

# Diễn giải nghiên cứu tiên lượng: ROC (Receiver Operating Characteristic)

Nguyễn Văn Tuấn  
Viện nghiên cứu Y khoa Garvan, Sydney, Australia

Trong bài trước, chúng ta đã bàn qua cách diễn giải kết quả chẩn đoán và tiên lượng, với kết quả xét nghiệm mang tính “nhị phân” (tức chỉ có hai giá trị “dương tính” hay “âm tính”). Trong bài này, chúng ta sẽ tìm hiểu cách diễn giải kết quả tiên lượng của một yếu tố có đơn vị đo lường liên tục (còn gọi là *continuous variable*). Yếu tố tiên lượng mang tính liên tục bao gồm những biến như huyết áp, insulin, glucose, creatinine, prostate-specific antigen, v.v... Việc phân tích và diễn giải khả năng tiên lượng của các yếu tố lâm sàng và sinh hóa này tương đối phức tạp hơn là những biến nhị phân.

Chẩn đoán là một qui trình bất định. Tính bất định xuất phát từ phương pháp xét nghiệm, dù là xét nghiệm sinh hóa hay điện tử, ít khi nào có những kết quả chính xác tuyệt đối. Do đó, kết quả chẩn đoán thường được diễn giải bằng xác suất. Chính vì thế mà William Osler (1849 – 1919), một người thầy thuốc lớn của Mĩ từng nói một câu bất hủ: “*Medicine is a science of uncertainty, and an art of probability*” (y học là một khoa học về tính bất định, và là một nghệ thuật xác suất).

## Độ nhạy và đặc hiệu

Một cách để xem xét mối liên hệ bất định giữa kết quả xét nghiệm và bệnh trạng là qua biểu đồ 2x2 như sau:

		Bệnh trạng	
		Có bệnh	Không có bệnh
Kết quả xét nghiệm	+ve	True positive Dương tính thật	False positive Dương tính giả
	-ve	False negative Âm tính giả	True negative Âm tính thật

Nói cách khác, so sánh giữa kết quả xét nghiệm và bệnh trạng thực tế của bệnh nhân, chúng ta có 4 tình huống khả dĩ:

- Bệnh nhân mắc bệnh, và kết quả xét nghiệm là dương tính (*dương tính thật* hay *true positive*) – không có vấn đề gì;
- Bệnh nhân mắc bệnh, nhưng kết quả xét nghiệm là âm tính (còn gọi là *dương tính giả* hay *false positive*);
- Bệnh nhân không mắc bệnh, và kết quả xét nghiệm là âm tính (*âm tính giả* hay *false negative*);

- Bệnh nhân không mắc bệnh, nhưng kết quả xét nghiệm là âm tính (còn gọi là *âm tính thật* hay *true negative*;

Tỉ lệ dương tính thật, dương tính giả, âm tính thật, và âm tính giả có thể ước tính qua các nghiên cứu tiên lượng. Cách tính các chỉ số này đã được mô tả trong bài trước, nên chúng ta sẽ không lặp lại ở đây, chỉ xin nhắc lại một vài điểm chính: tỉ lệ dương tính thật được gọi là *độ nhạy* hay *sensititivity*; tỉ lệ âm tính thật được gọi là *độ đặc hiệu* hay *specificity*. Việc chọn phương pháp có độ nhạy hay độ đặc hiệu cao tùy thuộc vào tình hình thực tế của bệnh và nhất là tùy thuộc vào sự trả giá cho “sai lầm”.

**Sử dụng xét nghiệm có độ nhạy cao.** Phương pháp có độ nhạy cao cho phép chúng ta loại bỏ khả năng mắc bệnh. Do đó, nên chọn phương pháp xét nghiệm có độ nhạy cao nếu bỏ sót bệnh dẫn đến hệ quả nghiêm trọng cho bệnh nhân. Chẳng hạn như nếu chúng ta có lí do để nghi ngờ rằng một bệnh nguy hiểm (như lao phổi, chứng huyết khối tĩnh mạch sâu [deep vein thrombosis hay DVT], bệnh Hodgkin, v.v...), chúng ta cần phương pháp xét nghiệm có độ nhạy cao, để loại trừ khả năng bệnh nhân mắc bệnh với kết quả âm tính. Trong trường hợp khác, chúng ta nên chọn xét nghiệm có độ nhạy cao như kháng HIV để loại trừ nhiễm trùng liên quan đến AIDS. Nói tóm lại, một phương pháp xét nghiệm có độ nhạy cao rất có ích cho người thầy thuốc khi kết quả xét nghiệm là âm tính.

