行政院國家科學委員會專題研究計畫　成果報告

運用粗集合理論於乳癌病患資料之知識發現

計畫類別：個別型計畫
計畫編號：NSC93-2218-E-029-001-
執行期間：93 年 08 月 01 日至 94 年 07 月 31 日
執行單位：東海大學工業工程與經營資訊學系

計畫主持人：黃欽印

報告類型：精簡報告

處理方式：本計畫可公開查詢

中　華　民　國 94 年 9 月 28 日

運用粗集合理論於乳癌病患資料之知識發現

黃欽印
東海大學 工業工程與經營資訊學系

摘要

本研究使用 1764 筆從 1998 到 2002 發生於美國的華人婦女乳癌資料，並以粗集合理論萃取其中相關的法則。多數結果並未與一般的臨床或常識相左，因此證明粗集合理論確實能協助從醫療資料庫中萃取相關醫療知識。此外，本研究發現對於初期與末期並年輕與年長的華人婦女乳癌患者，粗集合理論發現與一般臨床相左的結果，該結果需進行進一步分析以確定是否有其他因素影響到華人婦女的預後結果，而使得華人婦女的乳癌死亡模式與一般美國大眾不同。

導論

惡性腫瘤於 91 年度繼續蟬連國人十大死因之首（衛生署，2003）。其中就女性而言，癌症死亡前四大原因分別為肺癌、肝癌、結腸直腸癌、以及乳癌。就美國的女性而言，乳癌是最常發生的癌症病例；其病例個數佔總體癌症病例數高達百分之三十二。此外美國婦女有 15% 的癌症死亡病例為乳癌。在女性乳癌的病患資料收集上，美國堪稱是全世界作得最完整的國家。其中 National Cancer Institute (NCI)的 Surveillance, Epidemiology, and End Results (SEER) Program 收集了美國癌症病患的病理與療程的基本資料以供學者與醫師作大幅的醫療資訊分析，作後能對一些療程的改善上作出貢獻。

然而以目前的乳癌資料分析而言，多數的研究在分析的主題上仍不偏醫事專業的思維。就現今的資料探勘技術而言，其實是有能力對於乳癌的資料作更大規模的資料挖掘，從而發現其中過去所未知或不確定的知識。粗集合理論（Rough Sets Theory）是現今在抽取關聯式資料庫的知識中極具功能的一項工具。就粗集合理論而言，少量而穩定的資料仍是它所選中的知識標的，這和醫學的觀點是一致的。換言之，在人命價值無限的概念下，為極少數的病例找出一些解決方案仍有無限的價值；而這正是粗集合在本計劃所要扮演的功能。

此外，我國的醫病關係逐漸惡化之中，由於病人和家屬的醫學知識漸增導致對於醫療的過程與決策的參與權的要求日漸升高。當病人不幸過世時常導致醫師與家屬之間在療程的看法有分歧的現象而引發糾紛。透過大規模的癌症病患的資料分析的結果，不僅可以提供醫師對於醫療處置新的資訊，更可以作為其療程的

說明而避免糾紛的發生。

本研究取自 SEER 的資料庫中自 1998 至 2002 的華人婦女乳癌案例共 1764 筆作為資料探勘的對象。透過粗集合理論確實找出一些原則與現有的臨床醫學的結論可以相呼應。

研究資料

本研究從 SEER 資料庫 1973-2000 的 482052 個案例中萃取出於 1989-2002 之間的 1764 筆華人婦女的個案作為資料探勘的對象。列入分析的資料欄位如表一所示。

表一 列入分析的資料欄位

欄位	說明
SEER registry	註冊地點
Year of birth	出生年
Age at diagnosis	診斷時年齡
Marital status at diagnosis	診斷時婚姻狀態
Primary site	主要部位
Laterality	發生側數
Reason no surgery	未受手術的原因
Radiation	放射線治療
Radiation sequence with surgery	放射線治療與手術的順序
Vital status recode	存活狀態
Histologic T ype ICD-O-3	組織型態
Behavior Code ICD-O-3	組織行為
Age recode <1 year olds	年齡區分
AJCC stage 3rd edition (1988+)	癌症期別
Tumor Marker 1	Estrogen receptor ER
Tumor Marker 2	Progesterone receptor PgR

