

比較靜態與動態人口資料應用於新冠肺炎熱區之預測能力

周敬棋¹ 林先和² 陳正誠³

陳潤秋^{4,5} 詹大千^{6,7,*}

目標：本研究探討運用戶籍人口資料與電信人流資料進行臺北市、新北市新冠肺炎熱區預警的建構，比較其預測準確度與應用上的限制，期提供未來疫情防治參考。**方法：**研究的時空範圍為2021年5月之雙北市村里，分別使用靜態、動態兩種資料建構人流網絡。前者以重力模型模擬人流，後者以電信數據測量人流變化，透過網絡空間的結構等位特性估計人流足跡相似度，進而計算村里感染風險，並以ROC曲線、羅吉斯迴歸檢驗在不同確診門檻值下的模型表現。**結果：**在研究區間內，電信人流之平均曲線下面積較高（AUC為0.75，重力模型之AUC為0.69），且傾向預測距離疫情爆發中心較遠的村里，適合用於預測空間上的傳染趨勢。羅吉斯迴歸的結果也顯示，使用電信人流資料所計算的未來一周確診人數是否高於門檻值的平均風險之勝算比為1.45；重力模型之平均勝算比為1.10。**結論：**在評估地區的感染風險時，除了參考疫情調查與累積病例數外，人流網絡的資訊可以協助辨識潛在的高風險區域並進行及早預警。（台灣衛誌 2022；41(6)：611-626）

關鍵詞：新冠肺炎、手機信令、人流、預警模式

前 言

新興傳染病的威脅逐年增加[1]，面對新興傳染病的流行如新冠肺炎的全球大流

行，在疫情初期並沒有有效的疫苗可以提供預防，也沒有有效的抗病毒藥物可供治療，因此除了即早篩檢與診斷病人外，非藥物性的介入措施，如降低人們彼此間的接觸[2]與做好自我健康防護的措施如戴口罩、酒精洗手[2]都是疫情初期延緩傳染速度的重要方法，新冠肺炎的威脅自2020年開始，台灣主要透過嚴格的邊境管制、個體防疫（早期的個案偵測、疫調追蹤、隔離措施）與團體防疫措施（社交距離、戴口罩）的介入進行疫情的控制[3]，另外在民眾的配合度也非常高的情況下，讓疫情可以保持在低點，僅有散發的個案發生，時至2021年中，因為境外移入個案所引發的本土疫情，才讓台灣真的經歷歐美國家所遭遇的大流行事件，而該波疫情的流行區域主要在北台灣的台北市、新北市，確診病例數中約有80%位在此

¹ 國立臺灣大學地理環境資源學系

² 國立臺灣大學公共衛生學院流行病學與預防醫學研究所

³ 臺北市政府衛生局

⁴ 新北市政府衛生局

⁵ 國立陽明交通大學生物醫學影像暨放射科學系

⁶ 中央研究院人文社會科學研究中心

⁷ 國立陽明交通大學醫學院公共衛生研究所

* 通訊作者：詹大千

地址：台北市南港區研究院路二段128號

E-mail: tachien@gate.sinica.edu.tw

投稿日期：2022年9月12日

接受日期：2022年12月12日

DOI:10.6288/TJPH.202212_41(6).111074



區[4]，而中央流行疫情指揮中心也史無前例的宣佈了三級警戒，包含自2021年5月15日起的雙北三級警戒與2021年5月19日起的全國三級，對比國外嚴格的人流限制，我國的管制措施相對彈性，主要包含關閉休閒娛樂場所、停止進香團與繞進活動、避免不必要的移動、活動或聚會、營業場所的人流管制、鼓勵企業啟用異地或遠距辦公、彈性上班時間、學校停課等措施，大眾運輸則是加強清消並沒有取消與管制，三級警戒也順利讓疫情在兩個多月期間就獲得有效控制，其中兩項重要的因素包含了人流的管制策略與社區加強篩檢的措施[4]，但也因為此波疫情的經驗，可以了解當地方衛生單位面對社區出現大量無症狀或輕症的感染患者時，將面臨基層公衛人力不足，難以針對所有的感染個案或疑似個案進行疫情調查，導致防疫決策資訊不及的問題，倘若能透過動態人流資訊來輔助疫情趨勢的預測，將可以提前進行疫情防治的部署。

在疫情快速升溫時，如何將有限的防疫資源進行早期有效的分配是非常重要的議題，在政策實務上，常會根據流行病學與空間區位的特性進行分析，例如以行政區的人口數、累計確診數，或是與感染區的空間鄰近性等資訊，劃列高風險感染區以加強防疫量能的投入，如新北市在2021年6月初，就劃列199個重熱里作為防疫工作的重點區域，然而這邊所用來決策的資訊大多為靜態的人流資訊，一方面會有戶籍在人不在的問題[5]，另一方面並不知道村里之間往來的動態情況，可能存在感染地與活動地、居住地之間往返的風險問題，因此若能掌握人的流動，理論上可以將社區中動態的感染風險掌握的更清楚。

在實務上可將人流數據區分成三種資料來源，靜態人口統計資料、普查資料與電信信令數據資料三種，可藉由不同的人口資料進行人流的建構。靜態人口統計資料會基於一些人口吸引的假設，進行區域間人流關係的建構，使用的模型如重力模型（gravity model）[6]或是輻射模型（radiation model）

[7]；普查資料則是以大規模的面訪調查民眾的居住地與上班、上學的行政區，以建構不同地區之間日間與夜間流動的關係[8]；電信信令數據可根據電信信令訊號，定義人群在地區之間的停留時間與旅次關係。隨著科技革新與計算效能的進步，電信信令數據的應用日漸蓬勃，由於這種類型的資料可捕捉人流的動態變化，國外學者也陸續將其應用於觀測傳染病的傳播趨勢[8,9]。

電信數據用於傳染病預測的研究在近年蓬勃發展。在流感症狀相關研究方面，曾有學者在義大利根據個體的電信數據移動軌跡，預測每一個受試者在未來是否會出現流感的病徵。在行為研究方面，有文獻在德國地區以電信數據建構人流在地理空間中的網絡關係，討論封鎖政策對於網絡結構的改變將會如何影響COVID-19的傳播途徑，並觀測政策前後的人流限制效果[10]。在傳染病研究方面，這幾年因新冠肺炎的疫情大幅推進電信人流數據使用於傳染病預測的相關應用，例如在瑞士與冰島的研究中，學者以電信通信紀錄（CDR）分別監測2020年COVID-19疫情與2009年新型流感疫情下的人流變化與疫情的傳播趨勢[9,11]。

一般而言，由於動態人流資料的取得較為困難、價格也較為昂貴，因此在資料較難以取得的地區中，人流資料大都以靜態人口模擬的方式取得。在巴基斯坦，曾對比靜態人口資料與電信數據對於登革熱疫情預測表現的差異，結論指出電信數據有較好的預測力、且能更精準地估計第一波疫情抵達的時間[12]。然而，如上述討論人流資料型態對於熱區預測表現的文獻少見，又動態人口資料與靜態人口資料的時空解析度有非常大的差異，因此不同人流資料類型對於傳染病熱區預測表現的差異仍需更多的討論。

本研究將透過2021年5月雙北市的新冠肺炎疫情的實證資料，搭配靜態戶籍資料與手機動態人流資料，評估使用不同人流資料對於疫情空間分布預測的表現與其可能的優缺點。

材料與方法

一、研究資料：本研究所使用的研究資料主要有以下三類：

(一) 電信信令數據資料

電信信令數據資料作為建構人口移動網絡使用，本研究以2021年5月10日、5月16日與5月20日之遠傳電信公司的信令數據資料估計人流，該資料為以村里為單位的加總數據，並無法知道任何特定個人的移動紀錄，依據國家通訊委員會2021年第二季的行動通訊用戶數統計顯示該公司的市佔率為23.9% (https://www.ncc.gov.tw/chinese/files/22010/3773_47007_220104_1.pdf)。計算電信使用者所停留的方式為電信公司透過基地台的三角定位方式所估計出來的訊號服務網格，當用戶在該網格中停留的時間大於所定義的時間就會被計算到，而由於網格可能會被不同的村里所交集到，此處是使用網格內面積佔比最大之村里進行計算，不同的網格加總後則為該村里、該時段的停留人數，並將用戶過去一個月夜間22:00至03:00停留時間總和最長之村里認為用戶的居住村里；用戶在各村里的停留時間若超過十五分鐘將會被紀錄，以上這些資訊的處理過程都是由電信公司進行資料的前處理，研究人員所取得都是加總後的統計數據，並無法識別任何的特定個人。另外，電信服務使用者將依年齡分組為18-21歲、22-29歲、30-59歲與60歲以上四個族群進行分別的加總計算。此外，電信業者為保護個人隱私，村里的各年齡層電信服務使用者人數少於五人時將無法取得該村里、該時段的電信人流數據，故電信人口少於五筆的資料將不納入分析。

