

利用決策樹方法分析探討糖尿病人血糖控制之成效

Study of Glycemic Control Efficiency for Patients with Diabetes Based on Decision Tree

湯宗泰	簡守維	賴仲亮	陳永福	高琳詠
國立雲林科技大學資訊管理系 副教授	國立雲林科技大學資訊管理系 博士生	衛生署台中醫院 研發中心主任	中國醫藥大學醫務管理學系 副教授	中國醫藥大學醫務管理學系 碩士生
tangdt@mis4k.mis.yuntech.edu.tw	sowejan@mail.taich.doh.gov.tw	laipeter@yahoo.com.tw	yungfu@mail.cmu.edu.tw	u9675015@apple.cmu.edu.tw

摘要

糖尿病已成為二十一世紀全世界最重要的公共衛生課題之一。近年來由於肥胖問題日益嚴重，在開發中以及新興工業化國家的發生率及盛行率不斷地快速增加。糖尿病通常會同時伴隨著不同的併發症，對於個人生活品質及醫療資源的使用有相當大的影響。本研究的主要目的為藉由了解血糖過多的病人之控制血糖方式，達到控制血糖的目標，探討罹病族群的特徵，提供醫師或衛教人員參考。我們使用決策樹方法分析 C4.5，利用糖化血紅素為決策屬性，分析位於台灣中部某教學醫院之 2888 位病人。研究結果發現參加糖尿病醫療品質改善計劃(DHQIP)是影響成效的最大因素；同時，病人看診的次數、病人診療的時間、病人與門診醫師的關係，病人的年齡、居住地區、性別等也是影響的相關因素。

關鍵詞：資料探勘、決策樹、糖尿病

Abstract

Diabetes has become one of the most important issues of public health problem worldwide in the 21st century. Recently, the obese population in the developed and newly industrialized countries increases significantly, which accompanies the prevalence of patients suffered from diabetes. Accompanied with different complications, diabetes has considerable influences on the quality of life and the use of medical resources. In this study, we investigate the factors which are efficient in control the level of blood sugar and the characteristics of those patients who have better control in diabetes. The glycohemoglobin (HbA1c) is used as the decision attribute for the decision tree to analyze 2888 patients with diabetes diagnosed in a hospital in Taichung. The

results show that whether patients have participated diabetes health quality improving program (DHQIP) is the most influential factor. Additionally, number of patient visits, diagnosing time, relationship between patient and physician, age, resident location, and gender are also found to be important factors.

Keywords: Data Mining、Decision Tree、Diabetes

1. 前言

隨著醫藥的發達，傳染性疾病的控制，世人生活型態的靜態化，肥胖之高度盛行，糖尿病已成為二十一世紀全世界最重要的公共衛生課題之一。在開發中以及新興工業化國家，它的發生率及盛行率正在快速的增加當中。2000 年時，糖尿病的盛行率估計約佔世界各地所有年齡層的 2.8%，預估在 2030 年會提升至 4.4%。患有糖尿病的人口，在 2000 年上升至 1.71 億人口，推斷在 2030 年會上升到 366 億人口 [26]。此外，糖尿病病患佔全球人口的死亡率，在 2000 年估計約為 5.2%，相當於 290 萬人 [23]。糖尿病有可能成為全球的死亡病因的第五位。另外，2005 年美國人患有糖尿病的比例約為 7%，約 2080 萬人。資料顯示，2002 年美國之醫療總費用約為 1320 億美元，其中用於治療糖尿病的直接醫療費用為 92 億美元。

台灣於 2002 年之糖尿病盛行率為 3.35%，病患人數為 733,133 人，而 2003 年盛行率上升至 3.54%，人數為 784,475 人，病患人數呈現逐年增加的縣現象 [4]。在醫療資源的利用上，2003 年因糖尿病就醫的人數約為 60 萬人，花費的醫療費用約 120 億元，平均每年每人因糖尿病就醫的門診次數約為 10 次，每次平均費用約為 1,600 元；每年每人因糖尿病就醫平均住院次數約為 0.07 次，每次平均費用約 43,000 元 [5]。

