

運用資料探勘技術於六大死因慢性疾病之研究

吳瑞堯

世新大學資訊管理學系

周駿賢

世新大學資訊管理學系

摘要

心臟疾病、腦血管疾病、糖尿病、肝臟病變、腎臟病變、高血壓性疾病等慢性疾病一直名列我國國民重大死因，衛生署公布的「2008年國人十大死因」疾病中，慢性疾病佔總死亡人數的六成以上，顯示慢性疾病已成為當前危害國人身體健康的致命殺手。國家衛生研究院的全民健康保險研究資料庫包含了大量的病患門診就醫資料，這些資料中可能隱藏許多未知而有用的知識，因此本研究以該資料庫為基礎，經資料淨化與篩選後，進行統計分析，找出上述六大死因慢性疾病在不同生活圈、年齡層及性別上之盛行率，並利用資料探勘技術找出不同生活圈、性別、年齡之慢性疾病間的關聯性。

研究結果發現，高血壓性疾病盛行率男女性分別為11.37%、9.24%，均居6種慢性疾病之冠，65歲以上族群盛行率更高達37.61%，且高血壓性疾病不分生活圈、性別、年齡，均與其他慢性疾病密切關聯；男性在6種慢性疾病的盛行率均高於女性，但離島生活圈女性在高血壓性疾病、心臟疾病、糖尿病之盛行率皆高於男性；一般生活圈65歲以上高血壓性疾病及心臟疾病盛行率高達47.79%及30.11%，明顯高於都會生活圈；離島生活圈65歲以上高血壓性疾病及腦血管疾病盛行率達59.37%及21.87%，明顯高於另兩個生活圈；肝臟病變在40~64歲之盛行率高達10.88%，反而高於65歲以上的8.03%；心臟疾病在65歲以上族群中與其它慢性疾病間都有高度關聯；在40~64歲中患有腎臟病變及肝臟病變者很可能同時會有高血壓性疾病；女性心臟疾病與高血壓性疾病、糖尿病間的關聯較男性明顯。

關鍵字：慢性疾病、全民健康保險研究資料庫、資料探勘

The Study of the Six Major Causes of Death Chronic Diseases by Data Mining

Rei-Yao Wu

Department of Information Management,
Shih Hsin University

Chun-Hsien Chou

Department of Information Management,
Shih Hsin University

Abstract

Chronic diseases including Heart Disease, Cerebrovascular Disease, Diabetes, Liver Disease, Renal Disease, and Hypertension Disease are always listed in the top ten leading causes of death. Chronic diseases account for more than 60% of death toll according to the report of ten leading causes of death in 2008 published by the Department of Health. This fact shows that the chronic diseases have become the deadly health killer currently. The National Health Insurance Research Database that is maintained by the National Institutes of Health has accumulated a large volume of patient treatment information, and maybe hidden useful knowledge there. Applying data cleaning, analysis and mining, this research investigates the population rate of these chronic diseases and the association rules among them in each life circles, sexes, and ages.

The research reveals that the Hypertension Disease is the most popular disease among the six chronic diseases. The population rates are 11.37% in male and 9.24% in female, and 37.61% in the above 65 years old group. Despite of life circle, sex, and age, it is tightly associated with other chronic diseases. The population rates of male are higher than that of female for all the six chronic diseases in general. But the population rates of females in off-shore island life circle are greater than that of males there in Hypertension Disease, Heart Disease and Diabetes. The population rates of the above 65 years old group in the general life circle are 47.79% for Hypertension Disease and 30.11% for Heart Disease, highly beyond that in the metro life circle. The population rates of the above 65 years old group in the off-shore island life circle are 59.37% for Hypertension Disease and 21.87% for Cerebrovascular Disease, respectively, also highly beyond that in the other two life circles. The population rate of the 40~64 years old group

for Hepatitis and Cirrhosis is 10.88%, higher than 8.03% of the above 65 years old group. In the above 65 years old group, Heart Disease is highly associated with the other chronic diseases. In the 40~64 years old group, Renal Disease and Liver Disease are associated with Hypertension Disease. And, the association of Heart Disease with Hypertension Disease and Diabetes in female is more significant than that in male.

Key words: Chronic Disease, National Health Insurance Research Database, Data Mining.

壹、緒論

一、研究背景

隨著現代人生活水準的提高及飲食習慣的改變，「吃的 unhealthy、運動的少」似乎已成為大部分人的通病，相關的文明病也因此油然而生，其中最引人關注的疾病就是慢性疾病。內政部統計處2005年針對50歲以上「臺閩地區老人狀況調查」(內政部統計處 2008)發現「50~64歲國民患有慢性病或重大疾病者占38.46%，65歲以上老人患有慢性病或重大疾病者占65.20%」。2008年衛生署的統計資料(行政院衛生署統計資訊網 2009)顯示，台灣地區2008年十大死因疾病中，慢性疾病就占了六項，占總死亡人數的六成以上。行政院衛生署國民健康局統計(行政院衛生署國民健康局網站 2008)，2008年三高(高血壓、高血脂、高血糖)病患共吃掉323億元藥物，佔所有藥物費用的四分之一(1,250億元)。以上數據顯示，慢性疾病的預防與治療，已成為個人與衛生單位刻不容緩的重要工作。

二、研究動機

高血壓性疾病與其他慢性病有高度關聯性，糖尿病患者得到高血壓的機會是非糖尿病患者的1.5至2倍，相當多糖尿病患者會合併有高血壓(中華民國糖尿病衛教學會網站 2009)。許多慢性疾病間的關聯在既有醫學研究中都已有認知，但過去慢性疾病相關研究，大多以病理分析探討或只對小規模調查對象進行統計分析，鮮少運用大規模就醫資料進行研究。中央健保局自民國89年起委託國家衛生研究院製作「全民健康保險研究資料庫」，紀錄國人就醫紀錄，並且每年維護更新。其中的「抽樣歸人檔」以所有承保戶為抽樣母體進行隨機抽樣，記錄每個樣本的所有就醫記錄，是進行大規模疾病研究的最佳資料來源。

資料探勘是從大量的資料中找出未知、正確且有用資訊或知識的技術，常用於挖掘商業知識，近年來也被運用於挖掘醫學知識，姚志成(2005)運用資料探勘技術建立脂肪肝之預測模式；張俊郎等(民96)結合類神經網路與決策樹發掘出糖尿病前期診斷的預測規則。

「全民健康保險研究資料庫」擁有大量的就醫記錄，是從事挖掘醫學知識的最佳資料來源，因此本研究運用資料探勘技術，針對心臟疾病、腦血管疾病、糖尿病、慢性肝炎及肝硬化、腎臟病變、高血壓性疾病等國人六大死因慢性疾病進行疾病關聯分析，以過去未曾使用過的研究方法，找出醫學上未曾發現的新知識並佐證既有的醫學知識，提供個人疾病預警及醫師臨床診斷之參考。

三、研究目的

本研究以全民健康保險研究資料庫為基礎進行心臟疾病、腦血管疾病、糖尿病、肝臟病變、腎臟病變、高血壓性疾病等國人六大死因慢性疾病研究，以便獲得以下知識：

- (一) 各種慢性疾病的盛行率，及分年齡、性別及生活圈之盛行率。
- (二) 國人六大死因慢性疾病間之關聯性。
- (三) 個別生活圈、年齡層、性別族群中慢性疾病間之關聯性。

論文第一章說明研究背景、動機與目的；第二章進行慢性疾病、資料探勘相關文獻探討；第三章說明本研究所採用的研究方法與架構；第四章敘述資料探勘及分析結果；第五章對研究結果提出結論並對進一步研究提出建議。

貳、文獻探討

一、全民健康保險研究資料庫

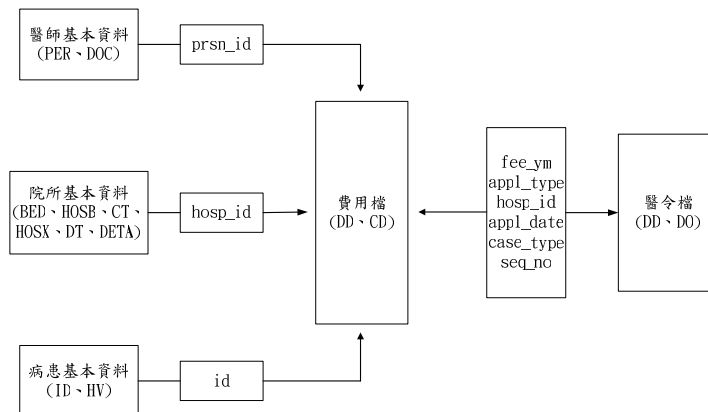


圖1：全民健保資料的三大子資料庫

資料來源：連賢明(2008)

