

## 以決策樹模型探討未開立 慢性病連續處方之影響因子

蔡佳玲

國軍左營總醫院資訊室

洪新原

國立中正大學資訊管理學系

袁繼銓

國軍左營總醫院資訊室

### 摘要

中央健康保險局為方便病情穩定之慢性病患者就醫，因而推廣慢性病連續處方箋制度，以提供病患做週期性取藥，藉以降低平均門診次數，並且避免非必要之醫療支出。本研究以健保局資料庫為研究材料，探討2007年全國醫院層級之門診慢性病案件，符合慢性病連續處方箋開立條件之開立與未開立慢性連續處方箋之記錄，運用C 5.0決策樹演算法之分類功能，將慢性病連續處方箋開立與否之影響因子生成決策樹模型與規則集。本研究以「是否開立慢性病連續處方箋」為分類欄位，來探討醫院權屬別、醫院層級別、醫事機構區域別、醫師性別、醫師年齡、科別、醫師平均每日看診量、患者性別、患者年齡、慢性病疾病範圍等十項因子之區別能力，期能找出慢性病連續處方箋開立與否之影響因素。

研究結果顯示決策樹模型整體正確率達79.31%，在開立與未開立慢性病連續處方規則集中，有六條符合開立條件而未開立慢性病連續處方箋之描述規則。分析結果發現：首先，有相關影響性之因子共八項（除了患者年齡與患者性別）。其次，有二十一條符合開立條件而選擇開立慢性病連續處方箋之描述規則，有相關影響性之因子共九項（除了患者性別）。最後，在醫院權屬別、醫院層級別、就醫科別等十項因子中，就醫科別與醫院層級別二項因子，對開立與未開立慢性連續處方箋之規則描述，最具有決定性影響。

**關鍵字：**慢性病連續處方箋、影響因子、資料探勘、決策樹

# Applying Data Mining Technique to Identify the Influencing Factors of Not Prescribing the Refilled Chronic Disease Prescriptions

Chia-Ling Tsai

Information Management Section, Zuoying Armed Forces General Hospital

Shin-Yuan Hung

Department of Information Mangement, National Chung Cheng University

Yuan-Chi Chuan

Information Management Section, Zuoying Armed Forces General Hospital

## Abstract

In order to make things easy for patients with stable chronic diseases, Bureau of National Health Insurance (BNHI) popularizes the refilled chronic disease prescriptions (RCDP). Thus, the patients can get the medicine periodically to decrease the average number of times of outpatient services and to avoid unnecessary medical expenses. This study used the database in Bureau of National Health Insurance to investigate the outpatient cases of chronic illness in 2007 which conform to the conditions of refilled chronic disease prescriptions but refilled prescriptions are not prescribed in various levels of hospitals in our county. The C 5.0 decision tree algorithm was taken to generate the decision tree model and rules. "Prescribing refilled prescriptions or not" is the classification outcome and ten factors including hospital ownership, level of hospital, region of medical institute, gender of doctor, age of doctor, division of medical care, average outpatient service per day, gender of patient, age of patient and chronic illness scope are used to predict "prescribing or not prescribing refilled prescriptions".

The findings show that the overall accuracy of the Decision Tree Model reaches 79.31%. Six useful rules were found. Eight factors including hospital ownership, level of hospital, region of medical institute, division of medical care, average outpatient service per day, gender of patient, age of patient and chronic illness scope, were also identified as influencing factors. Finally, implications from the findings are also provided.

**Key words:** Refilled Chronic Disease Prescriptions, Influencing Factors, Data Mining, Decision Tree

## 壹、緒論

慢性病在致病機轉上常具有多重病因、危險因子等造成對疾病影響程度作用不一，在加上具有長期潛伏、發病期難以確定等特性，故在慢性病致病機轉尚未明確界定前，有效預防工作益顯重要。然而預防工作大不同於傳染疾病，著重於初段預防如預防接種及環境衛生監測等策略。故中央健康保險局（簡稱健保局）於慢性病初段預防重點就在於廣泛性的生活形態健康促進及早期診斷、早期治療，慢性病連續處方箋之推動與推廣就是其中一項重要之策略。

根據健保局現行公告的慢性疾病共有98種，包括糖尿病、高血壓、痛風、肝硬化等，上述之疾病雖無立即致命之危險，但需定期就醫服藥控制，才能避免病情惡化及併發症的產生。如此長期累積之下對於患者和家人之心力與醫藥費用皆為極大負擔，所以健保局為方便病情穩定之慢性病患者就醫，提昇醫療利用可近性，因而推廣慢性病連續處方箋制度，經醫師確認屬於病情穩定之慢性病患者，可長期使用相同處方藥品治療時，醫師可開立慢性病連續處方箋，供病患做週期性取藥，藉以降低平均門診次數，避免非必要之醫療支出。慢性病連續處方箋推動之政策，使得醫療產業環境的複雜性與競爭程度大幅提升，醫院要在此多變的醫療產業環境中生存，則需透過良好的經營策略，才能有效回應醫療產業環境的改變，這些策略與醫療環境之變化對醫院經營管理與民眾就醫習慣之影響都值得進一步研究與探討。

針對國內目前以慢性病連續處方為主題之學術研究，發現其中探討慢性病連續處方箋開立與釋出之行為為模式，多以問卷或傳統統計之方法進行，而資料探勘可從資料庫中挖掘隱藏知識之技術，在醫療領域越來越受到重視。進一步深入了解，符合健保局公告之慢性病範圍，經醫師判定病情穩定，需長期使用同一處方藥品治療，卻無開立慢性病連續處方箋之影響因子。本研究將資料探勘技術運用於健保醫療資料庫，健保局醫療資料庫涵蓋全國民眾完整的就醫申報紀錄資料，可從中過濾出全國慢性病患者其符合開立慢性病連續處方箋條件之就醫資料，依有限的健保醫療資料找出病患、醫師、醫療院所的基本特性，建立決策樹模型，運用決策樹「特徵選取」（Feature Selection）之特性來引導選取最有意義之屬性，並以決策樹產生之規則集，找出未開立慢性病連續處方箋之隱含特徵，進一步探討醫師除了專業考量之外，選擇不開立慢性病連續處方箋之因素，以期能對未開立慢性病連續處方箋之行為有所解析。

## 貳、文獻探討

### 一、慢性病連續處方箋

由於慢性病患者需要定期就醫與長期服藥控制，才能避免病情惡化及併發症，這些不易治癒的慢性疾病，經長期門診治療，使得整體的醫療費用不斷地上漲。病患經常

在醫院、診所奔波往返，因長期看病而付出的掛號費、部分負擔等費用，經年累積下來，為一筆不小的開銷。很多慢性疾病俗稱「老人病」，反映出年齡是慢性病的高危因子，老年人隨著年齡的增加，導致各種生理機能的退化，罹患慢性疾病機率相對比年輕人高，老年人的經濟條件滿意程度仍是顯著影響老年人就醫與否的重要因素（張雅雯 2002）。對於不良於行又需長期服藥治療之慢性病患者，沒有家人協助就醫、服藥，亦造成不能持續用藥的問題。

慢性病連續處方箋每張效期從醫師開立處方日起3個月內有效，民眾可持「慢性病連續處方箋」前往醫療院所或健保特約藥局分次領藥，醫院則按病情需要，一次給予30日以內的藥量，無需繳納藥費負擔，經醫師確認屬於病情穩定的慢性病患，可長期使用相同處方藥品治療時，即符合開立慢性病連續處方箋的標準。

為鼓勵醫療院所開立慢性病連續處方箋，健保局自91年1月起調高西醫基層院所開立慢性病連續處方箋之診察費，醫院亦於91年10月起比照西醫基層院所，調高診察費支付點數，並將開立慢性病連續處方箋達成率列入各醫院總額管理指標內，透過宣導及資訊公開，提高醫師開立慢性病連續處方箋意願，更利用媒體工具加強民眾對慢性病連續處方箋之認識與使用之方便性及好處。

劉嘉年與楊志良（2006）醫師處方行為受到醫療與臨床因素、醫病互動、醫療組織因素、醫療保險與政策等因素影響。開立慢性病連續處方箋，對醫事機構、醫師、病人之影響如下，對醫事機構而言可以提高健保局實施的醫院總額管理指標，但會減少掛號費的收入。對醫師而言可以減少工作負擔控制門診量，但患者持慢箋至醫院、診所領藥時因不看診故健保局不給付診察費，開立3個月慢性病連續處方箋會減少診察費的收入。對病人而言可以減少掛號費、部分負擔的支出，以及待診與交通的時間。

健保局醫院總額專業醫療服務品質指標顯示，開立比例隨醫院規模與門診量有顯著之變化（中央健康保險局醫院總額專業醫療服務品質報告2007）。

何蘊芳（2004）以問卷調查台北市公立、財團法人醫學中心與市立醫院共203位醫師，得知九成五的醫師願意開立慢性病連續處方箋，其理由主要是「病人病情穩定」，其次是「病人領藥方便」，不願意開立慢性病連續處方箋的原因主要是「擔心病人病情的變化」，其次是「擔心病人服藥遵從性差」。

蔡瑞真（2005）也以問卷調查南部八縣市26家醫學中心與區域醫院共469位醫師，156位患者，有八成五的醫師表示同意為「符合慢性病連續處方箋開立之條件且用藥穩定之病患」開立慢性病連續處方箋。八成九的患者知道「醫師可以為已控制穩定的慢性病人開立連續處方箋」，只有二成八的民眾曾經主動請醫師開立慢性病連續處方箋，多數的民眾均認同「連續處方箋可以增加病患或家屬之方便性」及「可減少經濟負擔」。