糖尿病的死亡率逐年升高[6]，除了直接造成死亡之外，糖尿病所引起的急性或慢性併發症，例如糖尿病酮酸血症、高血糖高滲透壓症、低血糖症、心臟病、視網膜病變、尿毒症、神經病變和腳部截除等，造成難以估計損失及失能、殘障、生活品質低落等後遺症。由此可見，糖尿病所造成極大的社會負擔。糖尿病所引起的併發症可以透過良好的血糖控制和有效的衛生教育事先預防，延緩病狀的發生和惡化。糖尿病病人之照護通常以糖化血紅素或是鐮狀血球之值為控制標準[2,18]，因此了解病人控制糖化血紅素（A1C）的成效，區分危險族群與能夠達到目標之族群的特徵是非常重要的。醫師或衛教人員可以針對成效較低的族群，提供更適當的指導及幫助。經由資料庫探勘及分類是其中一種最有效的技術，以獲得有意義的知識。

本文的主要目的是依據糖化血紅素的控制結果，使用 C4.5 決策樹分析方法，建構病人特徵與控制結果之間的關係模式，發掘決策法則。

本文的架構如下。第二節介紹糖尿病病人之診斷、糖尿病的分類及糖尿病的危險因素。第三節描述決策樹的多級分類。第四節說明資料的收集方式和資料探勘。第五節進行結果討論與相關之發現。最後，第六節作一簡單的結論。

2. 糖尿病之相關知識

2.1 糖尿病的診斷與分類

標準的糖尿病診斷有三種方法，每個病例都必須被檢驗有明確的血糖過高症狀。診斷患有糖尿病的成因如下[2]：

- 1、隨時測得的血糖濃度(無論空腹與否的所謂隨機血糖) 大於或等於 200mg/dl，同時合併出現糖尿病之典型症狀，包括吃多、喝多、尿多、體重下降等症狀。
- 2、空腹血糖大於或等於 126mg/dl，其中，空腹的定義是指沒有攝取卡路里至少 8 小時。
- 3、口服耐糖試驗在第二小時之血糖值大於或等於 200mg/dl。這項測試的描述，由世界健康組織用葡萄糖負荷含有相當於 75 克無水葡萄糖溶解於水。

本研究根據「全民健康保險糖尿病專業醫療服務品質報告」中之定義[4]，糖尿病就醫案件之擷取條件為任一主、次診斷之 ICD-9-CM 前三碼為 250 之門、住診案件。2006 年美國糖尿病協會的臨床實踐中的建議指出，第一型糖尿病之原因為胰島素之絕對性缺乏，第二型糖尿病則可能是無法製造足夠的胰島素來滿足身

體的需要，或是患者身體中的細胞無法正確使用胰島素。妊娠糖尿病的定義為葡萄糖耐量異常或罹患其他類型的糖尿病。

2.2 血糖控制的風險因素

人口統計資料，如年齡、性別、種族等會，影響糖尿病的發展[3, 8]，而肥胖、缺乏運動、吸煙、糖尿病家族史等被指為糖尿病的危險因素[1, 9, 22]。社會經濟地位，例如教育程度、職業地位、收入，也和是否會患有糖尿病有關[1, 8, 22]，其他，如醫生診察病人的情況也可能會影響糖尿病的控制情況。

1993 年美國有一個臨床研究報告結果顯示，相對於傳統治療方式，採取加強胰島素治療方式將血糖儘可能控制於合理範圍內之密集治療，降低了發展成視網膜病變的風險達 76%，並可以減緩發展視網膜病變降低為 54%。密集式的治療同時也減少發生微量白蛋白尿 39% 的機率，同時也減少 54% 發生蛋白尿的機率及 60% 的神經病變。由此可知，加強胰島素治療的方式可以預防或延緩第一型糖尿病慢性病併發症的發生或進行。雖然此臨床研究是針對第一型糖尿病。英國前瞻性糖尿病研究(UKPDS) [25, a9] 以第二型糖尿病患者的研究結果也得到類似的結論。因此糖尿病的併發症可藉由良好的血糖控制來事先預防或延緩其併發症發。以上之研究均以糖化血紅素作為血糖控制的參考指標。