健保局自2000年起，委託財團法人國家衛生研究院建置全民健康保險研究資料庫，其內容包含了非常完整的國人就醫記錄。共分為三大子資料庫，連賢明(2008)列出該資料庫中三大子資料庫架構圖，如圖1，包括費用檔、醫令檔、基本資料檔。因資料內容龐大繁雜，所以國家衛生研究院針對不同研究需求，分別製作以下五種資料檔(全民健康保險研究資料庫網站 2008a)：

- (一) 基本資料檔：包含醫事機構基本資料檔、醫事機構診療科別明細檔、醫事機構病床主檔、專科醫師證書主檔、醫事人員基本資料檔、重大傷病證明明細檔，以及門診、住院費用總表等資料檔。

- (二) 系統抽樣檔：以每次門診或住院為抽樣對象的門診或住院資料檔。
- (三) 抽樣歸人檔：以被保險人為抽樣對象的被保險人就醫記錄。
- (四) 特定主題分檔：以不同疾病、病患身分及醫療層級區分之主題就醫記錄分檔。
- (五) 教學用資料檔：含85年至90年共6年間每年1000人的就醫資料，免費提供教師教學使用。

為保護個人隱私，各資料檔內之身分證字號、醫師代號及醫療機構代號欄位均已加密處理，研究者需利用國家衛生研究院提供之譯碼簿進行資料處理，譯碼簿提供健保資料庫中各種資料檔之欄位資訊，例如序號、英文欄位、中文欄位、資料型態、資料長度及資料起始位置。各種資料檔皆為文字檔，使用者可以將所需之資料檔匯入資料庫後，再依譯碼簿提供之欄位資訊進行資料淨化與篩選作業。

二、六大死因慢性疾病統計分析與相關性之研究

根據行政院衛生署公布的2008年國人十大死因，如表1，其中除肺炎、事故傷害、自殺外，惡性腫瘤、心臟疾病、腦血管疾病等慢性疾病就佔了6項，與1952年國人十大死因中慢性疾病僅佔3項相比，有顯著的增加，顯示這些疾病是當前威脅國人生命的致命殺手。慢性疾病需要經由長期的治療與藥物控制，加上生活形態與飲食習慣的改善，才能獲得控制，除了對個人造成生理與心理上的影響外，對家庭負擔、社會醫療成本支出亦產生深遠影響。

表1：2008年國人十大死因排名

排名	疾病名稱	排名	疾病名稱
1	惡性腫瘤	6	意外事故
2	心臟疾病	7	慢性下呼吸道疾病
3	腦血管疾病	8	慢性肝炎及肝硬化
4	肺炎	9	自殺
5	糖尿病	10	腎炎、腎徵候群及腎性病變

資料來源：行政院衛生署統計資訊網

心臟疾病涵蓋範圍廣，主要有心肌梗塞、瓣膜狹窄、板模閉鎖不全、心肌炎等疾病。徐敏耀等人(2007)利用統計方法針對某區域醫院1288位65歲以上老人進行研究，發現潛在冠狀動脈心臟疾病之危險因子依序為性別、高血壓、年齡及糖尿病等，其中有高血壓、糖尿病史且吸菸之男性，罹病機率高達98.2%；民國96年4月台北市政府衛生局發布之新聞稿揭露心臟疾病死亡率由民國90年之每十萬人中52人，增至94年之64人，增幅為21%。心血管疾病是心臟病與血管疾病的總稱，依年齡層來看，90年至94年因心血管疾病死亡者集中在65歲以上市民，約佔8成1，45-64歲約佔1成6，顯示心血管疾病死亡率隨著年齡增加而增加，若依性別區分，這五年因心血管疾病而死亡者，男性佔6成1，女性僅佔3成9。

腦血管疾病俗稱中風，危險因子有高齡、高血脂、糖尿病、心臟疾病、高血壓、吸菸、喝酒、缺乏運動等；腦中風類型可以分為缺血性中風及出血性腦中風兩種，缺血性中風為腦部的血管被血栓阻塞而造成局部血液循環不足，出血性腦中風為血管破裂而使得血塊累積在腦部造成，男性45歲以上，女性55歲以上為腦中風危險群，男性中風機率較女性高，家族史方面，有親人在55歲以前，或有女性親人在65歲以前發生心肌梗塞或猝死者，亦為高危險群。

糖尿病是一種慢性代謝異常疾病，主要原因是由於體內胰島素缺乏或功能不全，對醣類或葡萄糖的利用能力減低，造成血糖上升。糖尿病的併發症有視網膜病變、腎臟病變、末梢神經病變等疾病。糖尿病腎病變在第1型及第2型糖尿病病人都可能發生，大約有25~45%左右的糖尿病患者會併發腎病變，而且糖尿病腎病變及慢性腎絲球腎炎都是尿毒症的最主要原因。另外，依據流行病學之統計，糖尿病患者中大約40~50%的人同時患有高血壓，而高血壓患者中也約有10-14%的人患有糖尿病。糖尿病的發生率隨年齡增加而增高，40歲以上發生率顯著提升。盧美秀與林秋香(1991)指出女性患者血糖控制情形較男性患者差。賴世偉與黃金財(2004)發現整體糖尿病盛行率為11%，其中女性9.6%，男性為12.7%。糖尿病與心臟疾病有密切關係，簡國龍(2005)認為冠狀動脈心臟病是糖尿病患者重要的併發症及死因。

肝臟病變指慢性肝炎及肝硬化，慢性肝炎是指肝細胞持續發炎超過六個月以上，且無緩解及改善的跡象，而肝硬化是肝臟全面性纖維化合併結節形成，是慢性肝炎的結果。慢性肝炎及肝硬化都是肝癌的危險因子，受到國人飲酒及飲食習慣西化影響，酒精性肝病及脂肪肝問題日益嚴重，也成為形成肝癌的危險因子。在慢性肝炎中，B型肝炎是全世界罹患人口最多，病程最複雜，且與肝硬化及肝癌密切相關的肝炎。台灣地區是B型肝炎的嚴重感染區，全台灣約有15~20%的帶原者，遠比歐美地區不及1%嚴重，對國民健康構成了重大的威脅。劉如玲(民96)利用2004年4月至11月間台南縣社區健康普查資料進行第二型糖尿病與C型肝炎之相關性研究，患有C型肝炎的人，有較高的糖尿病盛行率。另外，肝硬化與心臟疾病亦有所關聯，吳宜珍等(2005)發現，肝硬化患者在出血時較容易併發心肌受損。范宜玲等(2009)綜整相關文獻指出，30%~50%肝硬化病患會出現心臟增大，稱為肝硬化性心肌症，若出現心力衰竭，稱為肝硬化性心力衰竭，在臨床上卻常被忽視。

腎臟病的種類繁多，包括腎炎、腎徵候群及腎性病變。糖尿病、高血壓及全身性紅斑性狼瘡也常併發腎臟病變。腎臟病人常併發高血壓，而高血壓本身又會造成腎功能惡化，兩者間成為一個惡性循環現象，因此高血壓的控制對腎臟病人極為重要(陳鴻鈞2001)。高血壓會升高血管壁的壓力，使得腎臟無法有效過濾廢物，導致慢性腎臟病的發生及惡化，而慢性腎臟病也會造成身體血壓上升，包括腎炎、多囊腎、水腎、糖尿病腎臟病變及腎動脈狹窄等腎臟疾病，皆會引起高血壓。腎臟疾病與心血管疾病有密切關係，簡國龍(2009)認為慢性腎臟疾病是心血管疾病重要的危險因子。

高血壓是指在靜止狀態時量度的血壓持續地高於或等於140/90毫米水銀柱(mmHg)的現象。高血壓性疾病雖未名列2008年國人十大死因疾病，但在我國盛行率非常高，且名列2007年國人第十大死因疾病，因此本研究亦將其納入研究。高血壓患者多數無明顯的病徵，只有部份高血壓患者會偶然感到頭痛、頭暈及疲倦等。患者會出現血管變厚、變