(二) 靜態人口資料

本研究亦使用靜態人口資料進行分析，使用重力模型 (gravity model) 模擬村里間的人流量作為參考基準，並與電信信令數據進行比較。人口資料使用的是內政部社會經濟資料庫中的2021年2月所統計之各村里單齡人口數資料 (<https://segis.moi.gov.tw/>

STAT/Web/Portal/STAT_PortalHome.aspx)，並依照電信數據相同的年齡層進行分組。

(三) 村里確診數資料

使用臺北市與新北市衛生局提供之每日各村里 (n=1,488) 確診數統計，時間範圍介於2021年5月8日至5月27日間，因涉及確診者隱私，村里確診數未以年齡進行分組統計。

(四) 研究倫理

本研究的研究設計經國立臺灣大學行為與社會科學研究倫理委員會核可 (核可編號：202007HM008)。

二、研究方法

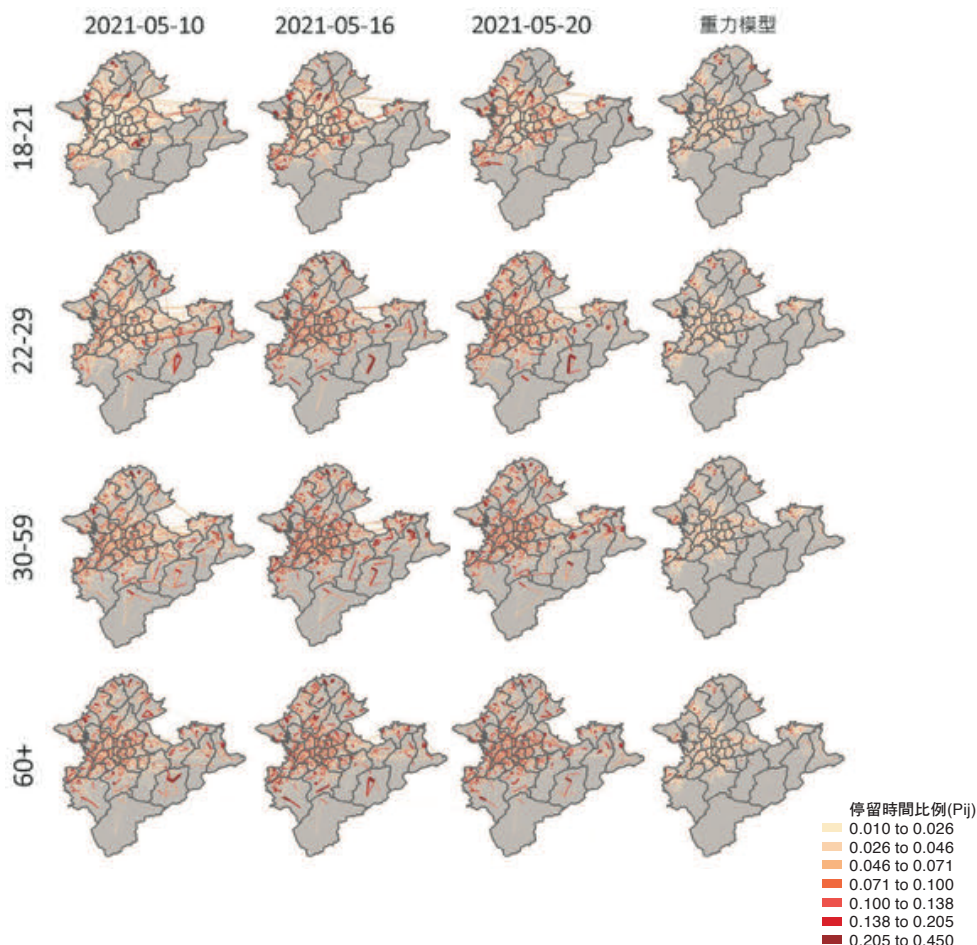
(一) 建構地理空間網絡

使用人流資料時，可透過居住村里與停留村里的關係，以村里作為節點 (nodes)，而連結 (edges) 的方向定義為居住村里 (起點，source) 指向停留村里 (終點，target)，並以人流量大小作為連結權重 (weight)，建構以年齡分組的雙北市人流有向網絡。在實務上，人流量愈大也表示兩區域的人口接觸機率較高，因此以表示區域間的人流量大小。

在電信數據中，以居住村里人口在其他村里的正規化停留時間比例 (*Normalized P_{ij}*) 衡量人流大小，使每個起點村里 (source，定義為i) 在所有停留村里 (定義為j) 的停留時間比例之和為1。

$$\text{Normalized } P_{ij} = \frac{P_{ij}}{\sum_{j \in J} P_{ij}}$$

另外，使用重力模型與戶籍人口資料模擬人口流動。在傳統文獻中，將靜態資料以重力模型的方式模擬人流大都採用 $P_{ij}^{\text{Gravity}} = \frac{POP_i \times POP_j}{d_{ij}^\sigma}$ 之模式估計城市間的移動量或流通量。根據Krings等人在2009年進行的研究結果[13]顯示，可利用此模型估計城市內部的通訊網絡、量化兩地人口互相吸引的程度，且經測量後發現參數為2。此外，Jung等人在南韓進行的城際間人流移動研究[14]也使用了此模型進行人口移動估



圖一 以電信信令資料、重力模型建構分齡人口流動網絡

註：由左至右三個欄位，分別代表2021/5/10、2021/5/16、2021/5/20當天的分齡人口流動網絡，第四個欄位為使用靜態村里戶籍人口數所建構的分齡人口流動網絡。

計，也得出的 $\sigma = 2$ 的結果。故在重力模型方面假定村里間之人流多寡與兩村里之人口數乘積成正比、兩村里之距離平方值成反比：

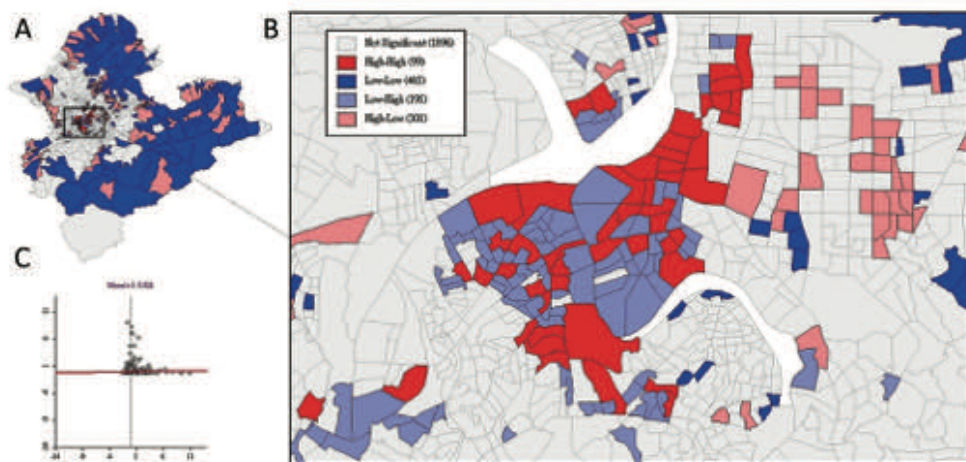
$$P_{ij}^{Gravity} = \frac{POP_i \times POP_j}{d_{ij}^2} \text{ Normalized } P_{ij}^{Gravity}$$

$$= \frac{P_{ij}^{Gravity}}{\sum_{j \in J} P_{ij}^{Gravity}}$$

其中 $P_{ij}^{Gravity}$ 表示人流量大小，被視為村里i之居民停留於村里j之機率； POP_i 、 POP_j 分別表示村里i、村里j之人口； d_{ij} 表示兩村里地理中心點之直線距離，最後將該數值進行正規化處理，使每個起點村里（source）在所有村里的停留時間比例之和為1。

