

怎「抉擇」一詞了得

水面下的冰山(一)

施惠德

前言

假設有這麼一天，你已是個大醫師了，你的診察室裏來了一個人。在這個人未說明來意之前，基於你的職業，你心中一定有個問號：他是不是來看病的？你可以猜他有病，也可以猜他没病。不論如何，你答對的可能性是 50%。

在沒有提供任何資料給你的情況下，猜是無可厚非的；不過猜中了可別得意，這並不表示你高明，只是你的運氣好。只要給了你第一個資料——你所看到來人的整體外觀 (general appearance) ——，你答對 (或答錯) 的可能性就再也 不是 50%。如果來人西裝革履，精神抖擻，還遞給你一張上書「X 大藥廠外務員」的名片，你當然不太可能說他有病；要是這人苦著臉進來，好像不勝痛苦的样子，你也不會迷糊到說這人一定沒病。由於醫師診斷疾病的方法就是病史詢問 (history taking) 和各種檢查，依著你想法的偏向，接下來你對要問來人的事情，一定有不同的決定。對那個 propa，你可能只是和他寒暄幾句，問他生意如何；但對那似乎十分痛苦的人，你就無法掉以輕心。

當你問了這一臉愁容的人有什麼

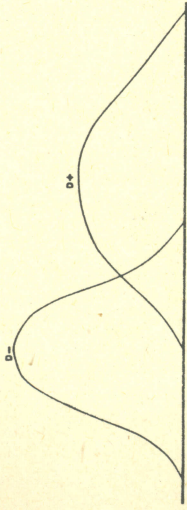
不舒服後，他可能告訴你，他的名字叫孫情，是個歌王，就住在附近，他只是來走走，拜訪一下。於是你只好一面暗罵這「痛苦歌王」浪費了你寶貴「一寸光陰一寸金」的時間，一面問他近來得意與否。但是，這人也可能告訴你，他的右側腰痛。這一來，你的決定必然不會是只和他 social 一番，而是繼續問他的病史。

你們之間的問題可能是這樣的：痛了多久？將近一天。怎麼一種痛法？持續的刺痛，很厲害。會不會延伸到其他部位？有的，右下腹和陰囊裏面。發作以前有沒有受傷、或受到其他刺激？沒有。痛的時候，身子不動比較舒服還是想動？噢，痛得無法翻滾身體。以前也這樣痛過嗎？兩個月前痛過一次。上次共痛了幾天？差不多一星期。但是，他的回答也可能是：將近一天。持續性的酸痛，不延伸到其他地方。只要身體不動就不太痛。以前也痛過，記不得多久以前，好像就是上次考鉛球以後，痛了大約四天。這次開始痛的前一天也考了鉛球。接下來，你決定要問的事情會一樣嗎？前者，可能你會再問問他平常喝的是井水還是自來水，家裏有沒有其他人也患相同毛病；後者，你大概會問他，是不是醫學院的學生？平常是不是都躲在屋子裏看書，不出來運動？因為，你會懷疑，前者有輸尿管結石，後者只是運動過度的肌肉酸痛，而你答對的概率是大於 50%。

後者的身體檢查，你可能發現，他的右側腰部肌肉僵硬，壓了會痛，身體歪向右邊；前者，身體也歪向右邊，右側腰部肌肉也僵硬，但壓了不會痛。此外，你還可能發現，敵敵前者腰椎右側的某一處，會使他像殺豬般地大聲嚎叫。於是，你會決定不同的下一步。後者，你大概不會繼續檢查下去，你可能就叫他回去休息；前者，你會想到讓他作 IVP (intravenous pyelography；靜脈內注射腎盂造影術)，看看尿路是不是真的結石。IVP 沒有使整個尿路顯影，你會考慮作逆行腎盂造影術 (retrograde pyelography)，看看尿路末端是不是還有結石。檢查的結果，會使你決定不同的療法。

沒看到結石，你可能只給這人止痛劑、肌肉鬆弛劑，叫他多喝水，隔段時間再來看；結石不大，腎盂未擴大，IVP 顯影時間正常，你也不大可能叫他去開刀；石頭大，腎盂擴大 (即水腎 hydronephrosis 現象)，IVP 顯影時間延長，你大概會決定讓他去動手術。