3. 研究方法與材料

資料探勘為發現有趣的知識，把大量的數據存儲在資料庫中的過程[16]。Dunham[15]定義資料探勘為透過演算法將資料進行分析與應用，找出其特徵、模式的過程。知識發掘是在知識發現的過程中找到有用的訊息和模式數據。知識發掘的過程包括以下幾個步驟：資料清理、資料集成、資料選擇、資料轉換、資料挖掘、模式評估和知識呈現[16]。

資料探勘是整個知識發掘過程中的一部份，它主要用於發現有用的隱藏模式(Hidden pattern)，也是整個過程中最重要的步驟。分類是利用機器學習的方式，從大量的歷史資料中萃取隱含在資料中潛在有用的知識[15]。目前在處理分類問題上，決策樹方法是其中一種有效的技術。

由上而下的誘導決策樹，大概是最廣泛被研究的機器學習方法[27]。決策樹是以樹狀資料結構為基礎的分類分析方法，其主要的優點為

能夠產生易於瞭解與運用的決策法則。決策樹的建構是一種監督式的學習法，從訓練範例的屬性中挑選可以作為建構決策樹根節點及內部節點的屬性，建構分析樹。

最為廣泛使用的決策樹學習法為c4.5 [20]，其預設分類標準所採用的是c4.5獲利比。這是一個以信息為基礎的措施，要考慮到不同的測試結果[15, 21]。獲利比被用於選擇測試屬性，每一個節點位於一棵決策樹中。屬性與最高獲利比率是獲選為測試屬性為當前節點。

3.1 資料收集

本研究分析台灣中部某教學醫院從 2005 年 1 月 1 日至 2005 年 12 月 31 日，一年間糖尿病病人之門診及住院之資料。糖尿病之病人。資料之篩選原則如下：

- 1、門診及住院申報案件之 ICD-9 前四碼中為 250 且出現一次以上之診斷。
- 2、空腹血糖大於或等於 126mg/dl，其中空腹之定義是指沒有攝取卡路里至少 8 小時以上。
- 3、隨機之血漿葡萄糖濃度大於或等於 200mg/dl (11.1mmol/l)，其中隨機之意思為不論何時或飯後多久進行檢測。
- 4、病患報名參加糖尿病醫療品質改善計劃 [5] 符合 a, b, c, d 之一者，且該病人在 2005 年間在檢驗室數據中含有糖化血色素之檢測的資料。
- 6、排除只有就診一次之病人。

篩選得到的結果顯示，門診及住院申報案件之 ICD-9 前四碼中出現 250.1-250.9 且出現一次以上之診斷資料，總計為 16383 筆，可歸納為分屬 3559 個病人。空腹血糖之資料總計有 37714 筆，其中空腹血糖大於或等於 126 mg/dl 有 6848 筆，分屬 4730 個病人。隨機血漿葡萄糖濃度之資料總計 1404 筆，其中隨機的血糖濃度大於或等於 200mg/dl 之資料有 593 筆，分屬 39 個病人。全院參加糖尿病醫療品質改善計劃(DHQIP)共有 1539 人，其中有糖化血色素檢測資料之病人有 1248 人。糖化血色素之檢測資料共 7741 筆，歸納為 3710 個病人。綜合以上結果，符合上述條件之有效樣本數共有 2888 個病人。

資料庫收集的資料包含：病案代碼、姓名、出生日期、性別、住址、郵政編碼、保險類別、就診科別、就診日期、就診之 ICD-9 碼、就診處方是否有治療糖尿病之藥物、是否有參加糖尿病醫療品質改善計劃、是否有併發症和糖化血色素之檢測資料。我們將併發症分為六個類別：急性併發症，腎臟併發症，眼科併發

症，神經系統併發症，血管併發症及足部併發症等。本研究之研究限制為一些影響糖尿病的重要因素，如體重史，家族史，運動史，抽煙和酒精的使用，以及其他因素，因不包含在目前資料庫中，因此並未將這些因素列入考慮。