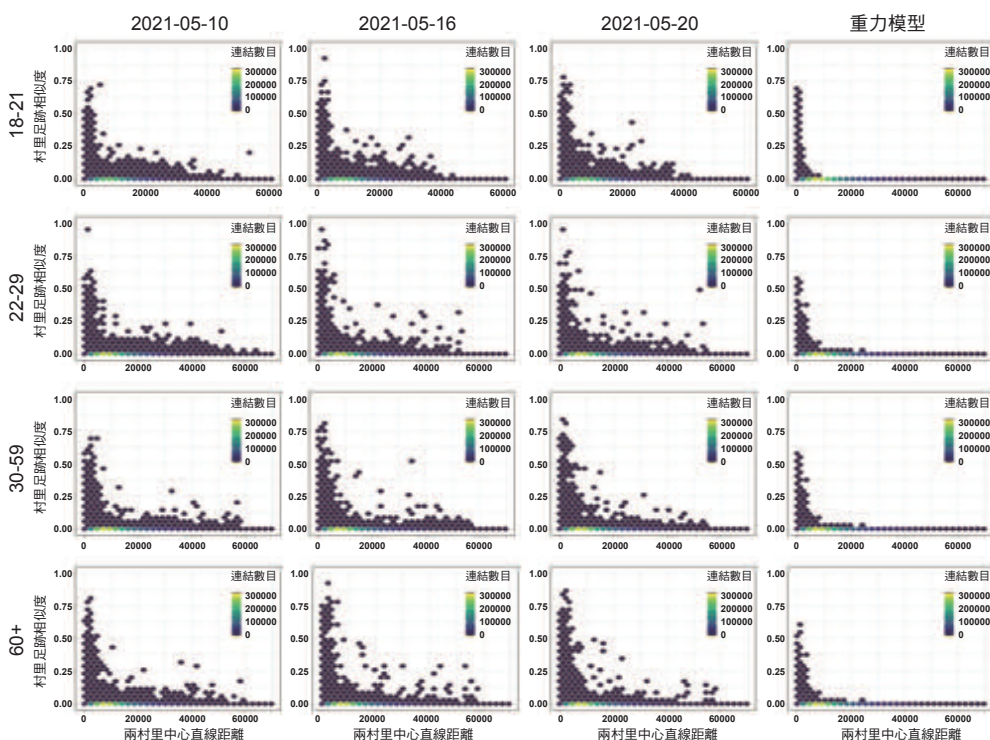
(二) 網絡中心性與確診數熱區分析

在傳染病數理模型中，接觸行為是造成疾病傳染的主要因素，我們可以透過計算村里間的往返停留關係描繪傳染病模型中的接觸過程。在人流流入量較大的地區，如商業區、重要交通轉乘站等所在村里，成為了雙北市足跡最為密集之處。為了衡量地區作為終點村里所吸納的人流量，本研究計算各村里的加權度中心性（weighted in-degree）觀測空間中的人流熱區，並透過雙變量空間自相關分析（Bivariate Moran's I）檢視人流所集中之熱區與確診數的熱區是否呈現空間自相關。



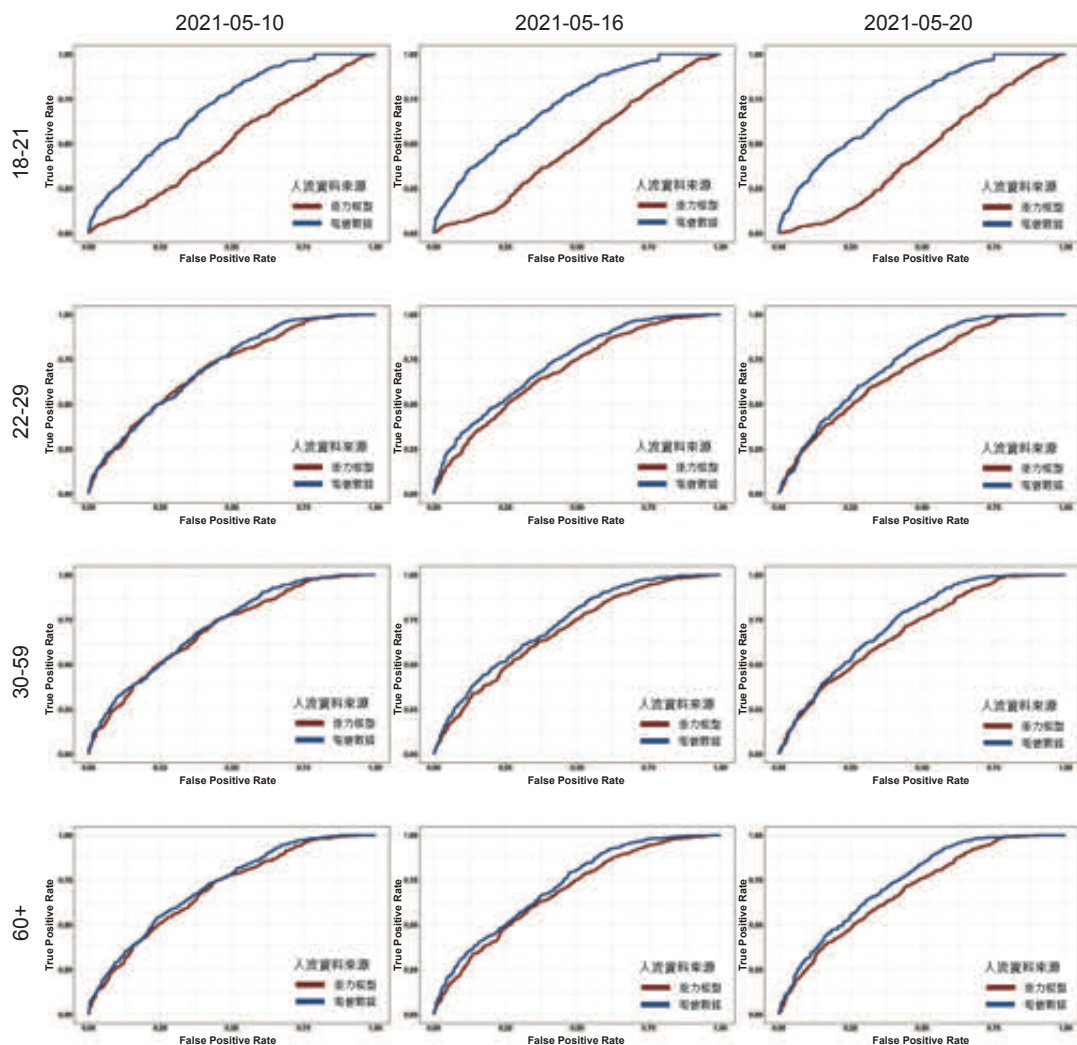
圖二 研究區間內新冠肺炎確診數與人流中心性之相關性

註：A：以雙變量局部空間自相關（Bivariate Local Moran's I）檢定確診數（y）與人流中心性（x）之相關性；
B：為A圖在台北市萬華區附近的局部放大圖；C：Moran's I 散佈圖。



圖三 村里足跡相似度（similarity）與村里中心點之歐幾里得距離（Euclidean distance）之六邊形分箱圖（hexplots）

註：X軸表示兩村里中心點直線距離，Y軸表示兩村里之足跡相似度。由左至右三個欄位，分別代表2021/5/10、2021/5/16、2021/5/20的關係圖，第四個欄位為使用重力模型所建構的關係圖。