像這一段假設，你在現實診斷疾病和治療病人的過程中，時時要決定



圖一 假想的檢查(T)在有病(D+)的群體和正常(D-)的群體中的結果。因為檢查並不完美而使其結果有所重疊。
橫軸：檢查結果。縱軸：人數

下一步該怎麼走。然而，你所想的、決定的正確嗎？那個 propa，萬一突然因顛底 Willis 氏環的動脈瘤 (aneurysm) 破裂、蛛網膜下腔出血而死在你面前，你還說他沒病嗎？那位痛苦歌王，如果不理會你問的話，只一再地重複「我來拜訪你」這句話，你還會埋怨他誤了你看病的時間，而不懷疑他是否真的孫情，不把他轉診到精神科去嗎？那個推了鉛球腰痛的人，若是照了 X 光發現右側腰部好像有出血現象，你會不會只叫他休息，而不建議他去作物理治療？那個像是有輸尿管結石的人，要是對碘過敏，你還要不要他作 IVP (其造影劑含碘)？如果你是不正確的，你用什麼來支持你的抉擇？

用數學方法說明這些，可以讓大家了解，一個抉擇後面，隱含著多少東西。為了方便，以下均以檢查 (test) 的結果來說明。

抉擇標準 (Decision Criterion)

當我們用一種檢查來分別正常和異常時，通常有兩種情形：一是對一群體建立一個診斷策略 (diagnostic strategy)，另一是對一個體作個抉擇 (decision)，到底他是正常或是異常。同一檢查的各個診斷策略間，可能有很大出入。

群體的抉擇策略

用一種檢查來分別正常和有病 (D) 的人時，我們必須有一個確切的

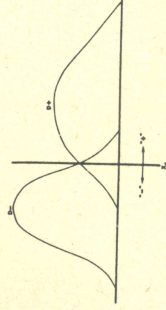
有病 (D) 的人，其檢查結果高低的，將被正確地歸類為真陽性 (true positive; TP)；沒病而檢查低於 X_c 的，將為真陰性 (true negative; TN) (圖三)。有病但檢查結果低於 X_c 的，會被誤認為有病，而為假陰性 (false negative; FN)；沒病而有高於 X_c 的檢查結果的，則為假陽性 (false positive; FP) (圖四)。

在上面這個情形中， X_c 的位使得 FN 和 FP 的數目大略相等，結果誤判的害處 (penalty) 對有病和病的人同樣嚴重的話，這策略是可接受的。如果漏掉一個病例比把該的人當做有病要糟得多，決策者 (strategist) 將會盡量把他的 X_c 向左移，以便少漏失一些病例。只要把所有病人都當做是「陽性」的 (這只要把那些檢查結果高於具有最優查值的病例的人都當做是「陽性」並)，假陰性的問題很簡單地就可以全消除 (圖五)。這時，標準線已去了所有的假陰性結果。但是，在策略下，半數的正常群體都會被當成陽性 (FP) 了。在幾乎所有的實際情

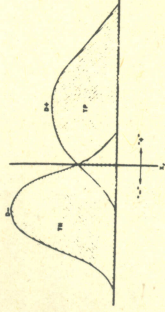
形 (其中陽性或危險性的中，這是不 FN 和 FN X_c 向右移者處與 FP 者處與 FN 檢查的因案而定，罹病率 (mortality))。比較兩者會有一點主者會考慮各人不同價值保持一定 (lence) 增著增加。

個體的

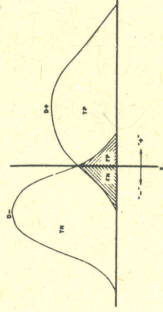
在這標準線的第一個別病人將會找到。那麼。



圖二 標準線 X_c 與兩個群體之交點。X 是右邊的真陽性率 (TP) 在總的為檢者 (T)。



圖三 真陽性 (TP) 區域在 X_c 左邊，真陰性 (TN) 區域在 X_c 右邊。



圖四 真陽性 (TP) 區域在 X_c 左邊，假陽性 (FP) 區域在 X_c 右邊。

形(其中陽性結果促成牽涉到代價或危險性的更進一步的檢查或行動)中,這是不能被接受的。為了平衡FP和FN的害處,決策者只好再把X_c向右移回(圖六),直到FP的害處與FP出現頻率的乘積和FN的害處與FN出現頻率的乘積達成平衡。檢查的害處依一些有形和無形的因素而定,其中包括代價、危險性、罹病率(morbidity)、死亡率(mortality)、輻射劑量、病人的痛苦等。比較兩種不確結果的害處,最少也會有點主觀的成分,因而不同的決策者會考慮各種害處的不同比重,使用各人不同的X_c值。

