

第三章 诊断试验

彭继茂（葛勤敏 译）

一、诊断不确定性

我们总是面临着诊断的不确定性。通过诊断试验进行决策是临床医师、尤其是面对危重患者的一项重要任务。举个例子，一位男性凌晨 2 点胸痛。心脏导管介入术作为判断是否存在冠状动脉阻塞是金标准，仍然是目前最好的选择，但是它是侵入性的，可能心脏病专家和实验室无法获得。但分秒必争。此时我们在诊断不确定的状况下，需要一个可靠的诊断试验来明确诊断或排除心肌梗塞。

诊断试验研究不同于「临床预测规则」研究。临床预测规则使用规则来预测结果，而不是诊断结果。对于诊断试验研究，试验结果可能是二分法，也可能是多层次的（连续的）。对于二分法检查，如床边超声诊断肺炎，我们可以建立一个计算表。敏感性和特异性只有一套，对于连续的数据，如胰腺炎的淀粉酶水平和心肌梗死的肌钙蛋白水平，我们可以进行多层面切割，构造一个 ROC 曲线。

二、临床指向性

评价诊断试验始之「人群」P，然后「诊断试验」（即暴露 E），然后得出「结果」O（用于诊断的疾病） \Rightarrow P E O。

三、有效性

（一）金标准：可被接受、独立于诊断试验并在所有受试者中完成

为了检测试验的诊断准确度，我们需要与金标准比较，金标准用来确认是否存在疾病，理论上是 100%敏感度和 100%特异度。金标准对有否疾病进行判断，应独立于诊断试验。可能需要盲法避免偏倚。所有受试者必须通过金标准进行验证。

（二）偏倚风险：「疾病谱偏倚」和「验证偏倚」

1. 疾病谱偏倚：一个无法识别（但可能非常真实）的问题是疾病谱偏倚。这是一个试验的敏感度和/或特异度的现象，随受试人群的性别比、年龄或疾病严重程度不同而变化。当然如果分析恰当，亚组异质性并不是一种偏倚。敏感度、特异度和似然比可以根据亚组分层。我们应该解决分析和应用中的异质性。

若纳入无病和严重疾病的受试者，我们会期望诊断试验的高准确度，这是疾病谱偏倚。如果存在疾病谱偏倚，敏感度和特异度结果可能不稳定。因此，我们应该将一系列的受试者，即没有疾病的受试者，以及轻度、中度和重度疾病的受试者放在一起。

2. 验证偏倚：当检测结果为阴性的患者未用金标准试验进行评估时，会出现验证偏倚。选择性研究包括被纳入通过金标准试验确认疾病的患者、只是初步试验或诊断试验本身阳性结果的患者。如将冠状动脉造影应用于肌钙蛋白水平升高的受试者，但不包括其他受试者，那就是验证偏倚。为了避免这种情况的发生，研究应该包括对某一特定疾病的连续风险所有患者，而不仅仅是接受确认试验的一个亚组。可以说，所有的实验对象都应该经过金标准的验证。

诊断试验也应该独立于金标准，反之亦然。以心肌梗死为例，根据世界卫生组织（WHO）标准，受试者出现 1. 心肌酶升高、2. 胸痛和 3. 心电图异常，以上任何 2 种同时发生，可作为诊断心肌梗死的金标准。当我们选择肌钙蛋白作为诊断试验，以及金标准诊断之一时，可能会出现问题。

四、结果

使用 2×2 表的二分数据：

	疾病/金标准 合计	疾病/金标准 合计
诊断试验 +	a (TP) a+b	b (FP)
诊断试验 -	c (FN) c+d	d (TN)
合计	a+c a+b+c+d	b+d

a: 真阳性 (TP) ; a+b: 试验阳性

b: 假阳性 (FP) ; c+d: 试验阴性

c: 假阴性 (FN) ; a+c: 有病

d: 真阴性 (TN) ; b+d: 无病

患病率：具有特定条件或特征的人口比例：例如某一城市中患有特定疾病的人口比例 = $(a+c) / (a+b+c+d)$ 。你可以称之为“验前概率”。

准确度 = $(a+d) / (a+b+c+d) = (TP \text{ 和 } TN) / \text{总体}$

敏感度 = 真阳性率 (TPR) : 所有阳性检测结果中患该疾病的比例 = $a / (a+c)$

特异度 = 真阴性率 (TNR) : 无病患者阴性比例 = $d / (b+d)$

SnNout 对于一个敏感度极高的试验, 假阴性便趋向于零。当检测结果为阴性时, 可有效做出排除该疾病的诊断。我们需要高敏感度试验对危险的疾病进行筛选, 如宫颈癌筛查的巴氏涂片。也就是说, 子宫颈涂片结果正常, 表示患者极不可能患上宫颈癌。