研究方法

Support 與 confidence

粗集合理論為資料探勘中關聯式法則萃取（association rule abstraction）的一種。

在關聯式法則萃取的二項重要指標分別為 support 與 confidence。Support 指的是被萃取的法則的案例與總資料庫的比例，confidence 則指法則左右二項都成立的案例與只有法則左項成立的案例的比例 (Han and Kamber, 2001)。資料探勘中最容易發現 support 高且 confidence 高的案例，然而因為其容易發現，價值通常也不高。就醫學而言，support 低而 confidence 高的案例也極具價值，因為少數穩定的案例是累積醫學知識的來源。

粗集合簡介

一個知識表示系統包含屬性集合 *A*（在此稱為條件屬性；Conditional attribute）與決策屬性（Decision attribute）集合 *D*，稱之為決策表（Decision table）。表二（a）為一個決策表，其包含三個條件屬性{a1，a2，a3}與一個決策屬性 d。另外，決策表亦可包含多的決策屬性，如表二（b）所示。

所有 *A* 中的 *D*-不可省略屬性的集合，稱為 *D*-core；另外，能夠辨別所有單位集合的最小條件屬性子集合，稱為 *D*-reduct。

表二 決策表（a）一個決策屬性；（b）兩個決策屬

(a)

U	a1	a2	a3	d
x1	2	1	3	1
x2	3	2	1	2
x3	2	1	3	1
x4	2	2	3	2
x5	1	1	4	3
x6	1	1	2	3
x7	3	2	1	2
x8	1	1	4	3
x9	2	1	3	1
x10	3	2	1	2

(b)

U	a1	a2	a3	d1	d2
x1	2	1	3	2	3
x2	3	2	1	3	1
x3	2	1	3	2	3
x4	2	2	3	3	1
x5	1	1	4	1	3
x6	1	1	2	1	3
x7	3	2	1	3	1
x8	1	1	4	1	3
x9	2	1	3	2	3
x10	3	2	1	3	1

決策表大致的分析步驟與資訊系統相同，主要有：

（1）建構決策表的單位集合，

（2）計算決策表中屬性的核與簡化，

（3）計算決策表中屬性值的核與簡化。

以表二（a）為例，若以決策屬性 d 找尋單位集合，可得到{x1，x3，x9}、{x2，x4，x7，x10}、{x5，x6，x8}三組單位集合，如表三所示。為了找尋 *D*-core 與 *D*-reduct，首先建構 *D*-可識別矩陣。根據表三，物件 x1、x3 與 x9 屬於同一類別，無法找出能夠識別這三個物件的條件屬性集合，故不考慮其個別間的關係。當然，{x2，x4，x7，x10}與{x5，x6，x8}這兩個單位集合也是同樣的情況。根據以上說明，建構 *D*-可識別矩陣，如表四所示。

表三 以決策屬性 d 為底的單位集合

U／D	d
{x1, x3, x9}	1
{x2,x4,x7,x10}	2
{x5,x6,x8}	3

表四 *D*-可識別矩陣

	1	2	3	4	5	6	7	8	9	10
1										
2	a1,a2,a3									
3	－	a1,a2,a3								
4	a2	－	a2							
5	a1,a3	a1,a2,a3	a1,a3	a1,a2,a3						
6	a1,a3	a1,a2,a3	a1,a3	a1,a2,a3	－					
7	a1,a2,a3	－	a1,a2,a3	－	a1,a2,a3	a1,a2,a3				
8	a1,a3	a1,a2,a3	a1,a3	a1,a2,a3	－	－	a1,a2,a3			
9	－	a1,a2,a3	－	a2	a1,a3	a1,a3	a1,a2,a3	a1,a3		
10	a1,a2,a3	－	a1,a2,a3	－	a1,a2,a3	a1,a2,a3	－	a1,a2,a3	a1,a2,a3	

