

# 國立交通大學

管理學院(資訊管理學程)碩士班

## 碩士論文

運用資料倉儲及資料探勘分析捐血行為

**Analyzing the Behaviors of Blood Donor Based  
on Data Warehousing and Mining**



研究生：黃莉雯

指導教授：劉敦仁 博士

中華民國九十四年六月

運用資料倉儲及資料探勘分析捐血行為

**Analyzing the Behaviors of Blood Donor Based  
on Data Warehousing and Mining**

研究生：黃莉雯


Student : Li-Wen Huang

指導教授：劉敦仁 博士

Advisor : Dr. Duen-Ren Liu

國立交通大學  
管理學院資訊管理學程  
碩士論文

A Thesis  
Submitted to Institute of Information Management  
College of Management  
National Chiao Tung University  
in Partial Fulfillment of the Requirements  
for the Degree of  
Master of Business Administration  
in  
Information Management



June 2005

Hsinchu, Taiwan, the Republic of China

中華民國九十四年六月

# 運用資料倉儲及資料探勘分析捐血行為

研究生：黃莉雯

指導教授：劉敦仁 博士

國立交通大學管理學院資訊管理學程

碩士論文

## 摘要

企業的生存隨時面臨來自經濟、政治、市場景氣、競爭對手等各方面地挑戰，故須無時不刻地戰戰兢兢，因應市場變化，求新、求變，瞭解競爭對手的動向，更須瞭解顧客在哪裡及顧客的需求。為達此項目的，從內部存在的企業資料及外部資料中均可挖掘出顧客或潛在顧客的相關資訊，從而瞭解其行為模式，進而預測購買行為。企業藉此化被動為主動，提升顧客滿意度，亦提高企業獲利能力。

捐血中心為非營利機構，雖非企業，其存在目的非追求獲利及最大利潤，但存在類似問題—如何瞭解捐血人行為模式以妥善管理血量的調度，包括捐血人募集及血品的供應調整。捐血人在哪裡？捐血人的重複捐血率、不適合捐血者有哪些特性？第一次捐血人特性為何？如何保持各種不同效期血品對醫院的供應無虞？

本研究以建置之捐血資料庫，架構資料倉儲，提供捐血中心管理人對捐血人捐血行為之瞭解，進而做為決策之參考，以提升對捐血人服務品質及血液供應品質，以達捐血人、捐血機構、用血人參贏之目標。

關鍵字：捐血人、捐血行為、資料倉儲

# Analyzing the Behaviors of Blood Donor Based on Data Warehousing and Mining

Student : Li-Wen Huang

Advisor : Dr. Duen-Ren Liu

Institute of Information Management

College of Management, National Chiao Tung University

## Abstract

Enterprises may at any time face a variety of challenges that affect their business viability. These challenges may come from various aspects such as the state of the economy, politics, market prosperity and competitions from counterparts. Therefore, enterprises must be always readily, according to market change, to invent and innovate, understand counterparts' moves, and most of all, must know where customers are and what they need. In order to achieve the above objectives, enterprises may obtain customers or potential customers' relevant information from internal and external data, to understand their behavior mode so as to predict their purchase ability. By means of this proactive strategy, rather than a passive one, increases the degree of customers' satisfaction and enterprise profitability.

Taipei Blood Center is a non-profit organization, this means, the purpose of its existence is not for pursuing profits or a high profitability. Nevertheless, it faces similar challenges as a profit-driven organization—how to understand blood donors' behavior mode in order to appropriately manage the level of blood stock, including the supply dispatch of blood products and how to promote for more blood donors; where are blood donors? Blood donors' Repeat Blood Donation Rate? What are the traits for those who are not suitable to be a blood donor? What is the reason for first-time blood donor? How to maintain a range of different blood expiry dates to secure hospital blood supply?

The research conducted is based on the donor database established in the Blood Center and the data warehouse structure to provide the Blood Center management a more understanding in blood donors' donation behaviors. The results are to be used as strategic

references for the purpose of enhancing service quality to blood donors and blood supply quality, and to achieve an all-win objective for the blood donor, blood donation organization and blood receiver.

keyword : Blood Donor, Donation Behaviors, Data Warehousing



## 誌 謝

感謝指導教授劉敦仁博士悉心指導與鼓勵。劉老師的認真教學、關懷學生，使我在知識領域受益匪淺。在論文撰寫過程中，不厭其煩地從旁指導，使本論文得以順利完成。

承蒙 陳安斌所長與 陳恭教授在口試期間不吝指正，並提供了許多寶貴的意見，使本論文更為完整。在論文進度審查時的林妙聰教授，認真仔細的態度亦令學生感佩。

能在踏入社會後又重回校園繼續學業，實為有幸，但更幸運的是能受教在交通大學資訊管理研究所的諸位老師指導下，讓我感受到學問的追求也可以是種生命的享受，也體會到師者及學者的風範。

在此向您們致上我最誠摯的敬意與謝意



黃莉雯

# 目 錄

中文摘要	.....	i
英文摘要	.....	ii
誌謝	.....	iv
目錄	.....	v
表目錄	.....	vii
圖目錄	.....	viii
一、	緒論.....	1
1.1	研究背景及動機.....	1
1.2	研究目的.....	4
1.3	論文架構.....	5
二、	文獻探討.....	6
2.1	資料倉儲 (Data Warehouse) .....	6
2.2	線上分析處理 (OLAP) .....	8
2.3	資料探勘 (Data Mining) .....	9
三、	系統分析與設計.....	16
3.1	研究範圍.....	16
3.2	需求分析.....	16
3.3	系統架構.....	18
3.4	資料倉儲資料表設計.....	18
3.5	軟、硬體設備需求.....	22
四、	系統實作與展示.....	23
4.1	建立捐血維度資料表.....	23
4.2	Cube 模式分析.....	24
4.3	樞紐分析表建立.....	25
4.3.1	捐血人選擇以 email 寄發資料分析.....	25
4.3.2	捐血人捐血習性分析 (時間、職業、年齡區間、性別、捐血 方式) .....	30

4.3.3	分離術捐血人特性（職業、年齡區間）.....	35
4.3.4	不適捐血人與永久拒捐者分析.....	37
4.4	建立決策樹採擷模式.....	38
4.4.1	捐血方式與性別、首捐、年齡決策樹.....	38
4.4.2	C肝炎表面抗原（Anti-HCV）與首捐、年齡決策樹.....	39
4.4.3	捐血方式與性別、首捐、年齡變數網路依存性.....	40
4.5	建立叢集採擷模式.....	41
五、	結論與後續研究方向.....	43
5.1	結論.....	43
5.2	後續研究方向.....	43
參考文獻	.....	45





## 表目錄

表 1	資料倉儲與傳統資料庫的比較表.....	7
表 2	OLAP 與資料探勘的差異.....	8
表 3	居住地（隸屬中心別、桃、竹、苗地區）與職業別使用 email 願意比率表.....	27
表 4	居住地（隸屬中心別、桃、竹、苗地區）與職業別捐血人比率表.....	28
表 5	捐血人性別、職業別與捐血方式數量表.....	33
表 6	分離術捐血人性別、職業別與捐血方式數量表.....	36



## 圖目錄

圖 1	血液捐入處理流程.....	3
圖 2	客戶關係管理建立模式及預測的過程.....	12
圖 3	捐血資料應用資料倉儲系統架構圖.....	18
圖 4	以 DTS 做資料轉換.....	20
圖 5	DTS 轉換後資料表.....	22
圖 6	捐血資料維度資料表結構描述.....	24
圖 7	捐血方式與使用 email 意願捐血人關係.....	24
圖 8	年齡、性別與使用 email 意願捐血人關係.....	24
圖 9	年齡、性別與使用 email 意願捐血人次直條圖.....	26
圖 10	年齡、性別全部捐血人次直條圖.....	26
圖 11	居住地與使用 email 願意比率圖.....	29
圖 12	居住地與佔全部捐血人次數量圓形堆疊圖.....	29
圖 13	居住地（桃竹苗）、職業別與使用 email 願意直條圖.....	29
圖 14	時間—職業別堆疊直條圖.....	31
圖 15	時間—年齡區間堆疊直條圖.....	32
圖 16	年齡區間—職業別比例直條圖.....	32
圖 17	性別、職業別—捐血方式關係堆疊直條圖.....	34
圖 18	分離術捐血人年齡區間、職業數量直條圖.....	36
圖 19	非首捐檢驗不合格列“永久拒捐”之性別、年齡區間與職業捐血人次直條圖.....	37
圖 20	首捐檢驗不合格列“暫緩捐血”之性別、年齡區間與職業捐血人次直條圖.....	38
圖 21	捐血方式與性別、首捐、年齡決策樹.....	39
圖 22	Anti-HCV 與首捐、年齡決策樹.....	40
圖 23	捐血方式與性別、首捐、年齡變數關聯性.....	40
圖 24	捐血方式與性別、首捐否、職業別、性別叢集模型.....	41

# 一、緒論

## 1.1 研究背景及動機

92 年 4 月至 6 月間，國內暴發 SARS，除引起全國人民的不安外，對於捐血中心也是一次嚴格地考驗[3]。在此期間，對於捐血量及用血量均呈現了有異於往年的狀況，在血液效期限制的條件下，有效控制變動性增大的捐入量，以調配亦產生變動的醫院用血量，在過去均是未曾有的經驗。

93 年 1 月起，健保實施「卓越計畫」後[7]，供血量減少 5%。在血液供應減少、檢驗、人事各項成本增加，捐血、供血變數增加下，捐血機構的生存也日漸困難。

企業中 80% 的利潤來自於 20% 的客戶，若能將客戶流失率減少 5%，利潤將 100% 的成長。當企業不斷地投下資金大量行銷活動時，卻未能獲得實際的利潤，因為其未能掌握何者才是產品訴求的正確對象。

何時行銷 (when)？對象是誰 (who)？什麼種類的行銷方式才能吸引目標對象(what)？於何地行銷 (where)？如何行銷 (how)？在今日，企業行銷的重心已逐漸由「產品」轉移至「顧客」，誰能掌握顧客的喜好及需求，誰就掌握市場。

即使是非營利機構，雖不訴諸於利，但為與時代潮流共進，亦須在產品品質與服務品質中提升，追求成長亦須是如此。

### 1.1.1 研究對象簡介

本研究對象為台灣血液基金會轄下新竹捐血中心。93 年全國全年捐血量達 230 萬袋；目前累計捐血人數有 450 萬人。新竹捐血中心為 81 年，由台北捐血中心獨立出之新中心，負責業務轄區包括桃園縣／市、新竹縣／市及苗栗縣／市。至目前為止，累計捐血人數有 80 萬人，93 年捐血量 33 萬袋、18 萬捐血人次，平均每人捐血次數 1.58 人/次，轄區內國民捐血率 10.29%[5]。

台灣血液基金會任務包括：

- 1.捐血事業之策畫與執行事項。
- 2.捐血制度之建立及用血安全之研究發展事項。
- 3.血液科技之研究事項。
- 4.各地公私立醫院病患用血之採集、檢驗、與供應事項。
- 5.捐血人健康維護研究事項。
- 6.不適輸用血液之利用及安全處理事項。
- 7.重大災變或戰時大量用血之籌劃供應事項。
- 8.國產血漿製劑之委託製造、儲存及供應事項。
- 9.其他有關捐(供)血事項。

台灣的血液事業為 63 年 4 月 19 日由“中華民國捐血運動協會”成立而發端。又於 79 年 1 月配合國家醫療法頒布施行，由協會捐助設立“中華民國捐血事業基金會”，81 年 7 月更名為“財團法人中華血液基金會”，93 年 11 月又更名為“台灣血液基金會”。目前按業務區域劃分，於全省分設有台北、新竹、台中、台南、高雄、花蓮等六捐血中心。

對於所捐來的血液，除了 ABO 及 Rh 血型檢驗外，尚有肝功能、梅毒血清、B 型肝炎病毒表面抗原、C 型肝炎病毒抗體、人類免疫缺乏病毒抗體、人類嗜 T 淋巴球病毒抗體及紅血球異體抗體等九項檢驗項目。[4]

維繫著全國的血液供應業務，各地捐血中心有著來自全國民眾無償捐血所建立的豐富捐血資料庫。目前資料庫的運用偏重於：

- 1.承辦捐血活動團體資料的建立（捐血日期、活動地點、團體相關資料、捐血量）。
- 2.血液捐入後各項資料的蒐集（捐血人基本資料：姓名、生日、電話、地址、年齡、職業、體重）。
- 3.捐血人血液各項檢驗結果（ABO 及 RH 血型、血清轉胺酶、梅毒血清、B 型肝炎病毒表面抗原、C 型肝炎病毒抗體、人類免疫缺乏病毒抗體；人類嗜 T 淋巴球病毒抗體、紅血球異體抗體）。

- 4.血液捐入後成分分離成品、瑕疵品（全血及成分血：血小板濃厚液、洗滌紅血球、減除白血球之紅血球濃厚液、冷凍去甘油紅血球、白血球濃厚液、分離術白血球、血小板濃厚液、分離術血小板、新鮮冷凍血漿、冷凍血漿、冷凍沈澱品、分離術血漿）。
- 5.完成品後之血液供應資料（供血日期、供應血品項目、領血醫院）。
- 6.用血後醫院回報資料（用血病患姓名、用血成品、疾病名稱）。
- 7.由上述各項資料產生之各捐血中心統計報表，及彙總至基金會合計之各式統計報表。

### 1.1.2 血液捐入處理流程

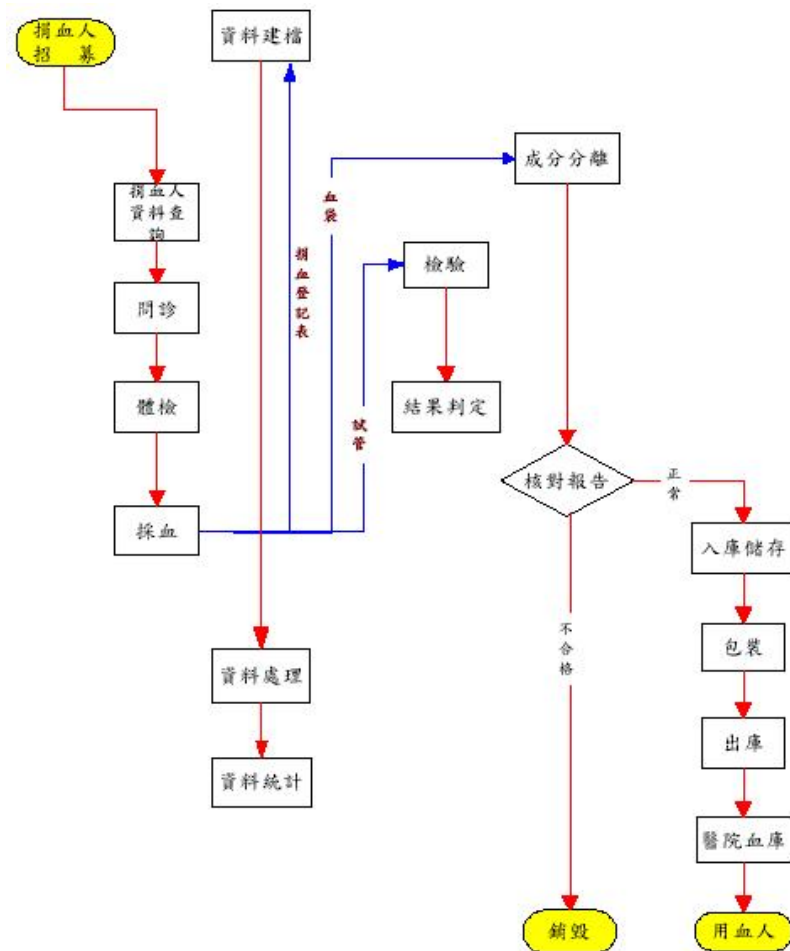


圖 1 血液捐入處理流程

資料來源：本研究整理

### 1.1.3 研究動機

除因 92 年 SARS 事件及 93 年 1 月起實施的「卓越計畫」因素影響外，因目前全國各醫療院所的用血，全數依靠台灣血液基金會分佈在全國六處捐血中心供應。各地捐血中心自控庫存，而業務轄區內的醫院直接向其訂血。血液的保存係有其期限的。在平時無特別事故時僅須保持一般安全庫存及各種血型比例的調配尚屬容易。一旦遇有特殊事故（如地震、SARS 等重大事故發生）及每年的寒、暑假、春節過年前後、春假時，易發生因習俗、人口流動或心理因素影響捐血意願，致捐血量分佈極為不均。各地捐血中心缺血或血液庫存過高，有報廢之虞時間亦多同時發生，故難以發揮共濟之效。

