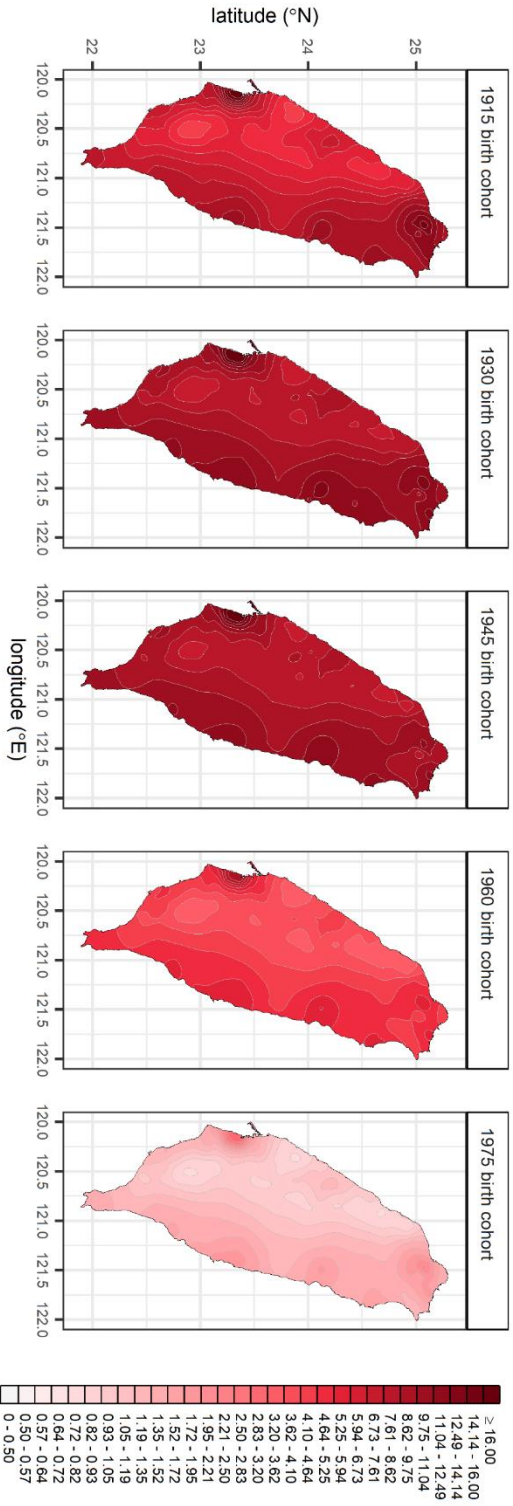
附錄二十一、1913-1977 出生世代 60-64 歲女性的肝細胞癌及小細胞與鱗狀細胞肺
癌發生率地理分布



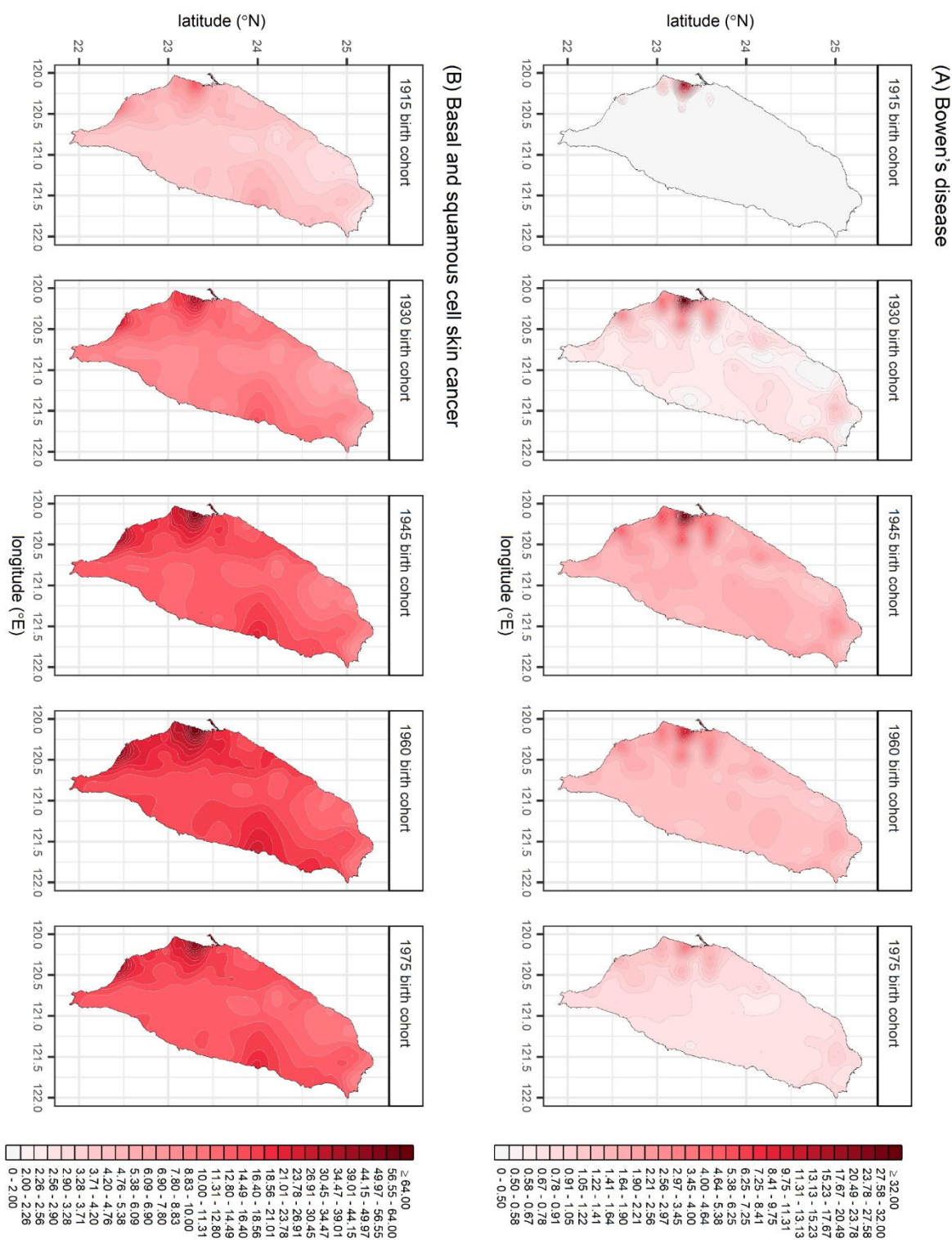
(A) Hepatocellular carcinoma



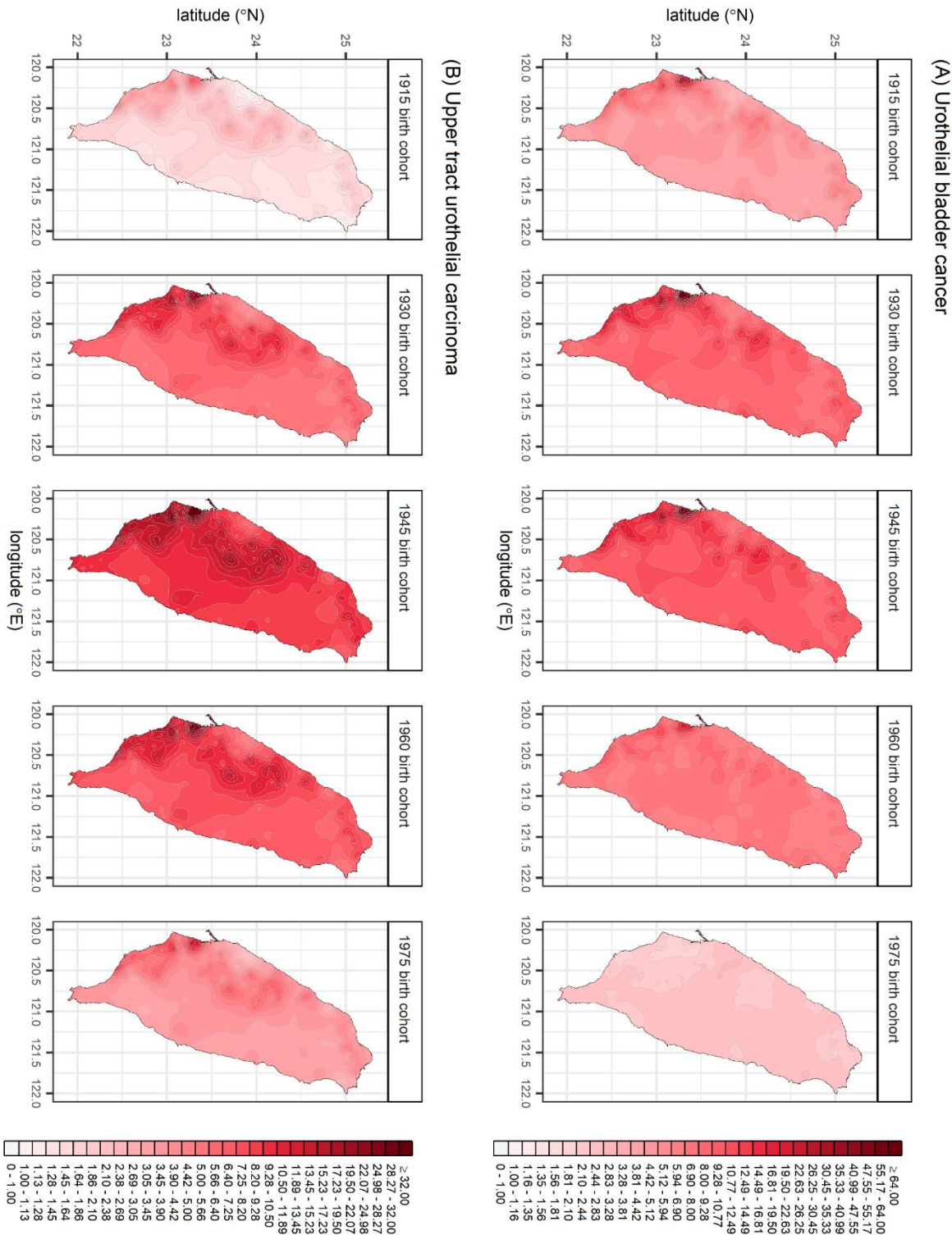
(B) Small and squamous cell lung cancer



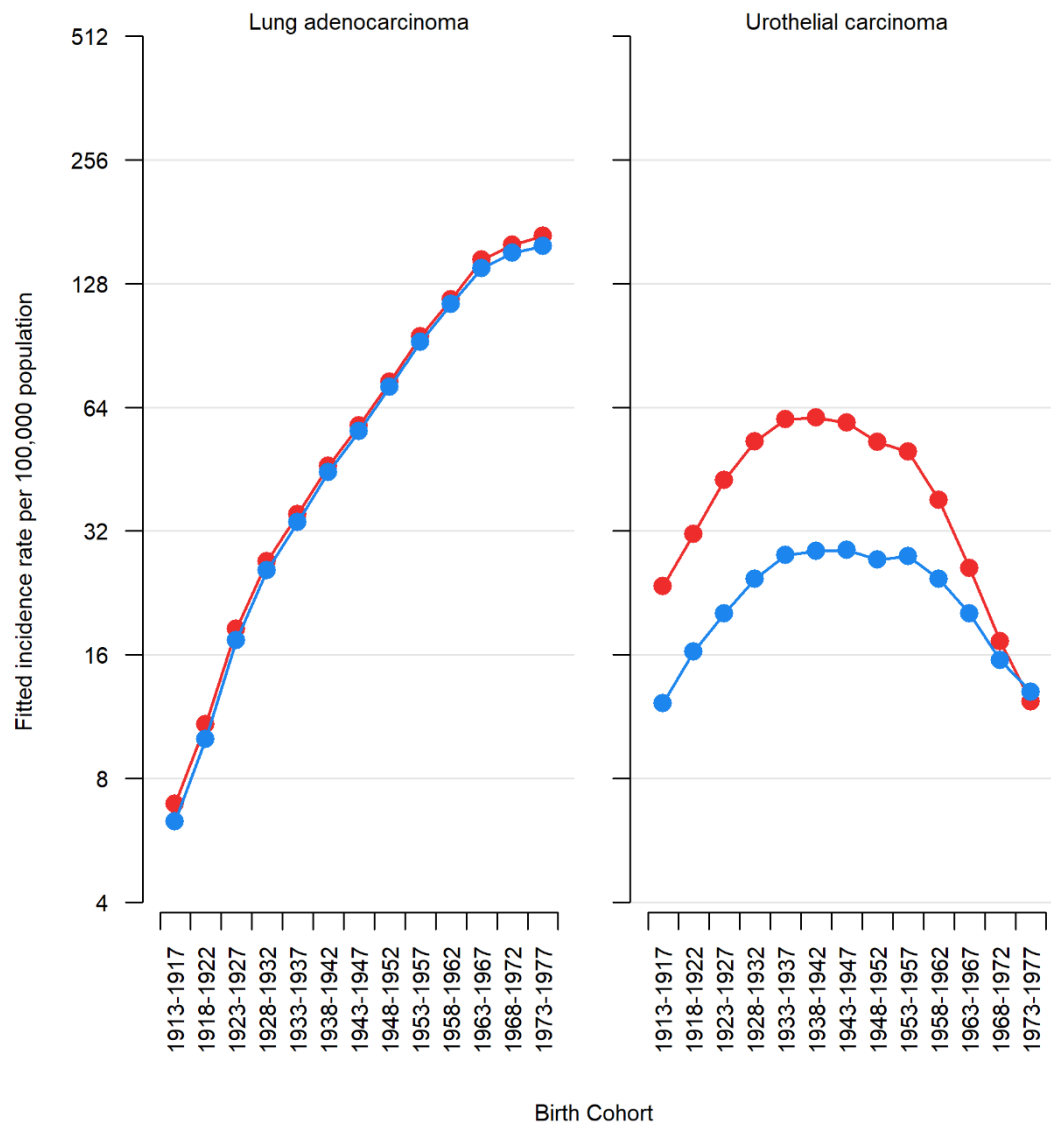
附錄二十二、1913-1977 出生世代 60-64 歲女性的鮑恩氏病及基底細胞與鱗狀細胞皮膚癌發生率地理分布



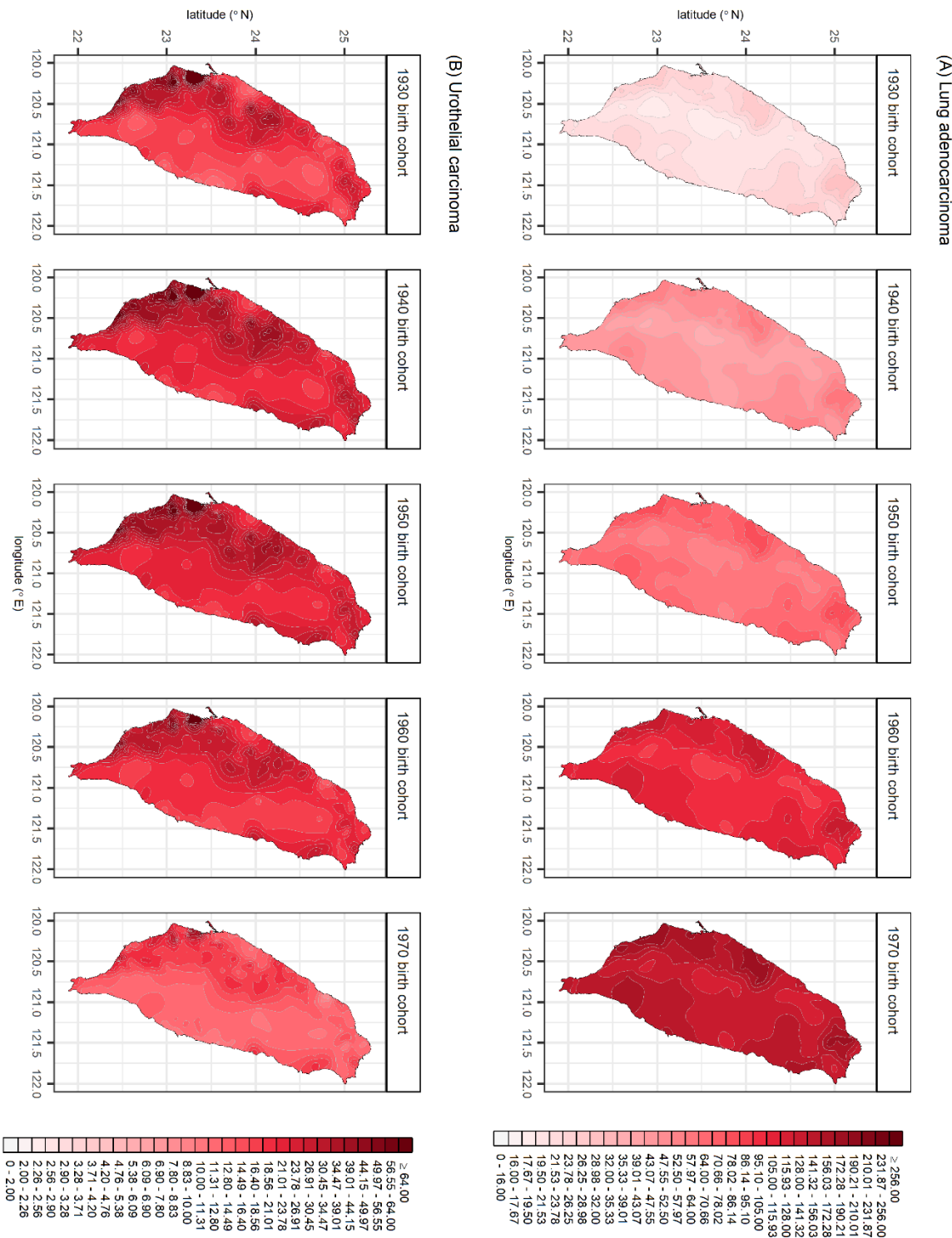
附錄二十三、1913-1977 出生世代 60-64 歲女性的膀胱泌尿上皮癌及上泌尿道泌尿
 上皮癌發生率地理分布



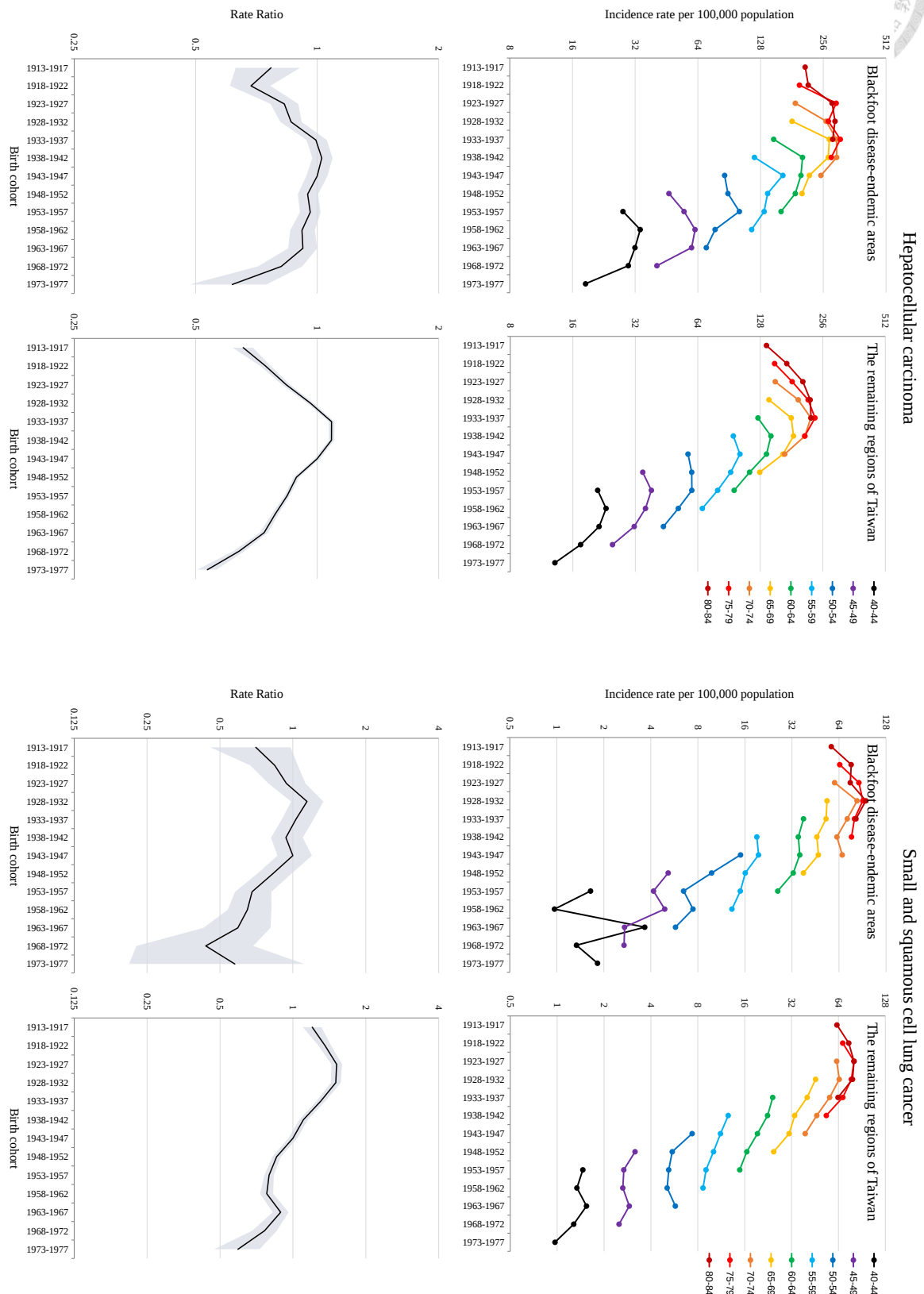
附錄二十六、1913-1977 出生世代 60-64 歲族群的肺腺癌與泌尿上皮癌發生率趨勢，紅色線為烏腳病流行地區，藍色線為非烏腳病流行地區



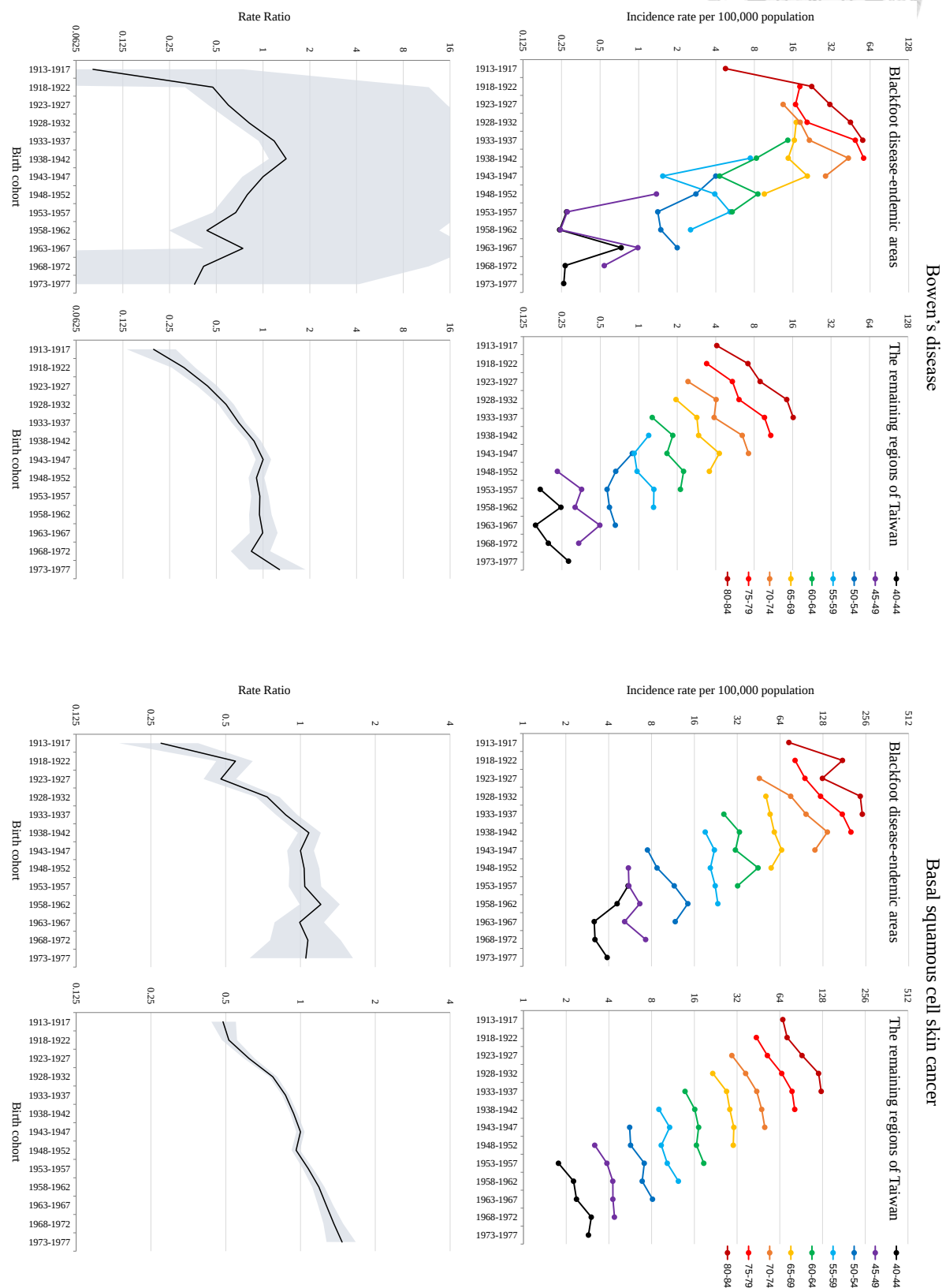
附錄二十七、1913-1977 出生世代 60-64 歲肺腺癌與泌尿上皮癌發生率地理分布



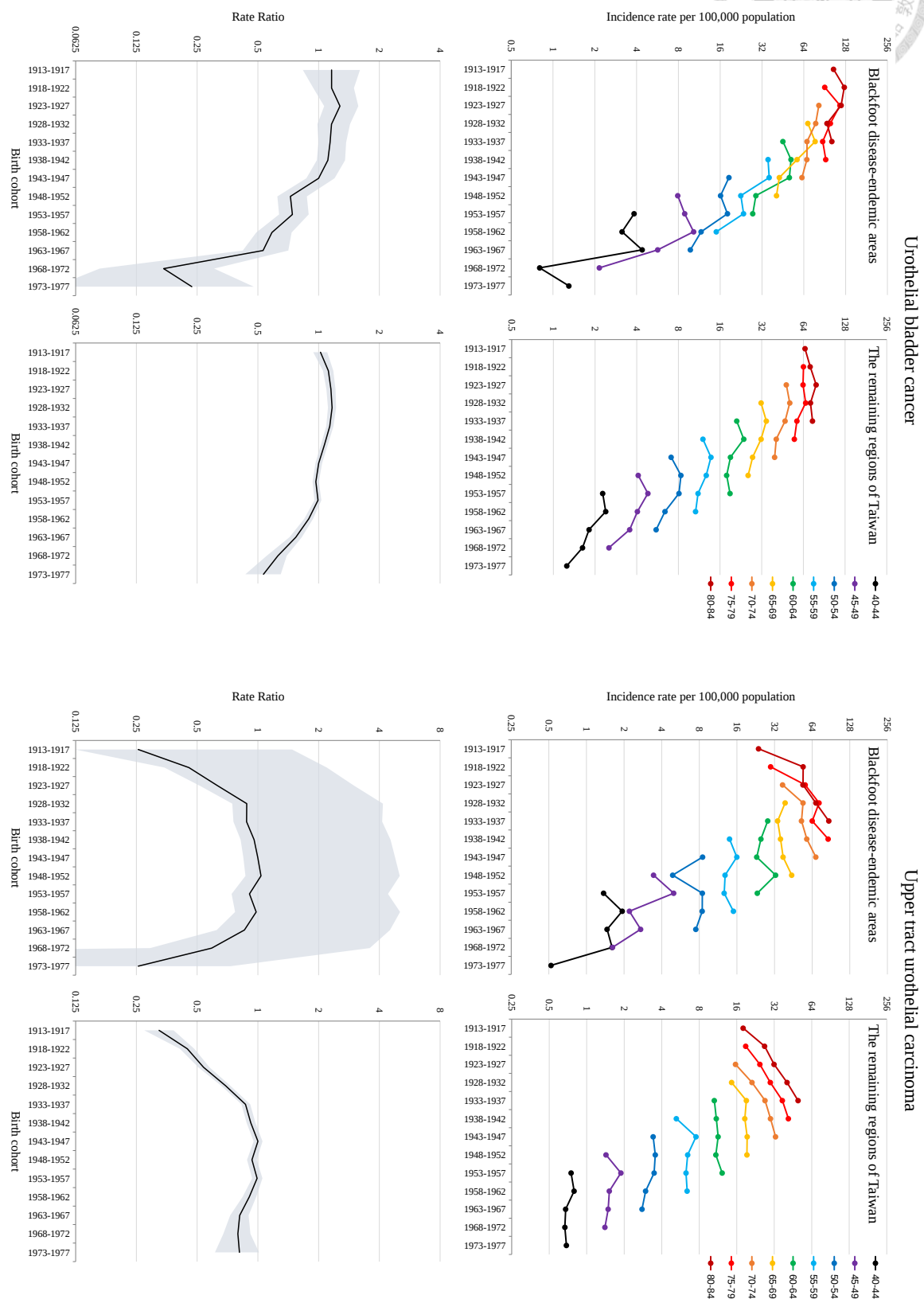
附錄二十八、肝細胞癌及小細胞與鱗狀細胞肺癌在烏腳病流行區與非流行區的年齡別發生率及世代效應趨勢，灰色陰影區為 95% 拔靴法信賴區間



附錄二十九、鮑恩氏病及基底細胞與鱗狀細胞皮膚癌在烏腳病流行區與非流行區的年齡別發生率及世代效應趨勢，灰色陰影區為 95% 拔靴法信賴區間



附錄三十、膀胱泌尿上皮癌及上泌尿道泌尿上皮癌在烏腳病流行區與非流行區的年齡別發生率及世代效應趨勢，灰色陰影區為 95% 拔靴法信賴區間



附錄三十一、電腦模擬之 R 程式碼



```
no.sim <- 1000
smape1 <- matrix(0,nrow=no.sim,ncol=8)
smape2 <- matrix(0,nrow=no.sim,ncol=4)

for(v in 1:no.sim){
  # Set parameters
  mu <- 5000 ; cv=0.5; sigma <- cv*mu
  brms.dat$POP <- rlnorm(nrow(brms.dat),
                        log(mu)-0.5*log((1+(1/cv)^2)),
                        sdlog = log(1+(1/cv)^2))

  p <- sample(1:ncol(gdis),2)
  d1 <- c(na.omit(gdis[p[1],]),na.omit(gdis[,p[1]])[-1])/1000
  wt1_int <- exp(-d1^2/200)
  wt1_coh <- exp(-d1^2/10000)

  # Intercept
  par1 <- log(50/100000) + wt1_int

  # Age
  par2 <- 0.40 + rnorm(length(par1),0,0.05)

  # Cohort
  par3 <- -0.10*wt1_coh

  par <- as.data.frame(cbind(par1,par2,par3,grid))