龔佩珍等（2007）依健保局特約醫事機構名冊篩選符合研究條件之西醫基層醫師為研究對象，從中隨機抽取4,000位，以郵寄結構式問卷的方式進行問卷調查，共計回收有效問卷813份。50.5%的基層醫師表示願意或非常願意釋出慢箋，但也有32.7%醫師表示不願意或非常不願意；不願意理由前三項主因為「醫療糾紛責任區分不明」、「領藥流程造成民眾不便」、「藥師可能私自更換藥品」。五成以上基層醫師表達對於慢箋釋出制度不滿意。在控制其他變項下，影響基層醫師釋出慢箋意願之因素為「醫師認為無法

即時掌握病情」、「增加醫師用藥困擾」、「對社區藥師協助用藥安全之認知」及「是否應制定社區藥局處方箋獨立總額之看法」、「對處方交付費滿意度」、「是否自聘藥師協助調劑」、「執業年資」、「平均每週病患人數」及「醫師主要診治之疾病」。何蘊芳等（2006）於某醫學中心門診藥局領藥窗口前，針對持慢箋領藥民眾進行訪查，共回收有效樣本1,073份，研究顯示，受訪民眾之慢箋使用資訊主要來源是醫師（62.1%）。

## 二、資料探勘

Fayyad et al. (1996) 對資料探勘的定義為「依據使用者需求，自資料庫中選擇合適資料，加以處理、轉換，探勘至評估的一連串步驟」。Berry and Linoff (1997) 定義「資料探勘是針對大量的資料，利用自動化或半自動的方式進行分析，以尋找出有意義的關係或法則」。Cabena et al. (1997) 定義資料探勘是將先前未知、有效的資訊從大型資料庫抽出的過程，並且將萃取出的有用資訊提供給主管做決定性的決策。Han and Kamber (2000) 則定義「資料探勘是從儲存在資料庫、資料倉儲或其它資訊儲存體的大量資料中，發現有趣特徵的過程」。

綜合上以學者的看法，資料探勘即是從大量的資料中，萃取出有意義的模式或規則，其應用廣範，只要是希望從歷史資料中找出規律或希望對未來進行預測，皆可利用資料探勘的技術。此以往需要領域專家和分析人員進行大量人工分析的問題，如今可以透過資料探勘技術迅速取出基於知識應用於決策。

## 三、決策樹

預測類別變數的過程稱之為分類，針對欲處理且未分類的資料集合，根據已知類別（Class）的物件（Instance）集合，將欲處理資料依據其屬性（Attributes）去完成分類的過程。分類的目的通常為尋找影響某一重要變數的因素，以及了解某一族群的特徵，和建立分類規則，而本研究之研究目標即在了解未開立慢性病連續處方箋之特徵，期望藉由過去的歷史資料，將記錄分類，找出隱含之規則並預測結果。

決策樹演算法（decision tree algorithms）屬於資料探勘技術中分類的方法之一，將資料依照演算法分割條件自動地將資料分類並以樹狀圖的方式來表示，適用於所有的分類問題，運用歸納方法找出資料來源的規則。在使用上可以用此模型分析已有的資料，也可用它來預測未來的資料，能夠建構明確且易於了解的決策模型，具有將複雜的決策判斷程序簡化的能力，對問題提供較易於解釋的答案（Safavian & Landgrebe 1991）。

Han and Kamber (2000) 指出目前決策樹的演算法大致有C5.0、CART、CHAID與QUEST四種演算法。C5.0是Quinlan 在1979年所提出的ID3所發展而來，而ID3的決策樹演算法，是最早出現的決策樹演算法之一，已經應用在推論系統和概念學習系統，C5.0經由ID3以及C4.5不斷的改良測試之後，是最具有效率以及準確性的演算法（Quinlan 2003），其演算的方法是先將每一筆資料都當成是同一群，接著運用亂度（Entropy）的概念，計算各屬性之資訊獲得（Information Gain），以挑選出最佳的屬性，對資料進行分類。C5.0並改進了ID3無法處理連續屬性的問題，適用在處理大資料集，採用Boosting方式，提高模型的精確度，這種方法按序列建立多重模型，第一個模型以一般的方式建

立，隨後建立第二個模型，聚焦於第一個模型錯誤分類的記錄，第三個模型聚焦於第二個模型的錯誤等等。而C5.0優點在於：（1）面對資料遺漏和輸入欄位很多的問題時非常穩健（2）通常不需要很長的訓練次數進行估計（3）比其他類型的模型易於理解，模型推出的規則有非常直觀的解釋（4）提供強大的增強技術提高分類的精度。

#### 四、資料探勘分類技術於醫療領域之相關研究

醫療產業採用資訊科技不僅是要提昇各項作業，更重要的是提昇醫護人員臨床相關決策的品質，進而提昇醫療照護品質。以資訊科技為主軸之醫院經營管理，其中資料之運用與管理通常為致勝之關鍵，如何有效運用資料探勘之技術，將資料轉化為知識，是發掘問題進一步解決問題之有效辦法。國內外多項研究已證實資料探勘從資料中所萃取潛在知識可用於醫療決策輔助，本研究大致將其應用分類技術於醫療決策領域之相關研究文獻彙整如下。

簡禎富與林國勝（2006）針對cDNA生物晶片之二元資料的特性，發展生物晶片資料挖礦（Data Mining）方法和模式藉以探索與尋找疾病與特定基因的關係，並建構其規則以作為醫療診斷決策支援參考。研究並採用史丹佛大學晶片資料庫中乳癌晶片資料以驗證研究效度，從四萬多個基因與64個樣本當中，使用顯著性分析（Significant Analysis of Microarray）與決策樹挖掘出具影響力的基因及診斷決策規則，從中萃取有價值之資訊。

戴建耘等（2007）以資料探勘、資料倉儲（Data Warehouse）以及分類演算法中之決策樹基礎分類法，為研究工具，以心臟科麻醉部為例，從醫療單位對抗生素進行導管菌落判斷感染監控的歷史案例中，建構出可用以預測心脈導管塗抹抗生素在病患身上的作用結果之分類模式，用以協助醫藥人員有效判斷感染情形與特徵，而提升抗生素之用藥醫療品質，降低可能的醫療資源浪費。

王偉驊等（2008）運用資料探勘分類演算法的多元區別分析（Multi-group Discriminant Analysis（MDA）），架構檢傷分類的預測模型，發現多群判別分析能夠有效的區辨出病患的危急程度達90.6%，並且透過資料導出一種規則，利用產出的判別函數來預測，一個新的病患應該歸屬檢傷分類的哪一等級，進而提高檢傷分類的一致性與穩健性。

Nattkemper et al.（2005）應用非監督式學習之K-means群集分析與類神經網路之自組織映射圖網路（self-organizing map,SOM）和最鄰近分類器（k-nearest neighbor classifiers,K-NN）、決策樹、支援向量機（support vector machines ,SVM）監督式學習之分類方法找出乳癌腫瘤X光片的特徵。

Park et al.（2006）使用資料探勘技術，探勘護理臨床資料，確定四個變量（患者年齡、皮膚完整性受損、疼痛感和傷口滲液狀況）可預測日後的照護情況，緩解病患疼痛；另五個變數（患者年齡、疼痛感、受感染的可能性、婚姻狀況及主要疾病）被確定是預測死亡率的重要指標。

Brossette and Hymel（2008）以資料探勘的技術運用從臨床實驗室以及醫院資訊系統裡所得來的資料，找到對病人、樣本的種類、地點、感染生物種類和對抗生素的敏感性之關聯規則，應用於醫院的感染控制。

## 參、研究方法

### 一、研究方法與步驟

本研究採用Hui and Jha (2000) 提出之資料探勘步驟進行研究，模型建立流程如下。

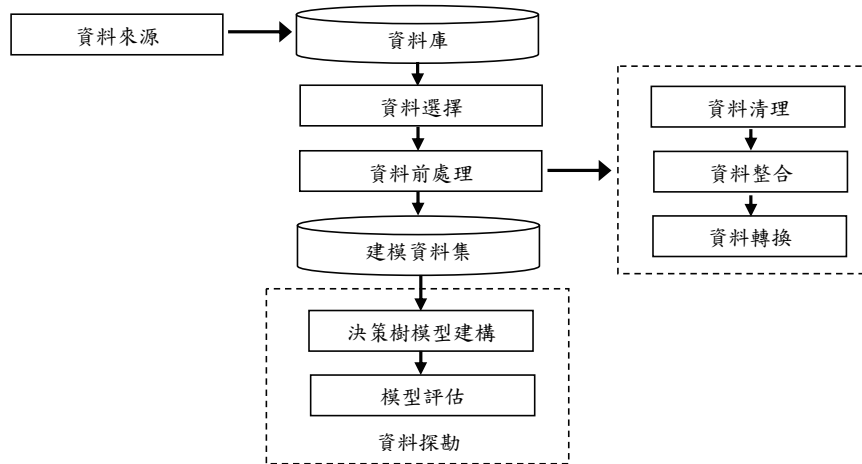


圖1：模型建立流程

### 二、研究因子定義

經由文獻探討，針對研究問題歸納出醫院、醫師、病患、疾病四個因素，接著透過專家學者訪談的方式，專家學者包含公共衛生領域之教授和區域教學醫院家醫科醫師、新陳代謝科醫師、血液腫瘤科醫師各一位及基層院所家醫科醫師一位共五位，訪談內容主要了解，在病患病情穩定之情況下，各科醫師選擇對慢性病連續處方箋開立與未開立之因素，最後彙總訪談結果後，提出之衡量因子如下，研究架構如圖2。