因捐血量的捐入不均，產生的問題有：

- 1.醫院訂血量依“經驗法則”訂血，致捐血中心血液庫存調度困難。在醫院下訂血通知單時，猜測捐血中心當時的庫存量。致會發生各依“經驗法則”預測訂購量：當 A.血量充足時期，會少訂些（即將捐血中心當作倉庫）。B.血量不足時期，多訂些（將倉庫轉回自家醫院）。
- 2.各捐血中心間庫存難發揮支援效果—血液來源不穩定，“多者恆多，少者恆少”，雖目前已有血液調撥機制，但因各地發生不足或過多的狀況多為同時，故提升調撥困難程度。

血液的來源，多來自學生，故當開學時期，血液有時會多到超過最高庫存量，有報廢之虞。在學生放假時期及特殊假期（如四月份春假及過年前後之長假），則常有低於最低庫存量，頻傳“血荒”。當捐血中心接獲醫院訂血通知單時，先查詢庫存，若足夠可全數供給，若不足時，會先向其他捐血中心調撥，但缺血時，各中心自顧不暇，無法支援。

## 1.2 研究目的

目前各項資訊系統作業均屬 OLTP(OnLine Transaction Processing)方式，仍偏重於資料輸入保存，若有統計或查詢需求時，以已設計好的程式功能讀取，或須依台灣血液基金會訂定之 ISO 流程會簽另提程式新增要求，時程費時，且

對於中心個別彈性之查詢需求往往較難滿足。

故本研究係以目前捐血中心的捐血資料庫內過去捐血人的歷史資料，架構資料倉儲，透過 OLAP(On-Line Analysis Processing)及利用 Data Mining 的技術，尋找出捐血人的行為模式，以提供捐血中心決策者採血作業的參考並找出潛力捐血對象目標，以維繫與捐血人間之關係，進而吸引其捐血，以達血源穩定為目標。

### 1.3 論文架構

本論文共分五章，其結構如下：

第一章「緒論」：說明本研究的研究背景、動機、研究目的及研究架構。目前在捐血機構的資料庫使用上仍以 OLTP 為主，資料倉儲與資料探勘技術運用在決策支援之應用仍為少見，故以此為主題，作為此篇論文的動機。

第二章「文獻探討」：主要介紹資料倉儲、OLAP 與資料探勘關係與技術的相關文獻，以建置出本研究的目標。

第三章「系統分析與設計」：本章實際以新竹捐血現行業務系統的資料，定義研究範圍、需求分析、系統架構圖、資料倉儲資料表設計,包括資料轉換及預處理與軟、硬體設備需求。

第四章「系統實作與展示」：將實作結果做一呈現與分析，以建立的資料倉儲，接著利用 Analysis Service 工具，建立各種分析模型，最後以 Microsoft 提供之 Data Mining 技術，將資料項目做預測。

第五章「結論與建議」：希望能將本論文的架構，推廣應用於其他作業程序項目及全國捐血機構，以達到資訊共享之目的，進而提供決策制訂之重要參考。

## 二、文獻探討

本研究係以建置資料倉儲(Data Warehouse)，使用 OLAP 與資料探勘 (Data Mining) 的技術來達成研究目標，相關文獻如下。

### 2.1 資料倉儲(Data Warehouse)

#### 2.1.1 資料倉儲(Data Warehouse)的定義

資料倉儲的正式定義係由 W.H.Inmon 所提供：指的是具有主題導向(Subject-Oriented)、整合性(Integrated)、時間差異性(Time-Variant)、不變動性(Nonvolatile)等特性的一種管理性資料庫，目的在於能快速支援使用者的管理決策[19]。

Ralph Kimball 將資料倉儲定義為：「人們可以存取其所需資料的儲存處」。而其需求如下[13]：

- 1.由資料倉儲中取得資訊，必須要夠快。
- 2.倉儲中的數值必須具有內部的一致性(consistency)。
- 3.使用者必須對倉儲中的資料，進行切割與比較的作業(萃取您所需要的資訊)。
- 4.資料倉儲必須包含易於使用的瀏覽工具。
- 5.倉儲的資料必須完整且可靠。
- 6.有正確的资料收集程序，才可能產生高品質的倉儲資料。

#### 2.1.2 資料倉儲與傳統資料庫的比較



表 1 資料倉儲與傳統資料庫的比較表

特色 \ 工具	傳統的資料庫儲存	資料倉儲儲存
資料的整合性	個別孤立，沒有整合；儲存在許多不同甚至不相容的平台	整合企業內外各種類型的資料；彙整儲存於單一平台上
資料的時間差異性	只有一個時點的資料記錄	包含目前與歷史資料（5~10年）
設計目的	支援交易作業	支援管理決策
資料設計導向	以程序工作流程來組織資料結構	以企業主要決策主題來組織資料結構
資料的變動性	資料隨時更新、互動	資料一旦存入則不能更新
資料的一致性	資料定義可能不一致	單一標準的定義
主要分析工具	資料庫結構化查詢語言(SQL)	OLAP

參考來源：[8]

資料庫以處理為中心，著重在用最短時間、最精確的方式，處理一個事先指定的動作（完成一筆交易、傳遞一份電子文件、印一份報表等）。資料倉儲以唯讀式分析(Read Only Analysis)為要務，著重在以最短的時間、最有彈性的方式，滿足決策者「知」的需求。

### 2.1.3 資料倉儲的應用

資料倉儲的設計目的是強調對大量資料具有分析能力，迅速地獲得資料。

資料倉儲是一種電子資訊集中儲存的所在。不同來源、不同型態的資料經過萃取、淨化、轉換、載入之後，以齊一的型態，有組織地排列儲存於 Warehouse 內，再以具前端使用者介面(Front User's Interface)，供使用者做查詢分析。

廣義的資料倉儲指得是一種 Solution，除資料集中儲存，亦包含了線上分析處理(OLAP)的功能。應客戶需要，有些資料倉儲也提供資料探勘(Data Mining)的應用。故資料倉儲可以應用的範圍就廣及決策支援系統及客戶關係管理及一些固定報表查詢功能等。

## 2.2 線上分析處理(OLAP)

### 2.2.1 線上分析處理目的及特色

OLAP 為 (On-Line Analytical Processing, 線上分析處理)。OLAP 為一種軟體技術，其目標為滿足決策支援或多維度環境特定的查詢和報表需求。它的技術核心為「維度」(Dimension)概念，正由於 OLAP 使用者可以從多維度來觀察和分析資料，故又稱「多維度分析」(Multiple Dimensional Analysis)。OLAP 亦可說是多維度資料分析工具的集合。

架構「線上分析處理」(OLAP)，可利用事先規劃的多種分析角度，有效地利用資料倉儲，進而轉換成決策支援分析的寶貴資訊。此種方式係統計來處理發現趨勢與關鍵性因素，而使用多維度結構，以提供快速存取資料，進行分析的技術，OLAP 的來源資料通常儲存在關聯式資料庫的資料倉儲中。

其特色為：快速(Fast)、分析(Analysis)、共享(Shared)、多維度(多種分析角度，Multidimensional)及資訊(Information)，此種技術可使資料倉儲有效地運用於線上分析，提供反覆的複雜分析查詢，而做快速地回應。

### 2.2.2 OLAP 與資料探勘的差異

OLAP 與資料探勘的最大差異點在於，OLAP 是透過事先定義的維度來檢視歷史資料，因此它屬於假說導向(Hypothesis driven)的方法學，我們只根據事先假設的維度下探(drill down)、積存(roll up)、切片(slice)來檢視資訊。而資料探勘則是透過資料導向(data driven)的方法學，將隱含在歷史資料中的規律以規則、方程式、類神經網路結構、函數等方式呈現，同時利用挖掘出來的規律預測未來。

表 2 OLAP 與資料探勘的差異

比較項目	OLAP	資料探勘
分析重點	會發生什麼?(What)	為什麼發生?(Why)
分析技術	剖分	自動發現

分析處理	客戶導向與控制	資料和系統導向
重 點	輪廓式資料／細節性資料	業務性資料／細節性資料
維 度	有限	許多
屬性數目	數十個	每一維度均有數百個

資料來源：[13]

## 2.3 資料探勘 (Data Mining)

資料探勘的技術正是運用企業中已存在的大量資料，挖掘出有益企業的資訊，找出客戶的行為通則以進行其他人的行為預測。當開發一個新客戶比維持一個舊客戶需花費更多的成本，資料探勘即是協助企業找出目標客戶及分析其喜好、習性，以最低的行銷成本創造最大的企業利益。

### 2.3.1 資料探勘於顧客關係之運用

資料探勘係利用過去資料，以挖掘出有用的資訊以辨識出誰是具有潛力的顧客→找出目標對象後提供其特殊的服務→維繫客戶關係，進而培養顧客忠誠度。且可提供企業獲得顧客資訊，使企業更瞭解客戶、做更正確的決策、發掘新的業務機會、對資訊的需求做更快速的回應及解決現有決策問題。

### 2.3.2 何謂資料探勘

Berry and Linoff 定義資料探勘：「資料探勘是為了發現有意義的模式或規則，以自動或半自動的方式，來勘查、分析大量資料所進行的流程」[23]。

Friedman 認為 "Data mining is a commercial enterprise that seeks to mine the miners "[22]。

所謂資料探勘 (Data Mining)，就是：

1. 「從大量資料庫中找到相關的模式(Relevant Patterns)並自動的萃取出可預期的資訊」。
2. 是從資料庫中發現知識，將隱含的、先前並不知道的，和潛在有用的資訊從資料庫中萃取出來的過程。

3.目的是要從龐大的資料庫中，將潛在有用的資訊及知識挖取出來。4.其是利用完備的統計與機器學習（Machine Learning）技術來建立自動預測顧客行為的模型。。

5.同時還能與商業資料倉儲（Commercial Data Warehouse）結合，發展出有價值的商業用途。2.3.3 資料探勘的六個應用方向

1.獲取新客戶(Customer Acquisition)：

- (1) 根據客戶的屬性來預測其對商品或通路計劃的反應。
- (2) 比照相對應的實際屬性與反應是否真如預期。
- (3) 從中挑選出尚未成為顧客者。

2.維繫客戶(Customer Retention)

- (1) 當資訊顯示企業的基本資料已開始流失或流失到對手陣營時，公司就該採取挽留措施。
- (2) 同時對那些還算穩定的顧客就該給些誘因使其更願意留下來。

3.放棄客戶(Customer Abandonment)當顧客資料出現「黑名單」，也就是企業投注於其上的費用遠超過他所回饋的，就該考慮是否停止為這些顧客付出努力與成本。

4.購物籃分析(Market Basket Analysis)

- (1) 購物籃指消費者所購買的商品種類及數量。
- (2) 分析消費者購買的產品對銷售產生的經濟效益稱購物籃分析或稱為聯合性分析(Association Analysis)。

5.需求預測與目標行銷

- (1) 處理過大量的資料，當再次收到一筆新資料時，電腦系統便會模擬完的結果。
- (2) 根據潛在顧客的特性去預測其需求。
- (3) 找出對提供商品最具有消費傾向的顧客。
- (4) 分析後可以加強對各種商品其主力顧客的促銷動作，而提高銷售量，且

可節省不必要的成本，如行銷與存貨的過剩或不足等。

#### 6.交叉銷售(Cross – Selling) 與主動銷售(Up-Selling)

##### (1) 交叉銷售:

A 如共同基金，不同之組合。

B 盡可能滿足顧客「一次購足」的需求。

##### (2) 主動銷售

根據不同族群的消費特性，向潛再客戶介紹適合的產品，但最有可能會對產品感興趣的人來。

### 2.3.4 資料採勘的五大模型

藉由五大模型，即分類、預測、分群、聯合性分析以及順序，資料採勘可以發揮強大的應用功能

#### (1).分類(Classification):

分類的目的是希望從資料的各種屬性值 (attribute values) 來判斷該資料的類別。例如根據信用卡申請資料來判斷申請者是好客戶、普通客戶、或者是屬於危險群。

#### (2).估計(Estimation):

藉由資料其他屬性來估計其某一個屬性的值例如根據不動產的各項屬性，如地段、樓層、面積....等來估計其市價。

#### (3).預測(Prediction):

根據相關資料的變化來預測某一現象是否將要發生，例如預測下個月將有哪些客戶會流失或預測股市的漲跌。

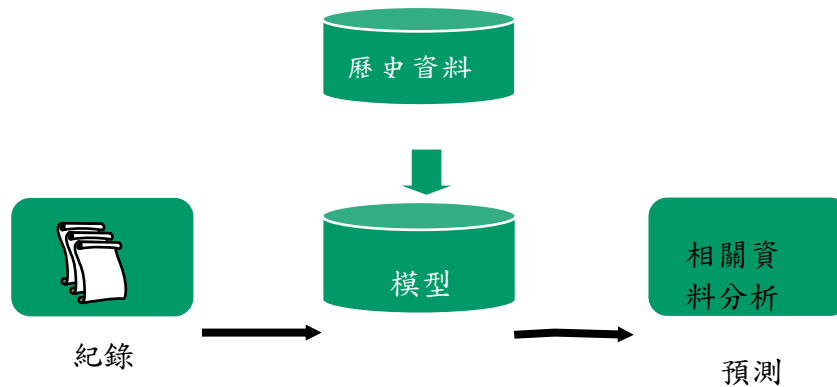


圖 2 客戶關係管理建立模式及預測的過程

資料來源：本研究整理

(4) 相關規律(Association Rule)：

找尋資料屬性間的關係，例如買鐵鎚的人通常也買釘子。

(5) 類聚(Clustering)：

資料依其相似程度分成數個類聚，事先可以知道有多少類聚，也可以不知，而對於每一筆資料屬於哪一個類聚事先通常未知，例如將網路使用者根據其上網行為分成數個群組。

以上的幾個目的有些也可以請專家根據經驗來設定規則，而資料探勘的目的就是希望可以透過機器學習(machine learning)的方式來使其自動化，並且可以彌補主觀的缺點及細微、隱藏無法由人察覺之處。

2.3.5 資料探勘處理步驟：

- 1.Data Cleaning :清除雜訊及前後矛盾的資料。
- 2.Data Integration :結合不同來源的資料。
3. Data Selection :挑選有意義的屬性或資料。
4. Data Transformation :將資料轉換成可以處理的格式。
5. Data Mining :抽出Pattern的步驟。
6. Pattern Evaluation :評估Pattern的可用性。
7. Knowledge presentation :將可用的知識表達出來。

