

# 信不信由你 信不信由你 信不信由你 信不信由你

醫十屆

曾凱元

本校附設醫院婦產科住院醫師  
曾任本校解剖學科助教  
現為美國約翰霍普金斯大學醫學院核子醫學碩士班研究生

惠德賢弟：

很抱歉，稿件一拖再拖，遲至今日方能付郵，雖未胎死腹中，也有點 *dysmature* 了。因此，如果不合需要的話，*please!* 丟到垃圾筒內；不過，在丟棄前別忘了把我名字塗掉。

前一陣子在電腦、醫院、期末考三重壓力之下，僅敢圖苟存，到了春假，才能喘一口氣。因此，這些東西是在苟延殘喘之下寫出來的一不要 *expect too much*。

這是一篇 *criticize* 的文章，裏面用了不少 *sensitivity*，*specificity*，*accuracy*，當我開始批評那篇批評的文章時，不斷地使用 *sensitivity*，*specificity*，*accuracy*，*false positive*，*true negative* 等，而這些 *term* 是以前在學時不太明瞭的，整理了幾

遍，略加說明之後，仍然怕同學們看不懂，（因為是小人之心度君子之腹）因此，又再改變主意，傾全力介紹這些東西，以期一次能讓同學們瞭解這些 *term* 之意義及應用，相信一定比我一個人在那裡自說自話好。同時，又為了引起興趣起見，用了一個莫名其妙的篇名，加上最後還有幾個題目請大家小試牛刀；一方面可看是否真正瞭解，一方面介紹這種美國 *practical* 的學習方式，其中第一、第二題是相當容易的一用以培養信心，第三題是考驗真正瞭解的，如果解不出的話，我可寄上解答。

關於裡面之用字，請你與憲宏大力斧正，使之 *readable*。若有你也看不懂之處，請儘快來信。

至於 *Receiver - operating - characteristic curve*，其內

容更深，更不易懂。不過，假如這篇能夠被接受—我是指同學們看得懂—的話，我則計劃在暑期間再寫一篇有關 *R. Q. C. curve* 之文章。

來此已近八個月了，回顧這三個 *Quarter* 以來，實在談不上什麼心得，即將來臨的第四 *Quarter* 更是忙碌，大約要在五月底才能 *relax*。這，或許就是 *typical* 的 *American life* 吧！

近來已溶了冬雪，春天的氣息已降臨此間了，台北還暖和吧？或是很熱了！

有空，有興趣的話，請歡迎來信。

祝  
學安，編安

愚兄 凱元上  
67. 4. 2 夜

## 前言

假如你在某一本醫學雜誌上看到一篇文章告訴你，目前已有一種新的檢查方法，其對於某種疾病之檢出率高達百分之九十九，而且，比傳統的方法不曉得好了多少倍！你相信嗎？

如果你去醫院照了一張胸部的X光片，一位醫師告訴你得了肺結核，你相信嗎？

假如上述那篇文章是刊登在新英格蘭醫學雜誌( The New England Journal of Medicine )上你就相信他，而刊登在一本不見經傳的雜誌上你就懷疑他嗎？

如果上述那位醫師是某大醫院的主任醫師你就相信他，而若出自一個不見經傳的醫師口中時，你就不相信嗎？

由於我們所接受到的醫學教育，甚或更進一步的說，一般的教育，使得我們相信了一件事，那就是權威——不論這個權威是那一方面的權威，他所說的是句句真言，字字珠璣；那就是教科書——不論那本書是那一方面的專書，只要書上寫的都是對的。這樣的接受是正確的嗎？

這些大膽的假設，若不經過小心的求証就進入我們的腦海裡，安全嗎？

本文的目的不在於向權威挑戰，亦不在於懷疑真理，而是在於討論如何判讀一篇醫學報導的文章，或者，更進一步的說，如果你有什麼見解之

時，應如何表達，方不致淪為笑柄。

## 方式

當一個醫師面對著一個病人之時，他所面臨的兩個問題是：第一，這個病人到底患了什麼病？第二，應該如何治療病人？這些問題要如何解決所憑藉的乃是那位醫師的知識與經驗。比如說：一個病人主訴在飢餓時會有上腹部疼痛的症狀，喝下牛奶之後就好了。那麼這個醫師一下就會聯想到他患的是消化性潰瘍，而且，他也能輕易地開些制酸的藥劑來治療他。但是，如果你要求這個醫師作一個百分之百的診斷，那麼，他可以根據這些症狀就遽下結論嗎？要不然的話，他該使用那一種檢查方法幫助他診斷呢？X光檢查？胃鏡檢查？活體組織切片檢查？還是「開開看」？另一方面，如果你要求這個醫師開一種最有效的(百分之百有效)制酸劑給你，他能夠嗎？這裡，就牽涉到如何來衡量那一種檢查方法最可靠，最準確以及那一種治療效果最徹底的問題了。再舉一個例子，假如由一張胸部X光片檢查已懷疑到一個病人患了肺癌，那麼，下一步應作何種檢查呢？那一種檢查先作，那一種檢查後作呢？檢查之後要不要開刀呢？如果開刀前沒有徹底瞭解情況，等到打開了胸腔一看，原來腫瘤已蔓延開了，這種開刀是必要的嗎？這一連串的問題，都