#### (一) 醫院因素包含：

##### 1. 醫院權屬別

依權屬別與經營類型區分為「公立醫院」、「財團法人醫院」、「私立醫院」共三類。

##### 2. 醫院層級別

依醫院評鑑等級及健保支付標準區分為「醫學中心」、「區域醫院」、「地區醫院」共三類。

##### 3. 醫事機構區域別

依健保分局可以區分為「台北」、「北區」、「中區」、「南區」、「高屏」、「東區」共六類。

#### (二) 醫師因素包含：

##### 1. 醫師性別：區分「M-男性」、「F-女性」。

## 2. 醫師年齡

利用健保資料檔中費用發生年，即2007年減掉醫師出生年份求得，以「1-10歲」、「11-20歲」、「21-30歲」、「31-40歲」、「41-50歲」、「51-60歲」、「61-70歲」、「71-80歲」、「81-90歲」、「91-100歲」，共十組做為醫師年齡分組方式。

3. 就醫科別：門診醫療費用清單明細檔中，患者看診之就醫科別。

## 4. 醫師平均日看診量

依醫師身份證字號by年度計算2007年，全部看診案件數與看診日，平均後求得，並根據基層院所每日門診合理量公式分割為5個級距，「1-30人」、「31-50人」、「51-70人」、「71-150人」、「150人以上」。

### (三) 病患因素包含：

1. 患者性別：區分「M-男性」、「F-女性」。

2. 患者年齡：利用健保局資料檔中費用發生年，即2007年減掉病患出生年份求得，依據衛生署統計資料分割「0歲=嬰兒」、「1-14歲=少年」、「15-24歲=青年」、「25-44歲=壯年」、「45-64歲=中年」、「65歲以上=老年」。

3. 藥理分類數：依據門診處方醫令明細檔中之患者用藥對應健保局版本之藥理分類系統，統計出患者服用之藥理分類數量。藥理分類為藥品藥理治療作用之分類，由患者服用之藥理分類數量的多寡，來判別患者病情的複雜度。

### (四) 疾病因素：

健保局公告之慢性病範圍歸類16種類別之98項疾病。本研究依門診醫療費用清單明細檔國際疾病分類號一，國際疾病分類號二，國際疾病分類號三，取得第一順位的慢性病疾病範圍。

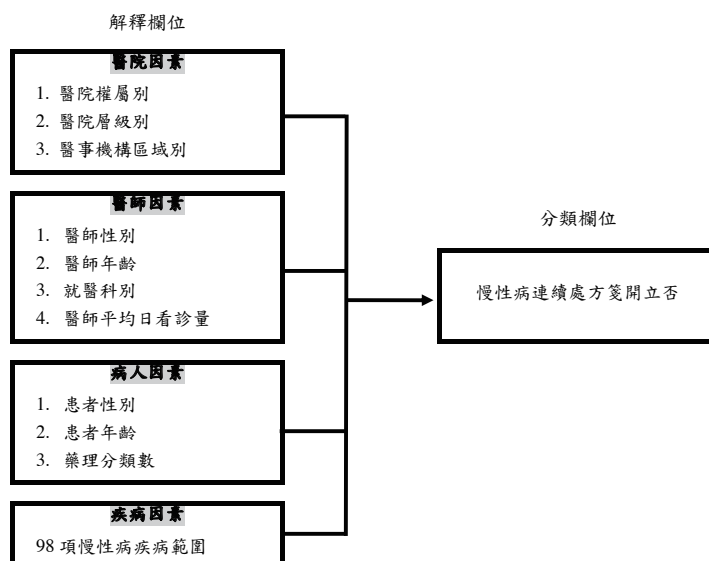


圖2：研究架構圖



### 三、資料來源與選擇

#### (1) 資料來源

本研究以健保局資料庫為研究材料，探討2007年近一年全國門診慢性病案件，符合開立慢性病連續處方箋條件之開立與未開立慢性連續處方箋之記錄。透過專案申請取得慢性病患資料相關檔案，醫事機構基本資料檔、醫事人員基本檔、醫師之看診統計、門診處方及治療明細、門診處方醫令明細檔、國際疾病分類檔等六個檔。另由全民健康保險研究資料庫網站下載代碼說明文件，取得醫事機構特約類別、權屬別及型態別其代號與型態別名稱、地區代碼、地區名稱和分局等，專案申請取得之慢性病患資料相關檔案如下：

1. 醫事機構基本資料檔：2006至2007年「04」慢性病案件有出現之院所資料，共9,684筆資料。
2. 醫事人員基本檔：共182,584筆資料。
3. 醫師看診統計檔：2006至2007年門診案件分類為「04」之醫師，依醫師身份證歸戶計算，當年全部看診件數及看診日數，加以平均。
4. 門診處方及治療明細：2006至2007年且案件分類為「04」慢性病案件之資料，共103,135,569筆資料。
5. 門診處方醫令明細檔：2006至2007年且案件分類為「04」慢性病案件之資料，共445,413,452筆資料。
6. 國際疾病分類檔。

#### (2) 資料選擇

本研究對象為，全國醫院層級之慢性病患者，其慢性疾病符合健保局公告之慢性病範圍，經醫師判定病情穩定，需長期使用同一處方藥品治療，卻無開立慢性病連續處方箋之案件，建立決策樹預測模型。本研究將資料條件分為二類，為開立慢性病連續處方箋與符合開立而未開立之資料，條件設定如下：

1. 開立慢性病連續處方箋
  - (1) 費用年為2007年度。
  - (2) 案件分類為「04」慢性病案件。
  - (3) 「給藥日份」欄位

婦產科給藥日份需 $\geq 21$ 天，因婦女賀爾蒙用藥規定服用21天應停用7天，故開立21天賀爾蒙用藥，免收取藥費負擔，其餘應開立28天以上，方可免除藥費負擔。

- (4) 診察費健保代碼為00170A、00171A、00172B、00173B、00174B、00175B、00176B、00177B。

2. 符合開立條件而未開立之資料
  - (1) 回溯時間定為四個月

因慢性病連續處方箋總調劑日份最多為92日，故本研究設定若患者連續四個月固定回診同一就醫科別，同一位醫師同診斷且服用同藥品時，表示病患病情穩定，可長期使用相同處方藥品治療，最少可以開立一次慢性病連續處方箋。

(2) 費用年為2007年度，但回推至2006年9月的就醫資料，回溯四個月的資料了解患者之就診與用藥是否符合本研究之條件。

(3) 案件分類為「04」慢性病案件。

(4) 「給藥日份」欄位

婦產科給藥日份需 $\geq 21$ 天，因婦女賀爾蒙用藥規定服用21天應停用7天，故開立21天賀爾蒙用藥，免收取藥費負擔，其餘應開立28天以上，方可免除藥費負擔。

(5) 診察費健保代碼不為00170A、00171A、00172B、00173B、00174B、00175B、00176B、00177B。

(6) 回溯連續四個月回診固定同一就醫科別同一位醫師之病患資料

固定回診同一就醫科別同一位醫師，醫師才可客觀判斷患者病情是否穩定。

(7) 回溯連續四個月看診之主診斷相同。

(8) 回溯連續四個月持續性拿相同藥品。

#### 四、資料前處理

資料探勘在探勘前需要經過資料整合、資料清洗、資料轉換等資料前處理，為資料探勘前的一項重要步驟，錯誤、不適合或不完整的資料將會影響探勘的結果，以下針對各步驟一一說明：

##### (一) 資料清理

資料清洗其目的是針對所收集到的資料進行資料的清理與過濾的動作，去除不必要的資料以獲得有用的資訊，確保資料的品質。本研究將逐一檢查各欄位是否有超出或不屬於該欄位的資料，資料清洗過程中，將移除有問題的紀錄，並檢視各筆紀錄的完整與否，如下：

1. 醫事機構基本資料檔：檢查醫事機構層級別、醫事機構縣市區碼、醫事機構代號權屬別是否為空白。
2. 醫事人員基本檔
  - (1) 檢查醫師性別是否有不為「M」或「F」、醫師出生日期是否空白。
  - (2) 檢查醫師年齡是否小於20或大於100歲。
3. 門診處方及治療明細
  - (1) 檢查患者性別是否有不為「M」或「F」、患者出生日期是否空白。
  - (2) 檢查門診處方及治療明細中，國際疾病分類號一是否可參照健保局公告之慢性病範圍進行歸類。

##### (二) 資料整合

由於資料的來源可能是來自於多個不同的資料格式或異質資料庫，亦或多個資料來源，因此必須先將這些資料來源做整合，將多個資料來源中的資料進行合併處理。

透過專案申請取得之慢性病患資料相關檔案，為文字檔，除資料龐大且為了方便進行資料處理，檔案一一轉入Oracle資料庫。另醫事機構特約類別、權屬別及型態別其代

號與名稱、地區代碼和名稱及分局是由全民健康保險研究資料庫網站下載之文件，需自行至Oracle資料庫建立資料表。

### (三) 資料轉換

資料轉換是為確保資料的格式或型態，適用於資料探勘之模型或方法，為配合研究需求，將全民健康保險研究資料原始檔的資料匯入資料庫後，進行資料轉檔，產生符合研究目的有效可探索之資料集。其中為了輔助資料探勘，產生衍生因子變項，例如為避免使用數值型態之資料進行決策樹分類，導致分類結果過度遷就且分類效率不加佳的問題，故將連續型態之資料轉換成類別型資料。衍生因子變項處理方式如下：