  colnames(par) <- c("Intercept","Age","Cohort","x1","x2")

  # log incidence rate
  input <- subset(brms.dat,brms.dat$TOWNCODE=="10010100")[,c("A","C")]
  input$A_dev <- input$A^2
  input$P_dev <- (input$A+input$C)^2
  input$C_dev <- input$C^2
  input <- as.matrix(cbind(1,input))
  mu <- apply(par[laa_grid,1:3],1,
              function(x) input %*% matrix(c(x,-0.05,0.00,-0.015), ncol=1))
```



```
# Simulate cases
brms.dat$EXPCASE <- exp(c(mu))*brms.dat$POP
brms.dat$CASE <- rpois(length(brms.dat$EXPCASE),brms.dat$EXPCASE)

# Poisson mixed-effects model
model <- glmer(CASE ~ A + C +
               A_dev1 + A_dev2 + A_dev3 + A_dev4 + A_dev5 + A_dev6 + A_dev7 +
               P_dev1 + P_dev2 + P_dev3 +
               C_dev1 + C_dev2 + C_dev3 + C_dev4 + C_dev5 + C_dev6 +
               C_dev7 + C_dev8 + C_dev9 + C_dev10 + C_dev11 +
               (1 + A + C|TOWNCODE),
               data = brms.dat, family = "poisson", offset = log(POP),
               control = glmerControl(optimizer = "nloptwrap",
                                       calc.derivs = FALSE))

# Original rates, True rates, and Fitted rates
brms.dat$RAW <- 100000*brms.dat$CASE/brms.dat$POP
brms.dat$TR <- 100000*brms.dat$EXPCASE/brms.dat$POP
brms.dat$FIT <- 100000*fitted(model)/brms.dat$POP

# Empirical Bayes
for(j in unique(brms.dat$A)){
  for(k in unique(brms.dat$C)){
    input <- subset(brms.dat, A==j & C==k)
    if(nrow(input)!=0){
      brms.dat[brms.dat$A==j & brms.dat$C==k,31] <- EBest(input$CASE, input$POP,
                                                           family="poisson")[,-1]*100000
      brms.dat[brms.dat$A==j & brms.dat$C==k,32] <- EBllocal(input$CASE, input$POP,