首先，我們將資料庫收集的資料轉換為適當的格式。預測屬性包括年齡、性別、居住地區、保險型、就診次數、就診的時間、診療糖尿病的時間、病人與門診醫師的關係、有無併發症、是否有參加糖尿病醫療品質改善計劃等，決定屬性是糖化血色素。由醫療記錄代碼轉換為 ID No，出生日期轉換為年齡，以 2005.1.1 為基準。同時，我們將年齡區分為三個不同階層：21-55、56-70、>70。居住地可以由地址和郵遞區域進行轉換，分別為居住於都市區（人口超過百萬）及居住於非大都市（人口超過 300000）等兩種類型。台灣開始進行全民健康保險從 1995 年，根據中央健康保險局之資料顯示，台灣投保人口涵蓋 22 億美元(98.67%)，在 2003 年底，只有 1.33 % 人口沒有全民健康保險，因此醫療保險可區分為 3 種類別：

- 1、沒有全民健康保險(健保局)之自費病人。
- 2、有全民健康保險，需要支付部份付擔之病人。
- 3、有全民健康保險，且為低收入戶或退伍軍人之免支付部份付擔之病人。

經由就診時間計算，每次就醫的處方中含有糖尿病之口服、注射藥物或有監測血糖或糖化血色素者，判定該次就醫屬於糖尿病就醫之診次。門診醫師的關係則由病人就醫次數及就診同一個醫生的比率是否大於或等於 70% 來研判，若大於或等於 70% 則表示病人與門診醫師的關係是穩定可信賴的，否則表示不可信賴。另外，由 ICD-9 診斷碼判定是否有併發症，決策的屬性是糖化血色素，數值得自檢驗室之檢驗數據，依血糖的控制狀況可區分為 3 類：糖化血色素(A1C)大於 7.0，分類為達到目標層級 [7]；糖化血色素小於或等於 9.5，屬於中間層級；糖化血色素大於 9.5，屬於高危險層級 [18]。糖化血色素的資料若有數個值，我們採用加權平均的方式處理，失蹤的資料就忽略 [16]。研究樣本之統計特性如表 1 所示。

3.2 資料探勘

本研究使用的分類方式為決策樹方法 C4.5，對於就診的時間與治療糖尿病的時間，這兩個數字的屬性是透過演算法。我們用同樣的頻率組合法 [27,6,14,17] 分化連續屬性送入預定箱的數目，我們使用三個箱子。在驗證上，我們用 10 倍交叉驗證。分析樹如圖 1 所示、表

2 顯示混亂矩陣、決策規則產生 c4.5 表 3。

4. 研究結果

本次研究之樣本中，69.32 %病人之年齡大於 56 歲，住在都會區的比例為 78.25%，所以本研究樣本可以代表台灣中部都會區中老人糖尿病的追蹤治療狀況。從保險制度的角度而言，在台灣實施社會醫療保險制度之後，全國 95% 以上的民眾皆納入此制度，樣本中 96.85% 的病人屬於醫療保險的範圍，其中 6.06% 因為收入低所以在就診時不需支付費用，其他 90.99% 的民眾就診時需要支付部份費用。在台灣社會醫療保險制度下，病人可以自由選擇任何科別的醫師進行疾病治療，但對於糖尿病的治療與控制，一般民眾選擇內科醫生或是家庭醫生來就診。本研究樣本中有 97.51% 的民眾選擇內科醫生或家庭醫生就診。

有關就醫的習慣，依據社會醫療保險制度的規定，第二型糖尿病屬於慢性疾病，其病患門診追蹤與取藥的間隔，醫師可以依病患病情的狀況來調整，範圍為一週至三個月。本研究有 63.36% 的病患一年至門診追蹤三次以上，其中 60.32% 有比較固定追蹤的醫師。長期糖尿病以及血糖控制不佳的糖尿病的病患常會有相關的併發症。本研究顯示，44.11% 的病患已經產生一個或一個以上的併發症。因此，政府為了讓糖尿病的病患能更有效的控制血糖並減少併發症，鼓勵醫院辦理糖尿病的醫療品質改善計