#### 1. 醫院權屬別

醫事機構代號前2碼，依「權屬及型態別其代號與名稱」轉換成「1」公立醫院、「2」財團法人醫院、「3」私立醫院，資料對應如下：

- (1) 公立醫院：省立醫院、軍方醫院（民眾診療）、榮民醫院。
- (2) 財團法人醫院：財團法人醫院、宗教財團法人醫院、其他法人附設醫院。
- (3) 私立醫院：私立醫院。

#### 2. 醫事機構區域別

依「醫事機構基本資料檔」之縣市區碼關聯「醫事機構地區代碼和名稱及分局」轉換。

#### 3. 醫師年齡

病患看診年份－醫師出生年份，另依10歲為一級距，轉換成10個級距。

#### 4. 醫師每日平均看診量

根據基層院所每日門診合理量公式轉換成5個級距。

#### 5. 病患年齡

病患看診年份－病患出生年份，依據衛生署統計資料轉換成6個級距。

#### 6. 藥理分類數

依據門診處方醫令明細檔中之患者用藥對應健保局版本之藥理分類系統，統計出患者服用之藥理分類數量。

#### 7. 慢性病疾病範圍

依「門診處方及治療明細檔」之國際疾病分類號一與特殊治療代碼一，參照全民健康保險醫療辦法公告之16種類別和98項疾病之慢性病特殊治療項目代碼歸類。

經由資料前處理，產生開立慢性病連續處方箋筆數7,533,119，符合開立而未開立之慢性病連續處方箋筆數1,462,333。依醫事機構代號、醫師代號、就醫科別、患者身份證、國際疾病分類號一，進行歸戶後，得開立慢性病連續處方箋筆數3,090,848，符合開立條件而未開立之慢性病連續處方箋筆數360,739。為取得具代表性抽樣，以隨機抽樣方式從開立慢性病連續處方箋筆數母體中抽出與符合開立條件而未開立之慢性病連續處方箋相同的筆數，二群資料混合成為建模資料，建模資料共721,478筆。

## 五、決策樹模型模擬

本研究工具採用統計軟體SPSS Clementine，使用C5.0演算法生成決策樹與規則集。從健保局資料庫慢性病患者就醫申報紀錄中，過濾出全國醫院層級符合開立慢性病連續處方箋條件之就醫資料，經由資料前處理，產生建模資料，以「是否開立慢性病連續處方箋」為分類欄位（classes field），以醫院權屬別、醫院層級別、醫事機構區域別、醫師性別、醫師年齡、就醫科別、醫師平均日看診量、患者性別、患者年齡、藥理分類數、慢性病疾病範圍等十一項因子做為解釋欄位（explanatory field），來找出慢性病連續處方箋開立與否之特性。

並利用相同的建模資料集，試驗不同的參數集，建立多個不同的模型，依資料特性與研究目的找出最適當之模型。

## 六、評估方式

資料探勘得到的模型有可能是沒有真實的意義或價值的，因此需要對資料探勘的結果進行評估，評估模型是否正確有效，目前正確率評估的技術常用的為 holdout validation、cross-validation與bootstrapping（Letourneau et al. 1999），三種評估方法：

- (1) holdout validation：將資料集劃分成訓練集和測試集，訓練集用來訓練分類器，測試集則用來測試其分類結果，避免過於保守的評估。
- (2) bootstrapping：也是將資料分為訓練資料和測試資料，但其抽樣可以重覆，而且一般必須做到對估計量的抽樣分佈到深入了解為止，通常非常耗時。
- (3) cross-validation：可以避免因隨機抽樣產生誤差作法，是將資料集劃分成k個同等大小的子集，每一次用k-1個子集作為訓練集，最後剩下的一個子資料集為測試集，然後重複此步驟k次，直到每一個子資料集都被做為測試集為止。

為避免隨機抽樣產生誤差，本研究採用以10折Cross-validation（k=10）的方式，將資料集隨機分成10折，每一組輪流當作測試樣本，其它則當作訓練樣本，也就是共有10折不同的訓練樣本和測試樣本。如此執行10次後，得到10折分類正確率，再平均10折分類正確率則是該資料集的平均正確率。

分類效能的評估方面，本研究以錯差矩陣（Confusion Matrix），來顯示分類效能。由錯差矩陣可以得知，一個預測模型所做出的預測，有多少正確，有多少錯誤，見表1所示，錯差矩陣表之分類資料共分為四型：

表1：錯差矩陣

實際發生 \ 分類結果	未開立慢性病連續處方箋	開立慢性病連續處方箋
	未開立慢性病連續處方箋	TP (True Positive)
開立慢性病連續處方箋	FP (False Positive)	TN (True Negative)

- (1) 真肯定 (True Positive)：實際資料為未開立慢性病連續處方箋，且分類模型判斷未開立慢性病連續處方箋，屬於正確分類。
- (2) 誤否定 (False Negative)：實際資料為未開立慢性病連續處方箋，但分類模型判斷開立慢性病連續處方箋，屬於錯誤分類。
- (3) 誤判 (False Positive)：實際資料為開立慢性病連續處方箋，但分類模型判斷未開立慢性病連續處方箋，屬於錯誤分類。
- (4) 真否定 (True Negative)：實際資料為開立慢性病連續處方箋，且分類模型判斷開立慢性病連續處方箋，屬於正確分類。

本研究採用的評估指標如下：

$$(1) \text{ 整體分類正確率 (overall accuracy) } = (TP+TN)/(TP+FN+FP+TN)$$

整體分類正確率意味著整體分類發生的正確率，即TP（實際資料為未開立慢性病連續處方箋，且分類模型判斷未開立慢性病連續處方箋）與TN（實際資料為開立慢性病連續處方箋，且分類模型判斷開立慢性病連續處方箋）之比例。

$$(2) \text{ 遺漏性誤差 (omission error) } = FN/(TP+FN)$$

該分類錯誤率表示實際為未開立慢性病連續處方箋但卻預測是開立慢性病連續處方箋之分類錯誤比率。

$$(3) \text{ 過度預測誤差 (commission error) } = FP/(FP+TN)$$

該分類錯誤率表示實際為開立慢性病連續處方箋但卻預測未開立慢性病連續處方箋之分類錯誤比率。

此外區別函數的分類正確率 (Hit Ratio) 也常被用來表示此區別函數的區別能力，而且其值必須高於機會正確率 (Chance Accuracy,  $C_{pro}$ ) 方可接受 (Hair et al.1995)。故本研究也將比較整體分類正確率與機會正確率，了解分類模型之區別能力，機會正確率  $= p^2 + (1-p)^2$ ， $p = (TP+FN)/(TP+FN+FP+TN)$ 。

## 肆、研究結果

### 一、研究樣本資料描述性統計

本研究依研究目的，經資料前處理，得符合開立慢性病連續處方箋條件而未開立之筆數360,739，得符合開立慢性病連續處方箋條件之開立筆數3,090,084。為取得具代表性抽樣，以隨機抽樣方式從符合開立慢性病連續處方箋條件之開立筆數母體中抽出360,793，二群資料混合成為建模資料，建模資料共721,478筆。整理研究樣本資料描述性統計如下所述，由研究樣本資料描述性統計可了解研究資料的分佈情形。

#### (一) 醫院屬性：

表2顯示，符合開立慢性病連續處方箋條件之資料中，醫事機構權屬別之未開立比例以財團法人醫院43.50%最低，私立醫院61.61%最高，醫院層級別以醫學中心之未開立比例40.13%最低，區域醫院65.56%最高，醫事機構區域別之未開立比例以南區46.04%最低，高屏地區53.61%高於其他。

表2：樣本數醫院屬性描述性統計

因子名稱	類別	開立人次 (A)	未開立人次 (B)	開立百分比A/ (A+B)	未開立百分比B/ (A+B)
醫事機構 權屬別	公立醫院	125,487 (34.79%)	139,463 (38.66%)	47.36%	52.64%
	財團法人醫院	187,179 (51.89%)	144,124 (39.95%)	56.50%	43.50%
	私立醫院	48,073 (13.33%)	77,152 (21.39%)	38.39%	61.61%
		360,739 (100.00%)	360,739 (100.00%)		
醫院層 級別	醫學中心	157,564 (43.68%)	105,604 (29.27%)	59.87%	40.13%
	區域醫院	141,897 (39.34%)	138,477 (38.39%)	50.61%	49.39%
	地區醫院	61,278 (16.99%)	116,658 (32.34%)	34.44%	65.56%
		360,739 (100.00%)	360,739 (100.00%)		
醫事機構 區域別	台北	143,564 (39.80%)	126,897 (35.18%)	53.08%	46.92%
	北區	44,536 (12.35%)	51,128 (14.17%)	46.55%	53.45%
	中區	58,142 (16.12%)	67,095 (18.60%)	46.43%	53.57%
	南區	52,262 (14.49%)	44,594 (12.36%)	53.96%	46.04%
	高屏	53,266 (14.77%)	61,565 (17.07%)	46.39%	53.61%
	東區	8,969 (2.49%)	9,460 (2.62%)	48.67%	51.33%
	360,739 (100.00%)	360,739 (100.00%)			