值得注意的是在圖六中,如果X_c保持一定,當D的盛行率(prevalence)增高時,TP和FN也都會隨著增加。

個體的抉擇策略

在這裏,必須強調的是,決定標準的策路只適用於群體而不是用於個別病人身上的。作臨牀抉擇的醫師將會找到最合乎他個人喜好的標準線。那麼,在圖七中,對檢查結果是X₁

的人,他會怎麼做?因為X₁大於X_c,這人會被歸類到陽性的一群裏。然而,這人真的生病的機會如何?如果這醫師能看看在X₁點上的兩條曲線,他可以發現,事實上這人正常(D-)的機會比異常(D+)的還大。那麼,為何說這人是異常的呢?為了萬一認為這人是正常但結果卻是漏失了的病例所帶來的害處。醫師可能會告訴這人,他有67%的概算(probability)(或2:1的勝算(odds))是正常的,但為了他的利益著想,在沒有用其他方法證明他沒病之前,必須把他當做有病的人來看待。同樣的,檢查結果是X₂的人有50%的概算(1:1勝算)是正常的,而X₃的人是33%的概算(1:2勝算)。

以上這種分析適用於已知正常和異常檢查結果分布的任何檢查-疾病組合(test-disease combination)。

抉擇矩陣 (Decision Matrix)

受檢人的最終結果(outcome)可以和檢查的結果(result)一起,放在一個2×2的表(最簡單的情形

下)裏來比較:

		最終 診斷	
		D+	D-
檢查 結果	T+	TP	FP
	T-	FN	TN

這種表就是所謂的抉擇矩陣(綠杏三十期「信不信由你」一文中譯為決定模式,意思較不相近)。

一項檢查的靈敏度(sensitivity),就是有病的人裏檢查結果為陽性的,和所有真正有病的人之間的比例,也就是所謂的眞陽性比例(true positive ratio; TPR):

$$\text{靈敏度} = \frac{\text{TP}}{\text{實際上的病例數}} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

特異性(specificity)是沒病的人裏檢查結果為陰性的,和所有真正沒病的人之間的比例,也就是眞陰性比例(true negative ratio; TNR):

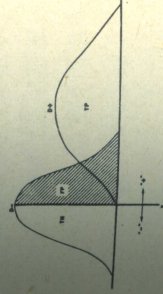
$$\text{特異性} = \frac{\text{TN}}{\text{實際上的非病例數}} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

陽性檢查結果的期望值(predictive value),是所有檢查結果為陽性的人中,真正有病者所佔的比例:

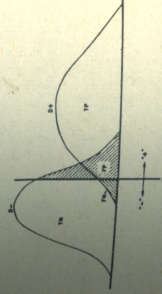
$$\text{期望值} = \frac{\text{TP}}{\text{所有陽性報告例數}} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

同理,陰性結果的期望值為TN/(TN+FN)。

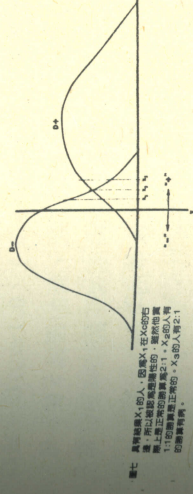
準確度(accuracy)則定義為所有檢查結果是正確的人數和所有接受檢查者的比例:



圖五 顯示X_c移回D-曲線左邊時, FN和FP的變化。



圖六 顯示X_c移回D-曲線右邊時, FN和FP的變化。



圖七 顯示X_c移回D-曲線右邊時, FN和FP的變化。

$$\text{準確度} = \frac{TP + TN}{\text{所有檢查例數}}$$

$$= \frac{TP + TN + FP + FN}{TP + TN}$$

如果用靈敏度和特異性來表示準確度，那麼：

$$\text{準確度} = \frac{TP + TN}{\text{所有檢查例數}}$$

$$= \frac{TP}{\text{所有檢查例數}} + \frac{TN}{\text{所有檢查例數}}$$

$$= \frac{\text{實際上的病例數}}{\text{實際上的病例數}} \times \frac{\text{實際上的病例數}}{\text{所有檢查例數}}$$

$$+ \frac{TN}{\text{實際上的非病例數}} \times \frac{\text{實際上的非病例數}}{\text{所有檢查例數}}$$

= 靈敏度 × 檢查例中實際病例百分比 + 特異性 × 檢查例中實際非病例百分比

假陽性比例 (false positive ratio; FPR) 為沒病的人裏，檢查結果陽性的和所有有病的人之間的比

$$\text{假陽性比例} = \frac{FP}{\text{實際上的非病例數}}$$

$$= \frac{FP}{FP + TN}$$

假陰性比例 (false negative ratio; FNR) 為有病的人裏，檢查結果陰性的和所有有病的人之間的比

$$\text{假陰性比例} = \frac{FN}{\text{實際上的病例數}}$$

$$= \frac{FN}{TP + FN}$$

$$= 1 - \text{靈敏度}$$

似真比例 (likelihood ratio; LR) 為真陽性比例對假陽性比例的比：

$$\text{似真比例} = \frac{\text{真陽性比例}}{\text{假陽性比例}}$$

$$= \frac{TP / (TP + FN)}{FP / (FP + TN)}$$

$$= \frac{TP}{FP} \times \frac{(FP + TN)}{(TP + FN)}$$

有關抉擇矩陣的探討，請參考綠杏三十期「信不信由你」一文，這裏不再贅述。

有條件概率 (Conditional Probability)

「有條件概率」就是在某一條件 (Y) 為真的情形下，一件事 (X) 為真的概率。舉例來說：一副象棋有 32 只棋子。從整副棋子中摸出一只黑棋的概率是 $\frac{1}{2}$ ，也就是說 P (黑棋) 整副象棋) = $16 / 32 = \frac{1}{2}$ ；但是，從已被拿掉一只紅棋的一副棋裏摸出一只黑棋的概率卻是 P (黑棋 | 已拿掉一只紅棋) = $16 / 31$ 。注意這 P (X | Y) 中，X、Y 之間的是條垂直線而非斜線，它代表的並不是「除」。

回過頭來看看前面所說的幾個比例，我們可以發現，它們都是這種有條件概率：

$$TPR = P(T+ | D+) = \text{靈敏度}$$

$$FPR = P(T+ | D-)$$

$$TNR = P(T- | D-) = \text{特異性}$$

$$FNR = P(T- | D+)$$

其中， $P(T+ | D+) + P(T- | D+) = 1$ ， $P(T- | D-) + P(T+ | D-) = 1$ 。而且，考慮檢查結果為陽性的人在整個群體中的概率 P (T+) 和有病、沒病的概率 P (D+)、P (D-) 間的關係時，我們可以發現

$$P(T+) = P(D+)P(T+ | D+) + P(D-)P(T+ | D-) \quad (1)$$

明瞭了這幾種比例的有條件表示法後，我們就可以進一步看這面這定理。

Bayes 氏定理 (Bayes' Theorem) 英國數學家 Thomas Bayes (1702 ~ 1761) 的研究成果，在他後兩年才被發表出來。Bayes 氏提供了一種依據新證據重新計算的方法，但是先期概率 (prior probability) 和做為新證據的一些數目必須是已知的。醫學上的 Bayes 氏定理最簡單的形式可以這樣表示：