根據表十，計算可識別函數 f_A（*D*）為：

f_A（*D*）=（a1+a2+a3）a2（a1+a3）（a1+a3）×
（a1+a2+a3）（a1+a3）（a1+a2+a3）×
（a1+a2+a3）（a1+a2+a3）×（a1+a2+a3）（a1+a2+a3）×
（a1+a2+a3）a2（a1+a3）（a1+a3）×
（a1+a2+a3）（a1+a3）（a1+a2+a3）×
（a1+a2+a3）（a1+a2+a3）×（a1+a2+a3）a2（a1+a3）×
（a1+a2+a3）（a1+a3）（a1+a2+a3）×
（a1+a2+a3）（a1+a3）（a1+a2+a3）×
（a1+a2+a3）（a1+a2+a3）（a1+a3）×（a1+a2+a3）（a1+a2+a3）
=a2（a1+a3）= a1a2+a2a3

由可識別函數 f_A（*D*）得到兩組 *D*-reduct，分別為{a1，a2}與{a2，a3}。由 *D*-可識別矩陣或{a1，a2}與{a2，a3}兩組 *D*-reduct 的交集，得知 *D*-core 為 a2。接著去除決策表中不必要的條件屬性值，首先分別建構以{a1，a2}與{a2，a3}為主的 D-可識別矩陣，以 reduct{a1，a2}為例，其 *D*-可識別矩陣，如表五所示。計算過程如下：

f_1（*D*）=（a1+a2）a2a1a1（a1+a2）a1（a1+a2）= a1a2

f_2（*D*）=（a1+a2）（a1+a2）（a1+a2）（a1+a2）（a1+a2）（a1+a2）= a1+a2

f_3（*D*）=（a1+a2）a2a1a1（a1+a2）a1（a1+a2）= a1a2

f_4（*D*）= a2a2（a1+a2）（a1+a2）（a1+a2）a2 = a2

f_5（*D*）= a1（a1+a2）a1（a1+a2）（a1+a2）a1（a1+a2）= a1

f_6（*D*）= a1（a1+a2）a1（a1+a2）（a1+a2）a1（a1+a2）= a1

f_7（*D*）=（a1+a2）（a1+a2）（a1+a2）（a1+a2）（a1+a2）（a1+a2）= a1+a2

f_8（*D*）=a1（a1+a2）a1（a1+a2）（a1+a2）a1（a1+a2）= a1

f_9（*D*）=（a1+a2）a2a1a1（a1+a2）a1（a1+a2）= a1a2

f_{10}（*D*）=（a1+a2）（a1+a2）（a1+a2）（a1+a2）（a1+a2）（a1+a2）= a1+a2

由以上的計算結果可得到最終的決策表，如表六所示。值得注意的是，當可識別函數計算的結果為 a1+a2 時，僅需考慮 a2 的值，原因在於 *D*-core 為 a2。

表五 以 reduct{a1，a2}為主的 *D*-可識別矩陣

	1	2	3	4	5	6	7	8	9	10
1	—	a1,a2	—	a2	a1	a1	a1,a2	a1	—	a1,a2
2	a1,a2	—	a1,a2	—	a1,a2	a1,a2	—	a1,a2	a1,a2	—
3	—	a1,a2	—	a2	a1	a1	a1,a2	a1	—	a1,a2
4	a2	—	a2	—	a1,a2	a1,a2	—	a1,a2	a2	—
5	a1	a1,a2	a1	a1,a2	—	—	a1,a2	—	a1	a1,a2
6	a1	a1,a2	a1	a1,a2	—	—	a1,a2	—	a1	a1,a2
7	a1,a2	—	a1,a2	—	a1,a2	a1,a2	—	a1,a2	a1,a2	—
8	a1	a1,a2	a1	a1,a2	—	—	a1,a2	—	a1	a1,a2
9	—	a1,a2	—	a2	a1	a1	a1,a2	a1	—	a1,a2
10	a1,a2	—	a1,a2	—	a1,a2	a1,a2	—	a1,a2	a1,a2	—