                                                            tw.nb)[,-1]*100000
    }
  }
}

colnames(brms.dat)[31:32] <- c("EB", "EL")
```



```
#####  
#--- Preparation for kriging ---#  
#####  
  
fix_ef <- fixef(model)  
ran_ef <- ranef(model)  
  
#####  
#--- Part II. Variogram & Kriging ---#  
#####  
ef <- ran_ef$TOWNCODE  
ef_var <- t(apply(cv <- attr(ran_ef$TOWNCODE,"postVar"),3,diag))  
external <- as.data.frame(cbind(ef, ef_var, as.character(rownames(ef))))  
colnames(external) <- c("ef_int", "ef_age_lin", "ef_coh_lin",  
                        "var_int", "var_age_lin", "var_coh_lin",  
                        "TOWNCODE")  
rownames(external) <- NULL  
  
mydata <- SpatialPointsDataFrame(laa, external)  
  
# Fit the variogram model  
emp_int <- autofitVariogram(ef_int ~ 1, mydata)  
emp_age_lin <- autofitVariogram(ef_age_lin ~ 1, mydata)  
emp_coh_lin <- autofitVariogram(ef_coh_lin ~ 1, mydata)  
  
fit_int <- emp_int$var_model  
fit_age_lin <- emp_age_lin$var_model  
fit_coh_lin <- emp_coh_lin$var_model  
  
# Exclude nugget effect  
nug_int <- fit_int$psill[1]  
nug_age_lin <- fit_age_lin$psill[1]  
nug_coh_lin <- fit_coh_lin$psill[1]  
  
fit_int <- fit_int[-1, ]  
fit_age_lin <- fit_age_lin[-1, ]  
fit_coh_lin <- fit_coh_lin[-1, ]
```



```
system.time({
  stkg_int <- krige2(ef_int ~ 1, mydata, SpatialPoints(grid), fit_int,
    errorVar = var_int+nug_int, computeVar = TRUE)
  stkg_age_lin <- krige2(ef_age_lin ~ 1, mydata, SpatialPoints(grid), fit_age_lin,
    errorVar = var_age_lin+nug_age_lin, computeVar = TRUE)