資料探勘的確能協助業者做好顧客關係管理，即做到「讓企業能適時、適地、經適當部門對適當人選作適當反應」

資料探勘的技術正是運用企業中已存在的大量資料，挖掘出有益企業的資訊，找出客戶的行為通則以進行其他人的行為預測。當以上述 Data Mining 的處理步驟，運用於捐血中心的資料庫，透過與 Data Warehouse 的結合，從龐大的資料庫中找出隱藏於其中的有關捐血人豐富資訊，並探討其捐血行為，以辨認出屬“加強維繫”，或是屬應“放棄客戶”（被列為永久不得捐血者）。在此 Data Mining 均可提供滿意答案。

### 2.3.6 決策樹

#### 1. 何謂決策樹

決策樹係採用樹狀分岔的架構來產生規則，它適用於所有的分類問題。決策樹是從訓練資料中，由上到下產生一個特定的方向，利用某項特性做為節點來分割方向。若所有樣本屬於相同類別且能獲得辨認，則完成決策樹的分類。新事務由決策樹樹根節點開始測試進行，直到遇到樹葉為止，則此樹葉就是一個類別。

決策樹產生的流程如下：

- (1) 資料母體做為根節點。
- (2) 根據最佳變數產生分岔，產生子節點。
- (3) 根據每個子節點案例分布狀況指派分類結果。
- (4) 決策樹持續生長，最後採用修剪技術修減去不必要的規則。

#### 2. 決策樹差異

一般常的決策樹有 CHAID、C4.5 及 CART。

CHAID：CHAID 是 J.A.Hartigan 在 1975 年率先提出的演算法，受到廣泛使用。CHAID 是從更早的一套自動互動偵測系統 AID 衍生而來，後者是 J.A.Morgan 與 J.N.Sonquist 於 1963 年提出。CHAID 原本的動機是偵測變數之間的統計關係。它藉著建構一個決策樹來完成此一目的[17]。

CART：CART(Classification and Regression Tree,分類迴歸樹)，是 Leo Breiman、Jerome Friedman、Richard Olshen 和 Charles Stone 所開發的資料探勘和預測演算法。他們在研究從光譜資料分析氯的成分時，發現這項新技術可被用在各種不同的問題[19]。

C4.5：1979 年由 J.Ross Quinlan 提出一種名為 ID3 (Iterative dichotomiser 3) 的決策樹演算，其係改良 CLS(Concept Learning Ssystem) [17]。

ID3 是使用資訊報酬來做為分岔準則，但又發現應用在實際案例上，資訊報酬會偏好選擇項數較多的變數來做為分岔變數，因此使用該分岔準則所建立的決策樹規則數目偏多，且容易造成過度學習的效應。為修正這項系統偏誤，J.Ross Quinlan 於 1993 年重新定義出「增益比值(gain ratio)」的計算公式來取代原有的分岔準則，這就是 C4.5 決策樹的由來[9]。

CHAID 和 C4.5 及 CART 兩種演算法的最大差異在於，後兩者先過度套用資料，再加以修剪，而 CHAID 嘗試在過度套用的情況發生之前就讓決策樹停止蔓生擴大。另一差異是 CHAID 只限於類別變數使用，連續變數必須被區隔成幾個區段範圍，或是以高、中、低等類別來取代。

在本研究中使用微軟的 C4.5，故此以 C4.5 做進一步介紹。

前所提在 C4.5 中 J.Ross Quinlan 提出的增益比值計算公式取代原有的分岔準則。分岔準則的目的是用來檢視使用某變數做為分岔變數時，母節點與子節點總和的純淨度變化量，能使純淨度提升多的變數就是有效變數，其增益比值的公式如下：

$$\text{gain ratio} = (\text{entropy}_{\text{before}} - \text{entropy}_{\text{after}}) / \text{split gains}$$

→ 增益比例 = (母節點亂度 - 子節點亂度) / 分岔程度的修正量

$$\text{entropy}_{\text{母節點}} = - \sum X \log_2 X \quad X: \text{預測變數選項 } i \text{ 的分率 } n_i/n$$

$$\text{entropy}_{\text{子節點}} = \sum (n_{\text{子節點}}/n_{\text{母節點}}) * \text{entropy}_{\text{子節點}}$$

entropy 表示亂度，此觀念係來自化學熱力學，用來表示物體分布的分散狀況，亂度愈高，則愈無規則，故決策樹演算法的目標就是希望能降低資料分類



結果的亂度[9]。

系統會計算分岔時產生的增益比值何者為高外，另根據自動隨機抽樣切割的鑑效組資料再重複計算適合的下一層分岔變數。

### 2.3.7 叢集(Clustering)

叢集就是將不同的資料數據加以分類，然後提供使用者一個較佳的資料觀察點來分析資料。一般來說，叢集分析法可以提供使用者，尤其是行銷部門，瞭解產品及顧客之特性，進而制訂出有效之行銷策略。叢集分類的方式可依照策略之需要來制訂相關資料變數，並無特定的方式或最佳之方式，由於叢集與最鄰近者分析方法類似，因此二者經常配合使用。

在微軟叢集演算法中提供兩種，分別是 K-means 以及 E-M 演算法。

叢集演算法的始祖 K-means 演算法是 J.B.MacQueen 於 1967 年正式發表，由於原理簡單、計算快速，很快就成為商用資料探勘軟體中的基本配備。它是屬於前設式的叢集演算法，也就是必須事前設定群集的數量，然後根據該設定找出最佳的群集結構。其因群集之間不可互相重疊，故稱之為剛性群集。

E-M 演算法(Expectation Maximum, 期望最大值演算法)則是每一個群集範圍是完全重疊地，故稱為柔性群集(soft cluster)。其計算流程與 K-means 演算法很像，但唯一不同是，E-M 演算法是使用高斯分配(Gaussian distribution)來做為距離函數[9]。

## 三、系統分析與設計

### 3.1 研究範圍

本研究係以新竹捐血中心（業務轄區包括桃、竹、苗地區）為實作研究對象，方向偏重於捐入者部份，捐血人即是顧客，如何吸引優質顧客上門及避免不適捐血人白跑一趟，均是維繫血液品質的起始點。

本研究實作範圍如下：

時間：92/01/01~93/06/30 捐血及複檢者。

對象：於上述時間內捐血及複檢者。

地點：桃園縣／市、新竹縣／市、苗栗縣／市。

### 3.2 需求分析

依據目前捐血中心數項存在的問題中，以捐血人募集對象最能影響血液捐入品質，及若能減少後續處理成本，亦可有效降低捐血中心的費用支出。故本研究選擇以捐血人捐入之分析為目標，架構出以下需求：

1. 捐血人選擇以 email 寄發資料分析，包括：

- (1) 捐血方式與使用 email 意願捐血人關係
- (2) 年齡、性別與使用 email 意願捐血人關係
- (3) 居住地與使用 email 意願捐血人關係

那些捐血人接受以 email 寄發檢驗報告的意願較高？由於捐血人捐血後，捐血中心須將其資料建檔，並將當次所捐血液之檢驗結果寄發予該捐血人。過去均以紙本寄發，除印刷紙張費用外，人工處理及郵費更形成一筆可觀的費用支出，故改以 email 寄發後，相對地降低成本。近年來，各捐血中心均積極地推動 email 寄發。

2. 捐血人捐血習性分析，包括：

- (1) 時間－職業別關係
- (2) 時間－年齡別關係

### (3) 性別、職業—捐血方式關係

此處分析有關捐血人的捐血習性。何種捐血人？習慣在何時、捐何種類的血液(250cc 或 500cc、分離術)。

血量的捐入受到時間因素的影響甚大，尤其以寒、暑假春節期間。如何對於各時期捐入者的特性做瞭解，有益於血源的適量調節。

### 3 分離術捐血人特性分析，

包括：捐分離術捐血人之職業及年齡區間。

目前推廣的分離術捐血主要有二種：分離術血小板和減除白血球分離術血小板，其血品優點有：

(1) 一位捐血人即足以供給病患一次輸用所需的血小板，可以減少病人因輸用多數人血液成分可能產生的不良反應。

(2) 使用減除白血球分離術血小板可降低輸用病患產生 A. 原發性人類淋巴球抗原同種異體免疫反應。B. 血小板輸注無效。C. 非溶血性發燒反應。D. 輸血後感染 CMV、HTLV 病毒。E. 因輸血引起的免疫抑制反應。

但捐分離術時間較一般血久，一般捐全血約 20 分鐘，任何捐血點都可以，分離術捐血需要兩個多小時，且必須在有儀器設備的場所才可以捐，且血小板的保存期限只有 5 天，故捐分離術捐血人的來源須力求穩定與充足，避免集中於假日[1]。

### 4 不適捐血人與永久拒捐者分析

包括：首捐與非首捐因檢驗不合格而列為“永久拒捐”者之性別、年齡區間與職業之分析。

捐血後捐血中心會對血液做各項檢驗，其中包含血型、Rh 血型、血清轉胺酶(ALT)、B 型肝炎表面抗原(HBsAg)、梅毒血清、HIV(愛滋病毒抗體)、C 型肝炎病毒抗體(HCV)、HTLV(人類嗜 T 淋巴球病毒抗體)、紅血球異體抗體等 (ABS)。若首次捐血檢驗不合格，會列為“D”值(不適捐血)，但可在規定時間後複檢，在規定的複檢流程後若仍為不合格(HBsAg、HCV、HTLV、HIV)，則

列為“R”值（永久拒捐）。

以維持血液品質及減少不適輸用血捐入以降低後續的處理成本而言，識別出高危險群的特徵，亦是極為重要地工作。

#### 5.提供彈性化統計查詢供決策參考

利用規劃出的系統功能，提供彈性化的多維度查詢：例：在不同的年、季、月在桃竹苗地區不同性別、年齡、職業的捐血統計。

### 3.3 系統架構

本研究係將捐血中心資料庫的資料透過 Microsoft SQL Server 資料轉換服務(Data Transformation Service,DTS)做資料轉換後建立 Data Warehouse ，配合 SQL 2000 Analysis Service 建置 OLAP 及 Mining 工具挖掘出本次研究所需資訊，再利用建置的功能提供於捐血中心管理人員做資訊分享，提供決策參考（如下圖）。

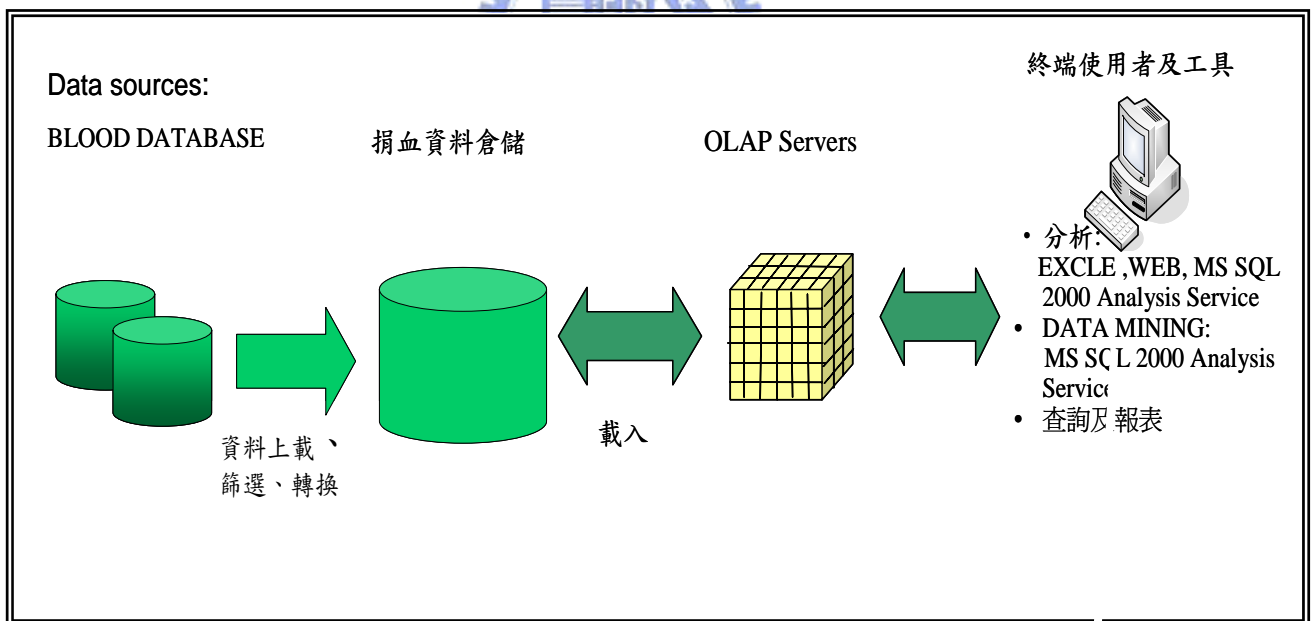


圖 3 捐血資料應用資料倉儲系統架構圖

資料來源：本研究整理

### 3.4 資料倉儲資料表設計

#### 3.4.1 資料蒐集

本研究係利用目前新竹捐血中心資料庫中儲存之：

1. 捐血人(donor)資料，包括：

- (1) 在新竹捐血中心業務轄區（桃園縣／市、新竹縣／市、苗栗縣／市）內捐血之捐血人【於本區捐血者，均擁有一獨立卡號（中心卡號）】及註記“R”、“W”值者。
- (2) 註記“D”值之不適捐血者：於其他中心捐血且曾任何一項檢驗不合格，且未經複檢合格者。
- (3) 註記“R”值之永久拒捐者：其於他中心捐血且經依規定流程複檢後仍不合格之部份檢驗項目，及依規定項目判為永久不得捐血者。
- (4) 註記“W”值之暫緩捐血者：於其他中心捐血或捐血前訪談發現屬於須於一段時間後始能捐血者（例：出國至屬疫區國家、手術後）。

在捐血人檔案中包含有捐血基本資料：捐血卡號、生日、體重、性別、是否願意收到 email、email、地址、電話、其他中心卡號、各項檢驗近值、近捐日期、下次可捐日期、拒捐碼、表揚日期、表揚次數....等共 127 欄位。

本案取其日期為 92/01/01~93/06/30 間曾於新竹捐血中心轄區內捐血或複檢之捐血人資料，且捐血人部份與本研究相關欄位，如后述。

- 2 捐血人血袋資料(blbag)：在新竹捐血中心業務轄區（桃、竹、苗）內之捐血或複檢者血袋或複檢試管均有一條碼編號。本檔包含血袋編號、捐血卡號、捐血日期、方式、袋別、血袋數量、年齡(捐血時年齡)、性別、各項檢驗結果、地址、電話、首捐碼...等共 106 欄位。

本案取其日期為 92/01/01~93/06/30 間曾於新竹中心轄區內捐血或複檢之血袋／試管資料，且血袋欄位部份與本研究相關欄位，如后述。

- 3 其他關聯檔案：例捐血方式、職業、捐血單位、地區別、血型等。

### 3.4.2 資料轉換

利用 Microsoft SQL 2000 DTS 轉換工具，將存於 SUN 主機上之 ORACLE 資料庫透過 ODBC 做資料轉換。因避免影響作業效率，故先將主要檔案轉換到