劃，希望藉由個案管理的概念，結合醫生、營養師、個案管理人等，達到上述的目的。本研究顯示 41.41% 的病患加入此照顧計劃。

有關病患血糖控制是否良好是以血中糖化血色素來做標準，其中糖化血色 7.0 定義為達到血糖目標管理，糖化血色素在 7.0 至 9.5 間定義為中等控制，而糖化血色素大於 9.5 定義為危險。本研究中 26.45% 達到管理目標，52.63% 為中間控制，20.91% 具有危險性。

5. 討論

由 C4.5 的分析，第一個分裂點(split node) 為病人是否參加糖尿病醫療品質改善計劃。收案病患之看診的醫師為持有糖尿病醫療品質改善計劃合格證書的內科專科醫生和家庭醫生，並且接受多重專業團隊共同照護，因此其血糖控制較不受其它因素的干擾，為一獨立影響因子，因此演算法可印證臨床之狀態。未收案之病人以看診科別為其最重要的影響因子，就該醫院執業狀態來看：

- 1、未收案的病人多分散在內科醫生看診，而內科醫師對糖尿病病人的診治看法相當的一致性，因為內科醫師的醫療養成教育中，對糖尿病患的治療有相關的基本認知。
- 2、家庭醫學科不同於其它科別，他們是提供醫療保險服務的把關醫生，因為在病患的組成特質中，一些隱藏性疾病或疾病的初期的病患較其它科別多，因而可見控制良好的狀態