  stkg_coh_lin <- krige2(ef_coh_lin ~ 1, mydata, SpatialPoints(grid), fit_coh_lin,
    errorVar = var_coh_lin+nug_coh_lin, computeVar = TRUE)

  kg1_int <- krige2(ef_int ~ 1, mydata, SpatialPoints(grid), fit_int,
    errorVar = rep(0.0000001, nrow(mydata)), computeVar = TRUE)
  kg1_age_lin <- krige2(ef_age_lin ~ 1, mydata, SpatialPoints(grid), fit_age_lin,
    errorVar = rep(0.0000001, nrow(mydata)), computeVar = TRUE)
  kg1_coh_lin <- krige2(ef_coh_lin ~ 1, mydata, SpatialPoints(grid), fit_coh_lin,
    errorVar = rep(0.0000001, nrow(mydata)), computeVar = TRUE)

  kg2_int <- krige2(ef_int ~ 1, mydata, SpatialPoints(grid), fit_int,
    errorVar = rep(nug_int, nrow(mydata)), computeVar = TRUE)
  kg2_age_lin <- krige2(ef_age_lin ~ 1, mydata, SpatialPoints(grid), fit_age_lin,
    errorVar = rep(nug_age_lin, nrow(mydata)), computeVar = TRUE)
  kg2_coh_lin <- krige2(ef_coh_lin ~ 1, mydata, SpatialPoints(grid), fit_coh_lin,
    errorVar = rep(nug_coh_lin, nrow(mydata)), computeVar = TRUE)
})

#####
#--- Part II. Kernel density estimation ---#
#####

bw_int <- bw.gwr(ef_int ~ 1, data=mydata, kernel = "gaussian",
  longlat = TRUE, adaptive=TRUE)
kw_int <- apply(pointDistance(grid, laa, lonlat = TRUE)/1000,1,
  function(x) gwr.Gauss(x^2, bandwidth = bw_int))

bw_age_lin <- bw.gwr(ef_age_lin ~ 1, data=mydata, kernel = "gaussian",
  longlat = TRUE, adaptive=TRUE)
kw_age_lin <- apply(pointDistance(grid, laa, lonlat = TRUE)/1000,1,
  function(x) gwr.Gauss(x^2, bandwidth = bw_age_lin))

bw_coh_lin <- bw.gwr(ef_coh_lin ~ 1, data=mydata, kernel = "gaussian",
  longlat = TRUE, adaptive=TRUE)