圖四 兩類人流資料來源（電信信令數據、重力模型模擬）於各期分齡之熱區預測ROC曲線——當確診數門檻大於1人（ $T=1$ ）

註：紅色表示重力模型結果，藍色表示電信信令資料結果。

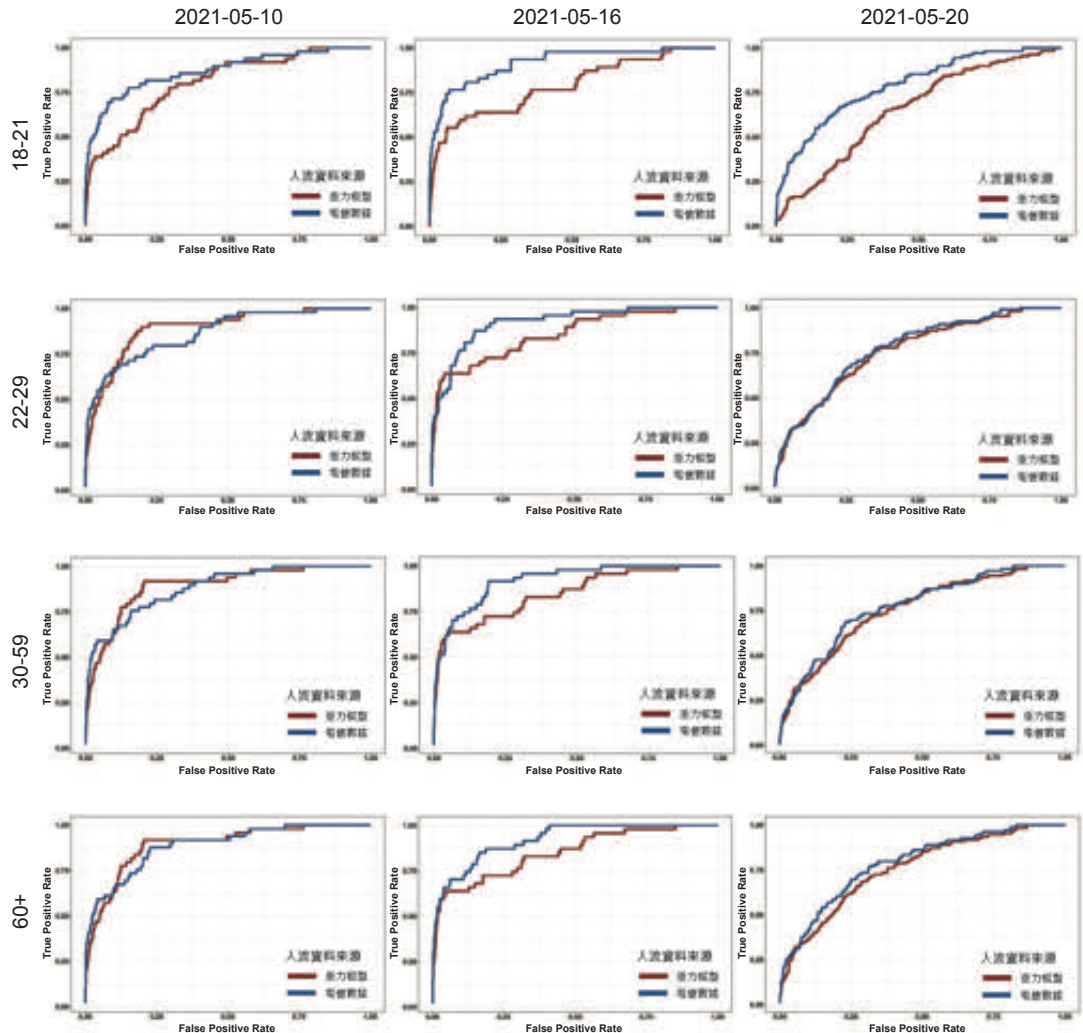
在計算村里 j 之加權度中心性 S_{jin} 時，將所有停留於該村里之停留時間比例進行加總：

$$S_{jin} = \sum_{i=1}^N \text{Normalized } P_{ij},$$

其中 $N = 1,488$ 為雙北市村里總數， $\text{Normalized } P_{ij}$ 表示人流量大小，被視為村里 i 在村里 j 之正規化停留時間比例。

(三) 結構等位、村里足跡相似度與相對風險計算

計算加權度中心性可得知雙北市中較為危險、人流較密集或來源複雜的區域。然而，中心性指標對於預測確診數不一定會有良好的準確度與參考價值，因確診數是以確診者的居住地作為統計，而非感染地，在分配防疫資源與篩檢量能時也應當參考確診者的居住地，因此，本研究以村里節點在地理



圖五 兩類人流資料來源（電信信令數據、重力模型模擬）於各期分齡之熱區預測ROC曲線——確診數門檻大於5人（ $T=5$ ）

註：紅色表示重力模型結果，藍色表示電信信令資料結果。

空間網絡中的位置（network position）估計確診者數量高低的相對風險。

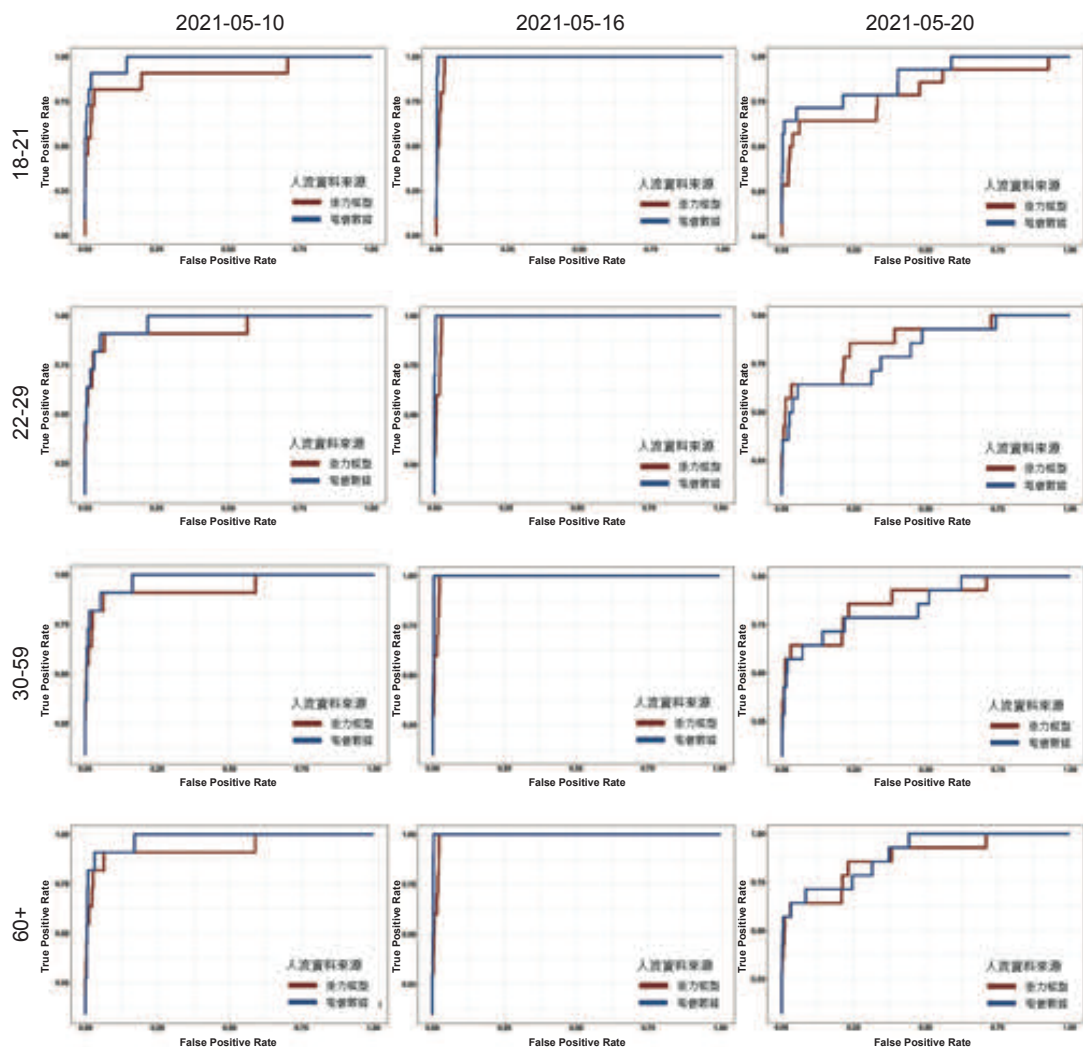
若兩村里擁有類似的人流流向，則可推測兩村里的居民有較高的機率互相接觸。換而言之，地理空間網絡中的兩節點共享相同的終點，且停留時間比例的分配也大致相同，當一個節點成為了感染區時，與其足跡相似的另一個節點也會暴露於較高的風險之中。