硬及心臟工作過勞等情形。患有高血壓會增加罹患冠心病、心臟衰竭、腎衰竭及中風等併發症的機率。根據2002年行政院國民健康局的統計，台灣地區15歲以上國民高血壓盛行率在男性為24.9%，女性為18.2%，整體盛行率為21.5%，65歲以上國民盛行率高達56.6%；台灣地區約有15.5%(男10.9%，女4.6%)的民眾不知道自己患有高血壓，整體高血壓控制的狀況並不理想(控制率：男20.79%，女28.50%)。林獻鋒等人(2006)利用台中市某醫學中心2000年1月至2002年12月民眾健檢資料(共3,392人)進行高血壓盛行率之描述性統計分析，發現整體高血壓的盛行率28.8%，其中女性為25.0%，男性為31.6%。高血壓不只是最普遍的慢性病，而且與心血管疾病有密切關係，陳明豐(2009)認為高血壓是心血管疾病重要的危險因子。

三、資料探勘

資料探勘(Data Mining)可從大量的資料中找出未知、正確且有用的隱藏知識，是知識發現的重要技術。Chen et al. (1996)認為資料探勘可從資料庫中萃取出重要、事先未知、潛在有用的資訊；Berry與Linoff (1997)認為資料探勘可針對大量的資料，應用自動或半自動的方式進行分析，以找出有意義的關係或法則。Fayyad(1996)和Fayyad et al.(1996)將知識發現流程分為資料選擇、資料前置處理、資料轉換、資料探勘、解釋與評估等步驟，如圖2。Han與Kamber(2000)將資料探勘方法分為四大類，分別為關聯分析(Association Analysis)、分類與預測(Classification and Prediction)、群集分析(Clustering Analysis)、推估與偏差分析(Evolution and Deviation Analysis)。

關聯分析(Association Analysis)又稱為購物籃分析(Basket Analysis)，因擅於分析購物籃中購物品項間的關係而得名，可從眾多交易中(多個購物籃)分析出哪些商品同時被顧客購買的頻率較高，分析結果可協助業者訂定相關商品促銷策略，可應用於市場行銷及客戶社群分析等許多應用上(Khan et al. 2008)。

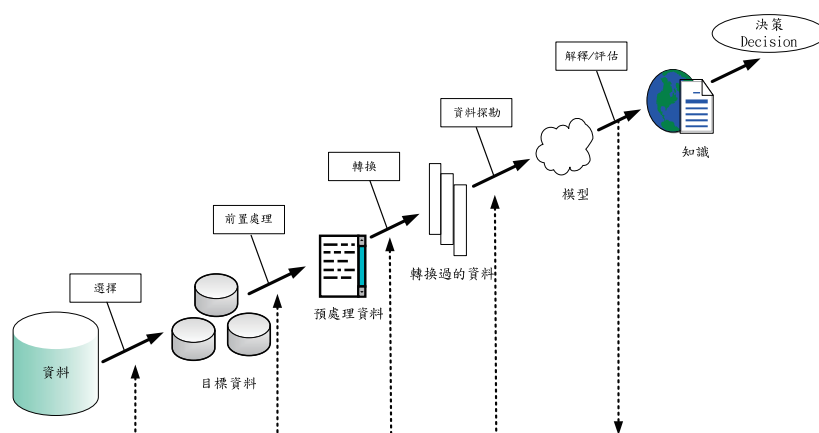


圖2：知識發現流程

資料來源：Fayyad (1996)

關聯分析由Agrawal等學者於1993年提出(Agrawal et al. 1993; Agrawal & Srikant 1994)，主要觀念建立在條件機率上，並且運用支持度(Support)與信賴度(Confidence)篩選關聯規則，以關聯規則 $X \rightarrow Y$ 而言，其支持度 $s = \text{包含項目X的交易數量}$ ，信賴度 $c = (\text{同時包含項目X、Y的交易數量}) / \text{包含X的交易數量}$ 。舉例而言，若在某賣場的交易中，60筆有購買香菸，40筆有購買啤酒，同時購買香菸及啤酒的交易有20筆，則對於買香菸就會買啤酒的關聯規則(以 $X \rightarrow Y$ 表示)來講，其支持度 $s = 60$ ，信賴度 $c = 20 / 60 = 0.33$ 。關聯分析的目的是找出具備足夠支持度與信賴度的關聯規則，所以一個關聯規則要成立必須同時滿足預先設定的最小支持度與最小信賴度。

Apriori演算法是最典型的關聯分析方法，若某賣場交易明細如表2，且設定最小支持度為2，則Apriori演算法將依下列流程進行：

- Step 1: 演算法先掃描所有交易紀錄，產生單一項目的候選項目集合(Candidate Item Sets)，如圖3步驟1所示之C1。
- Step 2: 將數量較少的項目集合排除，只保留滿足最小支持度的項目集合，稱為大型項目集合(Large Item Sets)，如圖3步驟2所示之L1。
- Step 3: 對L1中的各集合進行組合，成為具有兩個項目的集合，產生兩個項目的候選項目集合，如圖3步驟3的C2。
- Step 4: 排除C2中支持度小於2的項目集合，得到圖3步驟4的大型項目集合L2。
- Step 5: 對L2中的各集合進行組合，成為具有三個項目的集合，產生三個項目的候選項目集合，如圖3步驟5的C3。
- Step 6: 因為C3的兩個項目集合支持度都大於2，所以L3中的大型項目集合與C3相同，如圖3步驟6的L3。

表2：交易明細範例

交易編號	交易產品項目
001	A、B、E
002	B
003	B、C
004	A、B、D
005	A、C
006	B、C
007	A、C
008	A、B、C、E
009	A、B、C

資料來源：本研究整理

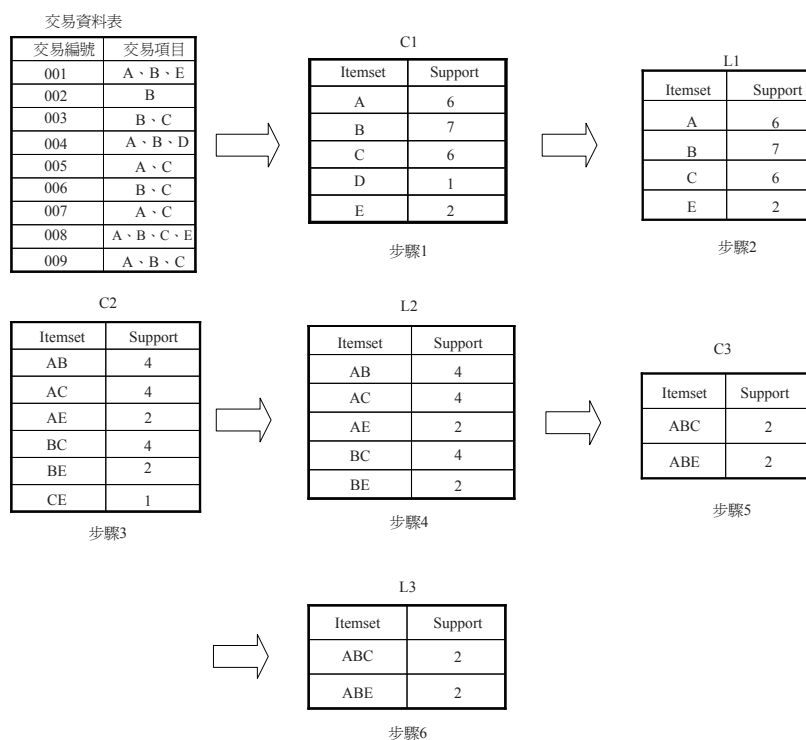


圖3：Apriori演算法範例

資料來源：本研究整理

Step 7: 因為無法再由L3產生四個項目的候選項目集合C4，結束項目集合產生程序。

Step 8: 從L2及L3中的項目集合產生符合最小信賴度的關聯規則。對每個項目集合 l 的子項目集合(非空集合) a 檢查 $a \rightarrow l-a$ 之信賴度是否高於最小信賴度，若是，則產生該關聯規則。例如對項目集合ABC，須檢視 $A \rightarrow BC$, $B \rightarrow AC$, $C \rightarrow AB$, $AB \rightarrow C$, $AC \rightarrow$, $BC \rightarrow A$ 等關聯規則是否高於最小信賴度。