## (二) 醫師屬性：

表3顯示，符合開立慢性病連續處方箋條件之資料中男醫師未開立比例50.43%，高於女醫師46.16%，醫師年齡介於21-30歲之未開立比例25.92%最低，年齡愈大未開立比例愈高，每日平均看診量之未開立比例以150人以上24.30%最低，介於1-30人54.37%最高，看診科別中以心臟血管內科未開立比例28.94%最低，精神科98.80%最高。

表3：樣本數醫師屬性描述性統計

因子名稱	類別	開立人次 (A)	未開立人次 (B)	開立百分比A/ (A+B)	未開立百分比B/ (A+B)
醫師性別	男性	321,511 (89.13%)	327,108 (90.68%)	49.57%	50.43%
	女性	39,228 (10.87%)	33,631 (9.32%)	53.84%	46.16%
		360,739 (100.00%)	360,739 (100.00%)		
醫師年齡	21~30歲	8,194 (2.27%)	2,867 (0.79%)	74.08%	25.92%
	31~40歲	110,893 (30.74%)	92,463 (25.63%)	54.53%	45.47%
	41~50歲	158,923 (44.05%)	166,843 (46.25%)	48.78%	51.22%
	51~60歲	70,497 (19.54%)	82,524 (22.88%)	46.07%	53.93%
	61~70歲	10,458 (2.90%)	13,474 (3.74%)	43.70%	56.30%
	71~80歲	1,570 (0.44%)	1,657 (0.46%)	48.65%	51.35%
	81~90歲	204 (0.06%)	886 (0.25%)	18.72%	81.28%
	91~100歲	0 (0.00%)	25 (0.01%)	0.00%	100.00%
	360,739 (100.00%)	360,739 (100.00%)			
醫師每日 平均 看診量	1~30人	99,141 (27.48%)	118,116 (32.74%)	45.63%	54.37%
	31~50人	108,596 (30.10%)	126,842 (35.16%)	46.13%	53.87%
	51~70人	66,533 (18.44%)	66,139 (18.33%)	50.15%	49.85%
	71~150人	81,548 (22.61%)	48,062 (13.32%)	62.92%	37.08%
	150人以上	4,921 (1.36%)	1,580 (0.44%)	75.70%	24.30%
	360,739 (100.00%)	360,739 (100.00%)			

表3：樣本數醫師屬性描述性統計（續）

因子名稱	類別	開立人次 (A)	未開立人次 (B)	開立百分比A/ (A+B)	未開立百分比B/ (A+B)
科別	心臟血管內科	89,032 (24.68%)	36,253 (10.05%)	71.06%	28.94%
	內分泌科	42,037 (11.65%)	24,175 (6.70%)	63.49%	36.51%
	內科	39,688 (11.00%)	44,602 (12.36%)	47.09%	52.91%
	神經科	37,918 (10.51%)	25,835 (7.16%)	59.48%	40.52%
	家醫科	26,644 (7.39%)	15,479 (4.29%)	63.25%	36.75%
	腸胃內科	23,170 (6.42%)	20,936 (5.80%)	52.53%	47.47%
	泌尿科	14,229 (3.94%)	13,929 (3.86%)	50.53%	49.47%
	腎臟內科	11,748 (3.26%)	10,319 (2.86%)	53.24%	46.76%
	胸腔內科	10,337 (2.87%)	8,020 (2.22%)	56.31%	43.69%
	骨科	9,638 (2.67%)	11,168 (3.10%)	46.32%	53.68%
	婦產科	8,817 (2.44%)	5,312 (1.47%)	62.40%	37.60%
	眼科	7,713 (2.14%)	7,543 (2.09%)	50.56%	49.44%
	風濕免疫科	7,441 (2.06%)	11,010 (3.05%)	40.33%	59.67%
	外科	6,215 (1.72%)	9,046 (2.51%)	40.72%	59.28%
	神經外科	5,013 (1.39%)	4,808 (1.33%)	51.04%	48.96%
	心臟血管外科	4,852 (1.35%)	1,310 (0.36%)	78.74%	21.26%
	小兒科	3,525 (0.98%)	5,000 (1.39%)	41.35%	58.65%
	血液腫瘤科	2,522 (0.70%)	4,380 (1.21%)	36.54%	63.46%
	耳鼻喉科	2,332 (0.65%)	2,731 (0.76%)	46.06%	53.94%
	復健科	2,279 (0.63%)	3,234 (0.90%)	41.34%	58.66%
	感染科	1,149 (0.32%)	983 (0.27%)	53.89%	46.11%
	精神科	1,072 (0.30%)	88,201 (24.45%)	1.20%	98.80%
	直腸外科	1,022 (0.28%)	1,729 (0.48%)	37.15%	62.85%
	皮膚科	936 (0.26%)	1,570 (0.44%)	37.35%	62.65%
	消化外科	385 (0.11%)	1,167 (0.32%)	24.81%	75.19%
	放射腫瘤科	303 (0.08%)	618 (0.17%)	32.90%	67.10%
	胸腔外科	192 (0.05%)	484 (0.13%)	28.40%	71.60%
	放射線科	100 (0.03%)	210 (0.06%)	32.26%	67.74%
	潛醫科	89 (0.02%)	105 (0.03%)	45.88%	54.12%
	疼痛科	64 (0.02%)	118 (0.03%)	35.16%	64.84%
	麻醉科	41 (0.01%)	71 (0.02%)	36.61%	63.39%
	核子醫學科	37 (0.01%)	183 (0.05%)	16.82%	83.18%
	整形外科	30 (0.01%)	55 (0.02%)	35.29%	64.71%
	職業醫學科	27 (0.01%)	0 (0.00%)	100.00%	0.00%
	放射診斷科	26 (0.01%)	30 (0.01%)	46.43%	53.57%
	小兒外科	24 (0.01%)	22 (0.01%)	52.17%	47.83%
	胸腔暨重症 加護	22 (0.01%)	10 (0.00%)	68.75%	31.25%
	脊椎骨科	16 (0.00%)	29 (0.01%)	35.56%	64.44%
	急診醫學科	15 (0.00%)	0 (0.00%)	100.00%	0.00%
	不分科	14 (0.00%)	3 (0.00%)	82.35%	17.65%
	洗腎科	11 (0.00%)	30 (0.01%)	26.83%	73.17%
新生兒科	9 (0.00%)	19 (0.01%)	32.14%	67.86%	
居家護理	4 (0.00%)	12 (0.00%)	25.00%	75.00%	
病理科	1 (0.00%)	0 (0.00%)	100.00%	0.00%	
		360,739 (100.00%)	360,739 (100.00%)		

### (三) 病患屬性：

表4顯示，符合開立慢性病連續處方箋條件之資料中，男性患者其未開立比例50.90%較女性49.04%高，患者年齡未開立比例以1-14歲71.36%最高，45-64歲46.65%最低。

表4：樣本數病患屬性描述性統計

因子名稱	類別	開立人次 (A)	未開立人次 (B)	開立百分比A/ (A+B)	未開立百分比B/ (A+B)
患者性別	男性	183,238 (50.80%)	189,923 (52.65%)	49.10%	50.90%
	女性	177,501 (49.20%)	170,816 (47.35%)	50.96%	49.04%
		360,739 (100.00%)	360,739 (100.00%)		
患者年齡	少年組1~14歲	3,044 (0.84%)	7,583 (2.10%)	28.64%	71.36%
	青年組15~24歲	4,770 (1.32%)	8,804 (2.44%)	35.14%	64.86%
	壯年組25~44歲	37,418 (10.37%)	55,831 (15.48%)	40.13%	59.87%
	中年組45~64歲	143,196 (39.70%)	125,211 (34.71%)	53.35%	46.65%
	老人組65歲以上	172,311 (47.77%)	163,310 (45.27%)	51.34%	48.66%
		360,739 (100.00%)	360,739 (100.00%)		

### (四) 疾病屬性：

表5顯示，符合開立慢性病連續處方箋條件之資料中，未開立比例以精神疾病89.51%最高，循環系統疾病34.75%最低。

表5：樣本數疾病屬性描述性統計

因子名稱	類別	開立人次 (A)	未開立人次 (B)	開立百分比A/ (A+B)	未開立百分比B/ (A+B)
慢性疾病 範圍	01癌症	11,148 (3.09%)	20,155 (5.59%)	35.61%	64.39%
	02內分泌及代謝疾病	77,339 (21.44%)	59,266 (16.43%)	56.62%	43.38%
	03精神疾病	11,290 (3.13%)	96,325 (26.70%)	10.49%	89.51%
	04神經系統疾病	14,497 (4.02%)	11,869 (3.29%)	54.98%	45.02%
	05循環系統疾病	150,849 (41.82%)	80,345 (22.27%)	65.25%	34.75%
	06呼吸系統疾病	12,469 (3.46%)	9,652 (2.68%)	56.37%	43.63%
	07消化系統疾病	24,248 (6.72%)	21,670 (6.01%)	52.81%	47.19%
	08泌尿系統疾病	5,540 (1.54%)	6,542 (1.81%)	45.85%	54.15%
	09骨骼肌肉系統及結締組織之疾病	17,522 (4.86%)	23,783 (6.59%)	42.42%	57.58%
	10眼及其附屬器官之疾病	7,291 (2.02%)	6,939 (1.92%)	51.24%	48.76%
	11傳染病	240 (0.07%)	463 (0.13%)	34.14%	65.86%
	12先天畸形	1,226 (0.34%)	1,220 (0.34%)	50.12%	49.88%
	13皮膚及皮下組織疾病	1,131 (0.31%)	1,792 (0.50%)	38.69%	61.31%
	14血液及造血器官疾病	1,997 (0.55%)	1,690 (0.47%)	54.16%	45.84%
	15耳及乳突之疾病	1,707 (0.47%)	2,407 (0.67%)	41.49%	58.51%
	16其他	22,245 (6.17%)	16,621 (4.61%)	57.24%	42.76%
		360,739 (100.00%)	360,739 (100.00%)		