SQL SERVER 後，再續行以 DTS 轉換。

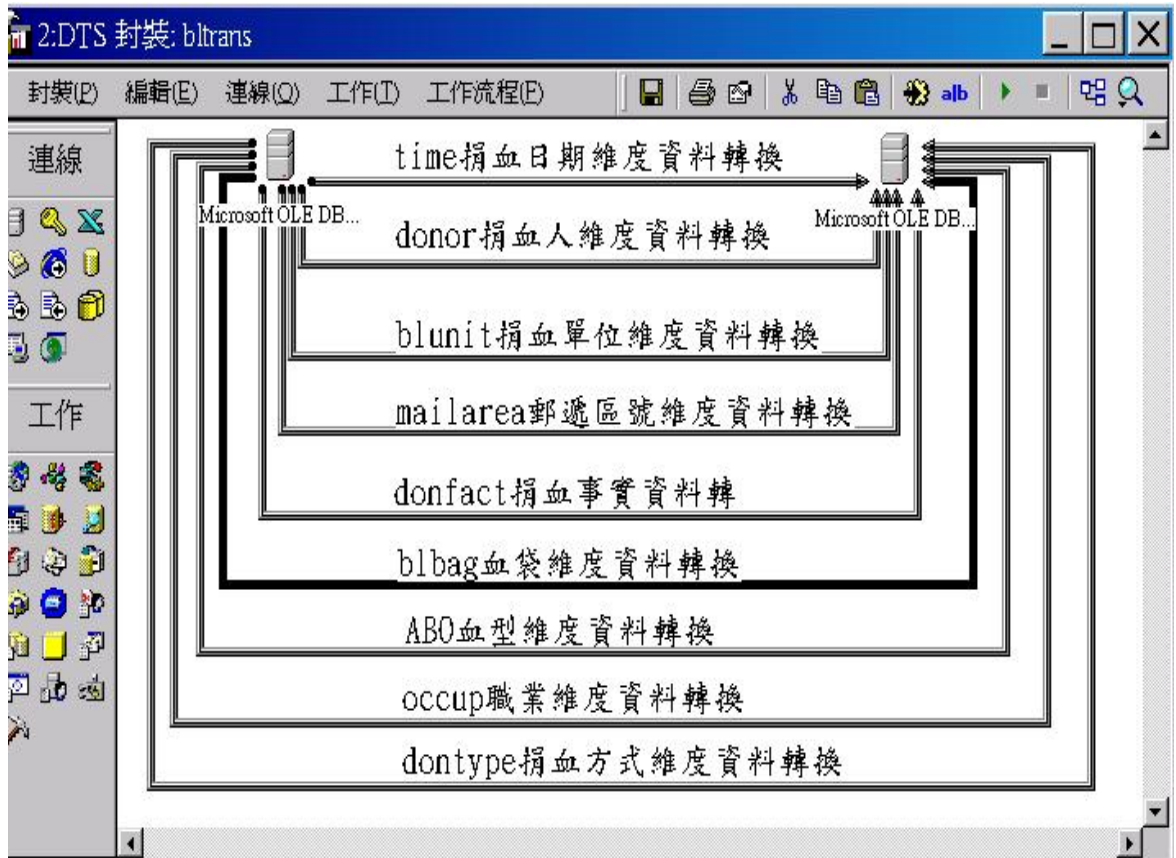


圖 4 以 DTS 做資料轉換

資料來源：本研究整理

### 3.4.3 資料預處理

在資料轉換過程，儲存在原 ORACLE 資料庫中部份資料，因原始資料庫、應用軟體設計或捐血後輸入人員輸入問題，致產生關聯錯誤，如下：

1. 年齡：因捐血人填寫資料時漏填生日，或有填寫為當日捐血日期，因無法及時連繫更正，第一線電腦查詢人員未即時核對，又因輸入時資料已寫入，致即使事後更正資料，值仍未齊步修正；或因應用軟體設計人員程式設計有誤，另可能為輸入人員輸入錯誤等問題，均致年齡出現 0 或負值。

對於本項問題，則因事後更正情況多，故另以 SQL 將年齡重計算後寫入年齡欄，餘因故而無法更正者，因筆數不多(19 筆，佔 0.005%)故刪除。另增年齡層分類檔。

2. 郵遞區號：部份捐血資料地址因輸入人員輸入錯誤或因行政區重編後捐血人未通知更正，造成未存在郵遞區號檔內，故無法產生關聯。

本項問題另以 SQL 將捐血人及血袋檔內郵遞區號欄位 update(11 筆)。

3. 身分證字號：應為唯一，但事實不然。原因有：

- (1) 部份捐血人身分證字號曾因與他人重號，更改後未通知資料修正。
- (2) 因早期戶政資料係以人工作業，故易產生錯誤，致確實存在二人重號。
- (3) 身分證字號空白：捐血機構成立之初，未要求驗證及填寫身分證字號，這些早期捐血者之後又未再來，故該欄位存在許多空白。
- (4) 無身分證字號：外籍人員，另以護照欄儲存字號。

因以上原因，故無法以身分證字號當 key。又捐血人在各中心均會有一唯一捐血卡號，故在此以捐血卡號做為 key。

4. “捐血日期”資料格式轉換：本欄位原以 ORACLE 資料格式 varchar2(8)儲存，透過 DTS 資料轉換後改為 datetime(8)。

### 3.4.4 DTS 轉換後資料表

本案以星狀架構模式設計，轉換後如下：

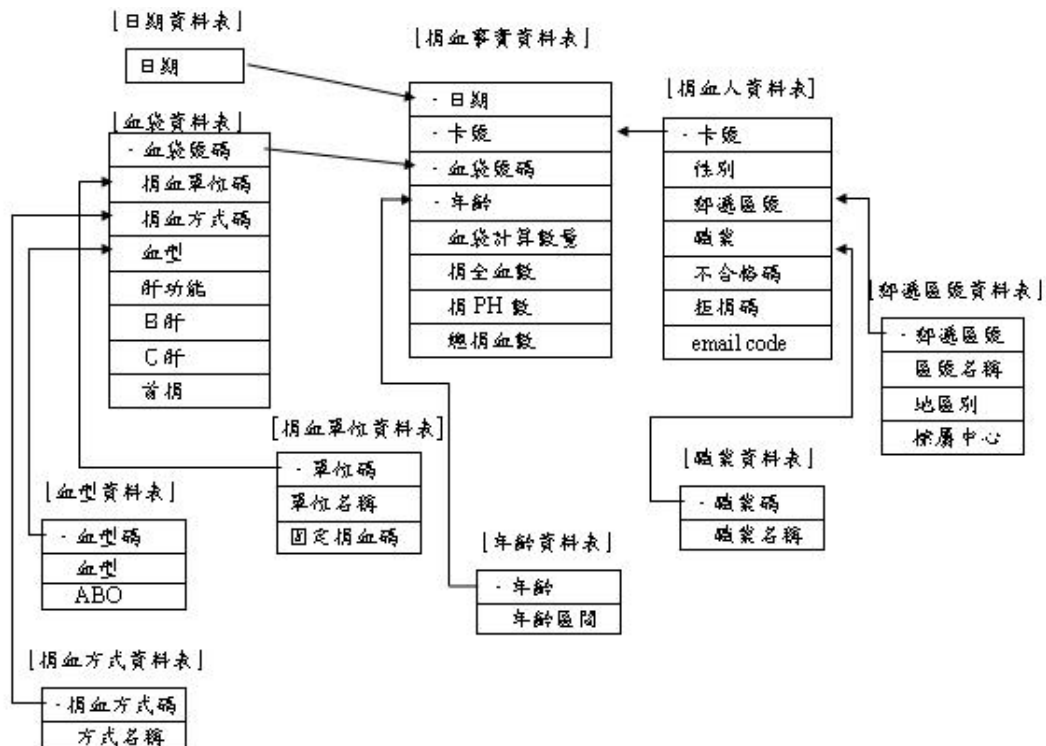


圖 5 DTS 轉換後資料表

資料來源：本研究整理

### 3.5 軟、硬體設備需求

- 1.個人電腦+Microsoft Windows 2000
- 2.電腦主機 (Sun V880) 、DBMS (Oracle 7.3.4)
- 3.資料轉換軟體：Microsoft Server SQL 2000 、Transformation Services (DTS) 、
- 4.線上分析工具：SQL 2000 Analysis Service 、Microsoft Excel 2002 、Internet Explorer
- 5.資料採擷工具：Microsoft SQL Server 2000 之分析管理員 (Decision Tree 及 Cluster )
- 6.圖形、報表工具：Microsoft Excel 2002





## 四、系統實作與展示

### 4.1 建立捐血維度資料表

利用 SQL 2000 Analysis Service Cube 編輯器建立如下：

1. 維度資料表：在此，加入年齡資料表。及建立了年齡別、職業別、捐血人血型、居住地（連絡地址）、捐血地點（各捐血車、點、站）、捐血方式(250cc、500cc、分離術捐血或複檢)、暫緩捐血人（D 值）、永久拒捐（R）、HbsAg(B 肝) 檢驗、時間、性別、email 寄發（是否接受 email 寄發之檢驗報告）、首捐值（是否第一次捐血）等 13 個維度資料。
2. 量值：包括有袋數（250cc 為 1 袋、500cc 及分離術捐血計 2 袋、分離術血小板兼收血漿計 3 袋）、人次、全血總捐血次數（本中心+其他中心捐全血次數）、PH 總捐血次數（本中心+其他中心捐分離術次數）、總捐血次數（全血總捐血次數+PH 總捐血次數）。

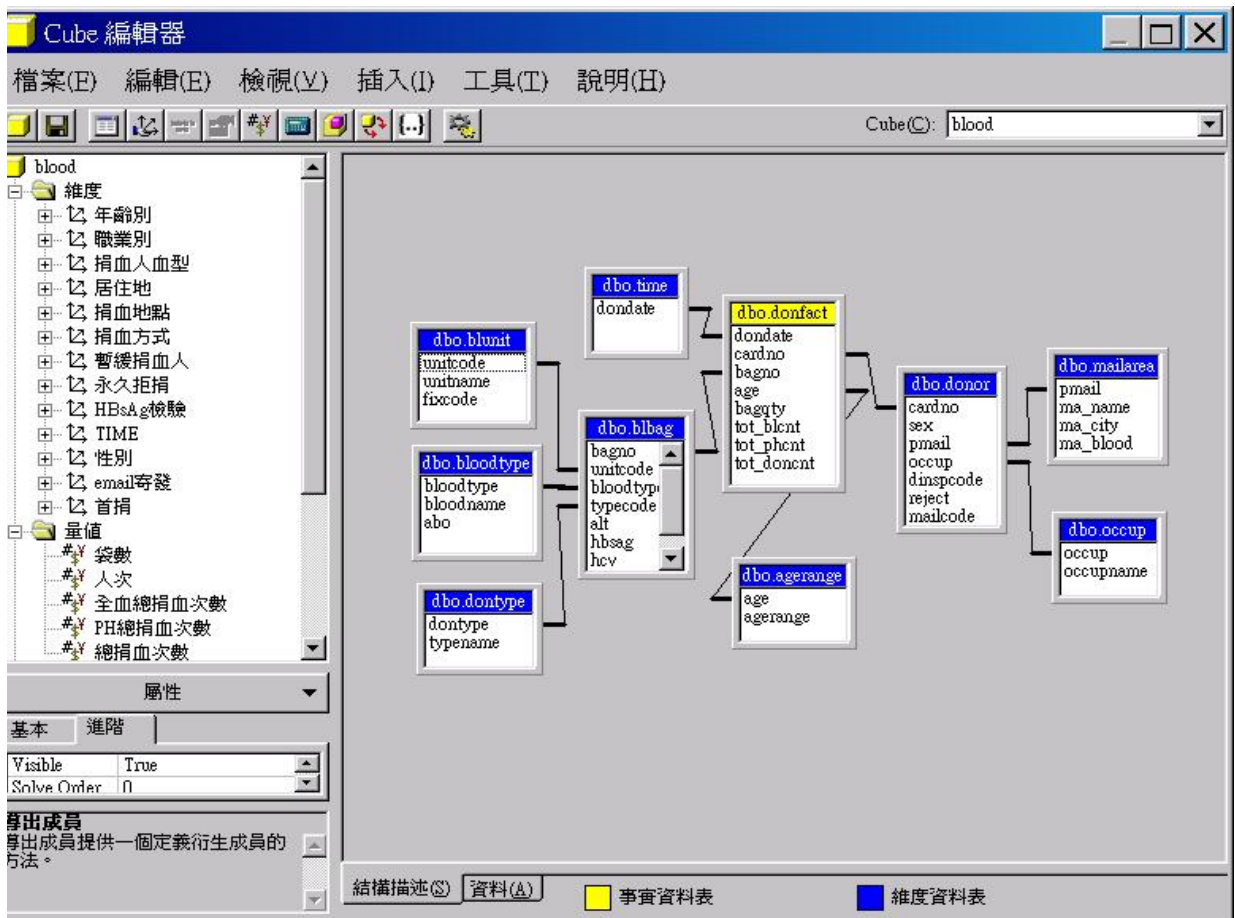


圖 6 捐血資料維度資料表結構描述

資料來源：本研究整理

## 4.2 Cube 模式分析

利用所建立的捐血維度，選擇所需 CUBE 觀察各項資料。



圖 7 捐血方式與使用 email 意願捐血人關係

資料來源：本研究整理

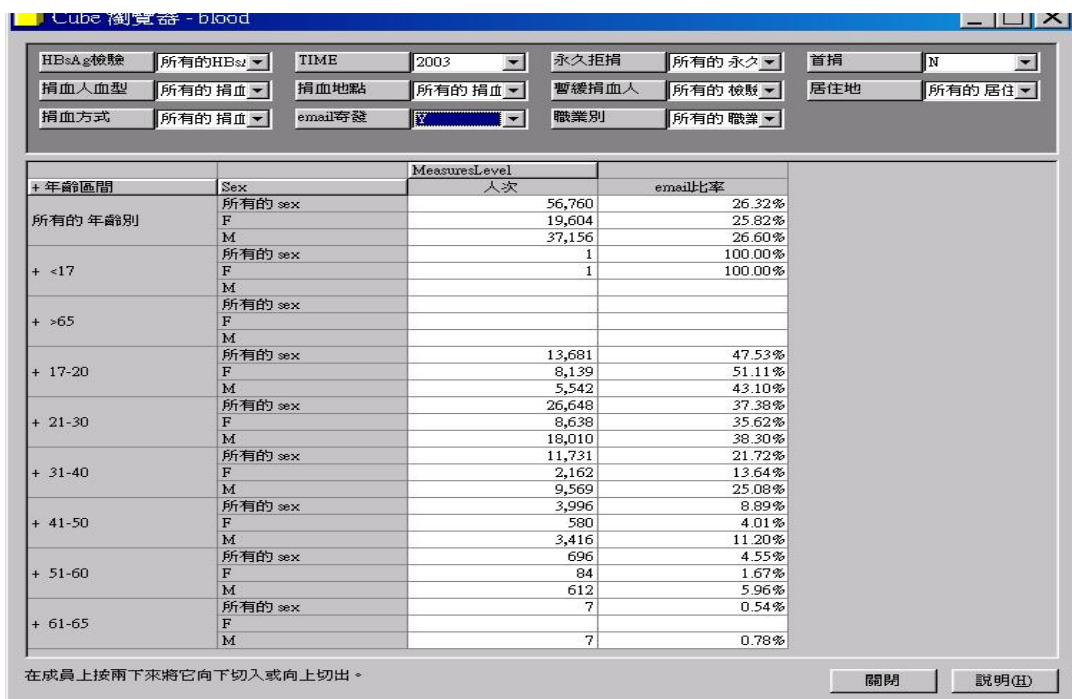


圖 8 年齡、性別與使用 email 意願捐血人關係

資料來源：本研究整理

### 4.3 樞紐分析表建立

利用 SQL 2000 Analysis Service 建立之維度、CUBLE 分析，因在轉換圖形及取得性上未若普遍使用的 EXCLE 方便，故續將前述建置的 CUBE 轉換到 EXCLE 上，以利資訊分享。