Apriori演算法在產生大型項目集合的過程中需要耗費大量的時間和空間，因此許多研究提出改良方法，Zaki(2000)將大型項目集合的搜尋空間區分成若干較小的子空間而提高效能，Han et al.(2004)則將交易資料建置成FP-growth樹，將資料庫的掃描降低到僅需兩次，Song與Rajasekaran(2006)運用交易樹壓縮交易項目，得到更佳的效能。

Apriori演算法只能分析單一屬性的關聯規則，Aggarwal et al.(2002)提出可以分析多維關聯規則的演算法。另外，Chu et al.(2008)和Li et al.(2008)則將交易中各項目的交易數量列入計算支持度之考量。

參、研究方法

一、研究架構

本研究分為資料準備及資料分析兩部分進行，如圖4所示。

(一) 資料準備

此部分分兩階段，第一階段旨在釐清並確認本研究所需資料來源並進行資料淨化與篩選，第二階段再根據健保局提供之譯碼簿轉譯建置本研究分析所需之資料庫。

1. 資料淨化與篩選：本研究向國衛院申請2005至2007年之抽樣歸人檔進行研究，抽樣歸人檔包含了門診處方及治療明細檔、門診處方醫令明細檔、住院醫療費用清單明細檔、住院醫療費用醫令清單明細檔、特約藥局處方及調劑明細檔、特約藥局處方調劑醫令檔及承保資料檔(全民健康保險研究資料庫網站 2008b)，本研究所需之資料檔為門診處方及治療明細檔、承保資料檔兩個檔案，所以擷取這兩個檔案中所需的欄位。
2. 分析資料庫建置：依據國家衛生研究院提供之譯碼簿欄位說明轉譯門診處方及治療明細檔、承保資料檔相關欄位資料，匯入資料庫中，建立本研究所需之分析資料庫，其中包含病患基本資料表及病患疾病明細表兩個資料表。

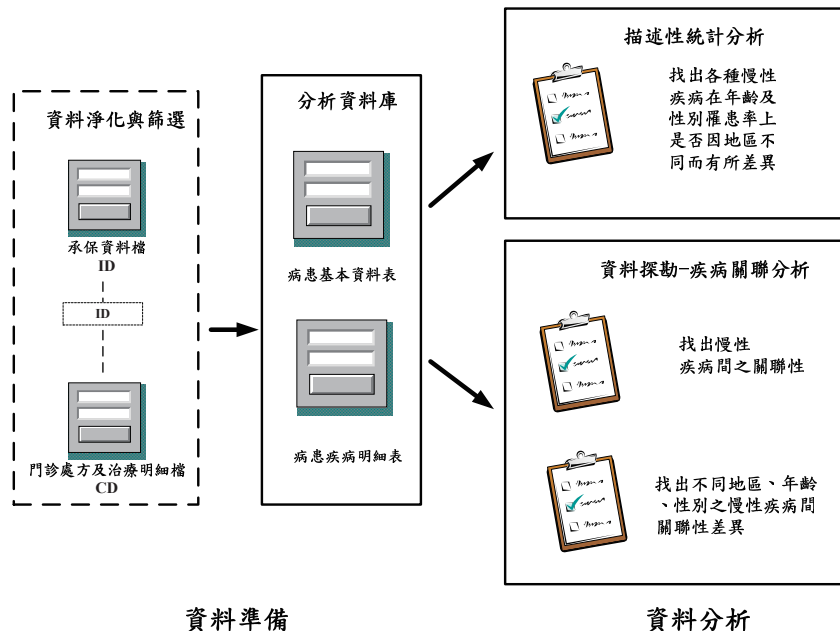


圖4：研究架構圖

(二) 資料分析

分析資料庫建立後，對病患基本資料表進行描述性統計分析，並利用病患疾病明細表進行慢性疾病關聯分析。

1. 描述性統計分析：進行統計分析，以找出各種慢性疾病在各年齡及性別盛行率上之生活圈差異。
2. 疾病關聯分析：利用病患疾病明細表找出慢性疾病間之關聯性以及分生活圈、年齡、性別之慢性疾病關聯性。

二、資料準備

(一) 資料淨化與篩選

抽樣歸入檔的產生方式是從全民健康保險在保人中隨機抽樣100萬人，而且每年將這100萬人該年度的就醫資料加入資料庫中，並將他們每4萬人分成1組，共分25組提供研究(全民健康保險研究資料庫網站2008b)，本研究所使用的就是其中1組4萬人自2005至2007年3年間的就醫資料。

抽樣歸入檔包含了門診處方及治療明細檔、門診處方醫令明細檔、住院醫療費用清單明細檔、住院醫療費用醫令清單明細檔、特約藥局處方及調劑明細檔、特約藥局處方調劑醫令檔及承保資料檔，本研究運用其中門診處方及治療明細檔、承保資料檔兩個檔案，在承保資料檔方面，2005、2006、2007年之承保資料分別有39,949、39,700及39,650筆，經整理分析後，不重複之承保戶共44,042筆。在門診處方及治療明細檔中，這三年原始資料筆數依序為601,579筆、566,825筆及570,587筆。

為取得研究用資料，本研究依據國家衛生研究院提供之譯碼簿逐列擷取承保資料檔中的身分證字號、出生日期、性別及投保單位所在地區代碼等欄位。門診處方及治療明細檔欄位相當多，但本研究僅需每次門診醫師診斷之病症，由於門診處方及治療明細檔中醫師所診斷之病症最多只有三項，所以擷取其中身分證字號、國際疾病分類代碼一、國際疾病分類代碼二及國際疾病分類代碼三等三個欄位資料。由於本研究僅對慢性疾病進行研究，非屬慢性疾病之記錄不記入資料庫中。

(二) 分析資料庫建置

在進行資料分析探勘前，必須先建立分析用資料庫。本研究所建立的分析用資料庫中包含了病患基本資料表及病患疾病明細表兩個資料表。

病患基本資料表中包含身分證字號、年齡層、性別、及生活圈等欄位，均可從承保資料檔擷取。因為部分資料不全，所以只能自44,042筆資料中擷取43,967筆納入資料庫。由於全民健康保險預防保健實施辦法規定提供40至64歲民眾每3年一次免費身體健康檢查，65歲以上每年一次免費身體健康檢查；而且涂宜均等人(2006)研究全民健保免費提供成人健檢的認知與使用情況時將年齡分為39歲以下、40至64歲及65歲以上，這兩份關

於成人健檢的資料都期望早期發現慢性病，與本研究相關，所以本研究也將年齡層分為這三個區段。

由於慢性疾病盛行率可能因城鄉生活型態差異而有不同，因此本研究依據經建會「國土綜合開發計畫」之規劃將台灣地區劃分為都會、一般及離島等3個生活圈進行研究。承保資料檔中的地區欄位記錄投保人之投保單位所在縣市，因此必須使用經建會所列生活圈劃分表轉入各投保人之生活圈資料，生活圈劃分表如表3。

表3：生活圈劃分表

生活圈別	包含縣(市)
都會生活圈	台北、桃園、新竹、台中、台南、高雄
一般生活圈	基隆、宜蘭、苗栗、南投、彰化、雲林、嘉義、屏東、花蓮、台東
離島生活圈	澎湖、金門、馬祖

資料來源：行政院經濟建設委員會網站

病患所患疾病記錄於病患疾病明細表，在該資料表中運用身分證字號、疾病兩個欄位即可記錄所有患者的所有疾病，資料來自門診處方及治療明細檔，該檔中每次門診醫師診斷之疾病最多雖然只有3種，但因民眾在2005~2007年間會有多次就醫記錄，任何一次就醫時醫師診斷出之慢性疾病都會列入資料庫中，因此每位患者所患慢性疾病記錄不限於3筆。由於本研究僅對慢性疾病進行研究，非屬慢性疾病之記錄不納入資料庫中。因門診處方及治療明細檔中所載疾病為國際疾病分類代碼ICD-9-CM，因此本研究依據表4之疾病名稱與國際疾病分類代碼(ICD-9-CM 1992)對照表將六大死因慢性疾病篩選出來，共得到14,224筆，彙整成病患疾病明細表，各種慢性疾病之罹患人數如表5。