**Sử dụng xét nghiệm có độ đặc hiệu cao.** Phương pháp xét nghiệm đặc hiệu cao có nghĩa là nếu đối tượng không có bệnh, phương pháp xét nghiệm ít khi nào cho ra kết quả dương tính. Điều này có nghĩa là các xét nghiệm có độ đặc hiệu cao rất có ích cho việc xác định bệnh trạng. Chẳng hạn như trong trường hợp DVT, xét nghiệm D-dimer không có độ đặc hiệu cao (khoảng 63%); do đó, khó mà quyết định điều trị cho dù kết quả xét nghiệm là dương tính. Tất cả các bệnh nhân có kết quả xét nghiệm dương tính thường được xét nghiệm tiếp tục bằng siêu âm (compression ultrasonography) – một phương pháp có độ đặc hiệu cao. Các xét nghiệm đặc hiệu cao rất cần thiết khi kết quả dương tính giả dẫn đến hệ quả nghiêm trọng cho bệnh nhân. Chẳng hạn như trước quyết định hóa học trị liệu cho bệnh nhân, các phương pháp xét nghiệm mô (tissue tests) cần phải được tiến hành. Nói tóm lại, phương pháp xét nghiệm đặc hiệu cao rất có ích khi kết quả xét nghiệm là dương tính.

## Cân bằng giữa độ nhạy và đặc hiệu

Một phương pháp xét nghiệm lí tưởng có độ nhạy và đặc hiệu gần 100%. Nhưng trong thực tế, ít khi nào chúng ta có một phương pháp xét nghiệm hoàn hảo như thế. Có những xét nghiệm có độ nhạy cao, nhưng độ đặc hiệu thấp (và ngược lại). Do đó, vấn đề là cần phải cân bằng giữa hai khía cạnh nhạy và đặc hiệu của một phương pháp xét nghiệm, nhất là trong trường hợp kết quả xét nghiệm là một biến liên tục.

Trong trường hợp kết quả xét nghiệm có nhiều giá trị (biến liên tục), việc xác định một giá trị tham chiếu (reference cut-off value) để phân biệt giữa “dương tính” và “âm tính” (hay giữa “bình thường” và “bất bình thường”) là điều rất quan trọng. Thông thường, các một xét nghiệm có nhiều giá trị liên tục, độ nhạy và độ đặc hiệu thường thay

đổi nghịch chiều. Khi độ nhạy tăng thì độ đặc hiệu giảm; ngược lại, khi độ nhạy giảm thì độ đặc hiệu tăng. Không có cách gì để gia tăng độ đặc hiệu và độ nhạy trong cùng một lúc cho một xét nghiệm.

Số liệu sau đây (**Bảng 1**) cho thấy mối liên hệ giữa xét nghiệm PSA (prostate-specific antigen) và ung thư tiền liệt tuyến ở đàn ông người Mỹ trong độ tuổi 70. Nếu giá trị tham chiếu được chọn là  $\geq 3$  ng/mL thì độ nhạy là 100%, nhưng độ đặc hiệu chỉ 60%. Điều này có nghĩa là dương tính giả lên đến 40%, tức rất nhiều người có kết quả “dương tính” (trên  $\geq 3$  ng/mL) nhưng họ không bị ung thư. Nếu chọn giá trị tham chiếu là  $\geq 13$  ng/mL thì độ đặc hiệu là 97% nhưng độ nhạy chỉ 23%; nói cách khác, xét nghiệm bỏ sót 77% người mắc bệnh! Các tác giả đề nghị lấy giá trị tham chiếu là  $\geq 5.5$  ng/mL vì ở ngưỡng này, độ nhạy là 95% và đặc hiệu là 78%, tức có thể chấp nhận được.