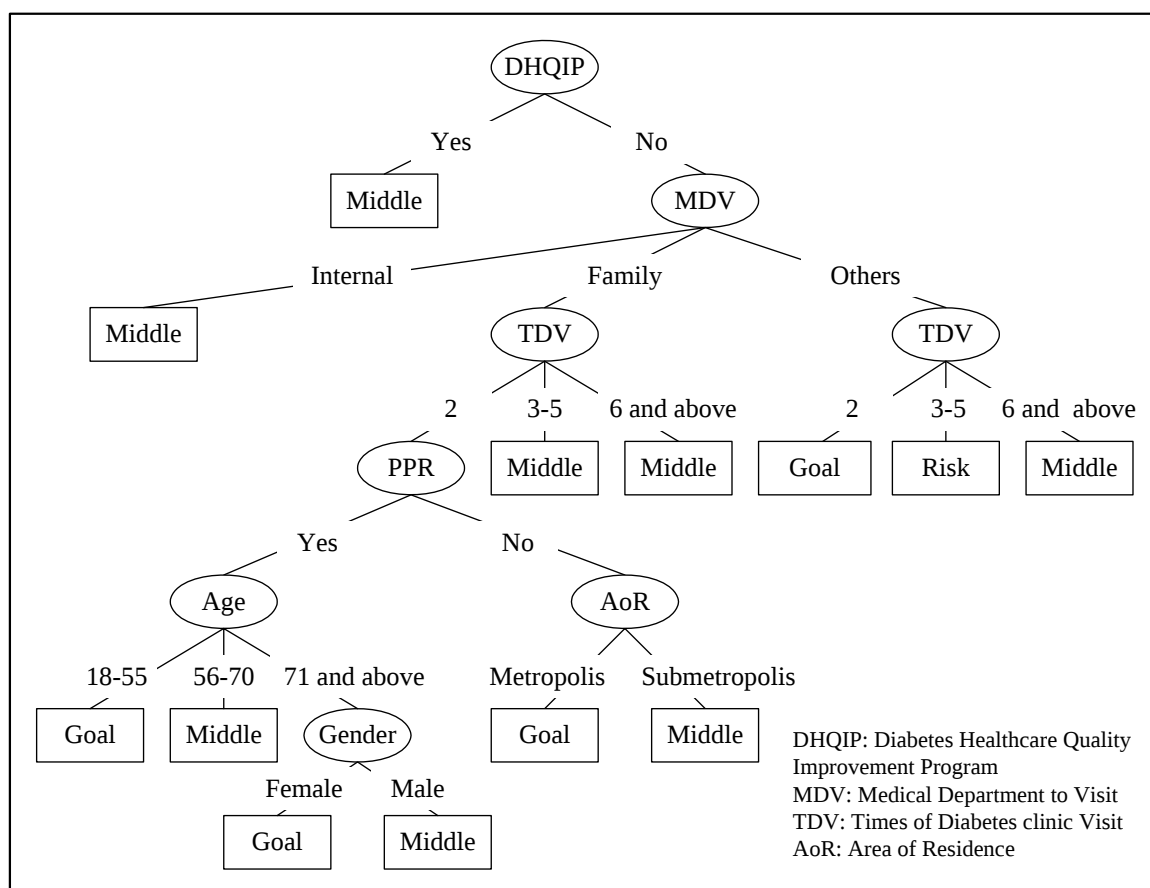


圖 1. 決策樹輸出與決策屬性 A1C<7.0 (目標), 7.0≤A1C≤9.5 (中等), A1C>9.5 (危險)

比例較高，也就是說，糖尿病初期只須飲食控制或簡單的藥物控制，控制良好的病患比例基本上就比其它科別高。

- 3、其它科別與上述二個科別之不同點為，控制良好的患者不會轉介，不像家庭醫學科與內科對糖尿病的控制都有一定的程度的認知，所以發生控制不良的比例較高，變化的程度也較大，也較有機會成為血糖控制不佳的高危險群。此結果亦和臨床現象相符。

從到家醫科門診看病的病人裏面，我們也發現看診醫師比較固定的病人中，21-55 歲的病人有比較好的血糖控制，應該與病患比較好的教育程序有關，而在 70 歲以上老年中，女性則有比較佳的血糖控制成效。此結果與台灣民情吻合，一般女性比較注意自己的疾病現況。而在非比較固定醫師看診的病人中，都會區的民眾有較佳的血糖控制，這與都會區民眾疾病資訊的獲得和教育程度較高皆有關係。

6. 結論

資料探勘的分類分析是利用機器學習的方式，從大量的歷史資料中，萃取隱含在資料中的潛在有用知識，藉以建構出特徵與事實結果間之關係模式。由此發現模式可以了解血糖控制不佳是危險群的特徵，此結果可以提供醫師或衛教人員參考，針對血糖控制不佳比例較高的族群，提供更適當的指導及協助，降低糖尿病併發症發生的比例，提高病人的生活品質、減少醫療資源的浪費。

我們發現病人是否參加糖尿病醫療品質改善計劃為血糖控制的一項強而有力的因素，其對血糖控制的影響遠大於其它因子，而多數收案的患者其血糖控制均落在控制中等這一區間，然而沒收案的病患之血糖控制則受到多個因素的影響。未收案病患則以看診科別為其最重要影響血糖控制的因子，其中，內科就診病患多為控制中等(70%)，而其它就診科別影響血糖控制的成因較多。

糖尿病一般為中老年之疾病，同時伴隨著不同的併發症，對個人生活品質及醫療資源的使用有相當大的影響。近年來由於肥胖的增加，發生的年齡逐漸的降低未來工作可以集中在不同的居住地區，年齡階層、參與醫療保險的狀況和社會經濟地位做進一步的研究，可以發現更多及有用的知識。

參考文獻

- [1].Agardh,E.E. and colleagues.: *Explanations of socioeconomic differences in excess risk of*