表4：疾病名稱與國際疾病分類代碼(ICD-9-CM 1992)對照表

疾病名稱	國際疾病分類代碼	疾病名稱	國際疾病分類代碼
心臟疾病	390-398、410-414、 423-429	慢性肝炎及肝硬化	571
腦血管疾病	430-438	腎炎、 腎徵候群及腎性病變	580-588
糖尿病	250	高血壓性疾病	401-405

表5：慢性疾病病患人數統計表

疾病名稱	病患人數	疾病名稱	病患人數
心臟疾病	2878	慢性肝炎及肝硬化	2647
腦血管疾病	1338	腎炎、 腎徵候群及腎性病變	762
糖尿病	2076	高血壓性疾病	4523

三、資料分析

(一) 慢性疾病盛行率分析

為了解各種慢性疾病在不同年齡層、性別及生活圈盛行率，本研究利用SPSS 12.0對資料庫資料進行描述性統計分析。包含慢性疾病盛行率統計、生活圈與性別交叉統計、生活圈與年齡交叉統計等。

(二) 慢性疾病關聯分析

為了解各種慢性疾病相互間的關聯性，本研究以SQL Server 2005為資料探勘工具對病患疾病明細表進行關聯分析，以便找出各種慢性疾病間的關聯規則。另外，為了解慢性疾病間的關聯規則是否因不同生活圈、性別、或年齡層而有不同，本研究也分別對不同生活圈、性別、及年齡層之病患疾病明細資料進行關聯分析。本研究運用SQL Server 2005進行關聯分析時最小支持度設為0.03，最小機率設為0.5。依據SQL Server 2005的參數意義，最小支持度設為小數時，代表佔總交易數的比率，本研究資料共含43,967人，相當於最小支持度設為1319；SQL Server的最小機率即為最小信賴度，也就是本研究關聯分析之最小信賴度設為0.5。

肆、研究結果分析

一、慢性疾病盛行率分析

本研究資料共含43,967人，其中都會生活圈33,318人，一般生活圈有10,431人，離島生活圈有218人；39歲以下26,301人，40至64歲13,436人，65歲以上4,230人；男性21,485人，女性22,482人。慢性疾病盛行率分析結果如表6，其中腎炎、腎徵候群及腎性病變以「腎臟病變」表示，慢性肝炎及肝硬化以「肝臟病變」表示，以下分析均以此表示法表示。

表6：慢性疾病盛行率統計表

疾病名稱	生活圈			年齡層			性別	
	都會	一般	離島	~39歲	40~64歲	65歲~	男	女
高血壓性疾病	10.82%	13.61%	16.51%	1.40%	19.07%	37.61%	11.37%	9.24%
心臟疾病	5.85%	8.78%	5.04%	1.75%	10.30%	24.39%	6.57%	6.52%
肝臟病變	5.63%	7.22%	6.88%	3.21%	10.88%	8.03%	7.32%	4.77%
糖尿病	4.29%	6.03%	7.33%	0.81%	9.02%	15.31%	5.18%	4.14%
腦血管疾病	2.59%	4.46%	3.66%	0.52%	4.32%	14.60%	3.72%	2.39%
腎臟病變	1.59%	2.22%	0.00%	0.52%	2.73%	6.02%	2.03%	1.44%

由表6可以發現6種慢性疾病在一般生活圈之盛行率均高於都會生活圈；從年齡層來看，除了肝臟病變外，其他慢性疾病盛行率都隨著年齡增高；另外，男性在各種慢性疾病的盛行率都比女性高。值得注意的是高血壓性疾病盛行率相當高，男性盛行率為11.37%，女性盛行率為9.24%，均居6種慢性疾病之冠，若以年齡區分，65歲以上族群盛行率更高達37.61%。

表7是生活圈與性別慢性疾病盛行率交叉統計表，由表中發現，無論男女，一般生活圈之所有慢性疾病盛行率均高於都會生活圈，沒有性別差異。普遍來講，男性無論在都會生活圈或一般生活圈之各種慢性疾病盛行率均較女性高，但一般生活圈的女性之高血壓性疾病及心臟疾病盛行率卻高於男性。離島生活圈比較特別，女性在高血壓性疾病、心臟疾病、糖尿病之盛行率皆高於男性，尤其離島女性高血壓性疾病盛行率高達19.32%。

表7：生活圈與性別慢性疾病盛行率交叉統計表

疾病名稱	性別	生活圈		
		都會生活圈 男：16289 女：17029	一般生活圈 男：5097 女：5334	離島生活圈 男：99 女：119
高血壓性疾病	男性	1,758 (10.79%)	673 (13.20%)	13 (13.13%)
	女性	1,309 (7.68%)	747 (14.00%)	23 (19.32%)
心臟疾病	男性	990 (6.07%)	421 (8.25%)	1 (1.01%)
	女性	961 (5.64%)	495 (9.28%)	10 (8.40%)
肝臟病變	男性	1,138 (6.98%)	427 (8.37%)	8 (8.08%)
	女性	740 (4.34%)	327 (6.13%)	7 (7.07%)
糖尿病	男性	811 (4.97%)	329 (6.45%)	4 (3.36%)
	女性	619 (3.63%)	301 (5.64%)	12 (10.08%)
腦血管疾病	男性	513 (3.14%)	283 (5.55%)	4 (4.04%)
	女性	351 (2.06%)	183 (3.43%)	4 (3.36%)
腎臟病變	男性	311 (1.90%)	127 (2.49%)	0 (0%)
	女性	219 (1.28%)	105 (1.96%)	0 (0%)

表8為生活圈與年齡慢性疾病盛行率交叉統計表，表中顯示除肝臟病變外，在所有生活圈中所有慢性疾病的盛行率都隨年齡增長而增加，肝臟病變在所有生活圈中也都以40至64歲盛行率居高，因此沒有生活圈差異。一般生活圈65歲以上高血壓性疾病及心臟疾病盛行率高達47.79%及30.11%，宜特別注意。離島生活圈65歲以上高血壓性疾病及腦血管疾病盛行率高達59.37%及21.87%，明顯高於其他兩個生活圈。

二、慢性疾病關聯分析

為獲得慢性疾病間的關聯規則，本研究以SQL Server 2005對病患疾病明細表進行關聯分析，分析時最小支持度設為0.03，最小信賴度設定為0.5，得到14條關聯規則，如表9所示。由表9分析發現，腎臟病變、腦血管疾病、心臟疾病、糖尿病、肝臟病變等五種慢性病與高血壓性疾病間都有關聯，顯示了高血壓性疾病與其他慢性疾病的密切關聯。

表8：生活圈與年齡慢性疾病盛行率交叉統計表

疾病名稱	年齡	生活圈		
		都會生活圈 39歲以下: 20,282 40至64歲: 10,196 65歲以上: 2,840	一般生活圈 39歲以下: 5,903 40至64歲: 3,170 65歲以上: 1,358	離島生活圈 39歲以下: 116 40至64歲: 70 65歲以上: 32
高血壓性疾病	39歲以下	280 (1.38%)	88 (1.49%)	1 (0.86%)
	40至64歲	1,864 (18.28%)	683 (21.54%)	16 (22.85%)
	65歲以上	923 (32.50%)	649 (47.79%)	19 (59.37%)
心臟疾病	39歲以下	356 (1.75%)	105 (1.77%)	1 (0.86%)
	40至64歲	976 (9.57%)	402 (12.68%)	6 (8.57%)
	65歲以上	619 (21.79%)	409 (30.11%)	4 (12.50%)
肝臟病變	39歲以下	634 (3.12%)	207 (3.50%)	4 (3.44%)
	40至64歲	1,047 (10.26%)	408 (12.87%)	8 (11.42%)
	65歲以上	197 (6.93%)	139 (10.23%)	3 (9.37%)
糖尿病	39歲以下	148 (0.72%)	65 (1.10%)	2 (1.72%)
	40至64歲	884 (8.67%)	320 (10.09%)	9 (12.85%)
	65歲以上	398 (14.01%)	245 (18.04%)	5 (15.62%)
腦血管疾病	39歲以下	104 (0.51%)	35 (0.59%)	0 (0%)
	40至64歲	402 (3.94%)	178 (5.61%)	1 (1.42%)
	65歲以上	358 (12.60%)	253 (16.93%)	7 (21.87%)
腎臟病變	39歲以下	107 (0.52%)	32 (0.54%)	0 (0%)
	40至64歲	276 (2.70%)	92 (2.90%)	0 (0%)
	65歲以上	147 (5.17%)	108 (7.95%)	0 (0%)