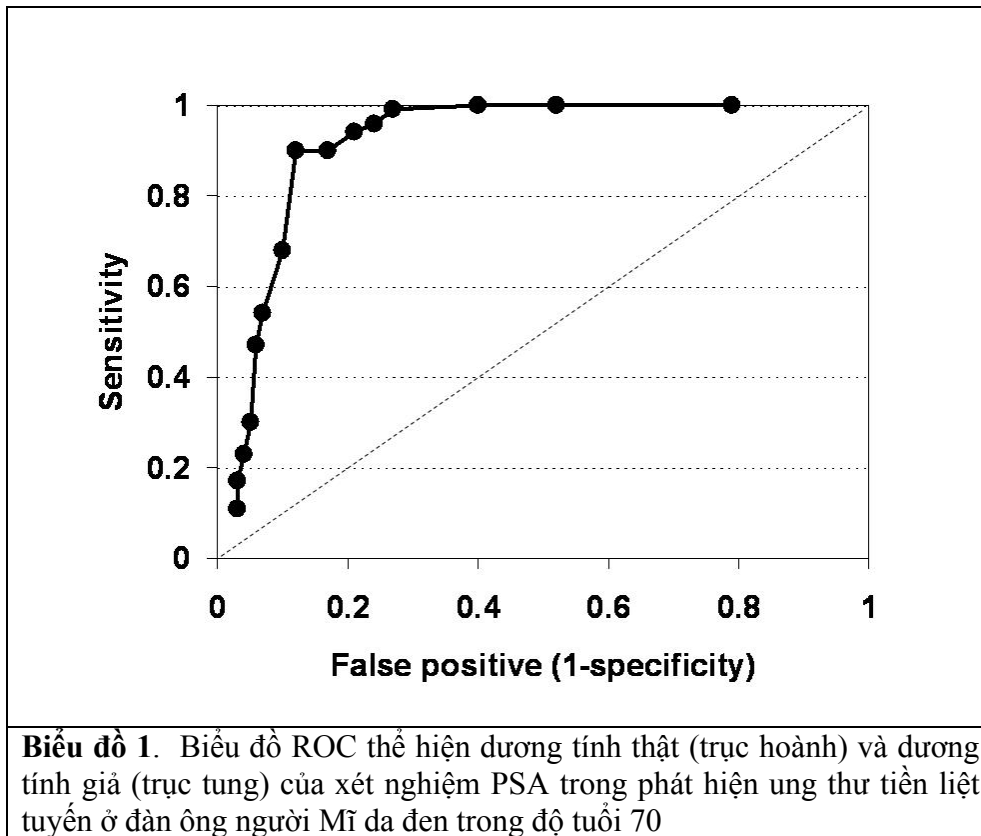
<b>Bảng 1. Độ nhạy và đặc hiệu của xét nghiệm PSA trong tiên lượng ung thư tiền liệt tuyến ở đàn ông người Mỹ da đen</b>		
<b>PSA (ng/mL)</b>	<b>Độ nhạy</b>	<b>Độ đặc hiệu</b>
$\geq 1$	1.00	0.21
$\geq 2$	1.00	0.48
$\geq 3$	1.00	0.60
$\geq 4$	0.99	0.73
$\geq 5$	0.96	0.76
$\geq 6$	0.94	0.79
$\geq 7$	0.90	0.83
$\geq 8$	0.90	0.88
$\geq 9$	0.68	0.90
$\geq 10$	0.54	0.93
$\geq 11$	0.47	0.94
$\geq 12$	0.30	0.95
$\geq 13$	0.23	0.96
$\geq 14$	0.17	0.97
$\geq 15$	0.11	0.97

Nguồn: xem tài liệu tham khảo [1]

## **Biểu đồ ROC**

Một cách khác để mô tả mối liên hệ giữa độ nhạy và đặc hiệu là biểu đồ ROC (receiver operating characteristic) [2]. Biểu đồ ROC có nguồn gốc quân sự, vì nó được ứng dụng trong việc phát hiện tàu của địch trên màn hình radar trong Thế chiến thứ II, nhưng trong thời gian 3 thập niên qua, biểu đồ ROC đã được ứng dụng trong chẩn đoán và tiên lượng y khoa rất thành công.