表六 最終之決策表

U	a1	a2	d
x1	2	1	1
x2	*	2	2
x3	2	1	1
x4	*	2	2
x5	1	*	3
x6	1	*	3
x7	*	2	2
x8	1	*	3
x9	2	1	1
x10	*	2	2

* 表示不用在意其值為何

規則的產生

最終決策表（表六），也可視為決策規則（Rule）的集合，以下列形式來表示：

$$a_{k_i} \Rightarrow d_j$$

其中 a_{k_i} 表示屬性 a_k 的值為 i，d_j 表示決策屬性 d 的值為 j。以表六為例，可得到下列三條規則：

$$a_{1_2} a_{2_1} \Rightarrow d_1$$

$$a_{2_2} \Rightarrow d_2$$

研究結果

本研究所萃取出的法則如下：

(1) 1764 筆華人女性乳癌患者中的 76 筆為台灣出生，其中有以下的法則出現：
ER 陰性 AND PgR 陰性 => 死亡 2 筆
此外，（ER 陰性 AND PgR 陰性）的案例有 10 筆
上述法則的 support = 2/76= 0.0263，confidence = 2/10=0.2。

(2) 泛中國出生（China, no other specified）共 393 筆。其中有以下法則：
ER 陰性 AND PgR 陰性 => 死亡 5 筆
此外，（ER 陰性 AND PgR 陰性）的案例有 51 筆
(support=5/393=0.0127, confidence=5/51=0.098)
又
ER 陰性 AND PgR 陰性 AND AJCC 第一期=> 死亡 2 筆
此外，（ER 陰性 AND PgR 陰性 AND AJCC 第一期）的案例有 22 筆
(support=2/393=0.0051, confidence=2/22=0.091)
ER 陰性 AND PgR 陰性 AND AJCC 第二期=> 死亡 1 筆
此外，（ER 陰性 AND PgR 陰性 AND AJCC 第二期）的案例有 19 筆
(support=1/393=0.0025, confidence=1/19=0.0526)
ER 陰性 AND PgR 陰性 AND AJCC 第三期=> 死亡 1 筆
此外，（ER 陰性 AND PgR 陰性 AND AJCC 第三期）的案例有 5 筆
(support=1/393=0.0025, confidence=1/5=0.2)
ER 陰性 AND PgR 陰性 AND AJCC 第四期=> 死亡 1 筆
此外，（ER 陰性 AND PgR 陰性 AND AJCC 第四期）的案例有 2 筆
(support=1/393=0.0025, confidence=1/2=0.5)
ER 陽性 AND PgR 陽性 => 死亡 10 筆
此外，（ER 陽性 AND PgR 陽性）的案例有 196 筆

(support=10/393=0.0254, confidence=10/196=0.0510)

又

ER 陽性 AND PgR 陽性 AND AJCC 第一期=> 死亡 4 筆

此外，(ER 陽性 AND PgR 陽性 AND AJCC 第一期) 的案例有 92 筆

(support=4/393=0.0102, confidence=4/92=0.0435)

ER 陽性 AND PgR 陽性 AND AJCC 第二期=> 死亡 5 筆

此外，(ER 陽性 AND PgR 陽性 AND AJCC 第二期) 的案例有 79 筆

(support=5/393=0.0127, confidence=5/79=0.0633)

ER 陰性 AND PgR 陽性 => 死亡 3 筆

此外，(ER 陰性 AND PgR 陽性) 的案例有 5 筆

(support=3/393=0.0076, confidence=3/5=0.6000)

又

ER 陰性 AND PgR 陽性 AND AJCC 第一期=> 死亡 1 筆

此外，(ER 陰性 AND PgR 陰性 AND AJCC 第一期) 的案例有 1 筆

(support=1/393=0.0025, confidence=1/1=1.0000)

ER 陰性 AND PgR 陽性 AND AJCC 第三期=> 死亡 1 筆

此外，(ER 陰性 AND PgR 陰性 AND AJCC 第三期) 的案例有 1 筆

(support=1/393=0.0025, confidence=1/1=1.0000)