$$P(D | S) = \frac{P(D)P(S | D)}{P(S)}$$

$$= P(D) \times \frac{P(S | D)}{P(S)}$$

其中，P (D | S) 為具有症狀 S 的人也有病 D 的概率；P (D) 為病 D 的先期概率 (給予有關 S 的證據前，D 的概率)；P (S | D) 為有病 D 的人也帶有症狀 S 的概率；P (S) 則為一個人 (不論有或無病 D) 有症狀 S 的概率。舉例來說：假設一個群體中甲狀腺機能低下症 (hypothyroidism) (即 D) 發生的概率為 5% (即 P (D) = 0.05)，一般人中只 10% 有失眠 (insomnia) 現象 (即 S) 但在甲狀腺機能低下的人中卻高達 30% (即 P (S) = 0.1, P (S | D) = 0.3)。那麼，P (D | S)，一個失眠的人甲狀腺機能低下的概率，就等於 $(0.05 \times 0.3) / 0.1 = 0.15$ ，或 15%。換句話說，新的證據 S 使一個人可能甲狀腺機能低下的概率提升為未知有 S 前的概率的 3 倍。這個 S，代表的不是症狀，它也可以是病徵 (sign)、實驗室檢查結果、或任何有

關病人的資料。
 從上面的公式，我們可以看出，若是 $P(S|D)/P(S)$ 大於 1，新的概率 $P(D|S)$ (也就是後生概率的 posterior probability) 將會大於先期概率 $P(D)$ 。這情形發生在有病的人有 S 的情形多於一般人時，也就是說，S 和 D 有關。反之， $P(S|D)/P(S)$ 小於 1 時， $P(D|S)$ 將小於 $P(D)$ 。

利用上節的表示法，Bayes 氏定理可以表示成

$$P(D+|T+) = \frac{P(D+)P(T+|D+)}{P(T+)} \quad (4)$$

$$= P(D+) \times \frac{P(T+|D+)}{P(T+)} \quad (5)$$

由於 $P(T+) = P(D+)P(T+|D+) + P(D-)P(T+|D-)$ ，且 $P(D-) = 1 - P(D+)$ ，上面的式子可以改寫

$$= \frac{P(D+)P(T+|D+)}{P(D+)P(T+|D+) + P(D-)P(T+|D-)} \quad (6)$$

$$= \frac{P(D+)TPR}{P(D+)TPR + (1 - P(D+))FPR} \quad (8)$$

此外，用比較簡單的方式來看 Bayes 氏定理，假設 TP 和 EP 的值為已知，那麼

$$P(D+|T+) = \frac{TP}{TP + FP} \quad (9)$$

值得注意的是，這個表示法剛好和期望值一樣。(由 TPR 及 FPR 的定義，諸位可自行試試把公式(8)轉換成公式(9))。

用似真比例來表示 Bayes 氏定理或，則因

$$LR = \frac{TPR}{FPR} \quad (10)$$

$$= \frac{P(T+|D+)}{P(T+|D-)}$$

(只要 FPR 不為 0)，把公式(10)右邊上下各除以 $P(T+|D-)$ ，我們可以得到

$$P(D+|T+) = \frac{P(D+)LR}{P(D+)LR + 1 - P(D+)} \quad (11)$$

$$P(D+|T+) = \frac{1}{1 + \frac{1 - P(D+)}{P(D+)}LR} \quad (12)$$

圖六中，如果曲線 D+ 和 D- 的性質已知，那麼 Bayes 氏定理就可以用積分的方式來表示：

$$P(D+|T+) = \frac{\int_{xc}^x [D+] dx}{\int_{xc}^x [D+] dx + \int_{xc}^x [D-] dx} \quad (13)$$

接受者操作特性曲線 (Receiver - Operating - Characteristic Curve)

一般人在「判定」一種檢查的好壞時，總會認為準確度越高，這個檢查就越好。但是，這並不是正確的觀念。例如，對一種極罕見的病(假設盛行率為 0.05%)，我只要不管是否有任何證據都說這病不存在，那麼在

99.95% 的時候都是對的。我的準確度是很高，然而我的診斷好嗎？準確度很容易受盛行率的影響？而且沒有一種數學上的修正能彌補這缺陷。假設一種檢查 A，它在盛行率為 17% 的一個 1200 人群體中，結果如下：

檢查結果	實際情形		檢查結果合計 (診斷)
	D+	D-	
T+	140 (TP)	100 (FP)	240 (P(T+))
T-	60 (FN)	900 (TN)	960 (P(T-))
實際情形合計	200 (P(D+))	1000 (P(D-))	總計 1200