type 2 diabetes in Swedish men and women. Diabetes Care. 27 (2004) 716-721

- [2].American Diabetes Association.:*Clinical Practice Recommendations* 2006. Diabetescare.Vol29,S1(2006)Retrieved March 8 ,2006,from http://care.diabetesjournals.org/content/vol29/suppl_1/
- [3].Breault, J.L., Goodall, C. R., Fos, P.J.: *Data mining a diabetic data warehouse. Artif Intell in Med.* 26 (2002) 37-54
- [4].Bureau of National health insurance, Department of Health, Taiwan.: *National health insurance Diabetes Healthcare Service Quality Report.* (2005) Retrieved March 8,2006,from http://www.nhi.gov.tw/webdata/AttachFiles/Attach_2095_1_940422.doc
- [5].Bureau of National health insurance, Department of Health, Taiwan.: *Diabetes Healthcare Quality Improvement Program (DHQIP).* (2005). Retrieved March 8, 2006, from http://www.nhi.gov.tw/webdata/AttachFiles/Attach_3078_2_w0950059032-a1.pdf
- [6].Catlett, J.: *On changing continuous attributes into ordered discrete attributes.* In Proceedings of the European Working Session on Learning. (1991) 164-178
- [7].Centers for Disease Control and Prevention, U.S. Department of Health and Human Services.: *National diabetes fact sheet: general information and national estimates on diabetes in the United States,* 2005. (2005) Retrieved March 8,2006,from http://www.cdc.gov/diabetes/pubs/pdf/ndfs_2005.pdf
- [8].Congdon, P.: *Estimating diabetes prevalence by small area in England. J. Public Health Med.* 28 (2006) 71-81
- [9].Connolly, V., N Unwin, P Sherriff, R Bilous,W Kelly.: *Diabetes prevalence and socioeconomic status: a population based study showing increased prevalence of type 2 diabetes mellitus in deprived areas.* J. Epidemiol Community Health. 54 (2000) 173-177
- [10].Dalstra, J., A. Kunst, C Borrell, E Breeze, E Cambois, G Costa, J. Geurts, E Lahelma, H Van Oyen, N. Rasmussen, E Regidor, T Spadea, and J. Mackenbach.: *Socioeconomic differences in the prevalence of common chronic diseases: an overview of eight European countries.* Int. J. Epidemiol. 34 (2005) 316 - 326

- [11].De Almeida, Mauricio A.; Lounis, Hakim; Melo, Walcelio L.: **An Investigation on the Use of Machine Learned Models for Estimating Software Correctability.** *Int. J. of Software Engineering & Knowledge Engineering.* 9 (1999) 565-593
- [12].Department of Health, Executive Yuan, Taiwan.: **White Paper on Health Indicators in Taiwan**, 2010. Taipei (2005)
- [13].**Diabetes Control and Complications Trial research group.: The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus.** New England Journal of Medicine. 329 (1993) 977-986
- [14].Dougherty, J., Kohavi, R., Sahami, M.: **Supervised and unsupervised discretization of continuous features.** In Proceedings of the 12th International Conference on Machine Learning. (1995) 194-202
- [15].Dunham, M. H.: **Data mining: introductory and advanced topics.** Prentice Hall, Upper Saddle River NJ (2003)
- [16].Han, J. and. Kamber, M.: **Data Mining: Concepts and Techniques.** Morgan Kaufmann Publishers, San Francisco CA (2001)
- [17].Kerber, R.: **Chimerge: Discretization for numeric attributes.** In National Conference on Artificial Intelligence (1992) 123-128. AAAI Press.
- [18].National Committee for Quality Assurance (NCQA).: **Diabetes Quality Improvement Project Initial Measure Set.** (2003). Retrieved March 8, 2006, from <http://www.ncqa.org/dprp/dqip2.htm>
- [19].Pabarskaite, Z.: **Decision trees for web log mining.** *Intelligent Data Analysis.* 7 (2003) 141-154
- [20].Quinlan, J. R.: **C4.5: Programs for Machine Learning.** Morgan Kaufmann, San Diego (1993)
- [21].Quinlan, J. R.: **Improved use of continuous attributes in C4.5.** *J. Artificial Intelligence Research.* 4 (1996) 77-90
- [22].Rathmann, W., B. Haastert, A. Icks, G. Giani, R. Holle, C. Meisinger, A. Mielck, and The KORA Study Group.: **Sex differences in the associations of socioeconomic status with undiagnosed diabetes mellitus and impaired glucose tolerance in the elderly population: the KORA Survey 2000.** *Eur J. Public Health.* 15 (2005) 627-633
- [23].Roglic, G., Unwin, N., Bennett, P. H., Mathers, C., Tuomilehto, J., Nag, S., Connolly, V., and King, H.: **The Burden of Mortality Attributable to Diabetes: Realistic estimates for the year 2000.** *Diabetes Care.* 28 (2005) 2130-2135
- [24].Rudolfer S. M., Paliouras G., Peers I. S.: **A comparison of logistic regression to decision tree induction in the diagnosis of carpal tunnel syndrome, Computers and biomedical research.** 32(5) (1999) 391-414
- [25].UK Prospective Diabetes Study (UKPDS) Group.: **Intensive blood-sugar control with sulphonylureas or insulin compared with conventional treatment in patients with type 2 diabetes.** *Lancet.* 352 (1998) 837-853
- [26].Wild, S., Roglic, G., Green, A., Sicree, R. and King, H.: **Global Prevalence of Diabetes: Estimates for the year 2000 and projections for 2030.** *Diabetes Care.* 27 (2004) 1047-1053