#### 4.3.1 捐血人選擇以 email 寄發資料分析

##### 1. 捐血方式與使用 email 意願捐血人關係

分離術捐血人（分離術血小板+減白血小板(1u)+減白血小板(2u)）使用 email 較高：圖 7 可看出，在 2003 年的非首捐捐血人，願意使用 email 收件者 (26.32%)，在捐血方式中以捐「分離術」捐血人 (42.79%，分離術捐血人次／全部分離術捐血人次)，高出一般捐血人(25.9%)。

這應是分離術捐血人多為固定且經常（二週可捐一次）的捐血人，對於檢驗報告的收件，注重時效；且此類捐血人檢驗結果多為”正常“，故對檢驗報告重視比重不若其他捐血人（帶有順道檢驗心理，看到紙本較放心）高。

##### 2. 年齡、性別與使用 email 意願捐血人關係

17-30 歲約有 4 成使用 email 收件：由圖 8、圖 9 可觀察到，在 2003 年的非首捐捐血人，願意使用 email 收件者(26.32%)。在各年齡層及性別中，以 17-20 及 21-30 年齡區間，10 人中就有近一半的人願意使用 email。當然，這應與 17-20 歲以學生為主有關，這可看出 email 對於新一代的重要性與普及性。21-30 歲間則部份為學生或為年輕的上班者，而就 21-30 歲中職業多為“工”及“學生”。

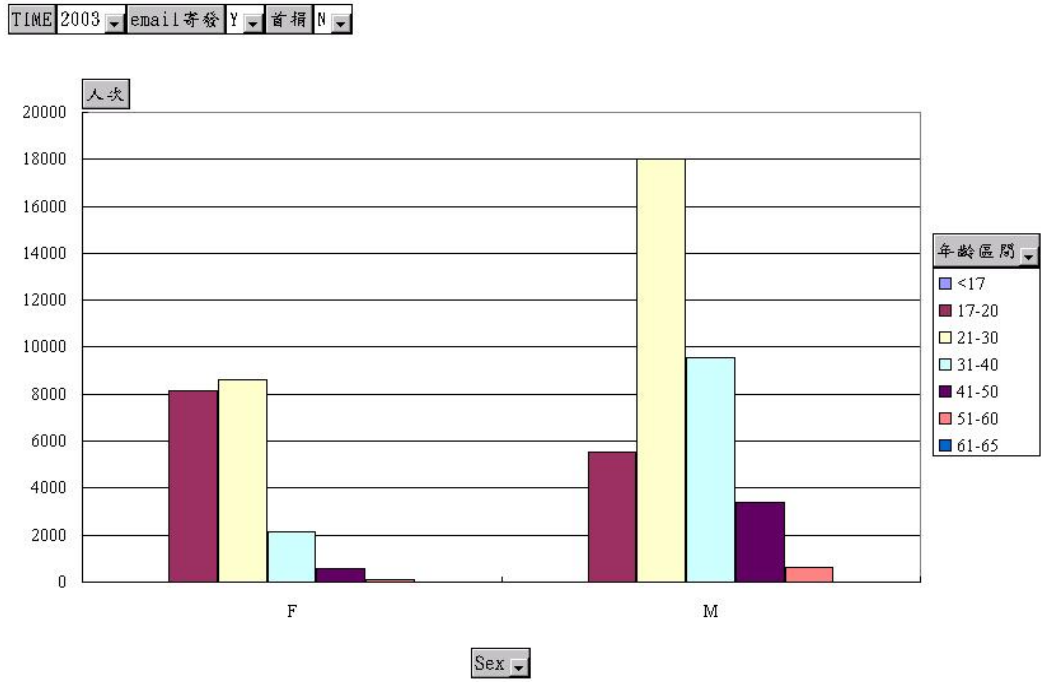


圖 9 年齡、性別與使用 email 意願捐血人次直條圖

資料來源：本研究整理

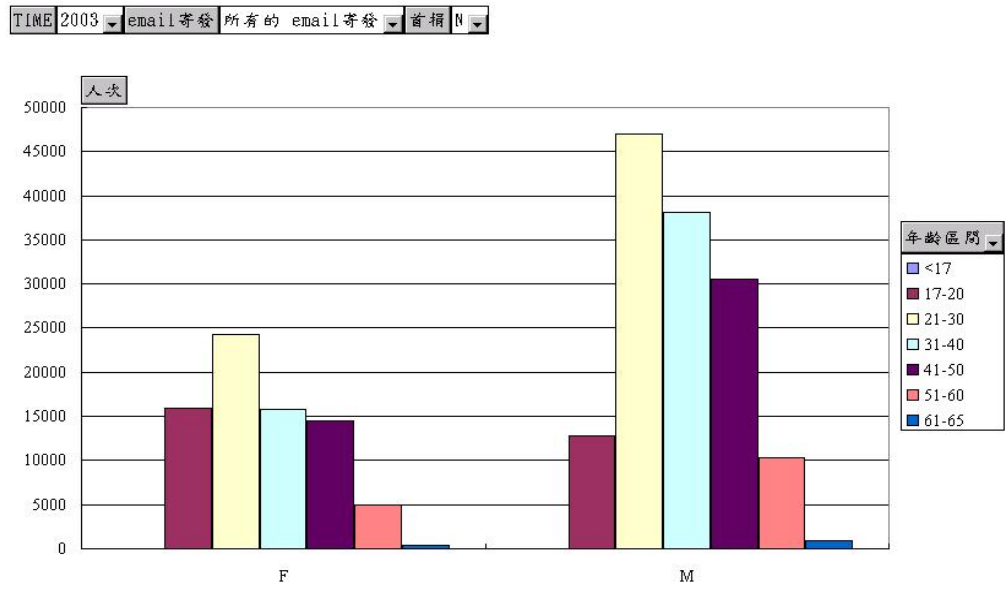


圖 10 年齡、性別全部捐血人次直條圖

資料來源：本研究整理

### 3. 居住地與使用 email 意願捐血人關係

桃園地區使用 email 收件人次最多：由圖 11 可觀察到，在 2003 年的非首捐捐血人，願意使用 email 收件者，在各居住地區中，以居住在新竹捐血中心業務轄區內尤以桃園縣／市人次為高，這當然因捐血量有 43% 來自於居住地在桃園縣、市之故，而新竹縣、市各 13%（圖 12）。

在桃、竹、苗地區，均以“工”及“學生”使用 email 高，桃園地區的“學生”使用人次又居三地之冠，這亦與桃園地區學生捐血人次多有關。若以桃、竹“學生”使用 email 意願，在比率上“學生”與“工”均為對於 email 使用的高使用群（圖 13）。

在居住地與所隸屬中心人次中，有 10% 在新竹中心業務轄區內捐血但居住地為台北地區（台北縣市），4% 屬台中中心（台中、彰化縣市），此類型捐血人較易屬跨區捐血之流動捐血人。可能因工作、就學、出遊，故跨區捐血。這亦可看出新竹中心與台北、台中中心捐血人的流動關係密切重要性。（圖 12）

表 3 居住地(隸屬中心別、桃竹苗地區)與職業別使用 email 願意比率表

email 寄發:Y		首捐 所有的 TIME 2003										
人次	居住地隸屬中心	職業類別										
	居住地城市	工	公教	其他	服務業	軍	家管	商	專技	農漁	學生	總計
	台中中心	4%	3%	4%	2%	7%	1%	2%	3%	2%	10%	6%
	台北中心	9%	10%	12%	13%	28%	10%	14%	13%	4%	17%	14%
	台南中心	2%	1%	2%	1%	5%	0%	1%	1%	1%	5%	3%
	其他	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	花蓮中心	0%	0%	1%	1%	2%	1%	0%	0%	0%	2%	1%
	高雄中心	1%	1%	2%	1%	6%	0%	1%	0%	1%	4%	3%
	新竹中心											
	苗栗縣	9%	12%	5%	7%	5%	14%	6%	7%	5%	8%	8%
	桃園縣	36%	45%	41%	49%	33%	51%	48%	39%	64%	32%	37%

	新竹市	22%	16%	20%	16%	9%	12%	16%	21%	6%	12%	16%
	新竹縣	17%	11%	12%	9%	5%	12%	13%	15%	18%	9%	12%
新竹中心 合計		84%	84%	78%	82%	52%	89%	83%	82%	92%	61%	73%
總計		100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

資料來源：本研究整理

表 4 居住地(隸屬中心別、桃竹苗地區)與職業別捐血人比率表

email 寄發	所有的 email 寄發	首捐		所有的		TIME		2003				
人次		職業類別										
居住地隸屬中心	居住地 城市	工	公教	其他	服務業	軍	家管	商	專技	農漁	學生	總計
台中中心		2%	2%	2%	2%	7%	1%	1%	2%	2%	9%	4%
台北中心		6%	7%	8%	10%	27%	5%	9%	11%	6%	15%	10%
台南中心		1%	1%	1%	1%	4%	0%	0%	1%	1%	4%	2%
其他		0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
花蓮中心		0%	0%	1%	0%	2%	0%	0%	0%	0%	2%	1%
高雄中心		1%	1%	1%	1%	4%	0%	0%	1%	0%	3%	1%
新竹中心	苗栗縣	15%	16%	8%	10%	7%	25%	10%	10%	23%	10%	13%
	桃園縣	44%	47%	48%	53%	35%	47%	52%	43%	42%	34%	43%
	新竹市	14%	14%	15%	12%	7%	10%	15%	18%	6%	12%	13%
	新竹縣	17%	12%	15%	11%	7%	12%	12%	14%	19%	11%	13%
新竹中心 合計		90%	89%	87%	86%	55%	94%	89%	84%	90%	66%	82%
總計		100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

資料來源：本研究整理

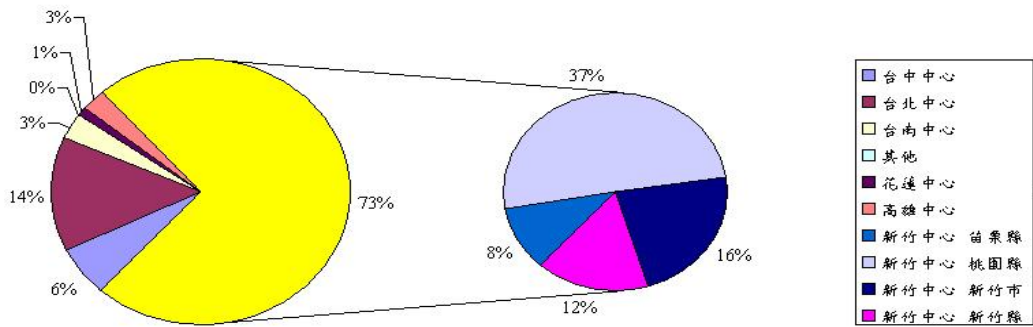


圖 11 居住地與使用 email 願意比率圖

資料來源：本研究整理

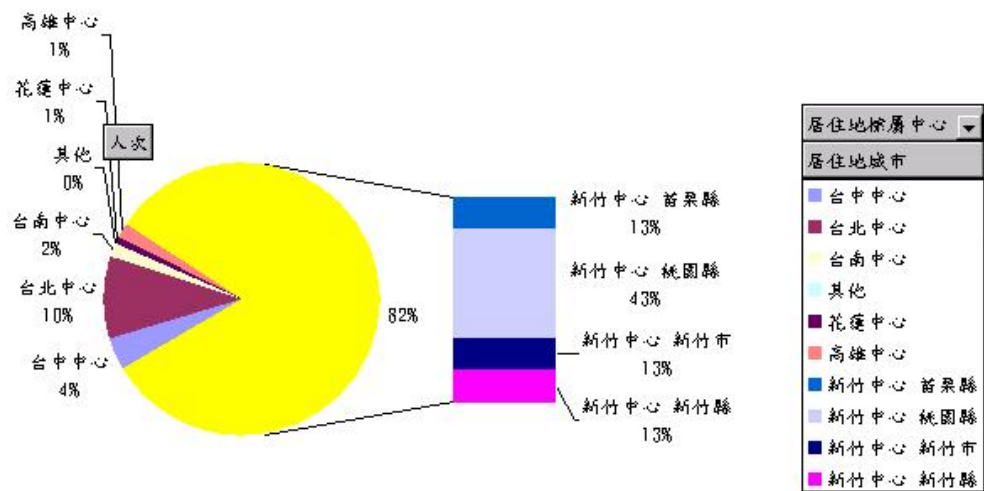


圖 12 居住地與佔全部捐血人次數量圓形堆疊圖

資料來源：本研究整理

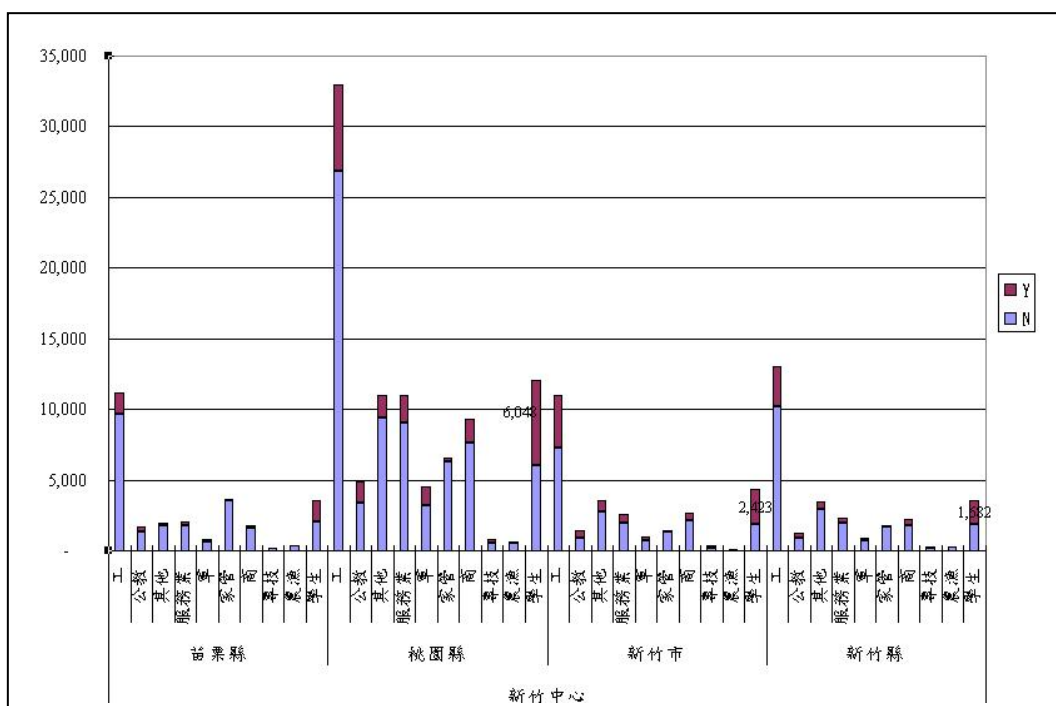


圖 13 居住地 (桃竹苗)、職業別與使用 email 願意直條圖

資料來源：本研究整理

### 【建議】

使用 email 收件，除可快速地得知捐血後檢驗結果外，亦可減少紙張浪費及人力處理成本。但目前以 email 收件者，約有 10% 退件，原因有下：

1. 信箱爆掉或退件：尤以免費帳號 (yahoo、pchome、msn、hotmail 等) 為多。
2. 帳號太多：一個人擁有多個帳號，致自己填了那一個也忘記了。
3. 防火牆阻攔：部份捐血人留的是工作單位帳號，但其因安全考慮設置防火牆或安全規則，致信件被阻攔。
4. 軟體誤判：部份 email 因防毒軟體誤判而退回。
5. 人為輸入錯誤：輸入人員可能因字潦草而誤判，亦可能為疏忽造成輸入錯誤。