靈敏度 = $TPR = 140 / 200 = 0.70$; $FNR = 1 - TPR = 0.30$
 特異性 = $TNR = 900 / 1000 = 0.90$; $FPR = 1 - TNR = 0.10$
 $P(D+) = 200 / 1200 = 0.17$; $P(D-) = 1 - P(D+) = 0.83$
 準確度 = $TPR \times P(D+) + TNR \times P(D-) = 0.87$

0.83 小。這是因為檢查 A 對真正有病的人來說較不準確(靈敏度小於特異性)，而真正有病的人在第二個群體中增加了。

準確度不能分別兩種檢查孰優

孰劣？假設檢查B對上面第一個群體
的結果為

檢查	實際情形		檢查結果合計 (診斷)
	D+	D-	
T+	80 (TP)	40 (FP)	120 (P(T+))
T-	120 (FN)	960 (TN)	1080 (P(T-))
實際情形合計	200 (P(D+))	1000 (P(D-))	總計 1200

靈敏度 = $TPR = 80 / 200 = 0.40$

； $FNR = 1 - TPR = 0.60$

特异性 = $TNR = 960 / 1000 = 0.96$

； $FPR = 1 - TNR = 0.04$

$P(D+) = 200 / 1200 = 0.17$ ；

$P(D-) = 1 - P(D+) = 0.83$

準確度 = $TPR \times P(D+) + TNR \times P(D-) = 0.87$

我們可以看出，就有病的人而言，A比B效果好（A的靈敏度較高）；對沒病的人，B是較佳的檢查（B的特異性較高）。然而，A和B的準確度卻是一樣的！A和B那個好？必須一檢查的靈敏度和特异性都高（或低）於另一種，或這兩個比例中有一個相同，才能比較阿！準確度顯然沒法分辨檢查間的優劣。

幾乎沒有一種檢查只具有正負兩種結果。如「抉擇標準」中討論的，我們必須在一串連續的數值中劃分正常和異常。問題是，這條標準線Xc該放在那裏？我們提高或降低TPR（靈敏度）時，FPR也會跟著升降阿！此外，對必須主觀判定結果的檢查（像X光、超音波、核子掃描、CT掃描等顯影法），檢查者心中的懷疑要強到什麼程度才能說一張圖片的結果是陽性？檢查者的「信心閾（confidence threshold）」無疑的和很多東西有關，他的診斷風格（style）、他對先期概率（盛行率）的推測、和他對許多可能正確和正確抉擇的結果的評定（即個人考慮的抉擇正確

的益處和正確的害處）等。「信心閾」在多數情形下是可變的。

因為抉擇標準可以任意選擇，我們可以有意的變動抉擇標準，觀察各個抉擇比例的變化。由於TPR和FPR一起就足以代表四個抉擇比例，我們就可用這兩者作圖。以TPR為縱軸，FPR為橫軸，得到的是一條平滑的雙曲線，這就是「接受者操作特性曲線（以下簡稱ROC曲線）」（圖八），因為它描述了檢查固有的偵檢「特性」，而且檢查資料的「接受者」使用適當的抉擇標準可以在曲線的任一點上「操作」。

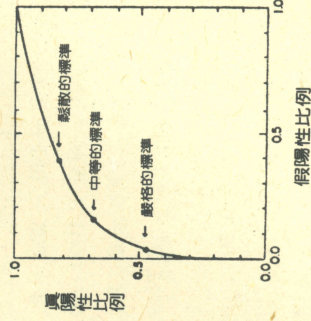
假設受檢者只有有病和沒病兩種實際情形而檢查結果只有陽性陰性兩種選擇，ROC曲線必然會通過圖的左下角（ $TPR = 0, FPR = 0$ ），因為所有檢查結果都可以被當做陰性；它也必然通過圖的右上角（ $TPR = 1, FPR = 1$ ），因為所有結果都可被當做陽性。此外，只要檢查提供了資料給檢查者，ROC曲線中間的點必會高於圖中左下到右上角的對角線

，因為這時對真正有病的人判成陽性結果的可能性要大於對真正沒病的人，也就是說， TPR 大於 FPR 。其次，只要檢查者適當地應用已有的資料，當曲線越到右上方時，它的斜率應該越來越小。最後，由於 TPR 和 FPR 不受疾病盛行率的影響，ROC曲線也不會因盛行率改變而有所變動。