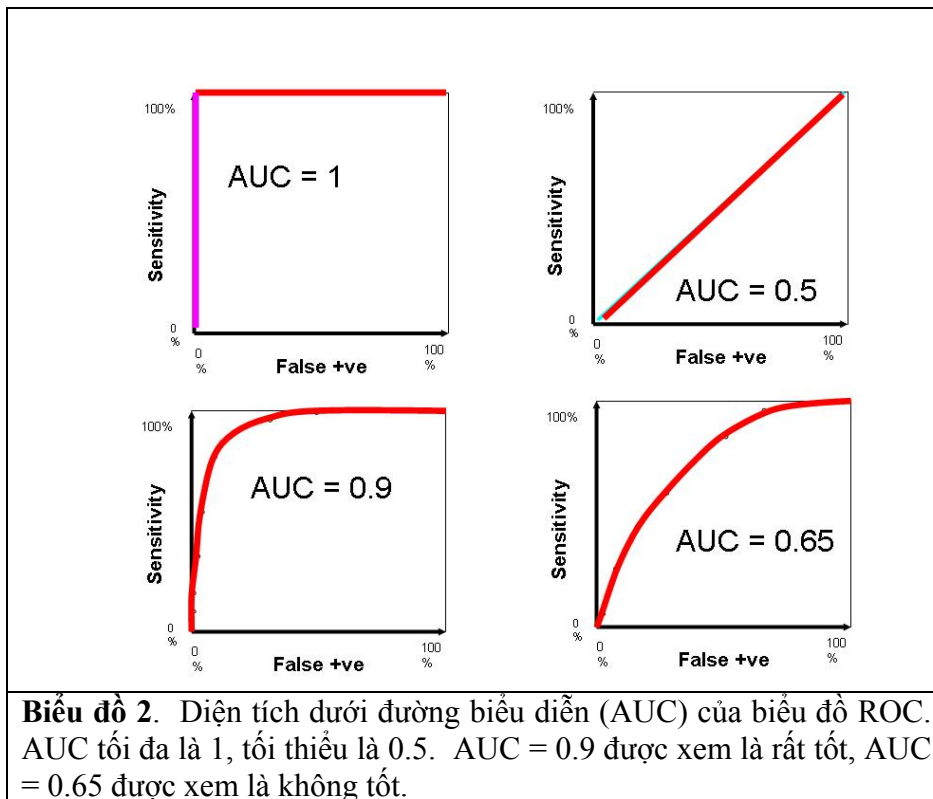
Biểu đồ ROC có trục tung (y-axis) là tỉ lệ dương tính thật, và trục hoành (x-axis) là tỉ lệ dương tính giả (tức 1 trừ cho độ đặc hiệu). Cả hai tỉ lệ dĩ nhiên có giá trị dao động từ 0 đến 100 (hay từ 0 đến 1, nếu dùng xác suất). Hai tỉ lệ này được ước tính cho từng giá trị tham chiếu. Phương pháp xét nghiệm tốt sẽ có những điểm tham chiếu tập trung vào khu vực “Tây Bắc”, tức là những điểm ở góc trái thuộc phía trên của biểu đồ. Những điểm này cho chúng ta biết đó là những giá trị tham chiếu có độ nhạy cao và độ dương tính giả thấp. **Biểu đồ 1** dưới đây thể hiện số liệu từ Bảng 1:



Bằng cách nối kết các điểm trên biểu đồ ROC, chúng ta sẽ có một biểu đồ ROC liên tục như trình bày trên. Nhưng ở đây chúng ta có hai chỉ số (độ dương tính giả và độ nhạy), và chúng biến thiên ngược chiều nhau. Do đó, chúng ta cần một “chỉ số dung hòa” cả hai chỉ số này. Một cách quân bình hóa tốt nhất là ước tính diện tích dưới đường biểu diễn ROC (còn gọi là *area under the curve* – AUC).

Nhìn vào biểu đồ trên, chúng ta thấy có diện tích tối đa của biểu đồ là 1 (vì biểu đồ hình vuông, với mỗi “cạnh” bằng 1). Diện tích dưới đường biểu diễn có thể tính toán bằng phương pháp tích phân, nhưng trong thực tế thì có những chương trình máy tính tính chính xác hơn.

Nếu phương pháp xét nghiệm vô dụng thì tất cả các điểm tham chiếu đều nằm trên đường thẳng nối hai điểm (0, 0) và (1, 1), tức đường 45 độ. Trong trường hợp này, diện tích dưới đường biểu diễn ROC phải bằng 0.5 (xem **Biểu đồ 2**).



**Giá trị AUC có nghĩa gì?** Không dễ diễn giải chỉ số AUC. Tuy nhiên, chúng ta có thể lấy một ví dụ để minh họa: nếu chúng ta chọn một cặp đối tượng (chọn một cách ngẫu nhiên từ một quần thể), một người có bệnh và một người không có bệnh. Nếu cả hai người đều được xét nghiệm bằng một phương pháp có AUC = 0.85, thì điều này có nghĩa là xác suất mà người có bệnh có kết quả xét nghiệm với giá trị cao hơn người không mắc bệnh là 85%.

Do đó, một phương pháp xét nghiệm có ích phải có diện tích AUC trên 0.5. Không có ngưỡng nào của AUC để xác định là một xét nghiệm tuyệt vời. Tuy nhiên theo qui ước thì một phương pháp xét nghiệm với AUC trên 0.8 được xem là tốt hay rất tốt; còn AUC dưới 0.6 được xem là không tốt và không thể áp dụng vào lâm sàng được (**Bảng 2**) [3].