kw_coh_lin <- apply(pointDistance(grid, laa, lonlat = TRUE)/1000,1,
  function(x) gwr.Gauss(x^2, bandwidth = bw_coh_lin))
```



```
X <- matrix(rep(1,n_laa),ncol=1)

kd_int <- apply(kw_int,2,function(y)
solve(t(X)%*%diag(y)%*%X)%*%(t(X)%*%diag(y)%*%mydata$ef_int))
kd_age_lin <- apply(kw_age_lin,2,function(y)
solve(t(X)%*%diag(y)%*%X)%*%(t(X)%*%diag(y)%*%mydata$ef_age_lin))
kd_coh_lin <- apply(kw_coh_lin,2,function(y)
solve(t(X)%*%diag(y)%*%X)%*%(t(X)%*%diag(y)%*%mydata$ef_coh_lin))

# True incidence
DM_A <- brms.dat[!duplicated(brms.dat$A),]
DM_A <- DM_A[,c(3,5:11)]
DM_C <- brms.dat[!duplicated(brms.dat$C),]
DM_C <- DM_C[order(DM_C[,4]),c(4,15:25)]

DM <- expand.grid(DM_A[,1],DM_C[,1])
DM <- merge(DM,DM_A,by.x="Var1",by.y="A")
DM <- merge(DM,DM_C,by.x="Var2",by.y="C")
DM <- DM[order(DM$Var2),]
DM <- cbind(DM[,2],DM[,1],DM[,c(1,2)])
colnames(DM)[1:2] <- c("A","C")

true_par <- as.matrix(cbind(par[,1:3],-0.05,0.00,-0.015))
temp <- apply(true_par,1,function(x) as.matrix(cbind(1,DM[,1:2],
                                                    DM[,1]^2,
                                                    0,
                                                    DM[,2]^2))%*%matrix(x,ncol=1))

par$IR1 <- apply(temp[which(c(DM$A+DM$C) %in% c(-2:2)),],2,
                function(x) mean(exp(x)*100000))
par$IR2 <- apply(temp[-which(c(DM$A+DM$C) %in% c(-2:2)),],2,
                function(x) mean(exp(x)*100000))

# Original rates
par$RAW <- merge(pts,aggregate(brms.dat,list(brms.dat$TOWNCODE),mean),
                by.x="region",by.y="Group.1"),["RAW"]
smape1[v,1] <- mean(abs(par$RAW-par$IR1)/(par$RAW+par$IR1))*100
```