不過要特別強調的是關聯規則只能了解患有某種疾病時有高度可能也同時患有另一種疾病，無法掌握疾病間的因果關係。例如「腦血管疾病→高血壓性疾病」之關聯規則表示患有腦血管疾病時同時也患高血壓性疾病的機率為66%，無法推論患有腦血管疾病會導致罹患高血壓性疾病，或是患有高血壓性疾病會導致罹患腦血管疾病。高血壓性疾病與其他慢性病的高度關聯性在既有醫學研究中都已獲認知，臨床上也有許多案例，例如醫學上已知高血壓可併發腦中風、心血管疾病，高血壓會造成腦血管壁增厚、血栓形成或血管壁完全閉塞，或因動脈粥狀硬化而導致腦血管阻塞。高血壓會對心臟血管產生

直接或間接影響，直接影響會使左心室負荷過重，間接影響會使血管動脈硬化，最終可能產生心絞痛、心肌梗塞、心臟衰竭或狹心症等併發症；糖尿病患者得到高血壓的機會是非糖尿病患者的1.5至2倍，相當多糖尿病患者會合併有高血壓(中華民國糖尿病衛教學會網站)。本研究結果強力佐證這些既有知識。

除了各種慢性病與高血壓性疾病的關聯外，「腎臟病變、腦血管疾病→心臟疾病」之關連規則也說明了同時患有腎臟病變、腦血管疾病者也有可能同時罹患心臟疾病。

表9：慢性疾病關聯規則

規則	信賴度
腎臟病變、腦血管疾病→高血壓性疾病	0.83
腦血管疾病、心臟疾病→高血壓性疾病	0.81
腎臟病變、心臟疾病→高血壓性疾病	0.79
腦血管疾病、糖尿病→高血壓性疾病	0.79
糖尿病、心臟疾病→高血壓性疾病	0.77
腎臟病變、糖尿病→高血壓性疾病	0.73
腦血管疾病→高血壓性疾病	0.66
肝臟病變、心臟疾病→高血壓性疾病	0.62
腦血管疾病、肝臟病變→高血壓性疾病	0.59
糖尿病→高血壓性疾病	0.58
心臟疾病→高血壓性疾病	0.58
臟病變、腦血管疾病→心臟疾病	0.57
糖尿病、肝臟病變→高血壓性疾病	0.55
腎臟病變→高血壓性疾病	0.54

為了了解不同生活圈、年齡層、性別之慢性疾病間關聯性是否有差異，本研究將病患疾病明細表中的資料依病患所屬類別分開分別進行關聯分析，最小支持度及最小信賴度同樣設定為0.03及0.5。

表10是各生活圈慢性疾病關聯分析所獲得的關聯規則，其中都會生活圈慢性疾病關聯分析及一般生活圈慢性疾病關聯分析都獲得14項關聯規則，離島生活圈因人數太少，不適合進行關聯分析，所以沒有分析結果。這兩項關聯分析仍然可以看出所有的慢性疾病跟高血壓性疾病無論在都會生活圈或一般生活圈都有非常密切的關聯，但非常特別的是在都會生活圈中信賴度最高的關聯規則「腎臟病變、腦血管疾病→高血壓性疾病」在一般生活圈分析中並未產生，「腎臟病變、腦血管疾病→心臟疾病」亦未產生。另一方面，「腎臟病變、高血壓性疾病→心臟疾病」及「腦血管疾病、高血壓性疾病→心臟疾病」兩個在一般生活圈分析獲得的規則，在都會生活圈分析中都未獲得。

表10：生活圈慢性疾病關聯規則

規則	都會生活圈 信賴度	一般生活圈 信賴度
腎臟病變、腦血管疾病→高血壓性疾病	0.86	
腦血管疾病、心臟疾病→高血壓性疾病	0.79	0.85
腎臟病變、心臟疾病→高血壓性疾病	0.78	0.81
腦血管疾病、糖尿病→高血壓性疾病	0.77	0.82
糖尿病、心臟疾病→高血壓性疾病	0.75	0.81
腎臟病變、糖尿病→高血壓性疾病	0.75	0.71
腦血管疾病→高血壓性疾病	0.62	0.69
肝臟病變、心臟疾病→高血壓性疾病	0.60	0.65
糖尿病→高血壓性疾病	0.57	0.63
心臟疾病→高血壓性疾病	0.56	0.64
腎臟病變→高血壓性疾病	0.54	0.54
腦血管疾病、肝臟病變→高血壓性疾病	0.53	0.71
糖尿病、肝臟病變→高血壓性疾病	0.53	0.60
腎臟病變、腦血管疾病→心臟疾病	0.52	
腎臟病變、高血壓性疾病→心臟疾病		0.55
腦血管疾病、高血壓性疾病→心臟疾病		0.51

表11是男性及女性慢性疾病關聯分析所獲得的關聯規則，其中男性慢性疾病關聯分析獲得15項關聯規則，女性慢性疾病關聯分析則獲得16項關聯規則。由表中可以看出所有的慢性疾病跟高血壓性疾病無論在男性或女性都有非常密切的關聯，比較特別的是在男性中信賴度達到0.78的關聯規則「腎臟病變、腦血管疾病→高血壓性疾病」在女性分析中並未產生，「腎臟病變、肝臟病變→高血壓性疾病」、「腎臟病變、腦血管疾病→心臟疾病」亦未產生。另一方面，「腎臟病變、糖尿病→心臟疾病」、「腎臟病變、高血壓性疾病→心臟疾病」、「腦血管疾病、糖尿病→心臟疾病」、及「腦血管疾病、高血壓性疾病→心臟疾病」等四個在女性關聯分析中獲得的顯示心臟疾病與腎臟病變、糖尿病、腦血管疾病、高血壓性疾病等密切相關的規則，在男性關聯分析中都未獲得。

表12是各年齡層慢性疾病關聯分析所獲得的關聯規則，其中65歲以上慢性疾病關聯分析獲得多達26項關聯規則，40~64歲慢性疾病關聯分析獲得13項關聯規則，39歲以下因罹患慢性病人數太少，不適合進行關聯分析，所以沒有分析結果。在這兩項關聯分析中仍然可以看出所有的慢性疾病跟高血壓性疾病無論在65歲以上或40~64歲兩個年齡層都有非常密切的關聯，另外也發現在65歲以上年齡層中心臟疾病與所有的慢性疾病間都有非常密切的關聯。

表11：男性及女性慢性疾病關聯規則

規則	男性 信賴度	女性 信賴度
腦血管疾病、心臟疾病→高血壓性疾病	0.81	0.81
腎臟病變、心臟疾病→高血壓性疾病	0.80	0.77
腎臟病變、腦血管疾病→高血壓性疾病	0.78	
腦血管疾病、糖尿病→高血壓性疾病	0.77	0.82
糖尿病、心臟疾病→高血壓性疾病	0.76	0.79
腎臟病變、糖尿病→高血壓性疾病	0.74	0.72
腦血管疾病→高血壓性疾病	0.65	0.67
心臟疾病、肝臟病變→高血壓性疾病	0.64	0.59
心臟疾病→高血壓性疾病	0.62	0.55
腦血管疾病、肝臟病變→高血壓性疾病	0.59	0.60
腎臟病變→高血壓性疾病	0.57	0.50
糖尿病→高血壓性疾病	0.56	0.61
腎臟病變、肝臟病變→高血壓性疾病	0.55	
糖尿病、肝臟病變→高血壓性疾病	0.52	0.61
腎臟病變、腦血管疾病→心臟疾病	0.50	
腎臟病變、糖尿病→心臟疾病		0.53
腎臟病變、高血壓性疾病→心臟疾病		0.53
腦血管疾病、糖尿病→心臟疾病		0.52
腦血管疾病、高血壓性疾病→心臟疾病		0.50