## 二、決策樹模型建構

### (一) 輸出入變數

輸出變數指的是我們要預測的變數，輸入變數則為我們用來預測輸出變數所使用之變數。本研究輸出變數為慢性病連續處方箋開立否，建構決策樹模型之輸入變數如下：

1. 醫院因素：醫院權屬別、醫院層級別、醫事機構區域別。
2. 醫師因素：醫師性別、醫師年齡、就醫科別、醫師平均日看診量。
3. 病患因素：患者性別、患者年齡、藥理分類數。
4. 疾病因素：患者疾病為健保局公告之慢性病98細項範圍。

本研究根據上列變數，將所生成之決策樹模型結果分析後，其中藥理分類數與98項慢性病範圍變數，兩類之屬性特質因存在相關性與多項性之關係，生成之決策樹會產生子樹複製的問題會導致生成之決策樹太大，其分支很多，當樹的規模很龐大時，會導致預測能力變差所產生的分類規則太過特殊難以詮釋，如此失去了分類探勘的意義，因此本研究為考慮分類法的品質，故調整輸入變數，捨去藥理分類數，健保局公告之慢性病98細項範圍改為健保局公告之16種慢性病項範圍類別。

### (二) 決策樹模型參數條件

本研究利用SPSS Clementine軟體上的C5.0決策樹模型，以經由資料前處理產生之721,478筆建模資料，進行分類模型建構，本研究經試驗不同的參數設定後，將其校正至最適值，模型建立參數條件如下：

1. 「使用增益功能」(Use Boosting)，試用數目設定為2組，藉由重複選取及組合樣本訓練決策模型，用以增加其準確率的同時也預防過度訓練。
2. 「選擇交叉驗證」(Cross-validate)，折疊次數設定為10折，使用9組資料當做訓練資料建立模型，另1組資料來評估模型精確度。
3. 「模式」(Model)為高級模式
  - (1) 修剪純度(pruning severity)，設定為80，提高純度值，將獲得更小更簡潔的決策樹。
  - (2) 子分支最少記錄數(Minimum records per child branch)，設為30，子群大小可以用來限制決策樹任一分支的折分數。
4. 選擇排除屬性(Winnow attributes)，C5.0在建立模型前檢驗預測欄位的有用性，被發現與分析無關的預測欄位將不參與建模過程。

本研究在生成決策樹模型時並同時產生重要因子變項圖表。從重要因子變項圖表可初步得知，影響開不開立慢性連續處方箋之主要因子前三項依序為就醫科別、醫院層級別、慢性病疾病範圍，各因子變項權重資料整理如表6所示。

表6：重要因子變項權重資料表

順序	說明	權重	順序	說明	權重
1	就醫科別	0.772	6	醫事機構區域別	0.009
2	醫院層級別	0.156	7	醫師性別	0.003
3	慢性病疾病範圍	0.028	8	患者年齡	0
4	醫師年齡	0.017	9	患者性別	0
5	醫院權屬別	0.015	10	醫師平均日看診量	0

### 三、決策樹模型評估

為避免因隨機抽樣產生誤差，並且對於新進來的資料能有較好的解釋力，本研究使用10折交互驗證法來提升分類模型的正確率。採用錯差矩陣進行模型評估，在模型建構後其整體正確率（Overall Accuracy）達79.31%，機會正確率（Chance Accuracy）50%，整體正確率大於機會正確率，顯示本研究決策樹模型對開立或未開立慢性病連續處方箋可以有效的區分，未開立慢性病連續處方箋錯差矩陣參見表7。

表7：未開立慢性病連續處方箋錯差矩陣表

實際發生 \ 預測結果	未開立慢性病連續處方箋	開立慢性病連續處方箋	合計
未開立慢性病連續處方箋	281,736 (78.13%)	78,873 (21.87%)	360,609
開立慢性病連續處方箋	70,343 (19.51%)	290,280 (80.49%)	360,623
合計	352,079	369,153	721,232

在降低不確定性方面，慢性病連續處方箋未開立模式預測誤差可分為兩種，一種為遺漏性誤差（omission error），即實際上未開立慢性病連續處方箋，預測結果卻顯示其開立；另一種為過度預測誤差（commission error），即實際上開立慢性病連續處方箋，預測結果卻顯示其未開立。本研究之遺漏性誤差率為21.87%，過度預測誤差率為19.51%，此兩種誤差代表不同的意義。遺漏性誤差代表模式無法詮釋現有資料，對本研究而言是較為嚴重的誤差，因本研究之精神主要是針對醫院、醫師、患者疾病等特徵預測出符合健保局開立慢性病連續處方箋條件，卻無開立慢性病連續處方箋之患者，並找出相關因子。若根據錯差矩陣分析，將其中符合開立慢性病連續處方箋條件之患者，預測為開立但實際上卻無開立，對於此比例較少但卻又是最值得注意之未開立慢性病處方之患者族群造成遺漏，對健保主管單位之策略制定與資源之運用方向造成錯誤之引導影響程度，將遠大於預測為未開立處方但實際上卻開立處方之患者，應該將此誤差降到最低以提高模擬的準確性。此外，由於研究模式採用之資料來自全民健保資料庫，就醫紀錄之完整性而言醫師對慢性疾病穩定程度之認定、患者之就醫習慣均無法完整紀錄，因此過度預測誤差則較可以容忍。

為了解容易造成遺漏性誤差之資料屬性，本研究進行錯誤分析，選擇分類錯誤之記錄再進行模型建構，建立分類規則，調整最小支持度為1000，最小信賴度75%，根據未開立慢性病連續處方箋預測為開立之規則集之研究發現，此類特徵醫院層級多屬於「地區醫院」之「私立醫院」且科別為「內科」及「神經科」。實際開立慢性病連續處方箋

預測為未開立之規則集研究發現，此類特徵醫院層級多屬於「醫學中心」及「區域醫院」且科別多為「心臟血管內科」。上述分析結果發現「內科」及「神經科」應屬於開立慢性病處方之族群反而未開立處方，對於此比例較少但卻又是最值得注意之未開立慢性病處方之患者族群造成遺漏，對健保主管單位之策略制定與資源之運用方向造成錯誤之引導影響程度，將遠大於預測為未開立處方但時際上卻開立處方之患者，應該將此誤差降到最低以提高模擬的準確性。

#### 四、決策樹規則集

本研究在試驗不同的最小支持度（Minimum Support）1000、2500、3000與最小信賴度（Minimum Confidence）75%、80%、85%，下所得到的規則數量現象，較符合本研究需求的最小支持度為2500，最小信賴度75%，得到六條描述符合開立而未開立慢性病連續處方箋之規則，整理如表8符合開立慢性病連續處方箋條件而未開立規則集，二十一條描述選擇開立慢性病連續處方箋之規則，整理如表9符合開立慢性病連續處方箋條件之開立規則集，關聯規則結果顯示，科別、醫院層級、醫事機構區域別、醫事機構權屬別、醫師年齡、醫師每日平均看診量、醫師性別、慢性病疾病範圍、患者年齡等9項因子與影響開立或不開立慢性病連續處方箋有關。

表8：符合開立慢性病連續處方箋條件而未開立規則集

編號	規則	支持度	信賴度百分比
1	若 科別 = 「精神科、整形外科、洗腎科、麻醉科、核子醫學科、新生兒科、疼痛科、居家護理、放射診斷科、脊椎骨科」	90,058	98.5%
2	若 科別 = 「內科」 且 醫院層級 = 「地區醫院」 且 醫事機構區域別 = 「北區」 且 醫事機構權屬別 = 「私立醫院」 且 醫師每日平均看診量 = 「31人-150人」	5,868	80.6%
3	若 科別 = 「泌尿科」 且 醫師性別 = 「男」 且 醫院層級 = 「地區醫院」 且 醫師年齡 = 「41-80歲」	4,514	79.2%
4	若 科別 = 「心臟血管內科」 且 醫院層級 = 「地區醫院」 且 醫師年齡 = 「51-60歲」 且 醫事機構權屬別 = 「財團法人醫院」	4,086	79.1%
5	若 科別 = 「內科」 且 醫院層級 = 「地區醫院」 且 醫事機構區域別 = 「高屏」 且 慢性病疾病範圍 = 「內分泌及代謝疾病」 且 醫師年齡 = 「41-60歲，71-80歲」	3,282	78.7%
6	若 科別 = 「腸胃內科」 且 醫院層級 = 「地區醫院」 且 醫師每日平均看診量 = 「51人-70人」 且 醫師性別 = 「男」	2,655	85.5%