要由病人的醫師來作決定( make a decision )。用來衡量，判斷這些檢查與治療的手段，有下列三種方法：

- 第一：決定模式( Decision matrix )。
- 第二：Receiver-operating-characteristic curve。
- 第三：資料定理( Information theory )。

本文不擬一下子將這三種方法論完畢，以免不但無法讓大家學到一樣觀念，卻反而搞亂了原來井然有序的思考途徑，故僅介紹決定模式( Decision matrix )一種如下。

## 決定模式 (Decision - Matrix)

利用決定模式之方法可以判斷一種檢查方法之靈敏度( sensitivity)以及特異性( specificity)。在尚未深入討論靈敏度與特異性前，必須先明瞭兩件事實：

- 第一：到底有多少人或那些人是患有此病的，而不患此病的又有多少人。
- 第二：凡是檢查結果異常者，皆劃入陽性( positive )反應群中，而正常者概劃入陰性( negative )反應群中。

再以上述消化性潰瘍為例說明如下：

假若目前有 100 個具有上述症狀之「病人」來就診，而其中真正患有消化性潰瘍者只有 60 人，而其餘 40 人並沒有潰瘍（假設我們可以將這些「病人」皆作開腹檢查以證明其是否患有此病）。倘若我們利用上消化道鋇灌 X 光系列攝影（upper G-I barium meal series）作為檢查之方法。在患有此病的 60 個病人當中，有 48 人其檢查結果為陽性（異常），那麼，這 48 個既患有消化性潰瘍而且檢查結果又為陽性者則列入真陽性（true positive）群中；至於另外 12 個患有此病而檢查結果為陰性（正常）者則列入偽陰性（false negative）群中。易言之，這項檢查無法將這 12 個病人之病診斷出來。再說那 40 個不患有此病的「病人」當中，有 4 個人其 X 光檢查是陽性（異常）那麼，這 4 個不患有消化性潰瘍但是檢查却為陽性者則列入偽陽性（false positive）群中，至於其他 36 個不患此病而且檢查又呈陰性（正常）者，即應列入真陰性（true negative）群中了。為了了解方便起見，將這些結果列表如下：

結果 \ 患者	是	否	計
陽性	(真陽性) 48	(偽陽性) 4	52
陰性	(偽陰性) 12	(真陰性) 36	48
計	60	40	100

這個結果有什麼意義呢？如果我們計算所有患有此病的病人當中，利用此項檢查到底可以檢出多少的比例的話，我們可以得到下列的結果。

$$\frac{48}{60} = 80\%$$

換句話說，在每 100 個患有此病的病人當中，可以利用此種檢查法發掘其中的百分之八十，此即該檢查法之靈敏度（sensitivity）。另外，如果我們計算所有不患此病之「病人」當中，利用此項檢查可以判定多少的比例是正常的話，我們可以得到下列的結果。

$$\frac{36}{40} = 90\%$$

換句話說，這種檢查只可以在 100 個不患此病的「病人」當中斷定其中 90 個人是正常的，此即該檢查法之特異性（specificity）。此外，如果我們計算到底此項檢查的結果有多少的比例是正確的，我們會得到下列的結果。

$$\frac{48 + 36}{100} = 84\%$$

換句話說，這種檢查的正確性（accuracy）為百分之八十四，百分之十六的檢查結果是錯誤的。

那麼，到底這種檢查好不好呢？如果你是醫師的話，你要不要採用另一種檢查法呢？

就最理想的狀況而言，一種檢查最好能將所有的病人檢出（百分之一百的靈敏度），並且將所有非病人排除（百分之一百的特異性）。但是，在現實的世界中是很難找到這種方法的。比如說，上面的這項檢查，其靈

敏度只有百分之八十，特異性只有百分之九十。假設另一種檢查法是利用胃酸 pH 值作為判定的標準，假定「病人」之胃酸 pH 值低於某一定值時即判定為患有消化性潰瘍，高於某一定值時則判定沒有，那麼可能會獲致下列之結果。