SpPin 同样, 对于一个特异度极高的试验, 假阳性倾向于零。当检测结果为阳性时, 我们基本可以确认疾病。当治疗疾病潜在危险时 (如剖腹产、长期抗凝) SpPin 很重要, 避免检测结果假阳性而导致伤害。也就是说, 结果阳性, 表示患者极不可能被误诊。

阳性预测值 (PPV) = $a / (a+b)$ 是指患者诊断试验阳性时, 患疾病的可能性。鉴于阳性结果, 我们想知道我们的病人有多大可能患上这种疾病。然而, PPV 随「患病率」变化 (敏感度和特异度不变情况之下)。

阴性预测值 (NPV) = $d / (c+d)$ 是指如果诊断试验为阴性, 患者无疾病的概率。鉴于阴性结果, 我们想知道我们的病人有多大可能没有这种疾病。PPV 和 NPV 受「患病率」影响, 即 PPV 和 NPV 随不同情景而变化 (如心脏 ICU 的 PPV 高于急诊室, 因为心脏 ICU 的心肌梗死患病率较高)。

阳性似然比 (PLR) : TPR / FPR

= (试验结果为阳性的受试者以及有疾病的受试者) / (试验结果为阳性但无疾病的受试者)

PLR = 真阳性率 / 假阳性率 = 敏感度 / (1 - 特异度)

PLR 值 10 意味着如果患者诊断试验阳性, 他患病的可能性是没有此病的 10 倍。

阴性似然比 (NLR) : FNR / TNR

NLR = 假阴性率 / 真阴性率 = (1 - 敏感度) / 特异度

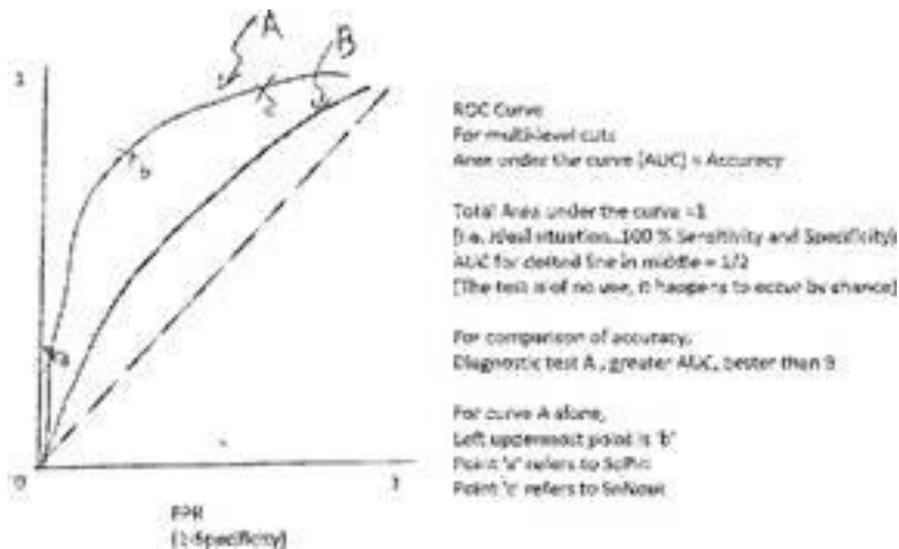
NLR 为 0.1 意味着如果病人诊断试验阴性, 他不患病的可能性是患病可能性的 10 倍。

PLR>10 或 NLR<0.1，对验前概率到验后概率产生很大的决定性变化；

PLR 5-10 和 NLR 0.1-0.2 在验前到验后的概率中产生适度的变化；PLR 2-5 & NLR 0.5-0.2 在概率方面产生小的（但有时很重要）变化；PLR 1-2 & NLR 0.5-1 概率的影响几乎可以忽略（很少重要）的水平。

ROC 曲线

我们之前提到了二分类诊断程序，只有一组敏感度和特异度数据。然而，试验可有数值尺度上的测量值。例如，肌钙蛋白水平是诊断心肌梗死的连续性指标。当试验连续性检测时，敏感度和特异度水平取决于在阳性和阴性结果之间设置的截止点。ROC 曲线是一个显示多层次切割关系的绘制图。ROC 曲线是通过在各种阈值设置下绘制真阳性率（TPR）与假阳性率（FPR）的对比来创建的。真阳性率也称为敏感度。假阳性率也称为（1-特异度）。敏感度和特异度之间存在权衡。敏感性与特异性呈负相关。当敏感度升高时，特异性下降。