此處以結構等位（structural equivalence）的方式計算兩村里間之足跡相似度：

$$Equivalence_{ij} = \sum_{k \in K} |P_{ik} - P_{jk}|$$

$$Similarity_{ij} = 1 - \frac{Equivalence_{ij}}{\max(Equivalence_{ij})}$$

其中 $Equivalence_{ij}$ 表示村里i與村里j之結構等位性，由此兩村里與所有村里的停留時間比例差距總和計算而得， K 表示村里的集合。而 $Similarity_{ij}$ 則是將 $Equivalence_{ij}$ 進行標準化的處理，將結構等位特性轉換為相對的足跡相似度，數值介於0與1之間，



圖六 兩類人流資料來源（電信信令數據、重力模型模擬）於各期分齡之熱區預測ROC曲線——確診數門檻大於10人（ $T = 10$ ）

註：紅色表示重力模型結果，藍色表示電信信令資料結果。

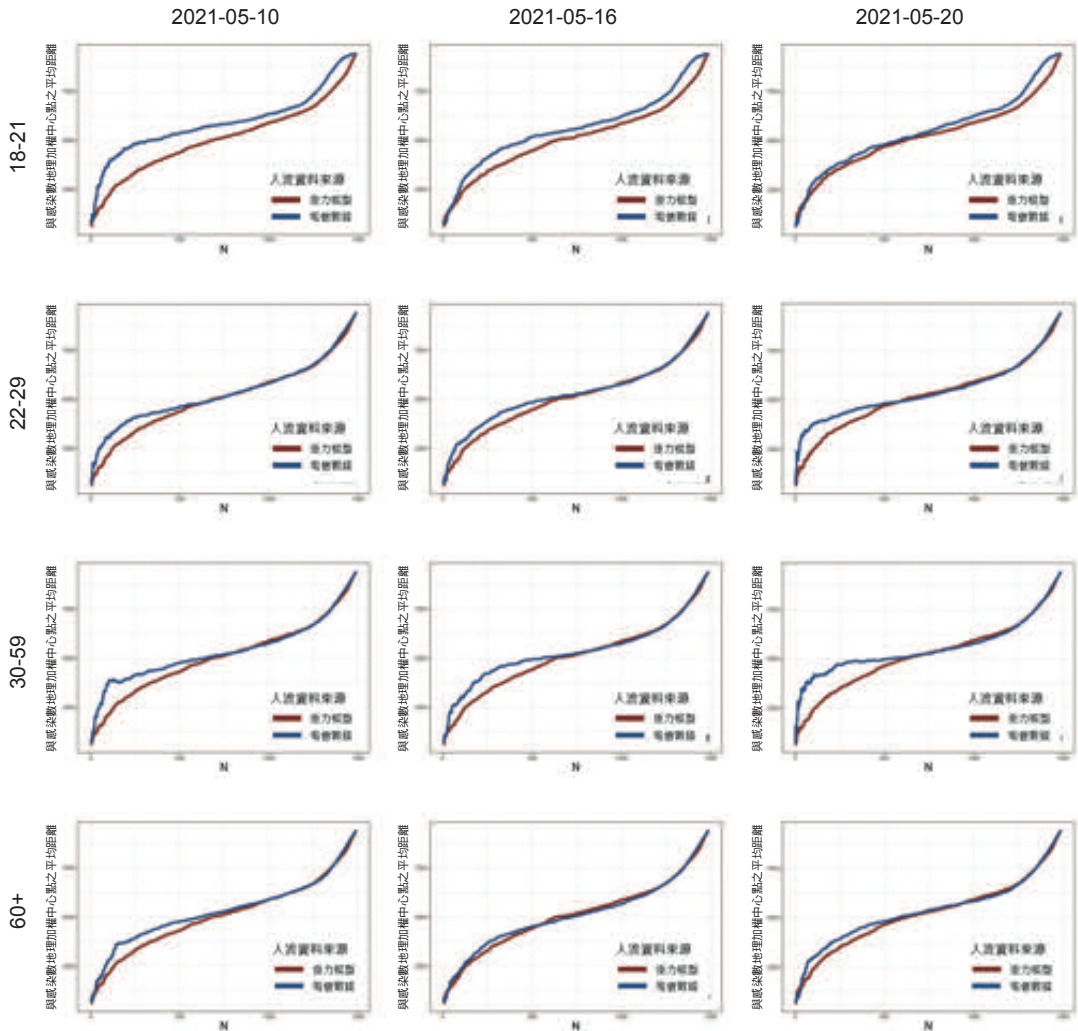
$Similarity_{ij}$ 為1表示兩村里之足跡相似度是所有村里中最高的；0則代表兩村里的足跡無任何重疊現象。

若進一步結合逐日確診數資料，可利用足跡相似度推估村里未來成為感染區的風險。假定相似度為1之村里代表兩村里足跡高度重疊，因此當其中一處受感染時，另一個村里將承受相同的風險；因此若a、b兩村里之相似度為 S_{ab} ，且a處新增 C_a 例確診個案（ C_a ），則另一個村里將承 $C_a \times S_{ab}$ 之風險。

$$Risk_j = \sum_{i \in I} C_i^{3days} \times Similarity_{ij}$$

其中， I 表示村里之集合、 C_i^{3days} 表示近3日之村里之累積確診者數量、 $Similarity_{ij}$ 表示村里 i 與 j 之足跡相似度，以此方式計算雙北市每個村里之相對風險高低，即可針對雙北市村里進行風險估計與確診個案數的預測。

新冠肺炎之可傳染期與症狀期開始的時間不同，一般而言，在感染後的2至3日會開始具有傳染性[15]，而患者大多在感染後5日開始出現症狀[16]。本研究假設患者在出



圖七 兩類人流資料來源（電信信令數據、重力模型模擬）於各期分齡之風險前N高之村里與確診數地理加權中心（mean center）之平均距離

註：紅色表示重力模型結果，藍色表示電信信令資料結果。

現症狀後即停止外出，因此具感染力、且容易透過移動傳播病毒的時間長度約為3日，故本研究取近3日累計確診數估計未來之各區域感染情形，舉例而言，針對5月11日至17日之村里相對風險，我們使用5月10日之人流足跡相似度與8日、9日、10日之村里確診數資料進行風險估計。

(四) 預測與迴歸分析

為探討靜態、動態兩種資料對於熱區預測的表現差異，本節定義三段研究區間，包

含2021年5月11日至17日、5月17日至23日、5月21日至27日，並衡量此三段區間中，相對風險對於累積確診數的預測能力。

考量到防疫資源有限，在分配時大都以地區疫情嚴重與否作為優先順序上的參考，例如新北市在2021年所劃列確診數較多的村里為重熱里。因此，本研究假定施政者依資源多寡，決定所需劃列的高風險感染區個數，被劃列到的區域將優先被分配到較多的防疫資源。

相對風險可用於預測未來7日是否需被劃列為高風險感染區，提前分配大量的防疫量能。若假設欲劃列的高風險感染區數量為N，則依據風險選出較高的N個村里劃列為高風險感染區，並觀測所有村里是否達到確診數門檻值T。若一村里被劃列為高風險感染區，且未來7日村里累積確診數大於T值，則屬於分類正確被歸類為真陽性（true positive），小於T值則為分類錯誤且歸類為偽陽性（false positive）；若一村里未被劃列為高風險感染區，且未來7日村里累積確診數小於T值，則分類正確被歸類為真陰性（true negative），大於T值則為分類錯誤被歸類為偽陰性（false negative）。