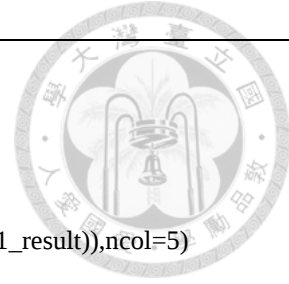
```
# Empirical Bayes
par$EB <- merge(pts,aggregate(brms.dat,list(brms.dat$TOWNCODE),
                                function(x) mean(x, na.rm=T)),
                by.x="region",by.y="Group.1"),"EB"]
smape1[v,2] <- mean(abs(par$EB-par$IR1)/(par$EB+par$IR1))*100

par$EL <- merge(pts,aggregate(brms.dat,list(brms.dat$TOWNCODE),
                                function(x) mean(x, na.rm=T)),
                by.x="region",by.y="Group.1"),"EL"]
smape1[v,3] <- mean(abs(par$EL-par$IR1)/(par$EL+par$IR1))*100

# Poisson mixed-effects incidence
par$FIT <- merge(pts,aggregate(brms.dat,list(brms.dat$TOWNCODE),mean),
                by.x="region",by.y="Group.1"),"FIT"]
smape1[v,4] <- mean(abs(par$FIT-par$IR1)/(par$FIT+par$IR1))*100

# Kd incidence
kd_result <- as.data.frame(cbind(kd_int,kd_age_lin,kd_coh_lin,grid))
kd_result <- kd_result + matrix(rep(c(fix_ef[1:3],0,0),
                                    each=nrow(kd_result)),ncol=5)
colnames(kd_result) <- c("Intercept","Age","Cohort","x","y")

kd_par <- as.matrix(cbind(kd_result[,1:3],
                          matrix(rep(fix_ef[-c(1:3,11:13)],each=nrow(kd_result)),
                                  ncol=length(fix_ef[-c(1:3,11:13)]))))
temp <- apply(kd_par,1,function(x) as.matrix(cbind(1,DM))%*%matrix(x,ncol=1))
kd_result$IR1 <- apply(temp[which(c(DM$A+DM$C) %in% c(-2:2)),],2,
                      function(x) mean(exp(x)*100000))
kd_result$IR2 <- apply(temp[-which(c(DM$A+DM$C) %in% c(-2:2)),],2,
                      function(x) mean(exp(x)*100000))
smape1[v,5] <- mean(abs(kd_result$IR1 - par$IR1)/(kd_result$IR1 + par$IR1))*100