Y 轴：真阳性率=TPR=敏感度

X 轴：假阳性率=FPR=（1-特异度）

斜率=敏感度/（1-特异度）=TPR/FPR=阳性似然比

曲线下面积（AUC）=诊断准确度

AUC 最大值=1

接近 1=优秀

0.8-0.9=相当不错

0.7-0.8=中等好

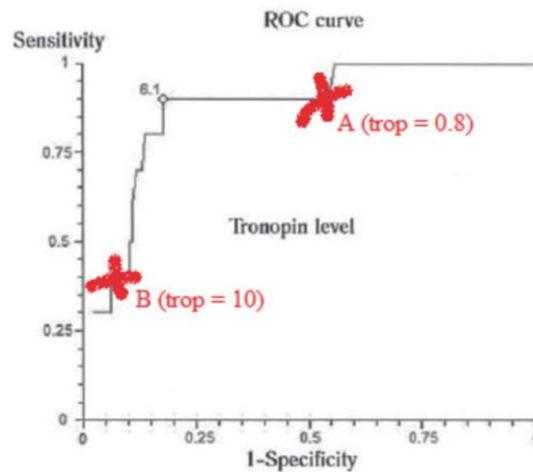
0.6-0.7=略微有用

0.5 或以下= 作为诊断试验无用，与掷硬币相同

对角线：对应于一个偶然出现阳性或阴性结果的试验。意味着 AUC 是 0.5，即一个无用的试验。最佳点是指敏感度和特异性最强化的点。最佳点是最靠近 (0, 1) 点的左上角（最西北点）。ROC 曲线还可以区分两个诊断试验之间的效能差异。上曲线具有较大的 AUC，因此准确度较高。

疾病筛查和确认的临界点

对于筛查试验，截断点往往更宽，敏感度增加，特异度降低，假阴性最小化。对于确认试验，临界点变得更为严格。特异度增加，敏感度降低，假阳性结果减少。例如，下图表示肌钙蛋白诊断 AMI 的 ROC 曲线（仅供说明）：



当截断标准更为宽松时（A 点）

- 敏感度高，
- 特异度低
- 假阴性率低
- SnNout 诊断试验为阴性时，可排除疾病

当截断标准更严格时（B 点）

- 特异度高，
- 敏感度低
- SpPin 诊断试验阳性时可确认存在疾病

概率与赔率的关系

如何从 概率 p 中得到 赔率 odds

	probability	odds
	1/4	1/3

取 概率= P , 则 赔率= $p / (1-p)$
odds= $p / (1-p)$

从上表中, 阴影框的概率为 $1/4$ 。
得出无阴影的盒子的概率是 $1-1/4=3/4$

有阴影框的赔率是 $p / (1-p) = 1/3$

再举一个例子。掷骰子后得到“1”的概率= $1/6$ 。
获得“1”的赔率= $(1/6) / (1-1/6) = 1/5$

如何通过赔率得到概率

进一步解方程,

$$\text{赔率} = p / (1-p)$$

$$\text{赔率} \times (1-p) = p$$

$$\text{赔率} - \text{赔率} \times p = p$$

$$\text{赔率} = p + \text{赔率} \times p$$

$$\text{赔率} = p (1 + \text{赔率})$$

$$\text{赔率} / (1 + \text{赔率}) = p$$

因此, 我们得到 概率 (p) = 赔率 / (1+赔率)

$$p = \text{odds} / (1 + \text{odds})$$

总结, 我们共有 2 个方程式

$$\text{odds} = p / (1-p) \quad p = \text{odds} / (1 + \text{odds})$$

应用似然比预测验后概率:

1 通过使用验前赔率和验后赔率计算验前概率:

验前概率

=患者诊断试验前患该疾病的临床可能性。

=患病率：在特定时间受目标疾病影响的人群比例。

验前赔率

=诊断试验前患者患待诊断疾病的赔率

=验前概率 / (1-验前概率)

验后赔率

=诊断试验后患者有待诊断疾病的赔率

=验前赔率 x 似然比

验后概率

=诊断试验后，试验结果中患该疾病个体的比例，取决于试验结果=验后赔率 / (验后赔率+1)