因上述原因，致 email 收件的方式產生部份瑕疵。其實捐血人除以 email 及傳統紙本寄發方式得知檢驗結果外，亦可主動由網站上註冊後立即得知“捐血(複檢)日期”、“最近捐血(複檢)中心”、“現有累計捐血次數”、“年度捐血量”、“年度捐分離術次數”、“檢驗結果”、“下次捐血(複檢)日期”等資料[1]，又較 email 快速。故在此建議，可善用此工具，於捐血人填寫資料時增加選擇可自行由網頁自動查詢，而不須透過 email 或紙本收得之選項，並廣為宣導，相信這對於經常捐血講求時效的分離術捐血人，及固定且經常的其他捐血方式捐血人，應具有高度選擇意願。另訓練捐血談時宣導捐血人填寫及使用 email 意願，亦可使仍習慣看到結果的捐血人仍可利用 email 見到檢驗報告結果，亦可降低捐血中心的處理成本。對於人為疏忽輸入錯誤，則可藉資料分析錯誤特性而加以更正。字潦草則訓練資料過濾時的更正作業。

#### 4.3.2 捐血人捐血習性分析 (時間、職業、年齡區間、性別、捐血方式)

##### 1. 時間－職業別關係

由圖 14：2003/01~2004/06 的捐血人職業別 (不包括複檢血)，在各月的捐血人次分佈上可看出：



以“學生”與“工”捐血比例為高，且兩者數量互補：

在學校開學期間（3~6、9~12月），因學校的鼓勵及協助，“學生”捐血意願提高，為血源重要提供者；但1、2月（寒假）及7、8月（暑假）期間，學生因假期或畢業，血量則由“工”補入；甚至“軍”比例亦提高。因在此期間，捐血中心會因“學生”捐血量的下降，隨之調整血液來源。積極連繫公司、工廠、軍隊協助捐血，故“工”、“軍”比例增加，後隨學校開學，“學生”比例又隨之上升。

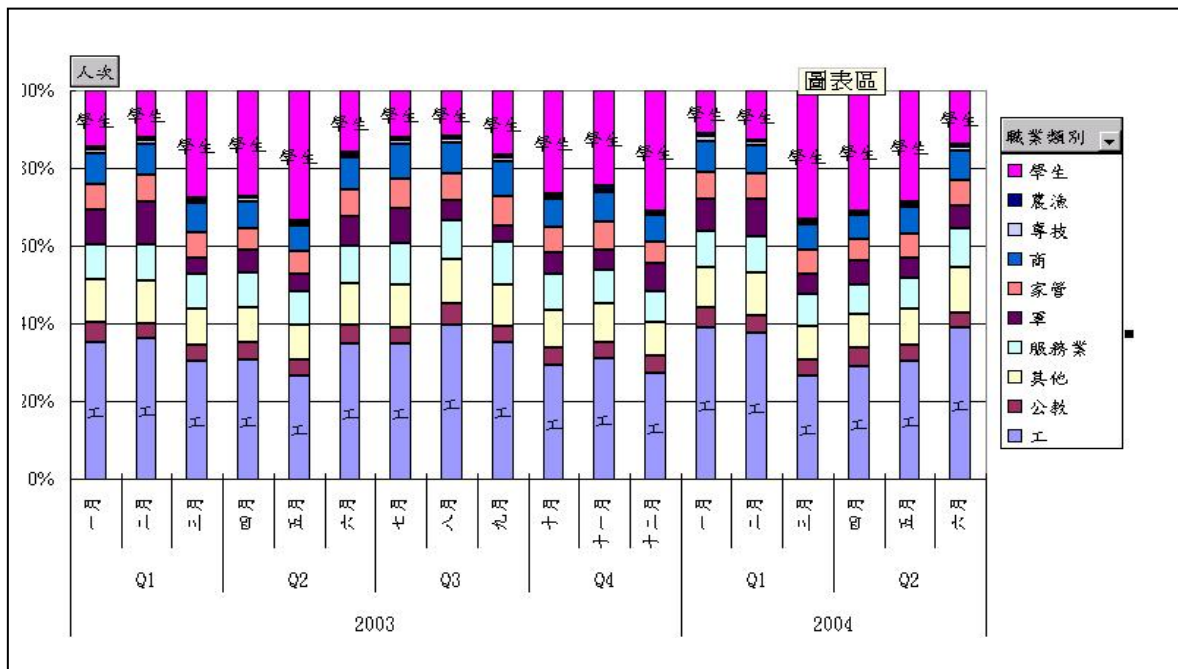


圖 14 時間－職業別堆疊直條

資料來源：本研究整理

## 2.時間－年齡區間關係

由圖 15：2003/01~2004/06 的捐血人各年齡區間（不包括複檢血），在各月的捐血人次分佈上，與時間－職業別類似，此為因 17-20 歲區間，多為“學生”（81%）（圖 16），及 21-30 歲區間佔 20%為學生之故。

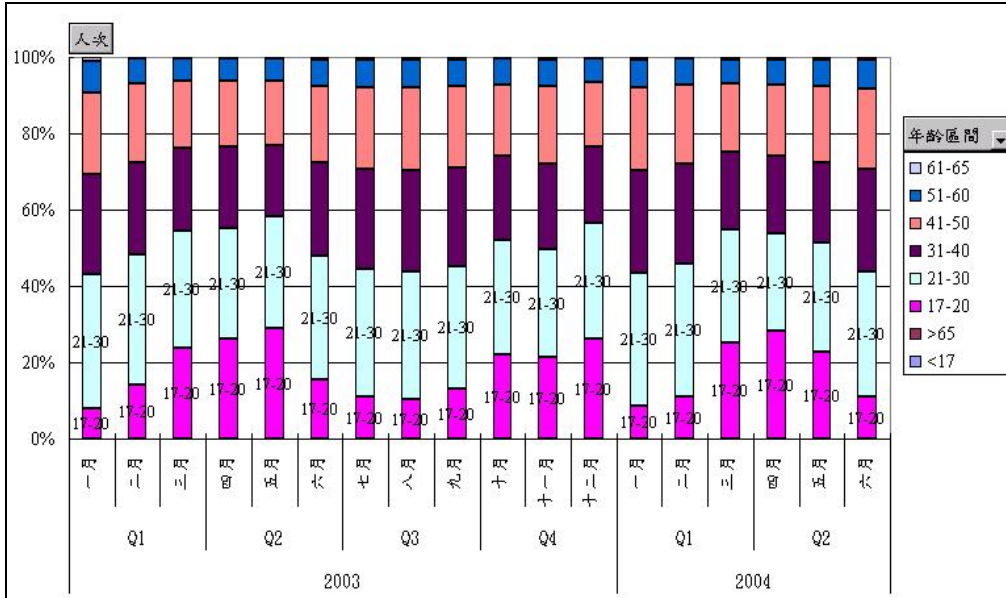


圖 15 時間—年齡區間堆疊直條圖

資料來源：本研究整理

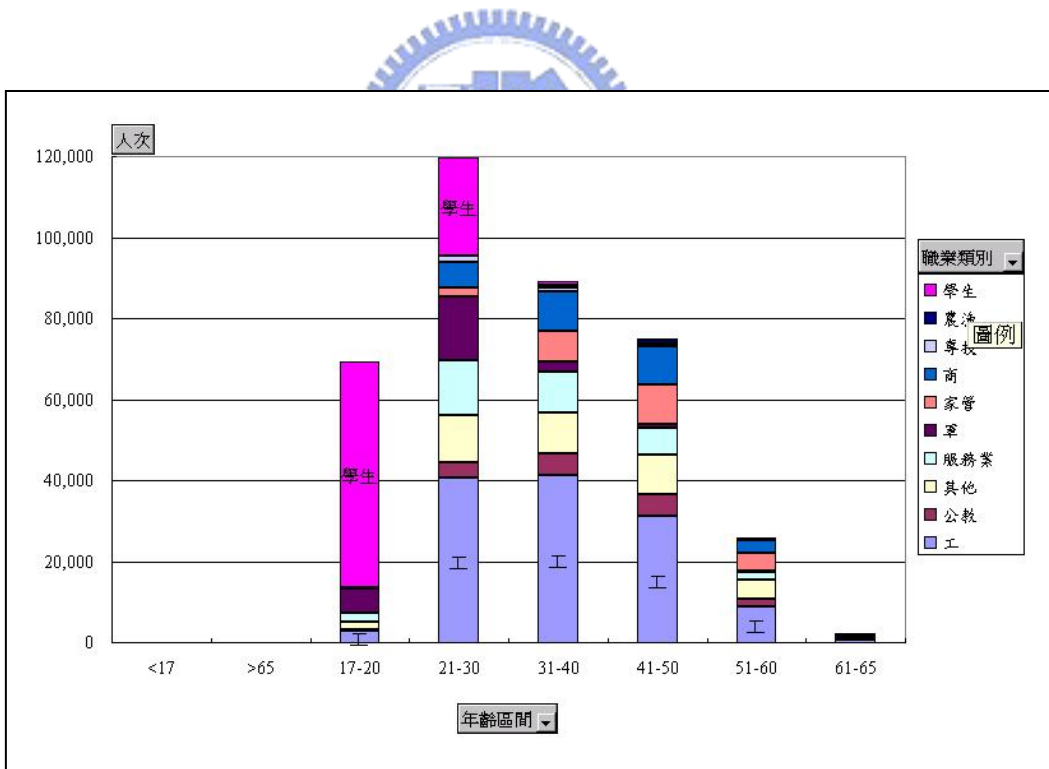


圖 16 年齡區間—職業別比例直條圖

資料來源：本研究整理

### 3.性別、職業－捐血方式關係

由圖 17：2003 年的捐血人中，在各項職業類別中，不同性別所捐的捐血方式亦有差異性。

500cc 及分離術捐血以男性為主，在男性捐血 250 cc、500 cc、分離術人次中，均以“工”為多(16.57%、42.67%、38%)；女性捐血 250 cc 以“學生”佔(16.43%)為最多，500 cc 以“家管”(0.91%)、分離術則以“其他”(1.26%)最高。

表 5 捐血人性別、職業別與捐血方式數量表

TIME	2003					單位：人次
人次		Typecode				
職業類別	Sex	250cc	500cc	5cc 複檢	分離術	總計
工	F	10.93%	0.86%	3.67%	0.76%	7.70%
	M	<b>16.57%</b>	<b>42.67%</b>	24.08%	<b>38.00%</b>	24.64%
公教	F	1.97%	0.10%	0.56%	0.09%	1.37%
	M	2.15%	4.70%	2.62%	8.77%	2.93%
其他	F	5.00%	0.47%	6.98%	<b>1.26%</b>	3.69%
	M	4.09%	9.50%	26.23%	11.73%	6.31%
服務業	F	6.44%	0.52%	1.98%	0.52%	4.53%
	M	3.25%	8.69%	4.39%	5.26%	4.92%
軍	F	0.28%	0.01%	0.03%	0.00%	0.19%
	M	5.45%	8.65%	5.33%	6.65%	6.42%
家管	F	9.47%	<b>0.91%</b>	3.66%	0.33%	6.73%
	M	0.08%	0.14%	0.05%	1.29%	0.10%
商	F	2.98%	0.20%	1.33%	0.32%	2.10%
	M	3.44%	9.70%	6.68%	14.22%	5.42%
專技	F	0.42%	0.03%	0.06%	0.05%	0.29%
	M	0.29%	0.72%	0.36%	3.47%	0.43%

農漁	F	0.10%	0.01%	0.08%	0.43%	0.07%
	M	0.32%	0.88%	0.86%	2.30%	0.50%
學生	F	<b>16.43%</b>	0.51%	3.18%	0.42%	11.27%
	M	10.33%	10.71%	7.84%	4.12%	10.38%
總計		100.00%	100.00%	100.00%	100.00%	100.00%

資料來源：本研究整理

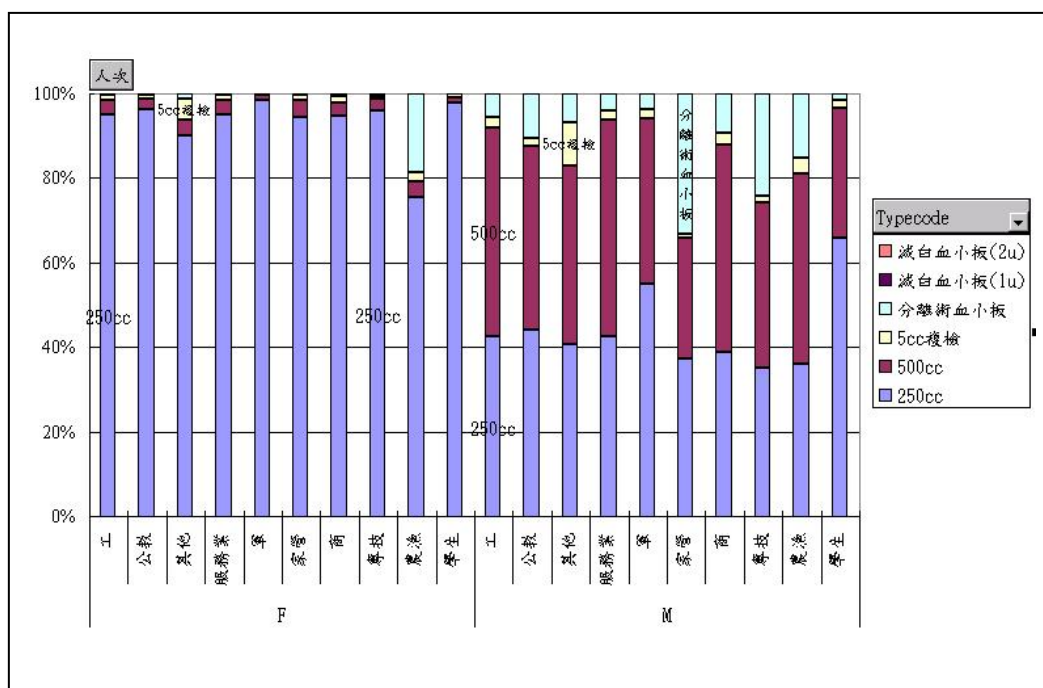


圖 17 性別、職業－捐血方式關係堆疊直條圖

資料來源：本研究整理

### 【建議】

血量來源佔多數的”學生”，因受限於寒暑假因素，致造成血源不穩定的原因之一。而學生的捐血動機多來自於學校的鼓勵及提供的激勵因素影響。一旦放長假，無學校的時間安排及捐血地點亦不若在學校安排的捐血車方便。故在假期間，建議在固定捐血地點多宣傳及企劃辦理吸引年輕人的捐血活動。一則宣傳捐血地點，二則以年輕、活潑化活動吸引年輕的學子回到家後亦能持續捐血活動。

除在假期間宣傳捐血地點及辦理活動，使回到家的學生持續捐血外，學生多為首

次捐血，故可在長假期間，藉捐血紀念品樣式的設計（例：系列收藏及活潑化），亦可吸引年輕的捐血人回捐。除捐血紀念的設計外，對於捐血地點亦可以配合加以設計以凸顯獨特性，即除以原社會公益、發揮愛心之訴求點，納入行銷方式手法連繫住舊捐血人，更引入新捐血人。