smape2[v,1] <- mean(abs(kd_result$IR2 - par$IR2)/(kd_result$IR2 + par$IR2))*100
```



```
# Kg1
kg1_result <- as.data.frame(cbind(kg1_int,kg1_age_lin,kg1_coh_lin))
kg1_result <- kg1_result[,c(1,3,5,7,8)] + matrix(rep(c(fix_ef[1:3],0,0),
                                                    each=nrow(kg1_result)),ncol=5)

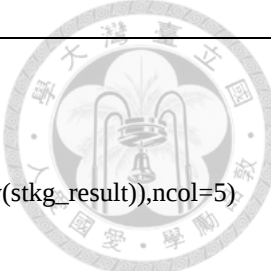
colnames(kg1_result) <- c("Intercept","Age","Cohort","x","y")
kg_par <- as.matrix(cbind(kg1_result[,1:3],
                          matrix(rep(fix_ef[-c(1:3,11:13)],each=nrow(kg1_result)),
                                ncol=length(fix_ef[-c(1:3,11:13)]))))
temp <- apply(kg_par,1,function(x) as.matrix(cbind(1,DM))%*%matrix(x,ncol=1))
kg1_result$IR1 <- apply(temp[which(c(DM$A+DM$C) %in% c(-2:2)),],2,
                        function(x) mean(exp(x)*100000))
kg1_result$IR2 <- apply(temp[-which(c(DM$A+DM$C) %in% c(-2:2)),],2,
                        function(x) mean(exp(x)*100000))

smape1[v,6] <- mean(abs(kg1_result$IR1 - par$IR1)/(kg1_result$IR1 + par$IR1))*100
smape2[v,2] <- mean(abs(kg1_result$IR2 - par$IR2)/(kg1_result$IR2 + par$IR2))*100

# Kg2
kg2_result <- as.data.frame(cbind(kg2_int,kg2_age_lin,kg2_coh_lin))
kg2_result <- kg2_result[,c(1,3,5,7,8)] + matrix(rep(c(fix_ef[1:3],0,0),
                                                    each=nrow(kg2_result)),ncol=5)

colnames(kg2_result) <- c("Intercept","Age","Cohort","x","y")
kg_par <- as.matrix(cbind(kg2_result[,1:3],
                          matrix(rep(fix_ef[-c(1:3,11:13)],each=nrow(kg2_result)),
                                ncol=length(fix_ef[-c(1:3,11:13)]))))
