• 系統編號	RN9308-3309		
• 計畫中文名稱	以階層式叢集研究多重外傷		
• 計畫英文名稱	Hierarchical Clustering of Multiple Trauma		
• 主管機關	行政院國家科學委員會	• 計畫編號	NSC91-2213-E038-004
• 執行機構	台北醫學大學醫學資訊研究所		
• 本期期間	9108 ~ 9207		
• 報告頁數	6 頁	• 使用語言	--
• 研究人員	蔣以仁 --		
• 中文關鍵字	資料探勘; 叢集分析; 醫學資料庫; 申報資料庫; 全民健保研究資料庫; 外傷; 燙傷		
• 英文關鍵字	Data mining; Cluster analysis; Medical database; Claim database; National Healthcare Insurance Research Database (NHIRD); Trauma; Burn injury		

• 中文摘要

醫學資料探勘是 2001 年台灣地區健保局釋出健保資料庫 (NHIRD) 供研究使用後的熱門題目。本團隊以多重外傷爲例，探討階層式叢集分析在此種大型資料庫上協助醫學研究及醫院管理之重要性與適用性。階層式叢集分析具有顯現叢集過程的優點，提供專家在互動環境下依研究目的來探討資料結構。我們以 Matlab 實作與資料庫聯結，產生資料倉儲的線上分析 (OLAP) 所需之維度資料表，供進一步資料分析。對於健保資料庫來源龐雜的主副 ICD 診斷資訊的整理，我們引入臨床分類系統 (CCS) 作資料的初步整理。我們在兩個主題上介紹在醫學研究及發現的細節── (1)以叢集排序後之樹狀圖展現台灣地區多重外傷之分布情況，可提供政策制定與監測之用 ;(2)以叢集分析各級醫院在燙傷住院時的不同治療類型，而分群可協助發掘同質性的病人族群，做更精細的統計分析，或用以發現異常的病例或醫院。

• 英文摘要

Medical data mining becomes more mandatory after the release of national healthcare insurance research database (NHIRD) in 2001. In multiple trauma injuries, our research demonstrated the feasibility and benefits of the hierarchical clustering methodology in medical research and in hospital administration. The process of clustering could be explicit shown in hierarchical cluster analysis. The domain experts could elucidate the structure of data in interactive setting. We implement in Matlab the realtime linkage with the data sources, and propose the methodology of dimensional table generation for further online analytical processing (OLAP) in a data warehouse. For the widely varied coding policy in the International Classification of Disease (ICD) diagnosis fields, we introduced Clinical Classification System (CCS) for more medically relevant data preparation. The medical findings were illustrated in details in two topics - (1) Visualization of the distribution and correlation of multiple trauma in Taiwan, which could provide abundant information for research needs from clinical and administrative fields; (2) To find the major treatment patterns of hospitalized burn patients in various hospitals, to facilitate sophisticated

statistical analysis of more homogeneous patient populations after clustering, or as the surveillance of normal pattern and to find patient or hospital/doctor outliers.