結果 \ 患者	是	否	計
陽性	57	24	81
陰性	3	16	19
計	60	40	100

那就是說，在 60 個患有消化性潰瘍的病人當中，有 57 人其胃酸 pH 值低於某一判定值，只有 3 人被判定在正常範圍內；但是，在 40 個非患有潰瘍之「病人」當中，却有 24 人其胃酸 pH 值低於判定值，而只有 16 人被判定是正常的。根據上述之計算法則其：

$$\text{靈敏度} = \frac{\text{胃酸 pH 值低於某一定值之病人數}}{\text{患病之病人總數}}$$

$$= \frac{57}{60} = 95\%$$

$$\text{特異性} = \frac{\text{非患者中 pH 值高於某一定值之人數}}{\text{所有不患此病之「病人」總數}}$$

$$= \frac{16}{40} = 40\%$$

$$\text{正確性} = \frac{\text{真陽性與真陰性之和}}{\text{所有檢查總數}} = \frac{73}{100} = 73\%$$

這兩種檢查方法一比較之下，很顯然的，胃酸之測定法可以抓住更多的病人；然而很不幸的，也使許多正常人被列入病人考慮之列了。易言之，此法之靈敏度很高，可是却沒有特異性可言了（non-specific）。同時，檢查之準確度也由百分之八十四降為百分之七十三。

推而廣之，任何之檢查或治療皆可用此種決定模式來計算其靈敏度，特異性與準確度，並且，利用這三種值也能比較各種不同方法之優劣了。

### 討論

然而，假如你作進一步分析時，你會發現，其中尚存留著兩個問題：

第一，靈敏度與特異性通常皆互為消長，一個靈敏度高的檢查法，多半其特異性也低；一個特異性高的檢查法，其靈敏度多半較差。那麼，你究竟應該要選擇靈敏度高的檢查呢，還是特異性高的檢查？

第二，檢查之結果如何稱為陽性（不正常）如何稱為陰性（正常）？如果你改變了這些判斷所根據的標準（criteria），那麼靈敏度與準確度都會隨之而改變，如此究竟應該怎麼訂定標準值呢？

在此先談第一個問題。假如對某種病的診斷，你寧可信其有而不可信其無的話，你就應該選取靈敏度較高的方式。比如說，你希望不要遺漏掉任何一個甲狀腺機能低下症（hypothyroidism）的病例（因為，即使正常人給予甲狀腺素治療也沒有太大的損害），那麼，你應該把甲狀腺素在血液中之濃度標準值訂得高些，那麼，所有甲狀腺低下症之病人以及一部分正常人皆可被認為是患有此病而加以治療了。另一方面，假如你希望只有絕對患有轉移癌症的病人才給予化學治療，其餘的皆要用手術療法時，你就應該把轉移癌症之標準訂得嚴格些，只有那些真正有癌細胞蔓延的病人才列入化學療法，否則可以開刀治療者即會失去了他治癒的契機了，此時，特異性必然要訂得高。

因此，判斷值之選擇應由臨床上各個角度來衡量才作決定，以使得每一種病之檢查法皆能得到其最適當之特異性與靈敏度。當然，如果將這些輔助診斷的檢查方法分為兩類，其一為靈敏度高者先將絕大多數之病人挑出，再利用特異性高的檢查方法將這些選出來的「病人」中正常者剔除出來，那麼，就再好不過了。

再談第二個問題，其所牽涉的範圍更為廣泛了。比如說，胸部X光照片之檢查並沒有一定之標準可言，何者謂之正常，何者謂之異常，端視個人所訂的標準而定。如果你訂的標準較為嚴格，那麼，被判定有問題的片子就少了。但是，反過來說，如果你所訂的標準較鬆，那麼，被認為有問題的片子就多了（真陽性與偽陽性之比率皆增高），特異性也就下降了。此外，這也與一般人口中患有此病之罹患率（prevalence rate）有關，為了克服此類偏差，可以用 receiver-operating-characteristic curve 來解決之，這方面以後有機會再詳加討論。

### 應用

#### 1. 問題一

在上述利用上消化道X光攝影之檢查中，如果有100個檢查結果異常（陽性）之病人，你期望其中有多少是患有此病的？易言之，如果你的檢查是陽性的話，你有多大之可能性患有消化性潰瘍？

$$\text{答：} \frac{48}{52} = 92\%$$

#### 2. 問題二

假設某種檢查對於某種疾病之靈

敏度為百分之九十九，特異性為百分之九十五，該病之罹患率由百分之一增到百分之五時，其偽陽性率增加還是減少？

答：由83%減為49%

#### 3. 問題三

某一臨床醫師作冠狀動脈疾病（coronary arterial disease）的探討，在一群為數五千人的團體（年齡介於45~60之間）中作休息時心電圖之檢查。此項檢查之靈敏度為百分之八十，特異性為百分之九十，凡是休息時心電圖的檢查，此項檢查之靈敏度為百分之九十，特異性亦為百分之九十。假定，這五千人中，此病之罹患率（prevalence rate）為百分之二十。

請問

①有多少人要作運動時心電圖檢查？

答：1200人

②此兩項檢查之總靈敏度（overall sensitivity）若干？

答：72%

③此兩項檢查之總特異性（overall specificity）若干？

答：99%

### 後記

本篇所記之這些觀念，目前已相當普及了，在醫學雜誌上時常可見到，但願這些說明能有助於大家對各項文獻之判讀，至於 receiver-operating-characteristic curve 之詳情，以後有機會再談。

最後之題目只是在加強一次各項觀念之運用而已，希望你能夠答對。