在給定T值代入不同N值作為防疫量能的閾值，以ROC曲線與ROC曲線下的面積（AUC）比較下列三種不同風險預測方式的成效：以過去三日各村里累積確診數多寡計算風險、以重力模型模擬之人流所計算之風險、以電信信令數據所計算之風險。

迴歸分析部分則使用Logistic regression 計算不同模型的勝算比（odds ratio）：

$$y = \sigma(f(x)) = \frac{1}{1 + e^{-\beta^T X}}$$

其中，Y為二元變量，表示未來一周該村里之確診人數是否高於門檻值T，X為該村里的感染風險大小。此式用以計算在不同的確診門檻值（T）之狀況下，各期村里感染風險預測確診數是否達門檻值之勝算比。

結 果

一、建構地理空間網絡

分別以電信信令數據與重力模型模擬之人流建構以年齡分組的地理空間網絡，依年齡、日期分組之網絡可見圖一。在電信人流網絡中，年齡的組間差異顯示年輕族群流動力較佳，連結之地理距離較遠；而老年人口之流動力較弱，隨距離的遞減率也較大，符合地理學第一定律的特性。在不同日期間的人流動態變化方面，5月15日實行雙北三級警戒政策後之村里間人流網絡連結變短、且

網絡連結的強度降低，說明村里間的人流量減少、且民眾多集中於居住村里附近活動。然而若使用靜態資料建構的重力模型之人流網絡，無法呈現疫情在不同流行階段、不同年齡層間的人流變化。

二、網絡中心性與確診數熱區之相關性

網絡加權度中心性分析結果在兩種資料中分別有不同的現象。電信信令數據的數值較高地區集中於市中心等交通樞紐；而重力模型模擬之人流網絡中，位於住宅區村里擁有較高的加權度中心性，市中心的村里數值較低，因為在使用重力模型模擬人流時並未考慮土地使用與交通網絡之型態，因此所得之人流中心大都與現實之中所認知之熱鬧區域不同。

此外，根據加權度中心性與確診數進行雙變數空間自相關分析，確診數熱區與人流中心性熱區並無明顯的空間自相關（圖二，Moran's I = 0.021），說明人流集中程度較高的區域不一定有較高之確診數，主要原因為確診數的統計是以確診者的居住地進行計算，而非感染地或是高度人群群聚的地區。綜合以上發現，在劃列進行高風險感染區時，人流之中心性不適合用於估計未來確診數。

三、結構等位、村里足跡相似度與相對風險計算

無論是使用電信數據資料，或是使用重力模型模擬人流，所得兩村里足跡相似度與村里間之距離關係在各年齡層、各日皆呈現顯著負相關（p-value < 0.05）。電信數據之四組年齡層（18-21、22-29、30-59、60+）之三日平均相關係數為-0.44、-0.49、-0.54、-0.55；重力模型各組之三日平均相關係數分別為-0.31、-0.31、-0.33、-0.33（圖三）。以電信數據而言，距離較為遙遠的兩村里，也有可能出現較高的足跡相似度，須依居民特性、兩村里在產業、交通上的依賴度而定；然而重力模型所得之結果卻說明距離遙遠的兩村里足跡相似度必定為低，突顯了靜態資料的限制。

四、預測與迴歸分析

根據三種不同確診數門檻值 $T = 1$ （圖四）、 $T = 5$ （圖五）、 $T = 10$ （圖六）分類熱區的ROC-AUC 統計圖表（表一），電信人流之AUC在三個研究區間中所有年齡層的平均為0.75，對比重力模型之平均為0.69。且在 $T = 1$ 、 $T = 10$ 兩門檻中，動態資料（ $AUC_{T=1} = 0.61$ ； $AUC_{T=10} = 0.94$ ）與靜態資料（ $AUC_{T=1} = 0.57$ ； $AUC_{T=10} = 0.92$ ）之表現較為接近，說明若要利用電信信令數據分類出1例確診數以上的區域較為困難；

而若以10例確診數做為區域是否感染的界線，則兩種資料都能有很好的效果。對於門檻值 $T = 5$ 而言，電信人流資料（ $AUC_{T=5} = 0.84$ ）的分類效果明顯較重力模型（ $AUC_{T=5} = 0.76$ ）好，說明電信人流適合預測確診數尚未飆升的新感染區。在本研究的觀察期中，高年齡層族群的預測效果較好，推測是因該波流行中高齡人口確診數較多所致。

在三段研究觀測區間中，重力模型傾向將高確診數村里之鄰近區域列為高感染風險區域，然而電信信令數據卻不完全依距離或鄰近關係進行風險高低的排序，傾向預測

表一 各期分齡之AUC數值

5/11-5/17			
年齡	確診數門檻	重力模型	電信信令數據
18-21	$T = 1$	0.51	0.62
	$T = 5$	0.50	0.62
	$T = 10$	0.91	0.98
22-29	$T = 1$	0.62	0.62
	$T = 5$	0.88	0.86
	$T = 10$	0.93	0.97
30-59	$T = 1$	0.61	0.63
	$T = 5$	0.88	0.87
	$T = 10$	0.93	0.97
60+	$T = 1$	0.61	0.63
	$T = 5$	0.88	0.87
	$T = 10$	0.91	0.98
5/17-5/23			
年齡	確診數門檻	重力模型	電信信令數據
18-21	$T = 1$	0.50	0.62
	$T = 5$	0.78	0.90
	$T = 10$	0.99	0.99
22-29	$T = 1$	0.60	0.62
	$T = 5$	0.84	0.91
	$T = 10$	0.99	1.00
30-59	$T = 1$	0.60	0.62
	$T = 5$	0.84	0.91
	$T = 10$	0.99	1.00
60+	$T = 1$	0.60	0.62
	$T = 5$	0.84	0.91
	$T = 10$	0.97	1.00
5/21-5/27			
年齡	確診數門檻	重力模型	電信信令數據
18-21	$T = 1$	0.48	0.60
	$T = 5$	0.64	0.77
	$T = 10$	0.80	0.88
22-29	$T = 1$	0.59	0.61
	$T = 5$	0.74	0.75
	$T = 10$	0.86	0.82
30-59	$T = 1$	0.59	0.60
	$T = 5$	0.74	0.75
	$T = 10$	0.87	0.85
60+	$T = 1$	0.57	0.61
	$T = 5$	0.71	0.77
	$T = 10$	0.84	0.89

出離疫情中心較遠的區域。平均而言，電信數據在依照風險排序後的第52個村里時，所有熱區與確診數加權重心之平均距離超過2,500公尺；然而重力模型的預測中，需要劃列超過126個村里，才能使所有熱區與確診數加權重心之平均距離超過2,500公尺（圖七）。換而言之，使用靜態資料預測得到的結果與傳統實務上劃列感染區的方法類似，主要是以鄰近的匡列為主，然而利用電信信令數據可以提前預測到離疫情中心較遠的區域，更能達成提前警戒的效果。

迴歸分析方面，根據Logistic regression之結果（表二）顯示電信人流在未來一周確診人數是否高於門檻值的風險中，電信人流在確診數門檻值 $T = 1$ 之各年齡層平均勝算比為1.16對上重力模型之1.03； $T = 5$ 之各年齡層平均勝算比為1.45對上重力模型之1.09； $T = 10$ 之各年齡層平均勝算比為1.71對上重力模型之1.16。此結果說明本研究中以電信人流估計風險的方法，能夠有效反映未來7日各村里的確診數分布情形。

討 論

本文以高空間解析度之電信信令數據資料建構地理空間人流網絡，比較其與靜態人口資料的區別，並應用於雙北市新冠肺炎疫情的實際預測，提出未來感染區劃列的計算模型。電信信令數據資料具備靜態人口資料難以捕捉的特性，如距離對於移動行為的影響等。在建構地理空間網絡後，除了可以使用社會網絡分析、空間資訊等方法分析人流資料，亦可探討節點在網絡的位置與結構等位性，引入傳染病模型中「接觸」的概念，估計區域與區域間的足跡相似度，了解區域間的風險依存關係，並結合逐日病例數資料計算風險，在預測未來感染區上的表現能夠較傳統的方法來得佳。