#### 4.3.3 分離術捐血人特性（職業、年齡區間）

因捐血人對於不同的捐血方式是有限制的。捐血 250cc 體重只要 45 公斤以上，捐血 500cc 則須 50 公斤以上；分離術捐血資格限制更為嚴謹（1.半年內曾捐血且各項檢查結果正常 2.體重 50 公斤以上 3.手臂上血管一向扎針順利 4.捐血時間約二小時）[2]，故一般以男性為多。女性則以 250 cc 為主。

分離式捐血人次之年齡區間分佈以 31-40 歲居首（表 6）：在 2003 年、非首捐的捐血人中，捐分離術者在年齡區間以 31-40、41-50 歲間為多。這是因對剛接觸捐血的 17-20 歲新鮮人，一般會先捐 250cc，隨著次數增多，在資格條件允許下（檢驗正常、體重足夠、體能良好、上次捐血無不良經驗、捐血時間須二小時及心理建設亦達到），逐漸會嚐試 500cc，再到分離術捐血。再者，體能上的考量，故 51-65 歲即逐漸減少。

分離術捐血人次之職業別以“工”為冠：由圖 18，在 2003 年、非首捐的捐血人中，捐分離術者在 21-30、31-40、41-50、51-60 年齡區間均以“工”為首，另在 31-40、41-50 年齡區間的“商”亦佔職業別中之次。

#### 【建議】

分離術捐血因時間長且須在特定地點，若欲廣泛推廣，則建議將捐血 250 cc→500 cc→分離術捐血的程序縮短，使 21-30 歲間更快速地接受分離術捐血。

因捐血過程較久，故於此段時間對於捐血人或陪同捐血的親友關心照顧及書報或節目的提供更應用心，避免捐血過程煩悶無聊而影響下次續捐的意願。另可設計特定分離術捐血樣式紀念品以為吸引。

表 6 分離術捐血人性別、職業別與捐血方式數量表

首捐	N						
TIME	2003						
職業類別	17-20	21-30	31-40	41-50	51-60	61-65	總計
工	0.12%	7.53%	14.52%	12.78%	3.78%	0.03%	38.76%
公教	0.05%	1.36%	2.78%	3.34%	1.31%	0.01%	8.86%
其他	0.05%	2.58%	4.12%	3.06%	2.08%	1.09%	12.99%
服務業	0.04%	1.26%	2.55%	0.79%	1.13%	0.01%	5.78%
軍	0.08%	2.04%	2.99%	1.08%	0.46%	0.00%	6.65%
家管	0.01%	0.12%	0.86%	0.58%	0.05%	0.00%	1.63%
商	0.11%	2.57%	5.34%	5.36%	1.05%	0.12%	14.54%
專技	0.00%	0.87%	1.23%	1.40%	0.02%	0.00%	3.53%
農漁	0.00%	0.02%	1.65%	1.01%	0.04%	0.00%	2.73%
學生	0.23%	2.57%	1.01%	0.68%	0.05%	0.00%	4.54%
總計	0.69%	20.92%	37.06%	30.10%	9.98%	1.26%	100.00%

資料來源：本研究整理

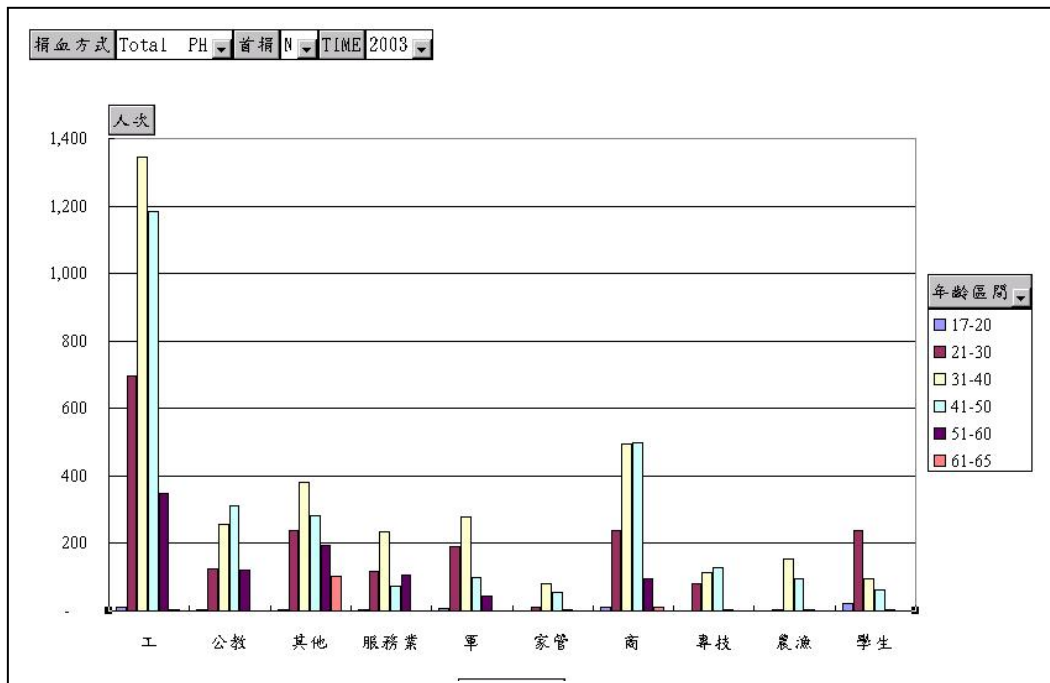


圖 18 分離術捐血人年齡區間、職業數量直條圖

資料來源：本研究整理

#### 4.3.4 不適捐血人與永久拒捐者分析

非首捐因檢驗不合格列為“永久拒捐”者，女性各年齡層的職業別“其他”最多。男性除 17-20 歲區間以“學生”居多外，其他年齡層的各職業別中亦以“其他”最多（圖 19）。

首捐因檢驗不合格列為“不適捐血”者，男、女性均以 17-20 歲間的“學生”為最多，但在 41-50 區間女性的“家管”卻高佔女性列“D”值之次高，男性則以 17-20 歲(10%)、21-30 歲(10%)“軍”比例均相對較捐入量(2%、6%)高（圖 20）。

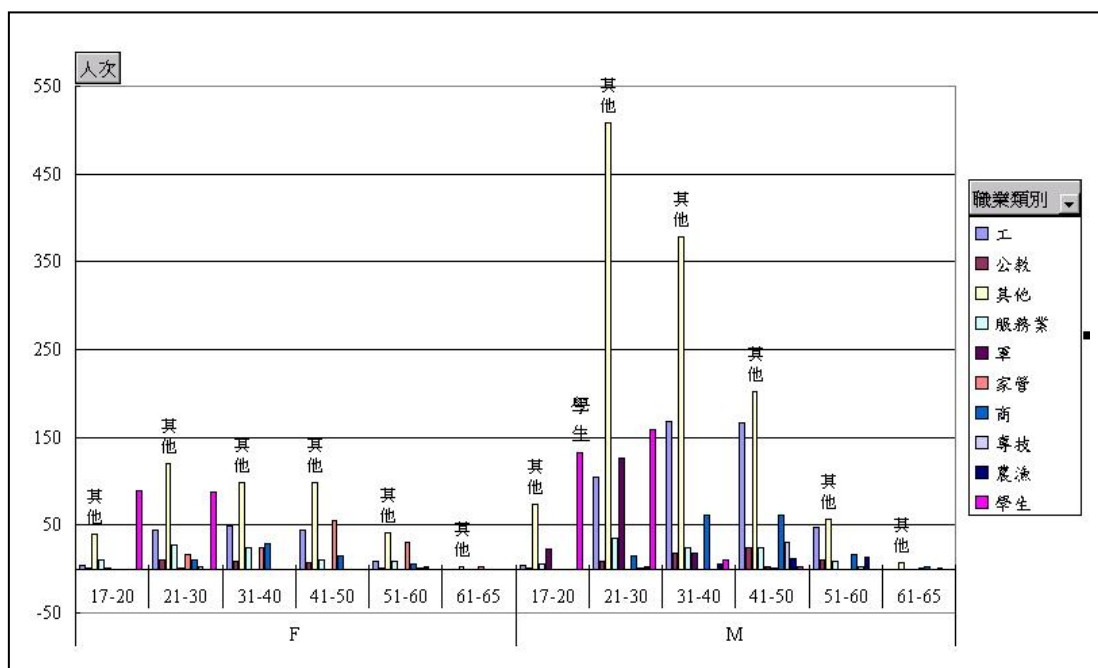


圖 19 非首捐檢驗不合格列“永久拒捐”之性別、年齡區間與職業捐血人次直條圖

資料來源：本研究整理

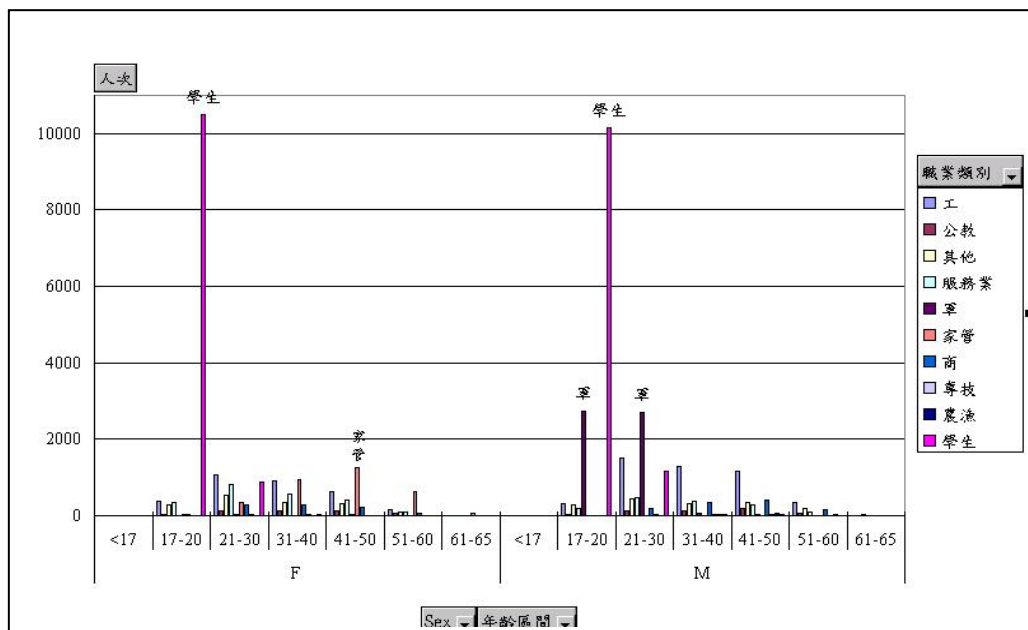


圖 20 首捐檢驗不合格列“暫緩捐血”之性別、年齡區間與職業捐血人次直條圖

資料來源：本研究整理

### 【建議】

對於首捐因檢驗不合格而列為“不適捐血”，女性“學生”、男性“軍”，建議可請學校及部隊利用組織資源，於捐血活動辦理前多加以宣導，對於曾有過相關檢驗不合格或高危險群不可捐血，亦不可明知不合格又希望藉捐血檢驗血液。

對於非首捐者，因各捐血地點均配備有連線設備或筆記型電腦，可查詢前次捐血的檢驗結果，可有效過濾不適捐血人，惟須注意離線電腦的更新時效。對於職業“其他”可利用捐血前面談多加以問診。

## 4.4 建立決策樹採擷模式

利用 Microsoft SQL Server Analysis Server 建立決策樹模型。

### 4.4.1 捐血方式與性別、首捐、年齡決策樹

在此以 93 年 1~6 月（不包含複檢血）為資料範圍，“血袋號碼”為案例索引鍵資料行、“捐血方式”為預測資料行、“性別”、“首捐否”、“年齡”、“捐血方式”為輸入資料行結果。

由圖 21 深色部份可看出：若為“男性”，“非首捐”，年齡在“>26 and <=34”則捐“分離術捐血”者會佔 7.85%；年齡在“>34 and <=43”會佔 8.9%；年



齡在”>43”則會佔 8.83%，均高於 92 年分離術捐血比率 3.4% [9]。在圖下方，可看出女性顯然不是分離術捐血的主要來源，即使是“女性”、“非首捐”，也只有佔 0.49%，95.92%會捐 250cc 一般捐血。

故可預測，當“女性”且“非首捐”之捐血人欲捐血時，可預知其捐分離術的可能極低，僅有 0.49%；有 99.51%的可能會做一般捐血，其中又以捐 250cc 可能為高。

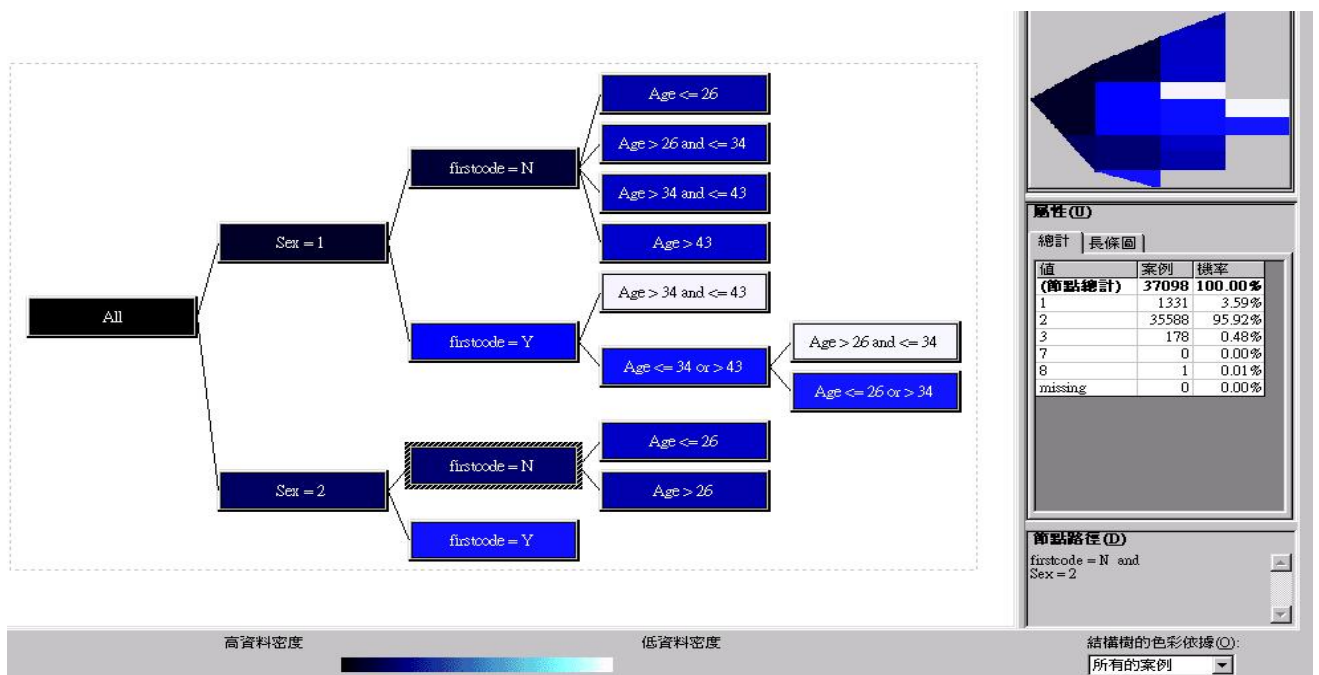


圖 21 捐血方式與性別、首捐、年齡決策樹

資料來源：本研究整理

#### 4.4.2 C 肝炎表面抗原 (Anti-HCV) 與首捐、年齡決策樹

以 93 年 1~6 月 (不包含複檢血) 為資料範圍，”血袋號碼”為案例索引鍵資料行、” Anti-HCV”為預測資料行、”首捐否”、”年齡 “、” Anti-HCV”為輸入資料行結果。