表 1. 典型的樣本(Characteristics of the sample)

特性		目標	中等	危險	TTL
年齡	21-55	199	412	226	837
	56-70	281	631	253	1165
	71 以上	284	477	125	886
性別	男性	387	741	297	1425
	女性	377	779	307	1463
居住地	大都市	622	1193	445	2260
	非都市地區	138	308	152	598
健保類型	沒有健保	21	40	12	73
	有健保且需共同支付	680	1391	551	2622
	有健保且免徵共同支付額	57	85	33	175
就醫科別	家庭醫生科別醫生	192	426	115	733
	內科醫生	543	1065	475	2083
	其他	29	29	14	72
就醫次數	2-3	315	458	201	974
	4-6	259	568	190	1017
	7 以上	190	494	213	897
糖尿病就醫次數	2	350	492	216	1058
	3-5	257	580	193	1030
	6 以上	157	448	195	800
與醫生關係	穩定	443	938	361	1742
	不穩定	321	582	243	1146
併發症	有	292	692	290	1274
	無	472	828	314	1614
參加 DHQIP	有	266	704	226	1196
	無	498	816	378	1692
A1C		764	1520	604	2888

Note: NHI: National Health Insurance 全民健康保險

DHQIP: Diabetes Healthcare Quality Improvement Program 糖尿病的醫療品質改善計劃

表 2.混淆矩陣 (Confusion matrix)

		以危險 因素為 前題	中等類別	目標	總數
現實 類別	危險	6	586	12	604
	中等	14	1458	48	1520
	不標	4	725	35	764
	總數	24	2769	95	

表 3. C4.5 決策規則(Decision Rules generated from C4.5)

1	如果病人參加 DHQIP 然後血糖控制是在中等水平。
2	如果病人沒有參加 DHQIP，並就醫於內科，然後控制血糖是在中等水平。
3	如果病人沒有參加 DHQIP 並就醫於家庭醫生門診和診療糖尿病次數>2，血糖控制是在中等水平。

4	如果病人沒有參加 DHQIP 並就醫於家庭醫生門診和診療糖尿病次數 ≤ 2 和門診醫師的關係是穩定且年齡是 56-70，然後血糖控制是在中等水平。
5	如果病人沒有參加 DHQIP 並就醫於家庭醫生門診和診療糖尿病次數 ≤ 2 和門診醫師的關係是穩定穩定年齡是 56-70，然後血糖控制是在中等水平。
6	如果病人沒有參加 DHQIP 並就醫於家庭醫生門診和診療糖尿病次數 ≤ 2 和門診醫師的關係是穩定且年齡 >70 的女性，然後血糖控制在目標水平。
7	如果病人沒有參加 DHQIP 並就醫於家庭醫生門診和診療糖尿病次數 ≤ 2 和門診醫師的關係是穩定且年齡 >70 的男性，然後血糖控制是在中等水平。
8	如果病人沒有參加 DHQIP 並就醫於家庭醫生門診和診療糖尿病次數 ≤ 2 和門診醫師的關係是不穩定的，住在都會區，然後血糖控制在目標水平。
9	如果病人沒有參加 DHQIP 並就醫於家庭醫生門診和診療糖尿病次數 ≤ 2 和門診醫師的關係是不穩定並且居住在非都會區，然後血糖控制是在中等水平
10	如果病人沒有參加 DHQIP 並就醫於其他科別和診療糖尿病次數 ≤ 2 ，血糖控制在目標水平。
11	如果病人沒有參加 DHQIP 並就醫於其他科別和診療糖尿病次數是介於 3~5 次，然後血糖控制是在風險水平。
12	如果病人沒有參加 DHQIP 並就醫於其他科別和診療糖尿病次數 >5 次，然後血糖控制是在中等水平。