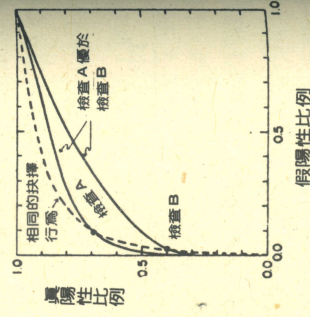
現在，我們來看看ROC曲線如何用來比較不同的檢查。圖九中的兩點分別是依照前面提到的檢查A和B的 TPR 和 FPR 畫上去的。如果我們改變兩者的抉擇標準來作圖，我們可能會得到圖中的那兩條實線。這時，我們就可以以下結論：A對病D比B有更好的偵檢力（detectability）。因為，對相同的 FPR 來說，A的 TPR 較大，而對相同的 TPR 而言，A的 FPR 較小。另一方面，我們也可能發現，兩者的ROC曲線大致相同，如圖中的虛線。這時，我們可以說，兩種檢查對病D的偵檢力相同。同樣的，兩個經驗不同的人判讀同一種檢查的能力，也可用ROC曲線來比較（圖十）。

總之，一項檢查的ROC曲線越向左、向上移，它的偵檢力越好（因為靈敏度、特异性都越高）。

如何建立一條ROC曲線？對結果是一些數值的檢查來說，問題較小



圖八 典型的ROC曲線，顯示有三個可能的操作點。



圖九 檢查A和B的結果在ROC曲線畫出的點，以及這兩點可能存在的ROC曲線。

觀成分，而且它無法說明在實驗的設
計上發生的一些變數。由於標準偏差
也無法說明這些變數（例如 ROC 曲
線是依照多次受統計左右的觀察而建
立的之時），它似乎低估了這種實驗
的 ROC 曲線分離的顯著性。另一方
面，標準偏差似乎也不盡說明了判讀
者生體或心理上的變化，或受測的顯
影系統的變化。然而，不論 ROC 曲
線在分離的顯著性的比較上還存有什
麼不足的地方，它仍然是個用來比較
檢查的有力工具。

以上討論的 ROC 曲線是只假設
受檢者有病或沒病且檢查者只有正負
兩種抉擇而畫出的兩度空間圖形。更
複雜些的情形將是如何？首先，我們
考慮檢查者不僅要判定實際有病的人
為陽性，還要說明陽性結果何在才算
完全真陽性的情形。例如，假設檢查
者還須指出病灶位於影像的那個象限
且病灶只出現在一個象限中，這時就
有五種實際情形和抉擇：沒病、病灶
在右上四分之一、……等，畫出的
ROC 曲線將是三度空間中的一條弧
線，而可在二度空間的圖中畫成兩條
線（一為陽性與陰性的抉擇，另一為
象限正確與錯誤的抉擇）。其次，如
果可能不只發生一個病灶而檢查者必
須算出病灶的數目，那麼我們得到的
ROC 曲線將是多度空間裏的一條弧
線，而可畫為一套二度空間的圖。

我們怎樣選擇一個最好的操作點
？一項檢查，判讀的結果正確時有什
麼利益？錯誤了有什麼代價？再者，
做一項檢查的利益和代價比起來計算
嗎？不做這檢查的利益和代價呢？當
然，我們希望的是花最少的代價得到
最高的利益，這就必須對這兩個做個
分析。

（待續——其餘部份將於下期刊出）

級、第 5 + 4 + 3 級、5 + 4 + 3 +
2 級、5 + 4 + 3 + 2 + 1 級反應作
陽性結果，畫出四點。以平滑曲線連
接各點，就可得到這位檢查者的 ROC
曲線。由於畫最後一點時是把所有結
果都當做陽性（即使幾乎絕對是陰性
），我們得到的點必然是最右上角、
TPR 和 FPR 都是 1 的那點，所以
我們實際上只有四個「操作點（oper-
ating point）」，也就是說，分成
N 個等級就有 (N - 1) 個操作點。

因為任何 ROC 實驗的受檢者數
目都是有限的，而且人的抉擇並不是
每次都能有相同結果，我們畫出來的
每一點都會有統計上的誤差。它的標
準偏差 (standard deviation) 可以
如下推算：