ER 陰性 AND PgR 陽性 AND AJCC 第四期=> 死亡 1 筆

此外，(ER 陰性 AND PgR 陰性 AND AJCC 第四期) 的案例有 1 筆

(support=1/393=0.0025, confidence=1/1=1.0000)

ER 陽性 AND PgR 陰性 => 死亡 1 筆

此外，(ER 陽性 AND PgR 陰性) 的案例有 35 筆

(support=1/393=0.0025, confidence=1/35=0.0286)

又

ER 陽性 AND PgR 陰性 AND AJCC 第二期=> 死亡 1 筆

此外，(ER 陰性 AND PgR 陰性 AND AJCC 第一期) 的案例有 13 筆

(support=1/393=0.0025, confidence=1/13=0.0769)

(3) 美國出生的華人共 295 筆，其中有 23 人死亡。23 人中除一人未進行腫瘤標記檢驗外，其餘 22 人的腫瘤標記及其結果如下。

ER 陰性 AND PgR 陰性 => 死亡 7 筆

此外，(ER 陰性 AND PgR 陰性) 的案例有 36 筆

(support=7/393=0.0178, confidence=7/36=0.1944)

又

ER 陰性 AND PgR 陰性 AND AJCC 第一期=> 死亡 3 筆

此外，(ER 陰性 AND PgR 陰性 AND AJCC 第一期) 的案例有 18 筆
(support=3/393=0.0076, confidence=3/18=0.1667)
ER 陰性 AND PgR 陰性 AND AJCC 第三期=> 死亡 1 筆
此外，(ER 陰性 AND PgR 陰性 AND AJCC 第三期) 的案例有 2 筆
(support=1/393=0.0025, confidence=1/2=0.5)
ER 陰性 AND PgR 陰性 AND AJCC 第四期=> 死亡 3 筆
此外，(ER 陰性 AND PgR 陰性 AND AJCC 第四期) 的案例有 4 筆
(support=1/393=0.0025, confidence=3/4=0.75)

ER 陽性 AND PgR 陽性 => 死亡 7 筆
此外，(ER 陽性 AND PgR 陽性) 的案例有 141 筆
(support=7/393=0.0178, confidence=7/141=0.0496)
又
ER 陽性 AND PgR 陽性 AND AJCC 第一期=> 死亡 1 筆
此外，(ER 陽性 AND PgR 陽性 AND AJCC 第一期) 的案例有 79 筆
(support=1/393=0.0025, confidence=1/79=0.0127)
ER 陽性 AND PgR 陽性 AND AJCC 第二期=> 死亡 4 筆
此外，(ER 陽性 AND PgR 陽性 AND AJCC 第二期) 的案例有 43 筆
(support=4/393=0.0102, confidence=4/43=0.0930)
ER 陽性 AND PgR 陽性 AND AJCC 第三期=> 死亡 1 筆
此外，(ER 陽性 AND PgR 陽性 AND AJCC 第三期) 的案例有 4 筆
(support=1/393=0.0025, confidence=1/4=0.25)
ER 陽性 AND PgR 陽性 AND AJCC 第四期=> 死亡 1 筆
此外，(ER 陽性 AND PgR 陽性 AND AJCC 第四期) 的案例有 3 筆
(support=1/393=0.0025, confidence=1/3=0.3333)

ER 陰性 AND PgR 陽性 => 死亡 2 筆
此外，(ER 陰性 AND PgR 陽性) 的案例有 8 筆
(support=2/393=0.0051, confidence=2/8=0.25)
又
ER 陰性 AND PgR 陽性 AND AJCC 第一期=> 死亡 1 筆
此外，(ER 陰性 AND PgR 陰性 AND AJCC 第一期) 的案例有 3 筆
(support=1/393=0.0025, confidence=1/3=0.3333)
ER 陰性 AND PgR 陽性 AND AJCC 第二期=> 死亡 1 筆
此外，(ER 陰性 AND PgR 陰性 AND AJCC 第二期) 的案例有 3 筆
(support=1/393=0.0025, confidence=1/3=0.3333)