一、人流與確診數之關聯

傳統上在估計基礎再生數（ R_0 ）或是感染率（ β ）時，皆會考量接觸特性、接觸次數等效果，說明人群間的接觸是疾病擴散

的動力來源。人流能捕捉區域間的接觸特性，進而與確診數產生關聯性，國外也有許多研究探討土地使用分區與確診數的關聯性[4,17]，如公園等休閒遊憩區域易造成較高的確診數等。然而，在不同的空間尺度下結果可能會有所不同。如本文現在使用更細的空間尺度下探討區域確診數時，因民眾之感染地與居住地不同，造成空間相關性不顯著的情形發生。

二、基於網絡結構等位的人流相似性分析與實務政策之應用

目前國外結合電信數據人流資料與傳染病的相關研究中，大都使用傳染病數理模型預測感染數、疫情抵達時間，且著重於分析感染數與人流量的相關性、比較政策前後所帶來之差異[9,11]，較少以網絡之結構性量化風險並預測熱區的研究。各國對於劃列高風險熱區、感染區的作法，主要還是採用距離、鄰近性、人口數與確診數等靜態資料作為指標進行劃列。舉例而言，有文獻使用Getis-Ord G_i^* 的統計量觀測各個靜態資料指標是否顯著呈現空間自相關，作為熱區畫列的參考指標，並指出在熱區中採取保持社交距離、避免外出等政策將能有效地預防COVID-19的傳播[18]。然而，在傳統空間分析方法中，區域間的鄰近性大都以行政區邊界是否接壤進行判定，在實際的地理空間中，相互接壤之兩行政區不一定會擁有較頻繁的接觸次數。使用電信數據人流資料可更完整且正確地捕捉區域間的接觸頻率關係，也能夠掌握資料的動態變化，隨著疫情、政策變化使用不同日期之電信資料進行分析，將能有更好的預警效果。

此外，國內外熱區劃列實務上，疫調與追蹤確診者足跡也是相當典型的方式[19]，然而確診者足跡並無法掌握人群中的接觸情形，且在空間尺度較小的情形也容易出現感染地與居住地不同的確診數統計問題，因此若使用人流網絡空間中的結構等位特性計算相似度，可以更完善地考量空間中人群的接觸關係。

表二 以各期村里感染風險預測確診數是否達門檻值之勝算比

5/11-5/17							
年齡	確診數門檻	重力模型			電信信令數據		
		95%CI (LOW)	OR	95%CI (UP)	95%CI (LOW)	OR	95%CI (UP)
18-21	T = 1	0.99	1.01	1.02	1.31	1.4	1.51
	T = 5	1.13	1.18	1.24	2.1	2.55	3.16
	T = 10	1.18	1.32	1.51	1.93	3.77	6.42
22-29	T = 1	1.08	1.1	1.12	1.18	1.22	1.25
	T = 5	1.21	1.27	1.34	1.62	1.85	2.15
	T = 10	1.25	1.44	1.68	1.67	2.63	3.85
30-59	T = 1	1.08	1.1	1.11	1.18	1.22	1.25
	T = 5	1.21	1.27	1.33	1.62	1.82	2.09
	T = 10	1.25	1.44	1.68	1.64	2.51	3.63
60+	T = 1	1.08	1.09	1.1	1.41	1.49	1.59
	T = 5	1.2	1.26	1.33	1.84	2.13	2.51
	T = 10	1.23	1.4	1.65	1.6	2.29	3.07
5/17-5/23							
18-21	T = 1	1	1	1	1.09	1.02	1.14
	T = 5	1.02	1.02	1.03	1.16	1.21	1.27
	T = 10	1.05	1.05	1.07	1.13	1.19	1.26
22-29	T = 1	1.01	1.01	1.01	1.06	1.07	1.09
	T = 5	1.02	1.03	1.04	1.12	1.15	1.19
	T = 10	1.03	1.06	1.09	1.11	1.26	1.49
30-59	T = 1	1.01	1.01	1.01	1.05	1.06	1.07
	T = 5	1.02	1.03	1.04	1.09	1.11	1.13
	T = 10	1.03	1.06	1.09	1.06	1.18	1.31
60+	T = 1	1.01	1.02	1.03	1.09	1.11	1.13
	T = 5	1.03	1.04	1.04	1.11	1.14	1.18
	T = 10	1.01	1.04	1.08	1.08	1.17	1.27
5/21-5/27							
18-21	T = 1	1	1	1	1.1	1.13	1.16
	T = 5	1.01	1.01	1.01	1.15	1.19	1.24
	T = 10	1.01	1.02	1.03	1.16	1.24	1.31
22-29	T = 1	1.01	1.01	1.01	1.05	1.06	1.06
	T = 5	1.01	1.02	1.02	1.06	1.08	1.1
	T = 10	1.03	1.04	1.05	1.08	1.13	1.18
30-59	T = 1	1.01	1.01	1.01	1.04	1.05	1.05
	T = 5	1.01	1.02	1.02	1.05	1.06	1.08
	T = 10	1.03	1.04	1.05	1.07	1.11	1.15
60+	T = 1	1	1.01	1.01	1.1	1.11	1.13
	T = 5	1	1.01	1.02	1.09	1.11	1.14
	T = 10	1.01	1.03	1.05	1.11	1.16	1.21
各年齡層平均	T = 1	--	1.03	--	--	1.16	--
	T = 5	--	1.09	--	--	1.45	--
	T = 10	--	1.16	--	--	1.71	--

註：OR代表Odds Ratio。

三、人口流動變化對預測效果之探討

COVID-19疫情在中國爆發的初期即有文獻指出人流與疾病的傳播息息相關，當傳染病擴散時，封鎖政策會使人流量降低，減少了民眾的接觸機會，因此能有效減緩感染數的成長，也降低周遭區域的風險[20]。在台灣2021年5月COVID-19疫情爆發時，人流也因政策介入如捷運減班、居家上班、禁止餐飲業內用等而受到影響，因此在電信信令數據上，停留時間比例也會有顯著變化，人們停留於居住村里的時間(P_{ii})更長。在本研究中，5月16日為雙北三級警戒實施之第二日、5月20日為全國三級警戒實施第一日，人流有大幅減少的現象，然而人流絕對量值雖然減少，但足跡相似度取決於村里人流流出比例的相似程度，絕對數值的差異未必能反映於比例，因此僅能依據村里是否成為感染區（未來7日確診數高於T值）之相對風險進行排序，在人流明顯下降的期間，勝算比將會較正常日來得低。

四、展望：運用機器學習模型進行分析

若使用機器學習演算法等較為成熟之技術建構風險熱區分類器，可能有機會更精準地掌握未來即將受感染的地區，然而受限於當前電信信令數據資料昂貴且稀少，難以取得大量多天的資料作為訓練使用。未來若電信數據資料的取得成本降低，能期待以更多電信數據資料結合機器學習技術，進行傳染病的動態熱區預測，並藉由模式訓練得到更好的預測表現。

五、研究限制

由於現階段台灣僅有遠傳電信願意提供電信數據作為研究用途使用，因此電信使用人口比例可能造成代表性不足的問題實為本研究之限制之一。

此外，目前的確診數資料統計都是以現居地作為確診數的統計基礎，居住地不完全等於感染地，也成為推估村里風險上的挑戰。在研究區域上，因為僅納入雙北市村里進行分析，在估計部分位於行政區邊界的村

里風險時容易有失真的現象。舉例而言，若感染者居住在外縣市但在雙北市工作，將有可能將疾病傳染給雙北市之民眾，然而，由於人流移動之起始節點(i)並不在雙北市範圍內，因此在計算風險時並不會納入考量，造成雙北市感染風險的低估。