「腎臟病變、腦血管疾病→高血壓性疾病」在65歲以上信賴度高達0.89，在40~64歲中卻不成立，而「腎臟病變、腦血管疾病→心臟疾病」、「肝臟病變、腦血管疾病→心臟疾病」、「腎臟病變、糖尿病→心臟疾病」、「腎臟病變、高血壓性疾病→心臟疾病」、「肝臟病變、高血壓性疾病→心臟疾病」、「腦血管疾病、高血壓性疾病→心臟疾病」、「腎臟病變→心臟疾病」、「腦血管疾病、糖尿病→心臟疾病」、「肝臟病變、糖尿病→心臟疾病」、「糖尿病、高血壓性疾病→心臟疾病」、「糖尿病、高血壓性疾病→心臟疾病」、「腦血管疾病→心臟疾病」、「高血壓性疾病→心臟疾病」、「肝臟病變→心臟疾病」等13項顯示在65歲以上年齡層中心臟病疾病與其他慢性病皆有密切關聯的規則在40~64歲年齡層的分析中都未檢出。另一方面，「腎臟病變、肝臟病變→高血壓性疾病」這個在40~64歲年齡層的分析中獲得的規則，在65歲以上年齡層中並未獲得，值得一提的是，此規則的意義是患有腎臟病變及肝臟病變者很可能同時會有高血壓性疾病，這樣的知識尚未見諸於文獻中。

表12：各年齡層慢性疾病關聯規則

規則	65歲以上 信賴度	40~64歲 信賴度
腎臟病變、腦血管疾病→高血壓性疾病	0.89	
腎臟病變、糖尿病→高血壓性疾病	0.88	0.68
糖尿病、心臟疾病→高血壓性疾病	0.87	0.71
腎臟病變、心臟疾病→高血壓性疾病	0.87	0.73
腦血管疾病、糖尿病→高血壓性疾病	0.86	0.74
腦血管疾病、心臟疾病→高血壓性疾病	0.84	0.79
糖尿病→高血壓性疾病	0.79	0.54
肝臟病變、腦血管疾病→高血壓性疾病	0.79	0.57
心臟疾病→高血壓性疾病	0.78	0.57
腦血管疾病→高血壓性疾病	0.78	0.61
肝臟病變、糖尿病→高血壓性疾病	0.78	0.53
腎臟病變→高血壓性疾病	0.78	0.52
肝臟病變、心臟疾病→高血壓性疾病	0.76	0.60
腎臟病變、腦血管疾病→心臟疾病	0.67	
肝臟病變、腦血管疾病→心臟疾病	0.65	
腎臟病變、糖尿病→心臟疾病	0.64	
腎臟病變、高血壓性疾病→心臟疾病	0.61	
肝臟病變、高血壓性疾病→心臟疾病	0.57	
腦血管疾病、高血壓性疾病→心臟疾病	0.56	
腎臟病變→心臟疾病	0.55	
腦血管疾病、糖尿病→心臟疾病	0.53	
肝臟病變、糖尿病→心臟疾病	0.53	
糖尿病、高血壓性疾病→心臟疾病	0.53	
腦血管疾病→心臟疾病	0.52	
高血壓性疾病→心臟疾病	0.51	
肝臟病變→心臟疾病	0.50	
腎臟病變、肝臟病變→高血壓性疾病		0.50

為了進一步了解年齡與各種慢性疾病間的關係，本研究將年齡與慢性疾病做關聯分析，其中年齡層分為~25、26~39、40~53、54~64、65~74、75~等6個年齡層，結果獲得與年齡相關的規則如表13。由表中可發現40歲以上就很可能罹患高血壓性疾病，54歲以上就很可能罹患腎臟病變、腦血管疾病、糖尿病、心臟疾病，而且這些疾病的患者很可能也都罹患高血壓性疾病，65歲以上罹患肝臟病變的患者很可能也罹患高血壓性疾病，75歲以上罹患腎臟病變、腦血管疾病、糖尿病、高血壓性疾病、肝臟病變的患者很可能也罹患心臟疾病。

表13：年齡與慢性疾病之關聯規則

規則
腎臟病變、75~ → 高血壓性疾病
65~74、腦血管疾病 → 高血壓性疾病
75~、糖尿病 → 高血壓性疾病
75~、心臟疾病 → 高血壓性疾病
65~74、糖尿病 → 高血壓性疾病
65~74、心臟疾病 → 高血壓性疾病
65~74 → 高血壓性疾病
75~ → 高血壓性疾病
75~、腦血管疾病 → 高血壓性疾病
腦血管疾病、54~64 → 高血壓性疾病
腎臟病變、65~74 → 高血壓性疾病
65~74、肝臟病變 → 高血壓性疾病
54~64、糖尿病 → 高血壓性疾病
54~64、心臟疾病 → 高血壓性疾病
54~64 → 高血壓性疾病
75~、肝臟病變 → 高血壓性疾病
腎臟病變、75~ → 心臟疾病
腎臟病變、54~64 → 高血壓性疾病
75~、腦血管疾病 → 心臟疾病
75~、高血壓性疾病 → 心臟疾病
75~ → 心臟疾病
75~、肝臟病變 → 心臟疾病
75~、糖尿病 → 心臟疾病
40~53 → 高血壓性疾病

伍、結論與建議

一、結論

心臟疾病、腦血管疾病、糖尿病、慢性肝炎及肝硬化、腎臟病變、高血壓性疾病等慢性疾病多年來一直名列我國重大死因疾病，而且相互間互有因果關係，這種相互關係的掌握對醫師檢驗、診斷及治療都非常重要。國家衛生研究院的全民健康保險研究資料庫包含了大量的病患門診就醫資料，這些資料中可能隱藏許多未知而有用知識，是發掘醫學知識的重要資源。因此本研究以該資料庫為基礎，經資料淨化與篩選後，進行統計分析，以便了解上述慢性疾病在不同生活圈、年齡層及性別上之盛行率，並利用資料探勘技術找出不同生活圈、性別、年齡之慢性疾病間的關聯性差異。獲得下列研究結果：

(一) 高血壓性疾病，男、女性盛行率均居6種慢性疾病之冠，65歲以上族群盛行

率更高達37.61%。而且不管哪種生活圈、性別、年齡，腎臟病變、腦血管疾病、心臟疾病、糖尿病、肝臟病變與高血壓性疾病間都有關聯，突顯高血壓性疾病照護之重要性。

- (二) 這6種慢性疾病在一般生活圈男女性之盛行率均高於都會生活圈，可能是一般生活圈民眾對慢性疾病防治較不積極，或醫療資源較為不足，無法早期醫療、早期控制，顯示一般生活圈民眾健康意識宣導及醫療資源均應加強。
- (三) 肝臟病變在40~64歲年齡層之盛行率高達10.88%，高於65歲以上的8.03%，提醒民眾肝臟保護應從年輕做起。
- (四) 男性之慢性疾病盛行率普遍較女性高，但一般生活圈的女性之高血壓性疾病及心臟疾病盛行率卻高於男性，一般生活圈的女性應特別注意這兩項疾病。離島生活圈女性在高血壓性疾病、心臟疾病、糖尿病之盛行率也高於男性，值得注意。
- (五) 一般生活圈65歲以上高血壓性疾病及心臟疾病盛行率明顯高於都會生活圈。離島生活圈65歲以上高血壓性疾病及腦血管疾病盛行率更明顯高於其他兩個生活圈。顯示政府應加強一般生活圈及離島生活圈老人此兩項疾病之照護。
- (六) 都會生活中信賴度最高的關聯規則「腎臟病變、腦血管疾病→高血壓性疾病」在一般生活中不存在。顯示不同生活圈間有疾病關聯的差異。
- (七) 在男性中信賴度達到0.78的關聯規則「腎臟病變、腦血管疾病→高血壓性疾病」並未存在女性中，可能是女性腎臟病變及腦血管疾病盛行率都遠較男性低。另外，女性心臟疾病與高血壓性疾病、糖尿病間的關係均較男性明顯。顯示性別間有疾病關聯的差異。
- (八) 年齡層愈高，慢性疾病間的關聯也愈明顯，顯示慢性疾病相互引發之關係密切，因此慢性病早期防治對醫療資源的有效運用相當重要。
- (九) 在40~64歲年齡層中患有腎臟病變及肝臟病變者很可能同時會有高血壓性疾病，這樣的知識尚未見諸於文獻中，值得醫界注意。
- (十) 關聯分析顯示高血壓性疾病、腎臟病變及腦血管疾病間有高度關聯；高血壓性疾病、腦血管疾病及心臟疾病間亦有高度關聯；而女性的高血壓性疾病、腦血管疾病及糖尿病有高度關聯。提醒民眾及醫生應注意其關聯性，一發現有其中一種疾病，應該確實檢查是否有相關疾病，或積極進行相關疾病防治。