TPR 的標準偏差

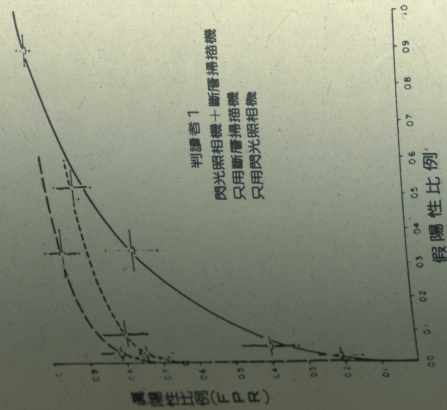
$$= \sqrt{\frac{\text{TPR} \times (1 - \text{TPR})}{(\text{實際病例數}) - 1}}$$

FPR 的標準偏差

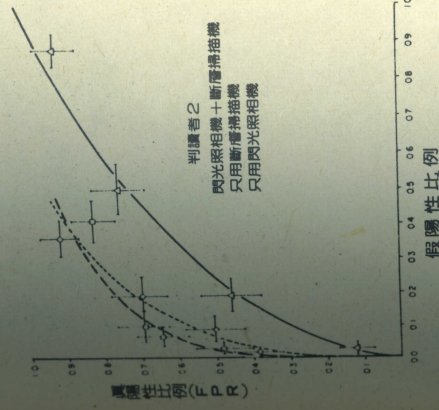
$$= \sqrt{\frac{\text{FPR} \times (1 - \text{FPR})}{(\text{實際非病例數}) - 1}}$$

標準偏差在圖上表示法可參考圖十。
值得注意的是(1)標準偏差值會隨點
位置變動，而以 TPR 或 FPR 接近
0.5 時最大（諸位可試以上列公式證
明。這表示「猶」的時候，誤差最大
）；(2) TPR 與 FPR 的標準偏差分
別和受檢者裏的實際病例數與非病例
數成反比。因此，作 ROC 實驗時，
通常會使用大略相等數目的病例和非
病例。

雖說標準偏差可以求出，但事實
上目前還沒有完全令人滿意的統計方
法可以定量地測出 ROC 曲線之間分
離程度 (separation) 的顯著性 (significance)。這是因為平常的目
測法（以肉眼觀察分離程度）帶有主



判讀者 1
閃光照相機 + 斷層掃描機
只用斷層掃描機
只用閃光照相機



判讀者 2
閃光照相機 + 斷層掃描機
只用斷層掃描機
只用閃光照相機

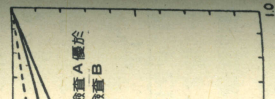
圖十
顯示較有經驗的判讀者有較好的成就。

；但是，那些結果是必須主觀地判讀
的圖像的檢查呢？正負兩種結果只能
得到一個點而不是一條線啊！要解決
這問題，我們通常用的是「分級法 (rating method)」。判讀的人，就
檢查所提供的資料（或其他補充資料
），選擇他的「信心等級」之一來表
示他的判斷。這些等級在性質上大致
分為：(1)絕對或幾乎絕對陰性，(2)很
可能陰性，(3)可能陽性，(4)很可能陽
性，(5)絕對或幾乎絕對陽性。檢查者
先以他的第五級反應做為「陽性」結
果（只有肯定確是陽性的才當做陽性
），對照病人的實際情形，算出他的
TPR 和 FPR，在 ROC 空間上畫
出一點。接下來，再分別用第 5 + 4

病的人判成陽性
對真正沒病的人
大於 FPR。其
地應用已有的資
方時，它的發
，由於 TPR 系
率的影响，ROC
改變而有所變

看 ROC 曲線如
查。圖九中的兩
判的檢查 A 和 B
上去的。如果我
準來作圖，我們
兩條實線。這時
：A 對病 D 比 B
ectability)。
來說，A 的
同的 TPR 而言
子一方面，我們
ROC 曲線大致
這時，我們可
D 的偵檢力相同
不同的人判讀同
用 ROC 曲線

ROC 曲線越
檢力越好（因
高）。
C 曲線？對結
說，問題較小



判
ROC 空間畫出的
存在的 ROC 曲線