結論

面對未來的傳染病時，如何在資源有限的情況下將醫療量能分配至最易受感染之區域，是至關重要的課題，人流資料可捕捉區域間的接觸模式，進而有效的監測傳染病，並可用於輔助防疫資源分配的決策。靜態資料在取得上相當容易，可根據地理學定律建構區域間的接觸模式；而動態資料雖較為昂貴，但在熱區預測上更為精準、能夠有效預測未來將受感染的區域，以建構更完善精準的預警系統。隨著科技日新月異的發展，資料的取得日益便捷、計算效能也更加卓越，未來可使用電信數據人流資料了解人口流動的趨勢，並且隨著疫情發展、政策執行，在不同的階段動態更新資料，進而精準掌握傳染病的發展，藉此更有效率地分配資源，將災害最小化，達到減低災害、增進社會福祉的效果。

致 謝

感謝衛生福利部疾病管制署「防疫新生活健康行為監測研究計畫」提供研究經費購買去識別化電信人流數據，也感謝遠傳電信公司在電信數據處理上的技術諮詢服務。

參考文獻

1. Nova N, Athni TS, Childs ML, Mandle L, Mordecai EA. Global change and emerging infectious diseases. *Annu Rev Resour Economics* 2021;**14**:333-54. doi:10.1146/annurev-resource-111820-024214.
2. Yen MY, Yen YF, Chen SY, et al. Learning from the past: Taiwan's responses to COVID-19 versus SARS. *Int J Infect Dis* 2021;**110**:469-78. doi:10.1016/j.ijid.2021.06.002.
3. Ng TC, Cheng HY, Chang HH, et al. Comparison of estimated effectiveness of case-based and population-based interventions on COVID-19 containment

- in Taiwan. *JAMA Intern Med* 2021;**181**:913-21. doi:10.1001/jamainternmed.2021.1644.
4. Chan TC, Chou CC, Chu YC, et al. Effectiveness of controlling COVID-19 epidemic by implementing soft lockdown policy and extensive community screening in Taiwan. *Sci Rep* 2022;**12**:12053. doi:10.1038/s41598-022-16011-x.
5. 廖培珊、蕭錦炎、楊雅惠：以大型抽樣調查評估戶籍人口與常住人口之可能差異。人口學刊 2018；(57)：1-39。doi:10.6191/JPS.201812_57.0001。
- Liao PS, Hsiao CY, Yang YH. Inconsistency between estimates of both registered and de jure population: evidence from a large-scale sample survey. *J Population Studies* 2018;(57):1-39. doi:10.6191/JPS.201812_57.0001. [In Chinese: English abstract]
6. Barthélemy M. Spatial networks. *Phys Rep* 2011;**499**:1-101. doi:10.1016/j.physrep.2010.11.002.
7. Simini F, González MC, Maritan A, Barabási AL. A universal model for mobility and migration patterns. *Nature* 2012;**484**:96-100. doi:10.1038/nature10856.
8. Barbosa H, Barthelemy M, Ghoshal G, et al. Human mobility: models and applications. *Phys Rep* 2018;**734**:1-74. doi:10.1016/j.physrep.2018.01.001.
9. Persson J, Parie JF, Feuerriegel S. Monitoring the COVID-19 epidemic with nationwide telecommunication data. *Proc Natl Acad Sci* 2021;**118**:e2100664118. doi:10.1073/pnas.2100664118.
10. Schlosser F, Maier BF, Jack O, Hinrichs D, Zachariae A, Brockmann D. COVID-19 lockdown induces disease-mitigating structural changes in mobility networks. *Proc Natl Acad Sci* 2020;**117**:32883-90. doi:10.1073/pnas.2012326117.
11. Vigfusson Y, Karlsson TA, Onken D, et al. Cell-phone traces reveal infection-associated behavioral change. *Proc Natl Acad Sci* 2021;**118**:e2005241118. doi:10.1073/pnas.2005241118.
12. Wesolowski A, Qureshi T, Boni MF, et al. Impact of human mobility on the emergence of dengue epidemics in Pakistan. *Proc Natl Acad Sci* 2015;**112**:11887-92. doi:10.1073/pnas.1504964112.
13. Krings G, Calabrese F, Ratti C, Blondel VD. Urban gravity: a model for inter-city telecommunication flows. *J Stat Mech Theor Exp* 2009;**2009**:L07003. doi:10.1088/1742-5468/2009/07/L07003.
14. Jung WS, Wang F, Stanley HE. Gravity model in the Korean highway. *Europhys Lett* 2008;**81**:48005. doi:10.1209/0295-5075/81/48005.
15. He X, Lau EH, Wu P, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat Med* 2020;**26**:672-5. doi:10.1038/s41591-020-0869-5.
16. Lauer SA, Grantz KH, Bi Q, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann Intern Med* 2020;**172**:577-82. doi:10.7326/M20-0504.
17. Zheng Z, Xie Z, Qin Y, Wang K, Yu Y, Fu P. Exploring the influence of human mobility factors and spread prediction on early COVID-19 in the USA. *BMC Public Health* 2021;**21**:615. doi:10.1186/s12889-021-10682-3.
18. Parvin F, Ali SA, Hashmi S, Ahmad A. Spatial prediction and mapping of the COVID-19 hotspot in India using geostatistical technique. *Spatial Inform Res* 2021;**29**:479-94. doi:10.1007/s41324-020-00375-1.
19. 柯貞好、徐士敏、洪敏南等：2021年台灣首起新冠肺炎Delta變異株社區群聚感染調查。疫情報導 2022；38：75-80。doi:10.6524/EB.202204_38(7).0001。
- Ke JY, Hsu SM, Hong MN, et al. Investigation of the first community cluster of SARS-CoV-2 Delta variant infection in Taiwan, 2021. 2022;**38**:75-80. doi:10.6524/EB.202204_38(7).0001. [In Chinese: English abstract]
20. Kraemer MU, Yang CH, Gutierrez B, et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* 2020;**368**:493-7. doi:10.1126/science.abb4218.

Comparing the performance between static and dynamic populations in COVID-19 hotspot prediction

CHING-CHI CHOU¹, HSIEN-HO LIN², KEVIN J. CHEN³, RAN-CHOU CHEN^{4,5}, TA-CHIEN CHAN^{6,7,*}

Objectives: This study aimed to set up the prediction model of COVID-19 hotspot areas by using the census data and human mobility from telecommunication data in Taipei and New Taipei City. The comparison between their accuracy and limitations can provide the relevant insights for future epidemic control. **Methods:** The spatio-temporal resolution is fixed at the village level in two cities in May 2021. The static and dynamic data are used to construct the mobility network. The former applies gravity model to mimic human flow, and the latter uses telecommunication data as the measure of mobility. We propose the footprints similarity by structural equivalence of spatial networks and integrate it with the number of confirmed cases for computing the risk level of the villages. The performance of the models is evaluated using ROC curves and logistic regression under different thresholds for the confirmed cases. **Results:** The mobility derived from the telecommunication data provided better prediction performance than that from the census data; they have an average AUC of 0.75 and 0.69, respectively. Besides, the telecommunication data had a tendency to identify a further village as high-risk zone compared to the gravity model. According to the results of logistic regression, the odds ratio (OR) of exceeding the cases' threshold estimated from the telecommunication data is 1.45 on average, while the one estimated from the census data is 1.10. **Conclusions:** Telecommunication data can be beneficial in identifying the potential high-risk areas and enhancing situational awareness in advance. (*Taiwan J Public Health*. 2022;**41**(6):611-626)

Key Words: COVID-19, mobile phone signal, human mobility, early warning

¹ Department of Geography, National Taiwan University, Taipei, Taiwan, R.O.C.

² Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University Taipei, Taiwan, R.O.C.

³ Department of Health, Taipei City Government, Taipei, Taiwan, R.O.C.

⁴ Department of Health, New Taipei City Government, New Taipei City, Taiwan, R.O.C.

⁵ Department of Biomedical Imaging and Radiological Sciences, National Yang Ming Chiao Tung University, Taipei, Taiwan, R.O.C.

⁶ Research Center for Humanities and Social Sciences, Academia Sinica, No. 128, Sec. 2, Academia Rd., Nankang Dist., Taipei, Taiwan, R.O.C.

⁷ Institute of Public Health, School of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan, R.O.C.

* Correspondence author E-mail: tachien@gate.sinica.edu.tw

Received: Sep 12, 2022 Accepted: Dec 12, 2022

DOI:10.6288/TJPH.202212_41(6).111074