<b>Bảng 2. Diễn giải ý nghĩa của diện tích dưới đường biểu diễn ROC (AUC)</b>	
<b>AUC</b>	<b>Ý nghĩa</b>
>0.90	Rất tốt (Excellent)
0.80 đến 0.90	Tốt (Good)
0.70 đến 0.80	Trung bình (Fair)
0.60 đến 0.70	Không tốt (Poor)
0.50 đến 0.60	Vô dụng (Fail)

Chỉ số AUC có rất có ích trong việc so sánh độ chính xác của 2 hay nhiều phương pháp xét nghiệm. Cổ nhiên, phương pháp nào có AUC cao hơn có nghĩa là phương pháp đó có độ chính xác cao hơn.

## Hiểu y văn

Ngày nay, có rất nhiều phương pháp xét nghiệm được phát triển và được quảng bá rầm rộ. Nhưng phần lớn những phương pháp mới được quảng bá không hẳn tốt hơn phương pháp cũ (qua đánh giá của chỉ số AUC), đó là chưa kể đến chi phí xét nghiệm của những phương pháp mới thường đắt tiền hơn những phương pháp cũ. Một đặc điểm khác là khi mới được giới thiệu, các phương pháp xét nghiệm đều có chỉ số AUC cao, nhưng khi kiểm định lại ở một hay nhiều quần thể độc lập khác thì độ chính xác không tuyệt vời như lúc ban đầu!

Tình trạng thiếu nhất quán này xảy ra vì vấn đề thiết kế nghiên cứu. Nhiều nghiên cứu tiên lượng dựa vào bệnh nhân ở bệnh viện và nhóm chứng thường là sinh viên y khoa khỏe mạnh. Đây là những đối tượng khá “tương phản” nhau về sức khỏe, cho nên kết quả thường “tốt” hơn khi xét nghiệm được áp dụng trong một quần thể đa dạng trong môi trường lâm sàng, nhất là ngoài cộng đồng.

Ngoài ra, một số người mặc dù không có bệnh nhưng có thể có vấn đề sức khỏe khác và do đó có kết quả xét nghiệm “dương tính”, và vì thế gia tăng tỉ lệ dương tính giả và giảm tỉ lệ âm tính thật. Chẳng hạn như CEA (carcinoembryonic antigen) – một xét nghiệm cho ung thư ruột – thường tăng cao ở những bệnh nhân xơ gan và viêm kết tràng; và nếu những bệnh nhân này được xét nghiệm như là nhóm chứng thì tỉ lệ dương tính giả sẽ tăng và độ đặc hiệu sẽ giảm.

Do đó, để đánh giá một phương pháp xét nghiệm mới, chúng ta cần phải trả lời cho được một số câu hỏi quan trọng như sau: Phương pháp xét nghiệm đã được kiểm định trong nhiều quần thể độc lập hay chưa; có chẩn đoán “tiêu chuẩn vàng” để làm tham khảo; phương pháp xét nghiệm đã được thẩm định trong một nhóm đối tượng thích hợp, như ngoài cộng đồng hay chưa; phương pháp được mô tả chi tiết và cụ thể; và độ nhạy, đặc hiệu, và chỉ số AUC? Chỉ khi nào các câu hỏi trên được trả lời thỏa đáng thì phương pháp xét nghiệm mới có thể ứng dụng vào lâm sàng.

## Trị số tiên đoán dương tính

Các chỉ số như độ nhạy và đặc hiệu phản ánh độ chính xác của một phương pháp xét nghiệm. Chẳng hạn như độ nhạy trả lời câu hỏi: nếu bệnh nhân mắc bệnh và nếu bệnh nhân đó đi xét nghiệm thì xác suất có kết quả dương tính là bao nhiêu. Tuy nhiên, trong thực tế, người thầy thuốc muốn có câu trả lời ngược lại: nếu một đối tượng đi xét nghiệm và có kết quả dương tính thì xác suất mà đối tượng đó mắc bệnh là bao nhiêu? Chỉ số trả lời câu hỏi này là *positive predictive value* (PPV, *trị số tiên đoán dương tính*).