圖 22 深色部份為“首捐”、“年齡>43”為 Anti-HCV”陽性(+)的高屬性機率 (3.64%)，較平均 0.27% 為高。而在“非首捐”部份，陽性機率為 0.12%，低於平均比率。

因在“非首捐”部份，已藉捐血前的電腦篩選而過濾曾有過 Anti-HCV”

十”而未經複檢程序檢驗合格者，故可看出捐血前的電腦查詢作業對於血液捐入品質的重要性。對於高屬性機率者，則因首捐者無歷史資料可供查詢，只能儘可能由捐血前衛教觀念宣導及面談以降低陽性率。

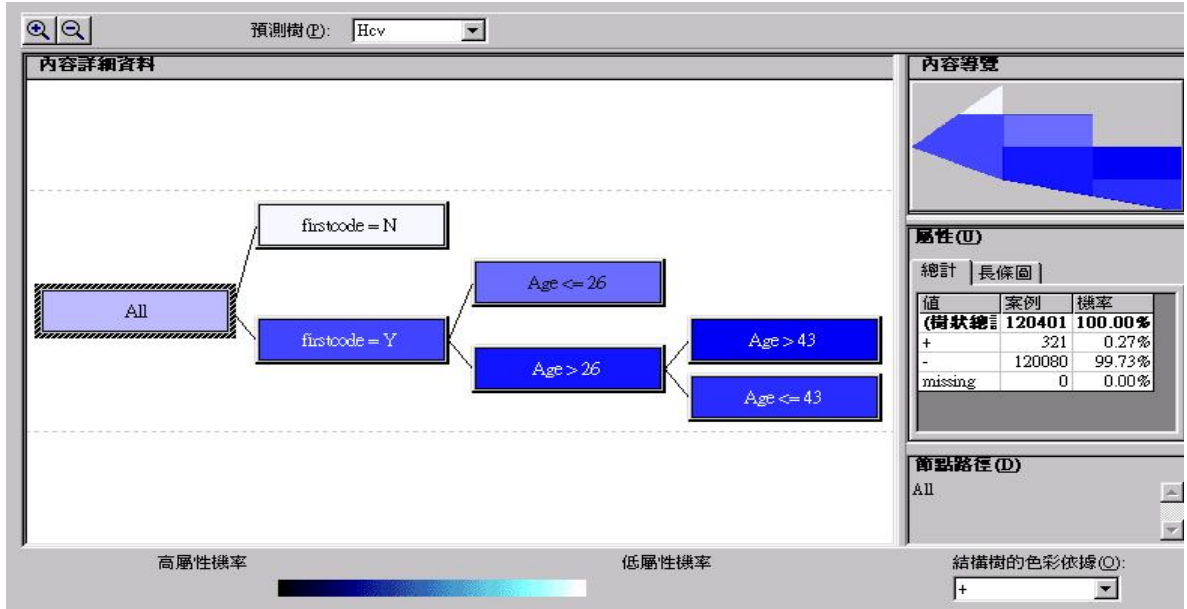


圖 22 Anti-HCV 與首捐、年齡決策樹

資料來源：本研究整理

#### 4.4.3 捐血方式與性別、首捐、年齡變數網路依存性

若就依存性網路來觀察變數之間的關聯狀態 (圖 23)，則可看出“捐血方式”與“性別”、“首捐否”有著很強的關聯性。

一般女性受限於體重原因，即使是非首捐，捐血方式多以 250cc 為主。而首捐者一般並不鼓勵捐 500cc 或分離術。故 500 cc 與分離術捐血以男性為主。

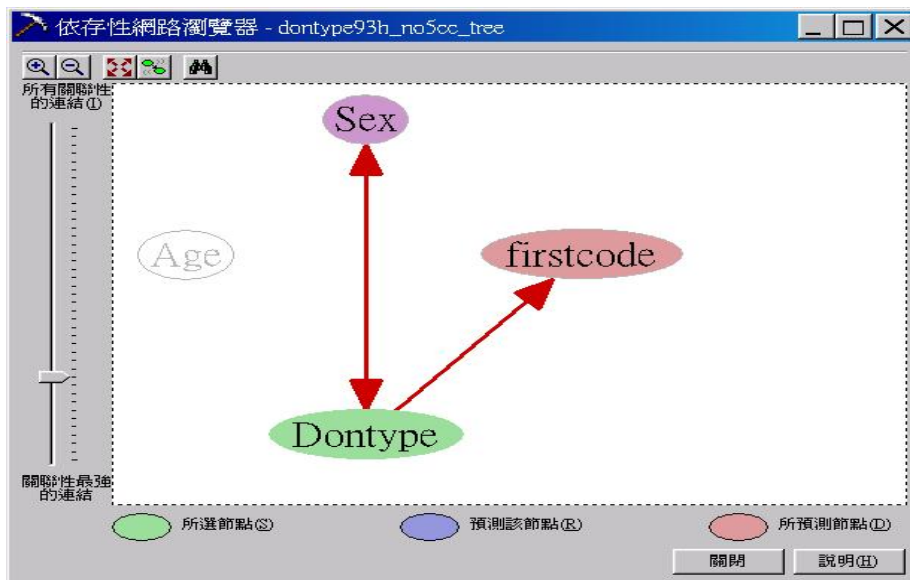


圖 23 捐血方式與性別、首捐、年齡變數關聯性

資料來源：本研究整理

#### 4.5 建立叢集採擷模式

同樣地，續利用 Microsoft SQL Server Analysis Server 建立叢集模型。

在此以 93 年 1~6 月（不包含複檢血）為資料範圍，”血袋號碼”為案例索引鍵資料行、”性別”、”首捐否”、”職業別 “、”捐血方式”為輸入資料行，共分為 8 個 Cluster 結果（圖 24）。

以”捐血方式”觀察，在所有的捐血方式中，250cc 佔全部的 65.57%；在各個 Cluster 中可觀察捐血方式（250 cc、500 cc、分離術捐血、減白血小板）、性別、職業別、首捐否的叢集特徵。

在 Cluster 1（捐血方式為 500cc、分離術血小板、減白血小板 2u）中，特徵為”男性”、”非首捐”、職業為”工”成一叢集，具該特徵捐血人其捐血方式 500cc 之捐血人次佔所有捐 500cc（53%），分離術及減白血小板則亦佔全部分離術及減白血小板 44%，此叢集所捐血液有 91.08% 為 500cc、8.87% 為分

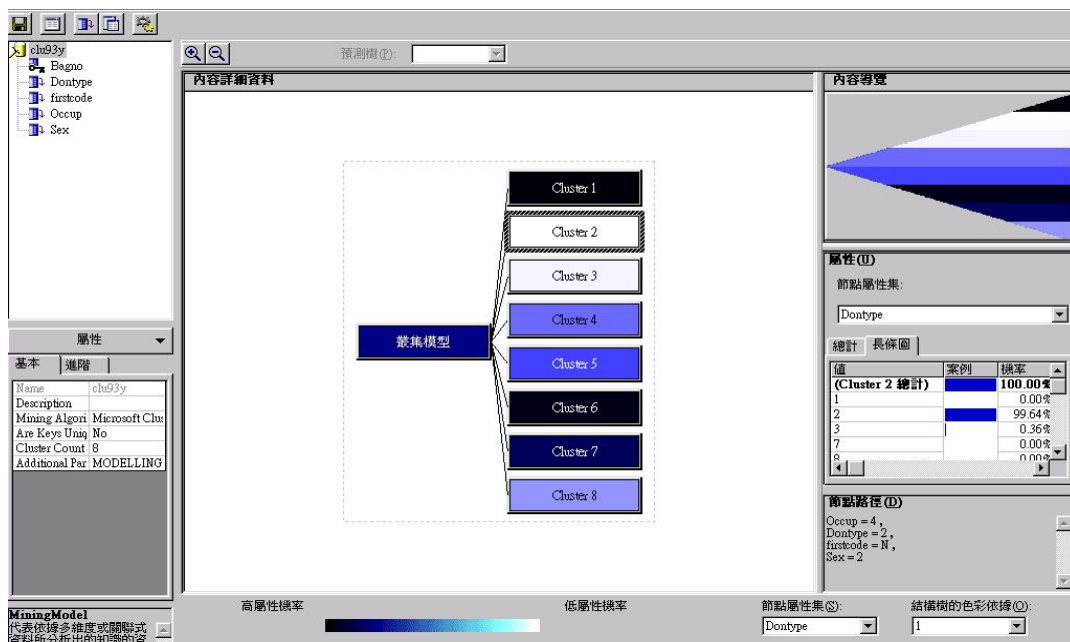


圖 24 捐血方式與性別、首捐否、職業別、性別叢集模型

資料來源：本研究整理

離術血小板，其屬捐血方式”500cc”及”分離術血小板”的高屬性 cluster。符合捐500cc 及分離術捐血人之特性，即習慣性及男性、職業特性。

Cluster2（捐血方式為 250cc、分離術血小板）中，特徵為“女性”、“非首捐”、職業為“工”成一 cluster，其屬捐血方式“250cc”的高屬性 cluster（99.64%為 250cc 捐血）。

經由 clustering 後，可觀察出在同一 cluster 中所具有高度的相似性（捐血方式），不同捐血方式的捐血人特徵提供辨識參考。



## 五、結論與後續研究方向

### 5.1 結論

捐血中心主管全國血液供應，負責用血病患所需血液之採集，而諸多因素造成各地捐血中心血量來源不穩定情形。

血液的工本材料費自 63 年成立至今已過 30 年，卻未隨經濟成長而提高，在各項成本的增加下經營更形艱難。幸有賴政府機構、各地學校、社團及社會普遍的熱心支持，雖艱難但在經費困窘期間，仍從 63 年的 5 項血液檢驗增加到今日的 9 項檢驗，並在 90 年 10 月基金會及各捐血中心全面通 ISO9001:2000 的驗證，用心地持續追求血液及服務品質的提升。

在 63 年草創之期，捐血人數有 3,817 人，至目前全國每年捐血人數達到近 1,700,000。大量的捐血人及血液作業資料，均須依賴各捐血中心的電腦化處理。目前電腦系統仍為 OLTP(OnLine Transaction Processing)處理方式，偏重於資料輸入保存，及制式報表的產出。

大量儲存於資料庫的資料若能妥為運用，由過去的歷史資料中挖掘出對明日事件的預測及因應，才不負眾多的熱心捐血人捐血資料。更能對目標對象主動地採取適合的服務。

在本研究中，利用軟體建置資料倉儲來分析以 email 寄發捐血人特徵為分離術捐血人、學生、工及 30 歲以下者。捐血人習性分析中，學生與“工”、“軍”血量的互相調節，及不同性別、職業、年齡的捐血人不同捐血方式。暫緩捐血人與永久拒捐者特性，另利用探勘工具決策樹預測不同捐血方式的捐血人及叢集分群，這對日後推展捐血事業有著參考意義及幫助。

### 5.2 後續研究方向

在本研究中，針對捐血人捐入部份的資料做分析，建置的資料倉儲及各式維度 CUBE 設計依資訊程度別提供予各部門主管，部份資訊則可置於內部網頁予中心工作人員或外部網頁供捐血人查詢。因各中心資訊系統架構均相同，故亦可提

供予各中心參考。

捐血人捐入資料及血液處理過程，最後供應至醫院，乃至醫院的用血回報，後續的程序均可以本研究方法納入分析、預測。在血液檢驗方面，可進一步分析各項檢驗值與捐血人關係及對列管的名單做分析；血品成分分離部份，則可對各成品及瑕疵、報廢、攔截血品做研究；血品供應到醫院方面，可研究各類血品與個人領血及醫院團體領血、逾期報廢情形及預測醫院用量。最後，醫院用血後的使用回報，有著用血人的各項用血資料，亦可以此進行用血人分析及預測。



## 參考文獻

- [1] 台灣血液基金會，<http://www.blood.org.tw>，「捐血答客問／分離術篇」。
- [2] 中華血液基金會，「捐血程序篇」，捐血問與答，5~7 頁，民國 85 年 10 月。
- [3] 中華血液基金會，「SARS 對捐血中心之影響」，九十二年年報，90 頁，2004 年 3 月。
- [4] 中華血液基金，「簡史」、「目的與任務」、「提昇血液品質，保障輸血安全」，台灣血液事業之發展，5-7,14 頁，民國 89 年 9 月。
- [5] 台灣血液基金會，「統計表」，九十三年年報，61-62 頁，2005 年 3 月。
- [6] 『飛躍 30—台灣捐血事業紀念特刊』，中華捐血運動協會、中華血液基金會，93 年。
- [7] 中央健康保險局，<http://www.realsun.com.tw/nhi.gov/inside4-1.93.94.asp>。
- [8] 中原大學林震岩教授，<http://www.rock-lin.idv.tw/>
- [9] 尹相志，SQL Server 2000 Analysis Service 資料採礦服務，維科圖書公司，台北，2003 年 7 月。
- [10] 麥可·裴瑞 (Michael J.A. Berry)，戈登·林諾夫 (Gordon S. Linoff) 著，資料採礦理論與實務—顧客關係管理的技巧與科學，吳旭智、賴淑貞譯，維科圖書公司，台北，2001 年 6 月。
- [11] 沈兆陽，資料倉儲與 Analysis Services SQL Server 2000 OLAP 解決方案，文魁資訊公司，台北，2001 年 3 月。
- [12] 林宏諭，SQL Server 2000 之決策分析—OLAP 建置與應用，博碩文化公司，台北，2001 年 2 月。
- [13] 林傑斌、劉明德，資料採掘與 OLAP 理論與實務，文魁出版社，台北，2002 年 5 月。
- [14] 施威銘研究室，Microsoft SQL Server 2000 設計實務，旗標公司，台北，民國 89 年 12 月。
- [15] 施威銘研究室，Microsoft SQL Server 2000 管理實務，旗標公司，台北，民國 89

年 11 月。

- [16] 陳俊源，SQL Server 2000 建置與管理手冊，知城數位科技公司，台北，2001 年 2 月。
- [17] 麥可·裴瑞 (Michael J.A. Berry)，戈登·林諾夫 (Gordon S. Linoff) 著，資料採礦顧客關係管理暨電子行銷之應用，彭文正譯，數博網資訊公司，台北，2001 年 1 月。
- [18] 新竹捐血中心，92 年捐血方式月報表，93 年 1 月。
- [19] Alex Berson, Stephen Smith, Kurt Thearling 著，CRM Data Mining 應用系統建置，葉涼川譯，麥格羅·希爾 (Mc Graw Hill) 出版社，台北，2001 年 1 月。
- [20] ISBT(International Society of Blood Transfusion), <http://www.isbt-web.org/isbt.html>
- [21] aabb(Advancing Transfusion and Cellular Therapies Worldwide), <http://www.aabb.org>
- [22] Friedman, J. Data Mining and statistics : What is the connection ? , The 29<sup>th</sup> Symposium on the Interface , Houston , TX , 1997 .
- [23] Berry, M.J. and Linoff, G. Data Mining Techniques: for Marketing, Sales and Customer Support , John Wiley & Sons, Inc , 1997 .