

①

出國報告（出國類別：研究）

②

美國舊金山 Palo Alto Medical Foundation 短期
研究心得(美國癌症登記資料庫相關研究)

③

服務機關：高雄榮民總醫院

姓名職稱：廖正義 主治醫師

派赴國家：美國

出國期間：2017年10月16日至2017年11月5日

報告日期：2017年11月10日

摘要

利用美國癌症登記資料庫 SEER (Surveillance, Epidemiology, and End Results)、NCDB (National Cancer Database)與 USCS (United States Cancer Statistics)來分析婦科癌症目前現況與未來可能的發展，同時開始接觸 National Health and Nutrition Examination Survey (NHANES)資料庫，嘗試尋找合併使用相關資料庫的方式與累積相關的經驗，藉由資料分析的過程可以了解資料庫的優缺點，可提供建立醫療資訊系統時的有用參考。

關鍵字：

癌症登記資料庫

SEER (Surveillance, Epidemiology, and End Results)

NCDB (National Cancer Database)

USCS (United States Cancer Statistics)

目次

一、 目的	P4
二、 過程	P4
三、 心得	P5
四、 建議事項(包括改進作法).....	P6

附錄

一、目的

先前赴美於 UCSF(加州大學舊金山分校醫學中心婦癌科)進修時，於 John K. Chan 教授的指導下對美國癌症登記資料庫之一的 SEER (Surveillance, Epidemiology, and End Results)有了初步的接觸與認識，回國後亦利用此資料庫發表海報於第 16 屆國際婦癌協會雙年會(16th biennial meeting of the international gynecologic cancer society)，後續亦陸續與 Chan 教授合作 SEER 的相關研究，才逐漸了解大數據的無限可能與不足之處；由於美國癌症登記資料庫中有另一項更完整的資料庫 NCDB(National Cancer Database)，但它需要為美國參加癌症登記計畫的醫院申請後方能閱覽，故於 Chan 教授的協助下，加入該研究後，啟程赴美進行癌症登記資料庫的相關研究。

二、過程

於今年 9 月獲知有短期出國的機會，便於醫院的積極協助下，於 2017 年 10 月 14 日赴美，於 10 月 15 日與 Palo Alto Medical Foundation 的 Chan 教授與 Stanford 大學的 Kapp 教授會面了解團隊目前的現況與研究方向；第一週先進行 SEER 資料庫中新的資料分析，主要是利用 USCS (United States Cancer Statistics) 2001-2014 Public Use Database 來分析婦癌的發生率變化；USCS 是由美國 CDC (Centers for Disease Control and Prevention 與 NIH(National Cancer Institute)整合 NPCR(CDC' s National Program of Cancer Registries)與 National Cancer Institute' s Surveillance, Epidemiology, and End Results program 所形成的資料，涵蓋全美 99%的人口，可充分反應美國的實際狀況。第二週 Chan 教授介紹 National Health and Nutrition Examination Survey (NHANES)資料庫與我，它是美國自 1959 年開始，每 2 年抽樣全美人口進行詳細的檢查、檢驗與問卷來反映當時美國人口的身體狀況，目前美國使用的生長曲線便是利用此資料研發而成的，它是屬於橫斷面的研究資料，於 SEER 及 NCDB 的世代研究資料不同。同時也利用時間回 UCSF 參加一年兩次的文獻選讀，獲益良多；第

三週則著重於 SEER 與 NCDB 資料的初步分析與探勘；目前已完成一篇研究，預計近期投稿；其他相關研究正在整理中。

三、心得

先簡介美國癌症登記資料庫：SEER(Surveillance, Epidemiology, and End Results)於 1973 年開始於美國 9 個地區收集癌症患者的診斷與治療相關資訊，後陸續於 1992 年擴展至 13 個地區，於 2000 年擴展至 18 個地區，目前約涵蓋美國 28% 的人口數；由於有標準化的資料收集設計與驗證機制，可確保所收集資料達 99% 以上的正確性，加上有完整的線上說明與教學課程，且相關資料已去連結不須申請 IRB，截至今日已有超過 8000 篇利用此資料庫的數據發表的文獻；但由於參加的地區大多位於美國西部，無法反映全美狀況，故美國外科協會(American College of Surgeons)於 1989 年開始 NCDB(National Cancer Data Base)計畫，增加 SEER 資料庫不足的項目，由各層級醫院依照標準回報癌症個案相關資料，目前涵蓋全美約 70% 癌症個案資料，但它為封閉型資料庫，須為參加 NCDB 計畫的醫療機構提研究計畫來申請資料，故藉由加入 Chan 教授的研究團隊來探索其資料。目前台灣的癌症登記資料庫便是綜合 SEER 與 NCDB 而發展的，三者擁有非常相似的結構，如何有效結合三項資料庫與分析，是我相當有興趣的議題。這些資料庫有一特性非常重要，就是資料的正確性非常高，美國設計了一套完善的制度來確保資料的正確，使得後續分析結果的代表性大幅提升；當擁有大量數據時，如何使用適當的工具與方法來分析與研究便是下一個重要的課題；目前在商業領域的研究模式可提供一個很好的建議。另外，藉由資料分析的過程可以了解資料庫的優缺點，可提供建立醫療資訊系統時的有用參考。

四、建議事項（包括改進作法）

目前美國癌症登記資料庫大部分是屬於開放性提供大眾使用，原因是它是公共財，且美國政府部門認為無足夠人力與物力可自行研究資料，故開放提供相關研究者進行研究，且提供相當豐富的軟體與參考資料來協助分析，且提供遠端運算，申請權限相當便捷，當日申請隔日便可得到授權；若須特殊資料來深入探討，亦可專案申請保密型資料，可謂相當便捷；反觀台灣相關資料如癌症登記資料，必須申請付費後至資料中心進行運算，無法進行遠端運算，一旦發現新問題或計算錯誤，必須重新申請並付費，會大幅降低資料的可近性，建議政府應將資料去連結後提供研究單位無償使用來促進研究發展。

附錄 (請附附錄：如議程、相片、簡報…)

SEER 資料庫提供的相關研究軟體與使用建議