(4) 未執行手術共 59 筆，其中有 20 個案例死亡。

未執行手術=>死亡 20 筆

(support=20/1764=0.0113, confidence=20/59=0.3390)

此外，

未執行手術 AND AJCC 第一期=>死亡 1 筆

其中 （未執行手術 AND AJCC 第一期）有 4 筆

(support=1/1764=0.0006, confidence=1/4=0.25)

未執行手術 AND AJCC 第三期=>死亡 3 筆

其中 （未執行手術 AND AJCC 第三期）有 7 筆

(support=3/1764=0.0017, confidence=3/7=0.4286)

又未執行手術 AND AJCC 第三期 AND 大於 80 歲=>死亡 3 筆

其中 （未執行手術 AND AJCC 第三期 AND 大於 80 歲）有 4 筆

(support=3/1764=0.0017, confidence=3/4=0.75)

未執行手術 AND AJCC 第四期=>死亡 10 筆

其中 （未執行手術 AND AJCC 第四期）有 14 筆

(support=10/1764=0.0057, confidence=10/14=0.7143)

(5) 有關組織病理方面

ER 陰性 AND PgR 陰性 AND 大於 55 歲 => 死亡 14 筆

其中（ER 陰性 AND PgR 陰性 AND 大於 55 歲）共 112 筆

(support = 14/1764=0.0079, confidence = 14/112=0.1250)

ER 陰性 AND PgR 陰性 AND 大於 55 歲 AND 8500/3 (Infiltrating duct carcinoma) => 死亡 10 筆

其中 （ER 陰性 AND PgR 陰性 AND 大於 55 歲 AND 8500/3）共 91 筆

(support = 10/1764=0.0057, confidence = 10/91=0.1099)

ER 陽性 AND PgR 陽性 AND 大於 55 歲 => 死亡 20 筆

其中（ER 陽性 AND PgR 陽性 AND 大於 55 歲）共 426 筆

(support = 20/1764=0.0113, confidence = 20/426=0.0469)

ER 陽性 AND PgR 陽性 AND 大於 55 歲 AND 8500/3 (Infiltrating duct carcinoma) => 死亡 18 筆

其中 （ER 陽性 AND PgR 陽性 AND 大於 55 歲 AND 8500/3）共 322 筆

(support = 18/1764=0.0102, confidence = 18/322=0.0559)

(6) 有關癌症期別方面

AJCC 第 0, I, II 期 => 死亡 52 筆

其中 （第 0, I, II 期）共 1091 筆

(support=52/1764=0.0295, confidence = 52/1091=0.0477)

AJCC 第 0, I, II 期 AND 年齡 >= 35 => 死亡 52 筆

其中 （第 0, I, II 期 AND 年齡 >= 35）共 1069 筆

(support = 52/1764=0.0295, confidence = 52/1069=0.0486)

AJCC 第 0, I, II 期 AND 年齡 < 35 => 死亡 0 筆

(support = 0, confidence = 0)

AJCC 第 III, IV 期 => 死亡 31 筆

其中 （AJCC 第 III, IV 期）共 94 筆

(support=31/1764=0.0176, confidence = 31/94=0.3298)

AJCC 第 III, IV 期 AND 年齡 >= 35 => 死亡 30 筆

其中 （AJCC 第 III, IV 期 AND 年齡 >= 35）共 91 筆

(support = 30/1764=0.0170, confidence = 30/91=0.3297)

AJCC 第 III, IV 期 AND 年齡 < 35 => 死亡 1 筆

其中 （AJCC 第 III, IV 期 AND 年齡 < 35）共 3 筆

(support = 1/1764=0.0006, confidence = 1/3=0.3333)