temp <- apply(kg_par,1,function(x) as.matrix(cbind(1,DM))%*%matrix(x,ncol=1))
kg2_result$IR1 <- apply(temp[which(c(DM$A+DM$C) %in% c(-2:2)),],2,
                        function(x) mean(exp(x)*100000))
kg2_result$IR2 <- apply(temp[-which(c(DM$A+DM$C) %in% c(-2:2)),],2,
                        function(x) mean(exp(x)*100000))

smape1[v,7] <- mean(abs(kg2_result$IR1 - par$IR1)/(kg2_result$IR1 + par$IR1))*100
smape2[v,3] <- mean(abs(kg2_result$IR2 - par$IR2)/(kg2_result$IR2 + par$IR2))*100
```



```

# Stkg incidence
stkg_result <- as.data.frame(cbind(stkg_int,stkg_age_lin,stkg_coh_lin))
stkg_result <- stkg_result[,c(1,3,5,7,8)] + matrix(rep(c(fix_ef[1:3],0,0),
                                                    each=nrow(stkg_result)),ncol=5)

colnames(stkg_result) <- c("Intercept","Age","Cohort","x","y")
stkg_par <- as.matrix(cbind(stkg_result[,1:3],
                            matrix(rep(fix_ef[-c(1:3,11:13)],each=nrow(stkg_result)),
                                    ncol=length(fix_ef[-c(1:3,11:13)]))))
temp <- apply(stkg_par,1,function(x) as.matrix(cbind(1,DM))%*%matrix(x,ncol=1))
stkg_result$IR1 <- apply(temp[which(c(DM$A+DM$C) %in% c(-2:2)),],2,
                        function(x) mean(exp(x)*100000))
stkg_result$IR2 <- apply(temp[-which(c(DM$A+DM$C) %in% c(-2:2)),],2,
                        function(x) mean(exp(x)*100000))

smape1[v,8] <- mean(abs(stkg_result$IR1 - par$IR1)/(stkg_result$IR1 + par$IR1))*100
smape2[v,4] <- mean(abs(stkg_result$IR2 - par$IR2)/(stkg_result$IR2 + par$IR2))*100

cat(v)
}

apply(smape1,2,function(x) mean(x, na.rm=T))
apply(smape2,2,function(x) mean(x, na.rm=T))

```