二、建議

前述研究結果不但可供醫生檢驗、診斷、醫治病人時作為重要參考，對於衛生公衛機關訂定公衛政策或推動疾病防治工作也提供相當好的依據，對民眾自身保健與養生更是相當實用的知識。為使研究發揮更大效益，建議後續可繼續從事下列研究：

- (一) 研究中發現一般生活圈各種慢性疾病之盛行率均高於都會生活圈，其原因是

否為一般生活圈民眾對慢性疾病防治較不積極，或醫療資源較不足，因而沒有早期醫療、早期控制，可以進一步研究。

- (二) 肝臟病變在40~64歲年齡層之盛行率較65歲以上高，是因為肝臟病變發生期較早，還是肝臟病變早期致命率高，或是肝臟病變發生期有年輕化趨勢，都值得進一步研究。
- (三) 「腎臟病變、肝臟病變→高血壓性疾病」關聯規則，顯示患有腎臟病變及肝臟病變者很可能同時會有高血壓性疾病，值得進一步就腎臟病變、肝臟病變、高血壓性疾病之病理關係研究探討。
- (四) 離島生活圈，65歲以上高血壓性疾病及腦血管疾病盛行率明顯高於其他兩個生活圈，女性在高血壓性疾病、心臟疾病、糖尿病之盛行率皆高於男性，異於其他生活圈，是環境還是文化因素，值得進一步研究。
- (五) 「腎臟病變、腦血管疾病→高血壓性疾病」在不同性別、年齡層、生活圈族群間都有差異，原因為何，可以進一步研究。

參考文獻

1. 內政部統計處，2008，老人狀況調查摘要分析 (available online at <http://sowf.moi.gov.tw/stat/Survey/94old.doc>)。
2. 中華民國糖尿病衛教學會網站，2009，糖尿病與高血壓 (available online at <http://www.tade.org.tw/05/web/糖尿病與高血壓.pdf>)。
3. 全民健康保險研究資料庫網站，2008a，資料庫內容說明 (available online at http://www.nhri.org.tw/nhird/date_01.htm)。
4. 全民健康保險研究資料庫網站，2008b，承保抽樣歸入檔 (available online at http://www.nhri.org.tw/nhird/date_cohort.htm)。
5. 行政院衛生署統計資訊網，2009，96年台灣地區主要死因分析(available online at <http://www.doh.gov.tw/statistic/index.htm>)。
6. 行政院衛生署國民健康局網站，2008，台灣地區高血壓、高血糖、高血脂盛行率調查報告 (available online at <http://www.bhp.doh.gov.tw/health91/study-2.htm>)。
7. 吳宜珍、吳登強、余方榮、盧建宇、王聖雯、林增記、周俊仁、陳漢文、郭富珍、詹昌明，2005，『肝硬化：急性上消化道出血時併發心肌傷害之危險因子』，台灣內科醫學會94年會員大會。
8. 林獻鋒、張靜梅、賴世偉、劉秋松、林璨，2006，『健檢民眾高血壓盛行率之描述性分析：以台中市一醫學中心為例』，台灣家庭醫學研究，第四卷·第三期：129~139頁。
9. 姚志成，民94，運用資料探勘技術建構脂肪肝預測模式，中原大學資訊管理學系碩

士論文。

10. 范宜玲、陳萬春、趙德強、陶春燕，2009，『肝硬化性心肌病和肝硬化性心力衰竭』，*Internat. J. Cardiovasc Med.*，第十卷·第一期：66~67頁。
11. 涂宜均、林宇旋、洪百薰、吳秀英，2006，『台灣地區國人對慢性疾病預防認知及預防保健服務利用情形初探—2005年國民健康訪問暨藥物濫用調查結果』，國家衛生研究院電子報，第一百七十八期。
12. 徐敏耀、劉夷生、馬作鑑、張木信、張丁權、賴昭宏、鍾國屏，2007，『冠狀動脈心臟病危險因子之老年人心導管檢查預測模型研究』，*台灣老年醫學雜誌*，第三卷·第一期：25~33頁。
13. 陳明豐，2009，高血壓面面觀(available online at <http://tw.myblog.yahoo.com/lovingleica/article?mid=389&prev=397&next=388&sc=1>)。
14. 陳鴻鈞，2001，『認識腎臟病』，*高醫醫訊*，第二十一卷·第五期。
15. 連賢明，2008，『如何使用健保資料進行經濟研究』，*經濟論文叢刊*，第三十六卷·第一期：115~143頁。
16. 張俊郎、陳啟浩、曾輝鈺，民96，『結合類神經網路與決策樹於糖尿病前期診斷之研究』，中華民國品質學會第43屆年會暨第13屆全國品質管理研討會。
17. 劉如玲，民96，第二型糖尿病與慢性C型肝炎之相關研究：以南部某縣健康篩檢為例，美和技術學院健康照護研究所碩士論文。
18. 賴世偉、黃金財，2004，『健檢民眾糖尿病之描述性分析—以台中市一醫學中心為例』，*Chunghua Journal of Medicine*，第九卷·第S期：267~271頁。
19. 盧美秀、林秋香，1991，『比較團體衛教和個別衛教對糖尿病病人知識態度、行為和血糖變化的影響』，*護理雜誌*，第三十八卷·第四期：101~112頁。
20. 簡國龍，2005，糖尿病對心血管疾病的危害與預防之道，(available online at <http://www.dmcare.org.tw/up3/2005-4-%E7%B0%A1%E5%9C%8B%E9%BE%8D.pdf>)。
21. 簡國龍，2009，初級心臟預防醫學領域研究新進展：慢性腎臟疾病發作之預測模型，*中華民國心臟學會會訊*，2009年5月號，3~7頁。
22. Aggarwal, C. C., Sun, Z., and Yu, P. S., "Fast algorithms for online generation of profile association rules," *IEEE Transactions on Knowledge and Data Engineering* (14:5), 2002, pp. 1017-1028.
23. Agrawal, R., Imielinski, T., and Swami, A. "Mining Association Rules between Sets of Items in Large Databases," in *Proceedings of the ACM SIGMOD Conference on Management of Data*, 1993, pp. 207-216.
24. Agrawal, R., and Srikant, R. "Fast Algorithms for Mining Association Rules," *Proceedings of the 20th International Conference on Very Large Data Bases*, 1994, pp. 487-499.
25. Berry, M. J. A., and Linoff, G. *Data Mining Techniques: For Marketing, Sales, and*

Customer Support, John Wiley and Sons, 1997.

26. Chen, M. S., Han, J., and Yu, P. S. "Data Mining: An Overview From a Database Perspective," *IEEE Transactions on Knowledge and Data Engineering* (8:6), 1996, pp. 866-883.
27. Chu, C. J., Tseng, V. S., and Liang, T. "An efficient algorithm for mining temporal high utility itemsets from data streams," *Journal of System and Software* (81:7), 2008, pp. 1105-1117.
28. Fayyad, U. M. "Data mining and knowledge discover: Making sense out of data," *IEEE Expert: Intelligent Systems and Their Applications* (11:5), 1996, pp. 20-25.
29. Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. "From Data Mining to Knowledge Discovery in Databases," *AI Magazine*, 1996, pp. 37-54.
30. Han, J., and Kamber, M., *Data Mining: Concepts and Techniques* (2nd ed.), Elsevier, 2006.
31. Han, J., Pei, J., Yin, Y., and Mao, R. "Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach," *Data Mining and Knowledge Discovery* (8:1), 2004, pp. 53-87.
32. Li, Y. C., Yeh, J. S., and Chang, C. C. "Isolated items discarding strategy for discovering high utility itemsets," *Data & Knowledge Engineering* (64:1), 2008, pp. 198-217.
33. Khan, M. S., Muyeba, M., and Coenen, F. "A Weighted Utility Framework for Mining Association Rules," in *Proceedings of European Modeling Symposium (IEEE)*, 2008, pp. 87-92.
34. Song, M., and Rajasekaran, S. "A Transaction Mapping Algorithm for Frequent Itemsets Mining," *IEEE Transactions on Knowledge and Data Engineering* (18:4), 2006, pp. 472-481.
35. Zaki, M. J., "Scalable algorithms for association mining," *IEEE Transactions on Knowledge and Data Engineering* (12:3), 2000, pp. 372-390.