Trong bài trước, chúng ta đã làm quen với PPV. Nói một cách ngắn gọn, PPV phụ thuộc vào 3 chỉ số: tỉ lệ hiện hành của bệnh (prevalence); độ nhạy; và độ đặc hiệu. Công thức của PPV là như sau:

$$PPV = \frac{prevalence \times sensitivity}{(prevalence \times sensitivity) + (1 - prevalence) \times (1 - specificity)}$$

Chẳng hạn như trong một nghiên cứu về tính chính xác của phương pháp xét nghiệm D-dimer để chẩn đoán DVT, các nhà nghiên cứu báo cáo rằng trong những người nghi có triệu chứng DVT, có 19% thật sự mắc bệnh DVT. Độ nhạy của xét nghiệm D-dimer là 97% và độ đặc hiệu là 63%. Do đó, nếu một đối tượng có kết quả xét nghiệm D-dimer dương tính, xác suất mà người đó mắc bệnh DVT là 38%:

$$PPV = \frac{0.19 \times 0.97}{(0.19 \times 0.97) + (1 - 0.19) \times (1 - 0.63)} = 0.38$$

**Một cách tính khác:** Một chỉ số rất phổ biến khác là *tỉ số khả dĩ* (likelihood ratio, LR) mà chúng ta đã làm quen trong bài trước. Tỉ số khả dĩ dương tính LR là hàm số của độ nhạy và đặc hiệu:

$$LR = \frac{sensitivity}{1 - specificity}$$

Và thay vì mô tả tỉ lệ lưu hành bằng *prevalence* chúng ta có thể mô tả bằng *odds*. *Odds* là tỉ số xác suất mắc bệnh trên xác suất không mắc bệnh:

$$odds_0 = \frac{prevalence}{1 - prevalence}$$

Thay thế LR và  $odds_0$  vào phân bên phải của công thức PPV, chúng ta có  $odds_1$ :

$$odds_1 = LR \times odds_0$$

Trong công thức này,  $odds_1$  là *odds* mắc bệnh sau khi đã biết kết quả xét nghiệm, còn  $odds_0$  là *odds* mắc bệnh trước khi biết kết quả xét nghiệm. Cố nhiên, xác suất mắc bệnh sau khi biết kết quả xét nghiệm là  $PPV = odds_1 / (1 + odds_1)$ .

## Diễn giải kết quả xét nghiệm

Trong thực tế lâm sàng, người thầy thuốc thường cần nhiều xét nghiệm, và mỗi xét nghiệm đều có độ nhạy và đặc hiệu thấp hơn 100%. Có hai loại xét nghiệm: xét nghiệm song song và xét nghiệm liên tục nhau. Xét nghiệm song song có nghĩa là nhiều xét nghiệm được tiến hành cùng một lúc. Xét nghiệm nối tiếp nhau có nghĩa là xét

nghiệm 2 theo sau kết quả xét nghiệm 1, và xét nghiệm 3 theo sau kết quả xét nghiệm 2, v.v... Vì nhiều kết quả xét nghiệm, vấn đề đặt ra là diễn dịch các kết quả xét nghiệm này như thế nào.

**Đối với xét nghiệm song song**, nếu 2 phương pháp xét nghiệm độc lập nhau (hiểu theo nghĩa kết quả của một phương pháp không có mối tương quan với phương pháp kia), thì phương pháp có độ nhạy cao nhất chính là độ nhạy của cả hai. Chẳng hạn như nếu xét nghiệm 1 có độ nhạy là 60%, và xét nghiệm 2 có độ nhạy 80%, thì độ nhạy của xét nghiệm song song là 80%. Trong thực tế, rất hiếm thấy trường hợp hai xét nghiệm hoàn toàn độc lập nhau, cho nên phương pháp ước tính *PPV* từ xét nghiệm song song cần đến mô hình thống kê như mô hình hồi qui logistic.

**Xét nghiệm nối tiếp nhau** (serial tests) thường có hiệu quả tăng độ đặc hiệu và *PPV*, nhưng giảm độ nhạy. Nói cách khác, nếu kết quả xét nghiệm theo chiến lược “nối tiếp nhau” dương tính, chúng ta tin tưởng hơn về khả năng mắc bệnh, nhưng có nguy cơ người mắc bệnh bị bỏ sót và kết quả âm tính. Xét nghiệm nối tiếp nhau có ích khi tất cả các xét nghiệm đều không có độ đặc hiệu cao.

Chiến lược xét nghiệm nối tiếp thường áp dụng khi yêu cầu xét nghiệm không quá gấp, hay có những xét nghiệm quá đắt tiền hay xâm phạm. Các xét nghiệm đắt tiền và xâm phạm chỉ thực hiện khi xét nghiệm đơn giản cho kết quả dương tính.