表9：符合開立慢性病連續處方箋條件之開立規則集

編號	規則	支持度	信賴度百分比
1	若 科別 = 「心臟血管內科」 且 醫院層級 = 「醫學中心」 且 醫事機構區域別 = 「台北」 且 醫師年齡 = 「21-50歲、71-90歲」	17,776	88.1%
2	若 科別 = 「心臟血管內科」 且 醫院層級 = 「區域」 且 醫事機構區域別 = 「台北」 且 醫師每日平均看診量 = 「1人-30人、71人以上」	10,798	88.2%
3	若 科別 = 「心臟血管內科」 且 醫院層級 = 「醫學中心」 且 醫事機構區域別 = 「高屏」 且 醫師年齡 = 「21-60歲」	7,589	97.7%
4	若 科別 = 「心臟血管內科」 且 醫院層級 = 「醫學中心」 且 醫事機構區域別 = 「北區、南區」	7,562	97.1%
5	若 科別 = 「神經科」 且 醫院層級 = 「醫學中心」 且 醫事機構區域別 = 「台北、北區、南區、高屏」	6,790	82.5%
6	若 科別 = 「心臟血管內科」 且 醫院層級 = 「區域醫院」 且 醫事機構區域別 = 「南區」 且 醫師每日平均看診量 = 「1-30人、51人以上」	6,656	90.4%
7	若 科別 = 「心臟血管內科」 且 醫院層級 = 「醫學中心」 且 醫事機構區域別 = 「台中」 且 醫師年齡 = 「21-50歲、61-80歲」	5,756	78.8%
8	若 科別 = 「家醫科」 且 醫院層級 = 「區域醫院」 且 醫師每日平均看診量 = 「1人-30人」 且 醫師年齡 = 「21-40歲、71-80歲」	4,839	83.4%
9	若 科別 = 「內分泌科」 且 醫院層級 = 「區域醫院」 且 醫事機構權屬別 = 「財團法人醫院」 且 醫師每日平均看診量 = 「51-70人、150以人」	4,750	75.6%
10	若 科別 = 「內分泌科」 且 醫院層級 = 「醫學中心」 且 醫師性別 = 「女」 且 醫事機構權屬別 = 「財團法人醫院」	4,569	92.6%
11	若 科別 = 「神經科」 且 醫院層級 = 「醫學中心」 且 醫師年齡 = 「41-50歲」 且 醫師性別 = 「男」 且 醫事機構區域別 = 「北區、南區、高屏、東區」	4,224	82.7%
12	若 科別 = 「內分泌科」 且 醫院層級 = 「醫學中心」 且 醫師性別 = 「男」 且 醫事機構區域別 = 「台北」 且 醫師每日平均看診量 = 「71-150人」	4,135	90.5%

表9：符合開立慢性病連續處方箋條件之開立規則集（續）

編號	規則	支持度	信賴度百分比
13	若 科別 = 「內分泌科」 且 醫院層級 = 「醫學中心」 且 醫師性別 = 「男」 且 醫事機構區域別 = 「高屏」	4,133	87.5%
14	若 科別 = 「神經科」 且 醫院層級 = 「區域醫院」 且 醫師年齡 = 「31-40歲」 且 醫事機構區域別 = 「台北、北區、東區」	3,868	80.7%
15	若 科別 = 「心臟血管內科」 且 醫院層級 = 「區域」 且 醫事機構區域別 = 「北區」 且 醫事機構權屬別 = 「財團法人、私立醫院」	3,651	86.6%
16	若 科別 = 「內科」 且 醫院層級 = 「醫學中心」 且 醫師年齡 = 「21-40歲、61-80歲」	3,556	91.2%
17	若 科別 = 「家醫科」 且 醫院層級 = 「醫學中心」 且 醫師每日平均看診量 = 「1人-30人」 且 醫事機構區域別 = 「台北、北區、南區、東區」	3,433	86%
18	若 科別 = 「心臟血管外科」 且 慢性病疾病範圍 = 「循環系統疾病」 且 醫院層級 = 「醫學中心」 且 醫師年齡 = 「21-60歲」	3,286	87.8%
19	若 科別 = 「眼科」 且 醫事機構權屬別 = 「財團法人醫院」 且 醫院層級 = 「醫學中心」 且 醫師年齡 = 「21-50歲、71-90歲」	2,824	80.2%
20	若 科別 = 「婦產科」 且 醫師每日平均看診量 = 「1人-30人」 且 慢性病疾病範圍 = 「其他」 且 醫院層級 = 「區域醫院」 且 患者年齡 = 「41-60歲」	2,805	83.1%
21	若 科別 = 「心臟血管內科」 且 醫院層級 = 「區域醫院」 且 醫事機構區域別 = 「東區」 且 醫事機構權屬別 = 「財團法人醫院」 且 醫師年齡 = 「21-50歲」	2,596	91.1%

## 五、分析與討論

### （一）全國醫院慢性病連續處方箋開立狀況

針對全國醫院慢性病連續處方箋開立狀況研究主要有四個構面，本研究透過描述性統計了解全國醫院慢性病連續處方箋開立狀況，各構面之相關分析討論如下：

#### 1. 全國醫院健保資料醫院屬性描述性統計顯示：

- (1) 就醫院層級別而言以醫學中心之未開立比例7.26%最低，區域醫院10.22%、地區醫院18.20%，與醫學中心核心任務提供急重、難症醫療服務，減少慢性病及

輕症患者就醫資源之目標與政策是相符的。

- (2) 醫院權屬別中之私立醫院開立比例84.24%，較財團法人醫院91.75%與公立醫院88.52%均低，未開立比例較財團法人醫院與公立醫院高。針對上述統計結果分析，就以醫院設立的精神來說，財團法人醫院與公立醫院其實是很重的公益性質，意味著其營收是不可分配，且其資金運用是有較嚴謹法規之規範。但在私人醫院方面，因醫院間競爭壓力下，醫療成本與收入營收之問題，往往造成醫院財務面極大影響，如慢性病連續處方之開立對於診察費用與其他檢驗檢查費用申報之變化，對醫院營運成本結構間接會有不同之營運策略產生，所以私立醫院較低之慢性病連續處方開立是直得深入研究另一個議題。

#### 2. 全國醫院健保資料醫師屬性描述性統計顯示：

- (1) 就醫師性別而言，因男性醫師開立之處方數遠大於女性醫師開立處方數，故整體而言男性醫師開立與未開立人次百分比均大於女性醫師。但就男女醫師開立與未開立百分比而言，男性開立百分比89.39% 與未開立百分比10.61% 及女性開立百分比90.89% 與未開立百分比9.11% 兩者較無顯著差異性。
- (2) 在醫師年齡部份，以21歲至30歲之醫師未開比例3.90% 最低。此項統計結果顯示，目前台灣醫師從醫學院7年畢業至醫院服務約30歲，算是醫界新鮮人行醫模式容易被塑造，自我意識較不強烈，健保政策配合度較高，故產生開立慢性病連續處方箋比率有較其他年齡層的醫師高之傾向。
- (3) 就醫師每日平均看診量而言，醫師每日平均看診量150人以上開立比例96.39% 最高，1-30人87.78%最低，此現象除符合統計學上的常態分配，當就醫人數增加產生機率增高的現象外，也有可能因為就醫患者多時，醫師為稀釋患者，故釋放慢性病連續處方箋給社區藥局，故有較高開立慢性病連續處方箋情形。另就醫師科別而言，其中精神科符合開立條件而未開立百分比高達98.8%，顯示該科存在異於其他科別隱含之資訊，例如人的心理變化比身體大，因此精神科醫師會比其他科系醫師更希望隨時常握患者病情，故開立慢性病連續處方箋比率較低。

#### 3. 全國醫院健保資料病患屬性描述性統計顯示：

- (1) 男性患者開立百分比89.22%與未開立百分比10.78%及女性患者開立百分比89.89%與未開立百分比10.11%兩者與醫師之性別特性類似，較無顯著差異性。
- (2) 在患者年齡部分，較多患者集中老人組65歲以上，開立處方人次占總開立處方人次47.64%，未開立處方人次也佔總未開立處方人次45.30%，反映出高齡是慢性病的高危因子，老年人隨著年齡的增加，導致各種生理機能的退化，罹患慢性疾病的機率相對比年輕人高，相對開立與未開立之慢性病連續處方箋比率會較其他年齡層高。

#### 4. 全國醫院健保資料疾病屬性描述性統計顯示，有關循環系統疾病就占有所有疾病別之開立處方人次41.88%，精神疾病未開立百分比占有所有疾病之未開立百分比49.55%，與各科資料描述性統計精神科應開而未開立百分比 98.8%同為最高比

例，顯示精神科與精神疾病，均存有異於其他科別及疾病隱含之資訊，值得未來進一步研究與探討。

## (二) 決策樹模型規則

本研究以上述全國醫院健保資料之描述性統計結果為基礎，再透過樣本抽樣，以資料探勘技術方式研究，進一步發掘隱含在描述性統計結果中之慢性病處方開立之相關規則，從決策樹模型歸納出之規則表8與表9中發現了一些現象，整理討論後分別敘述如下：

1. 表8符合開立慢性病連續處方箋條件未開立規則集，分析第1條分類規則，未開立之各類科別各有其特性：
  - (1) 精神科之精神病患本身疾病的複雜度極高，是極度需要受關懷的一群，藉由定期回診會談醫師可以了解用藥狀況與舒解病患的情緒。
  - (2) 而洗腎科之患者則是每週需固定回診2-3次。
  - (3) 麻醉科、核子醫學科、放射診斷科屬檢查科系。
  - (4) 整形外科、新生兒科、疼痛科、脊椎骨科以處理急性病為主。
  - (5) 居家護理為每2週或1個月即接受醫療和護理服務之科別故開立率低。