將上述結果依照 confidence 大小匯整如表 7 所示。

表 7 根據 Rough Set Theory 所萃取出的法則

No	法則	Support	Confidence
1	泛中國出生 AND ER 陰性 AND PgR 陽性 AND AJCC 第一期=> 死亡	0.0025	1.0000
2	泛中國出生 AND ER 陰性 AND PgR 陽性 AND AJCC 第三期=> 死亡	0.0025	1.0000
3	泛中國出生 AND ER 陰性 AND PgR 陽性 AND AJCC 第四期=> 死亡	0.0025	1.0000
4	美國出生 AND ER 陰性 AND PgR 陰性 AND AJCC 第四期=> 死亡	0.0025	0.7500
5	未執行手術 AND AJCC 第三期 AND 大於 80 歲=>死亡	0.0017	0.7500
6	未執行手術 AND AJCC 第四期=>死亡	0.0057	0.7143
7	泛中國出生 AND ER 陰性 AND PgR 陽性 => 死亡	0.0076	0.6000
8	泛中國出生 AND ER 陰性 AND PgR 陰性 AND AJCC 第四期=> 死亡	0.0025	0.5
9	美國出生 AND ER 陰性 AND PgR 陰性 AND AJCC 第三期=> 死亡	0.0025	0.5000
10	未執行手術 AND AJCC 第三期=>死亡	0.0017	0.4286
11	未執行手術=>死亡	0.0113	0.3390
12	美國出生 AND ER 陽性 AND PgR 陽性 AND AJCC 第四期=> 死亡	0.0025	0.3333
13	美國出生 AND ER 陰性 AND PgR 陽性 AND AJCC 第一期=> 死亡	0.0025	0.3333
14	美國出生 AND ER 陰性 AND PgR 陽性 AND AJCC 第二期=> 死亡	0.0025	0.3333
15	AJCC 第 III, IV 期 AND 年齡 < 35 => 死亡	0.0006	0.3333
16	AJCC 第 III, IV 期 => 死亡	0.0176	0.3298
17	AJCC 第 III, IV 期 AND 年齡 >= 35 => 死亡	0.0170	0.3297
18	美國出生 AND ER 陽性 AND PgR 陽性 AND AJCC 第三期=> 死亡	0.0025	0.2500
19	美國出生 AND ER 陰性 AND PgR 陽性 => 死亡	0.0051	0.25

20	未執行手術 AND AJCC 第一期=>死亡	0.0006	0.2500
21	台灣出生 AND ER 陰性 AND PgR 陰性 => 死亡	0.0263	0.2
22	泛中國出生 AND ER 陰性 AND PgR 陰性 AND AJCC 第三期=> 死亡	0.0025	0.2
23	美國出生 AND ER 陰性 AND PgR 陰性 => 死亡	0.0178	0.1944
24	美國出生 AND ER 陰性 AND PgR 陰性 AND AJCC 第一期=> 死亡	0.0076	0.1667
25	ER 陰性 AND PgR 陰性 AND 大於 55 歲 => 死亡	0.0079	0.1250
26	ER 陰性 AND PgR 陰性 AND 大於 55 歲 AND 8500/3 (Infiltrating duct carcinoma) => 死亡	0.0057	0.1099
27	泛中國出生 AND ER 陰性 AND PgR 陰性 => 死亡	0.0127	0.098
28	美國出生 AND ER 陽性 AND PgR 陽性 AND AJCC 第二期=> 死亡	0.0102	0.0930
29	泛中國出生 AND ER 陰性 AND PgR 陰性 AND AJCC 第一期=> 死亡	0.0051	0.091
30	泛中國出生 AND ER 陽性 AND PgR 陰性 AND AJCC 第二期=> 死亡	0.0025	0.0769
31	泛中國出生 AND ER 陽性 AND PgR 陽性 AND AJCC 第二期=> 死亡	0.0127	0.0633
32	ER 陽性 AND PgR 陽性 AND 大於 55 歲 AND 8500/3 (Infiltrating duct carcinoma) => 死亡	0.0102	0.0559
33	泛中國出生 AND ER 陰性 AND PgR 陰性 AND AJCC 第二期=> 死亡	0.0025	0.0526
34	泛中國出生 AND ER 陽性 AND PgR 陽性 => 死亡	0.0254	0.0510
35	美國出生 AND ER 陽性 AND PgR 陽性 => 死亡	0.0178	0.0496
36	AJCC 第 0, I, II 期 AND 年齡 >= 35 => 死亡	0.0295	0.0486
37	AJCC 第 0, I, II 期 => 死亡	0.0295	0.0477
38	ER 陽性 AND PgR 陽性 AND 大於 55 歲 => 死亡	0.0113	0.0469
39	泛中國出生 AND ER 陽性 AND PgR 陽性 AND AJCC 第一期 => 死亡	0.0102	0.0435
40	泛中國出生 AND ER 陽性 AND PgR 陰性=> 死亡	0.0025	0.0286
41	美國出生 AND ER 陽性 AND PgR 陽性 AND AJCC 第一期=> 死亡	0.0025	0.0127
42	AJCC 第 0, I, II 期 AND 年齡 < 35 => 死亡	0	0