**Ví dụ:** Trong quần thể có nguy cơ ung thư tiền liệt tuyến, tỉ lệ mắc bệnh là 3.3%. Xét nghiệm PSA có độ nhạy là 67% và độ đặc hiệu là 97%. Xét nghiệm tiếp theo là DRE (dùng ngón tay) có độ nhạy là 50% và độ đặc hiệu là 94%. Cách ước tính *PPV* cho chiến lược xét nghiệm liên tục như sau:

**Bước 1: tính PPV cho phương pháp PSA**

- Odds mắc bệnh trước khi xét nghiệm:  
 $odds_0 = 0.033 / (1 - 0.033) = 0.034$
- *LR* cho phương pháp PSA  
 $LR = 0.67 / (1 - 0.97) = 22.33$
- Odds mắc bệnh sau khi đã có xét nghiệm dương tính PSA:  
 $odds_1 = 0.034 \times 22.33 = 0.762$
- Xác suất mắc bệnh sau khi đã có xét nghiệm dương tính PSA:  
 $PPV = 0.762 / (1 + 0.762) = 0.433$

**Bước 2: tính PPV cho phương pháp DRE**

- Odds mắc bệnh sau khi có xét nghiệm PSA:  
 $odds_0 = 0.433 / (1 - 0.433) = 0.762$
- *LR* cho phương pháp DRE:  
 $LR = 0.50 / (1 - 0.94) = 8.33$
- Odds mắc bệnh sau khi đã có xét nghiệm dương tính PSA:  
 $odds_1 = 0.762 \times 8.33 = 6.35$
- Xác suất mắc bệnh sau khi đã có xét nghiệm dương tính PSA:



$$PPV = 6.35 / (1 + 6.35) = 0.86$$

Nói cách khác, trước khi có bất cứ kết quả nào, xác suất mắc bệnh ung thư tiền liệt tuyến là 3.3%, nhưng sau khi kết quả xét nghiệm PSA dương tính, xác suất tăng lên 43.3%. Sau khi đã có kết quả dương tính từ DRE, xác suất mà bệnh nhân mắc bệnh tăng lên 86%.

## Cỡ mẫu

Các nghiên cứu tiên lượng và chẩn đoán cần phải được thiết kế theo các chuẩn mực khoa học. Trong các chuẩn mực này có việc xác định cỡ mẫu cần thiết cho một nghiên cứu có ý nghĩa. Thông thường các nghiên cứu về xét nghiệm được thiết kế theo mô hình nghiên cứu bệnh chứng (case – control study), mà theo đó một nhóm bệnh nhân đã được xác định có bệnh theo một phương pháp chẩn đoán chuẩn (gold standard), và một nhóm đối tượng không có bệnh.

Công thức xác định cỡ mẫu cần thiết phụ thuộc vào 3 thông số: độ nhạy hay độ đặc hiệu cần ước tính; khoảng tin cậy của độ nhạy hay độ đặc hiệu; và thông số z từ luật phân phối chuẩn. Gọi độ nhạy hay độ đặc hiệu là  $\pi$ , khoảng rộng của tin cậy là  $w$ , và hệ số khoảng tin cậy từ luật phân phối chuẩn là  $z$ , công thức ước tính cỡ mẫu là:

$$N = \frac{4\pi(1-\pi)z^2}{w^2}$$

Chẳng hạn như nếu chúng ta muốn nghiên cứu về một phương pháp xét nghiệm mà chúng ta nghĩ rằng độ nhạy là 80% ( $\pi = 0.80$ ), và khoảng tin cậy 95% của thông số này dao động trong khoảng 0.75 đến 0.85 (tức  $w = 0.10$ ), với xác suất 95% (tức  $z = 1.96$ ). Số cỡ mẫu cần thiết là 246 người:

$$N = \frac{4 \times 0.8 \times (1 - 0.8) \times (1.96)^2}{(0.1)^2} = 246$$

Nhưng chúng ta cũng cần ước tính độ đặc hiệu, và giả dụ chúng ta cũng cần có độ đặc hiệu là 0.80, thì số cỡ mẫu là 246. Tổng cộng chúng ta cần 492 đối tượng (246 bệnh nhân và 246 người không mắc bệnh) để nghiên cứu.