在模型資料集中有90,058筆，98.5%的比率選擇不開立，以上規則集條件中之就醫科別，大多均不屬於內科系，與內科系之常見就醫特性描述慢性且病情較穩定之疾病特徵有所區隔，所以不屬於開立慢性病連續處方箋族群，此一規則結果描述是值得參考。

2. 表8符合開立慢性病連續處方箋條件而未開立規則集，規則2至6顯示，當醫院層級為地區醫院時有較低的開立情況，而這個結果跟表2樣本數醫院屬性描述性統計結果是一致的。
3. 表9符合開立慢性病連續處方箋條件之開立規則集顯示，若科別為心臟血管內科、內分泌科、神經科，且醫院層級分別為醫學中心和區域醫院時有較高的開立比率。
4. 綜整開立與未開立兩群規則分析，影響開立慢性病連續處方箋之各項因子中，未開立之六條分類規則，每條規則皆含醫院層級別與就醫科別，且醫院層級別均為「地區醫院」。在開立之二十一條分類規則中，每條規則也皆含醫院層級別與就醫科別，醫院層級之分類條件為「醫學中心」共計13項佔62%，醫院層級之分類條件為「區域醫院」共計8項佔28%，而「地區醫院」占率為0%。結果顯示醫院層級別與就醫科別佔有較重要之影響。其次較有影響之因子為醫事機構區域別及醫師年齡。

由上述之研究結果顯示，決策樹對開立慢性病連續處方箋開立與否之相關性能有具體之描述規則，提供醫療主管單位不同的資訊，以作為推廣慢性病連續處方開立之政策制定參考。

## (三) 敏感度分析

本研究透過決策樹模型得知醫院層級別與就醫科別兩因子對慢性病連續處方箋開立

否佔有重要之影響，故藉由敏感度分析比較決策樹模型其整體分類正確率、遺漏性誤差及過度預測誤差之變化，進而確認上述二因子對研究目標的影響程度，敏感度分析如表10。

由表10中發現，當建構決策樹模型時，若分別排除醫院層級別與就醫科別二因子，整體分類正確率均下降，遺漏性誤差與過度預測誤差均上升，代表醫院層級別與科別兩因子對於慢性病連續處方箋開立否的鑑別能力是有重要影響的。

表10：敏感度分析表

決策樹模型輸入因子	整體分類正確率	遺漏性誤差	過度預測誤差
如圖2研究架構圖之解釋欄位共計十項	79.31%	21.87%	19.51%
如圖2研究架構圖之解釋欄位剔除科別因子共計九項	76.13%	25.44%	22.29%
如圖2研究架構圖之解釋欄位剔除醫院層級別因子共計九項	76.95%	25.07%	21.02%

## 伍、結論與建議

本研究以健保局2007年全國醫院層級醫療院所之慢性病案件申報紀錄為研究材料，過濾出符合開立慢性病連續處方箋條件之開立與未開立慢性連續處方箋之記錄，運用C5.0決策樹演算法之分類功能，將慢性病連續處方箋開立與否之影響因子生成決策樹模型，使用10折交互驗證法來提升分類模型的正確率，採用錯差矩陣進行模型評估。在模型建構後其整體正確率達79.31%，機會正確率50%，整體正確率大於機會正確率，顯示本研究決策樹模型對開立或未開立慢性連續處方箋可以有效的區分。

透過決策樹模型規則集表8與表9發現，有六條分類條件描述符合開立條件而未開立慢性連續處方箋之規則，分析發現在醫院權屬別、醫事機構區域別、醫院層級別、醫師年齡、醫師性別、就醫科別、醫師平均日看診量、患者年齡、患者性別、慢性病疾病範圍十項因子中，有相關影響性之因子共八項（不包含患者年齡與患者性別）。有二十一條分類條件描述符合開立條件而選擇開立慢性連續處方箋之規則，有相關影響性之因子共九項（不包含患者性別），而十項因子中，就醫科別與醫院層級別二因子對開立與未開立慢性連續處方箋，最具有決定性影響。此外從描述性統計發現精神科符合開立條件而未開立的比率高於其他科系，顯示該科存在異於其他科別隱含之資訊，而心臟血管科、內分泌科、神經科為國人盛行之心臟病、糖尿病、腦血管疾病之看診科系，開立慢性連續處方箋之比率相對較高，醫院層級別未開立比率最高者依序為地區醫院、區域醫院、醫學中心。此決策樹模型產生以上之資訊透過系統化，可將資料轉化為知識，經由知識傳遞與分享，輔助決策者及全民健保相關業管單位做更精準之判斷。

由於健保資料庫，僅有保險給付相關資訊，不包括非保險給付之處方，亦不包含投保人自覺健康資訊、家族成員疾病史、重要社經、生命變數。外在因素方面，不同醫師對相同疾病之臨床專業判斷與診斷、病患病情之主觀描述方式不一，上述之情形無法於



健保局資料庫取得。另外，有部份藥品之藥理分類為未分類，所以藥理分類變數的資料並不完整，所以本研究未將藥理分類變數變數列入。最後，用藥金額區間為0~460,897元，決策樹模型建構加入用藥金額後，生成之決策樹太大，其分支很多，所以本研究也未將醫療費用相關變數列入。所以讀者在解讀與應用研究結果時應注意這些限制。

## 誌謝

感謝中央健康保險局（97年度獎勵大學院校碩、博士班研究生研究全民健康保險相關專題論文活動）經費補助本研究之進行。

## 參考文獻

1. 中央健康保險局，2007，『96年第4季醫院總額專業醫療服務品質報告』，6月12日（available online at [http://www.nhi.gov.tw/resource/Webdata/Attach\\_11606\\_2\\_96年第4季醫院總額970530.pdf](http://www.nhi.gov.tw/resource/Webdata/Attach_11606_2_96年第4季醫院總額970530.pdf)）。
2. 王偉麟、林文燦、賴政皓、陳慧敏，2008，『應用資料探勘技術提升急診醫學檢傷分類之一致性—以台灣某醫學中心急診醫學部為例』，品質學報，第十五卷·第四期：283~291頁。
3. 何蘊芳、林慧玲、蔡瑜珍、邱士峰、賴玉花、何富蕙、林芬如，2006，『門診患者對慢性病連續處方箋的認知』，臺灣醫學，第十卷·第五期：578~585頁。
4. 何蘊芳，2004，『慢性病連續處方箋釋出成效之探討與推廣（第一年）』，行政院衛生署委託研究計畫。
5. 張雅雯，2002，『醫療利用可近性—台灣老人之實證研究』，國立中央大學產業經濟研究所碩士論文。
6. 蔡瑞真，2005，『南部地區慢性病連續處方調劑現況及相關因素分析』，行政院衛生署委託研究計畫。
7. 劉嘉年、楊志良，2006，『門診醫師以抗生素治療上呼吸道感染症與急性支氣管炎的影響因素與介入策略』，臺灣公共衛生雜誌，第二十五卷·第五期：330~339頁。
8. 簡禎富、林國勝，2006，『建構cDNA生物晶片之二元資料挖礦模式及其實證研究』，資訊管理學報，第十三卷·第四期：133~159頁。
9. 戴建耘、盧治均、廖秋惠，2007，『應用Data Mining建置一分類模型』，Electronic Commerce Studies，第五卷·第一期：109~123頁。
10. 龔佩珍、呂嘉欣、蔡文正，2007，『基層醫師釋出慢性病連續處方箋之意願及相關因素』，臺灣公共衛生雜誌，第二十六卷·第一期：26~37頁。

11. Berry, M.J.A., and Linoff, G. "Data Mining Technique for Marketing, Sale and Customer Support," Wiley Computer, 1997.
12. Brossette, S.E., and Hymel, P.A. Jr. "Data mining and infection control," *Clin Lab Med* (28:1), 2008, pp.119-126.
13. Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., and Zanasi, A. "Discovering Data Mining From Concept to Implementation," Prentice-Hall Inc, 1997.
14. Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. "The KDD Process for Extracting Useful Knowledge from Volumes of Data," *Communications of the ACM* (39:11), 1996, pp.29.
15. Hair, J. F., Anderson, R.E., Tatham, R.L., and Black, W.C. "Multivariate Data Analysis With Readings," Prentice Hall International Editions, Fourth Edition, 1995.
16. Han, J., Kamber, M., and Pei, J. "Data Mining: Concepts and Techniques," Morgan Kaufmann, 2000.
17. Hui, S.C., and Jha, G. "Data mining for customer service support," *Information and Management* (38:1), 2000, pp.1-13.
18. Letourneau, S., Famili, F., and Matwin, S. "Data mining for prediction of aircraft component replacement," *IEEE Intelligent Systems and their Applications* (14:6), 1999 , pp.59-66.
19. Nattkemper, T. W., Arnrich, B., Lichte, O., Timm, W., Degenhard, A., Pointon, L., Hayes, C., and Leach, M.O. "Evaluation of radiological features for breast tumor classification in clinical screening," *Artificial Intelligence in Medicine* (34:2), 2005 , pp.129-139.
20. Park, M., Park, J.S., Kim, C.N., Park, K.M., and Kwon, Y.S." Knowledge discovery in nursing minimum data set using data mining," *Taehan Kanho Hakhoe Chi* (36:4), 2006, pp.652-661.
21. Quinlan, J.R. "C5.0 Online Tutorial," 2003 (available online at <http://www.rulequest.com>).
22. Safavian, S. R., and Landgrebe, D. "A Survey of decision tree classifier Methodology," *IEEE Transactions on Systems* (21:3), Man and Cybernetics, 1991, pp.660-674.