討論

法則 1, 2 與 3 儘管 confidence 高達 1.0000，然而由於筆數均僅一個，在實務的推論需要更多的資料來佐證該法則的普遍性。法則 4、8，與 9 說明 ER 陰性、PgR 陰性，與乳癌末期導致死亡的機率相當高。此結果與乳癌的醫學臨床結果相吻合(Hirshaut and Pressman, 2001)。法則 5 說明高齡、乳癌末期，與未接受手術切除易導致死亡，該項結果未與常識衝突。同時法則 6、10，與 11 亦強化乳癌末期與未接受手術切除易導致死亡。法則 20 亦強化手術切除的必要性，儘管乳癌處於早期。

法則 15、17、36，與 42 顯示末期患者年輕者較年長者易於死亡；然而初期

患者則年長者似乎較易死亡。其中初期患者則年長者似乎較易死亡的結果與一般研究認為年輕婦女的預後較差相左(Chia et al., 2004; Xiong et al., 2001; Yildirim et al, 2000)，因此需要進一步分析。

法則 26 與 32 說明同為 55 歲以上與 8500/3 的患者，荷爾蒙受體 ER 與 PgR 為陰性的患者較陽性患者易於死亡。這應與荷爾蒙受體 ER 與 PgR 為陰性的患者其缺少接受荷爾蒙療法的機會有關。

結論

本研究使用 1764 筆從 1998 到 2002 發生於美國的華人婦女乳癌資料，並以粗集合理論萃取其中相關的法則。多數結果並未與一般的臨床或常識相左，因此證明粗集合理論確實能協助從醫療資料庫中萃取相關醫療知識。此外，本研究發現對於初期與末期並年輕與年長的華人婦女乳癌患者，粗集合理論發現與一般臨床相左的結果，該結果需進行進一步分析以確定是否有其他因素影響到華人婦女的預後結果，而使得華人婦女的乳癌死亡模式與一般美國大眾不同。

參考文獻

1. Chia KS, Du WB, Sankaranarayanan R, et al. Do younger female breast cancer patients have a poorer prognosis? Results from a population-based survival analysis. Int J Cancer 2004; 108: 761-765.
2. Han, J. and Kamber, M., 2001, Data Mining: Concepts and Techniques, Morgan Kaufmann.
3. Hirshaut Y. and Pressman, P., 2001, *Breast Cancer: The Complete Guide*, 乳癌全書, 原水文化, 廖舜茹(譯).
4. Pawlak, Z., 1991, Rough Sets: Theoretical Aspects of Reasoning about Data, Kluwer Academic.
5. Xiong Q, Valero V, Kau V, et al. Female patients with breast carcinoma age 30 years and younger have a poor prognosis. Cancer 2001; 92: 2523–2528.
6. Yildirim E, Dalgic T, Berberoglu U. Prognostic significance of young age in breast cancer. J Surg Oncol 2000; 74: 267-272.
7. 衛生署，2003，民國 91 年國人主要死因統計資料，http://www.doh.gov.tw/statistic/index.htm