**Bảng 3** dưới đây trình bày số cỡ mẫu cần thiết cho một thông số theo từng  $w$  khác nhau, với khoảng tin cậy 95%.

<b>Bảng 3. Cỡ mẫu cần thiết để ước tính độ nhạy hoặc độ đặc hiệu (<math>\pi</math>)</b>							
$\pi$	<b>Độ rộng (width) của khoảng tin cậy</b>						
	<b>0.05</b>	<b>0.075</b>	<b>0.100</b>	<b>0.125</b>	<b>0.150</b>	<b>0.175</b>	<b>0.200</b>
0.70	1291	574	323	207	143	105	81
0.75	1152	512	288	184	128	94	72
0.80	983	437	246	157	109	80	61

0.85	784	348	196	125	87	64	49
0.90	553	246	138	89	61	45	35
0.95	292	130	73				

Nói chung, cỡ mẫu cần thiết cho một nghiên cứu đánh giá một phương pháp xét nghiệm càng cao nếu độ nhạy và đặc hiệu cao, và ở mỗi độ nhạy và đặc hiệu khi yêu cầu chính xác cao (thể hiện qua độ rộng của khoảng tin cậy hẹp).

## Kết luận

Tiên lượng và chẩn đoán đóng vai trò rất quan trọng trong thực hành lâm sàng, nhưng điều khó khăn là vấn đề diễn giải kết quả nghiên cứu và áp dụng cho một bệnh nhân. Vấn đề càng khó khăn hơn đối với các phương pháp xét nghiệm có kết quả là một biến liên tục (thay vì một biến nhị phân). Đối với các xét nghiệm có nhiều giá trị, phương pháp đánh giá độ chính xác của xét nghiệm là dựa vào sự quân bình giữa độ nhạy và độ đặc hiệu cho từng giá trị tham chiếu, và thể hiện bằng chỉ số AUC. Tuy AUC (diện tích dưới đường biểu diễn ROC) là một chỉ số quan trọng và có ích cho việc đánh giá sự hữu hiệu của một xét nghiệm, nhưng phần lớn diện tích này xuất phát từ các giá trị dương tính giả.

Như đề cập, AUC là một chỉ số có thể sử dụng để so sánh các phương pháp xét nghiệm. Nhưng vì một phần lớn của AUC chịu sự chi phối của tỉ lệ dương tính giả, cho nên nói chung AUC rất thiếu tính nhạy. Chẳng hạn như nếu xét nghiệm thứ nhất có AUC là 0.80, và nếu thêm xét nghiệm hai AUC tăng lên 0.82, thì điều đó không có nghĩa là xét nghiệm hai góp phần khiêm tốn vào tiên lượng, mà vì chỉ số AUC thay đổi rất nhỏ. Vài năm gần đây, các nhà nghiên cứu đã phát triển một số phương pháp phân tích mới để đánh giá các xét nghiệm mới so với các xét nghiệm chuẩn, và các phương pháp phân tích mới này đang dần dần được áp dụng rộng rãi trong y văn, nên chúng ta có lẽ sẽ quay lại một bài sau [4,5].

Cần ghi nhận rằng các chỉ số như độ nhạy, đặc hiệu, hay AUC phản ánh độ chính xác của phương pháp xét nghiệm; chúng không cung cấp xác suất mắc bệnh cho một cá nhân. Để ước tính xác suất này chúng ta cần một thông số quan trọng: đó là tỉ lệ lưu hành (prevalence) hoặc tỉ lệ phát sinh (incidence) của bệnh. Thông số này cần kết hợp với các chỉ số về độ chính xác của phương pháp xét nghiệm để cung cấp một chẩn đoán hay tiên lượng cho bệnh nhân. Nghệ thuật xác suất trong y khoa chính là ở khía cạnh áp dụng nhuần nhuyễn các chỉ số này nhằm nâng cao tính chính xác của chẩn đoán và tiên lượng.

## Tài liệu tham khảo và đọc thêm

[1] Morgan TO, et al. Age-specific reference ranges for serum prostate specific antigen in black men. *N Engl J Med* 1996;335:304-310

[2] Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982 Apr;143(1):29-36.

[3] Pepe MS. The statistical evaluation of medical tests for classification and prediction. Oxford University Press 2004.

[4] Michael J. Pencina, Ralph B. D' Agostino Sr, Ralph B. D' Agostino Jr, Ramachandran S. Vasan. Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. Stat Med 2008; 27:157-316.

[5] Cook NR, Ridker PM. Advances in Measuring the Effect of Individual Predictors of Cardiovascular Risk: The Role of Reclassification Measures. Ann Intern Med 2009; 150:11795-802.