进一步的阐述，比如诊断试验前的验前概率是 20%，PLR 是 10

步骤 1：估计或从健康数据中获得验前概率...假设为 20%

第二步：将验前概率转化为验前赔率

验前赔率= $p / (1-p) = 20/80 = 0.25$

第三步：将验前赔率乘以 PLR 得到验后赔率

验后赔率=验前赔率 x PLR= $0.25 \times 10 = 2.5$

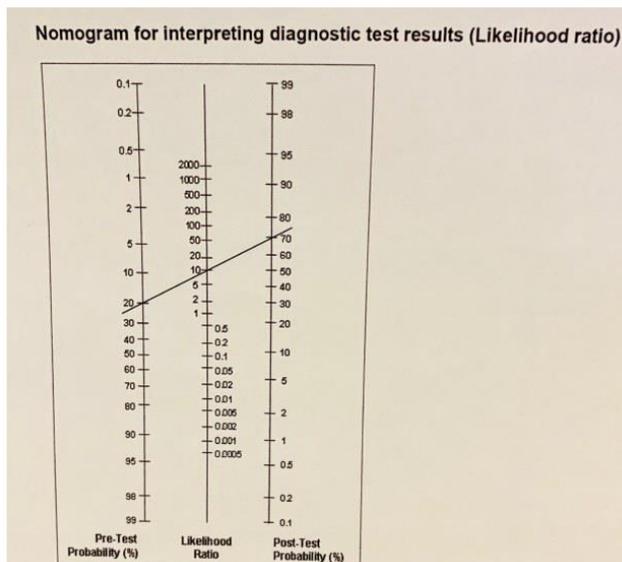
步骤 4：将验后赔率转换回验后概率（百分比）

验后概率=赔率 / (1+赔率) = $2.5 / 3.5 \times 100 = 70\%$

2 Fagan's 列线图（似然比列线图）

列线图可以简化验后概率的计算。如果我们知道患病率（验前概率）和诊断试验的似然比，我们从左侧画一条直线（验前概率）到试验的 LR；连线所指向的右侧即验后概率。

Fagan's 列线图：举例



似然比优点：

- 不受试验前概率（疾病患病率）影响
- 能够直接从验前概率得到验后概率
- 可应用于多水平的诊断试验结果，直接获得似然比

试验特性说明：

- (1) PPV 和 NPV 随不同疾病患病率而变化
- (2) 当存在疾病谱偏倚时，敏感度和特异度变得不稳定。
- (3) 当受试者的诊断试验结果为阳性时，与阳性预测值比较而言，验后概率更能估计患病的可能性。

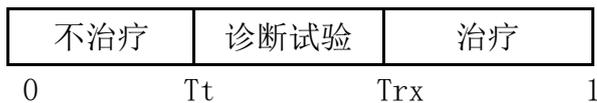
五、适用性

1. 5 Killer B's：生物学：SCRAP，代表性别、合并症、种族、年龄和病理学。权衡：在临床利益和伤害/副作用之间进行权衡；屏障：您所在地的诊断试验可用性/易获得性；信念：患者自己的信念/考虑；负担：患病率和涉及的经济负担。

2. 阈值模型：诊断阈值 (Test threshold, T_t) 和治疗阈值 (Treatment threshold, T_r)

医学是一门艺术。当面对诊断不确定时，我们可能对自己的患者发生疾病的可能性有自己的判断。我们可以将疾病的概率估计为一条从 0 延伸到 1 的线。诊断阈值 T_t 是概率线上的点，在该点上不治疗患者（不治疗）的价值和进行诊断试验（试验）的价值之间无差异。同样，治疗阈值在该点上，对患者进行诊断试验（试验）和不做诊断试验直接治疗（治疗）之间的价值无差异。

由于我们有自己的临床判断，所以我们之间的 T_t 和 T_r 是不同的。如果该病发生的概率太低 ($<T_t$)，临床医生不会行诊断试验来排除诊断或开始治疗。如果疾病的概率太高 ($>T_r$)，临床医生将不会行诊断试验来确认诊断，而是直接进行治疗。如果概率介于 T_t 和 T_r 之间，临床医师需要进行诊断试验来确认或排除疾病诊断。



影响治疗阈值的因素：

- ①疾病本身
- ②治疗费用
- ③治疗引发不良反应

影响试验阈值的因素：

- (1) 疾病本身
- (2) 试验费用
- (3) 试验侵入性

如果从验前概率到验后概率的结果超过其中一个阈值，则诊断试验非常有用。

3. 利用当地验前概率对结果进行个体化判断

疾病的当地患病率是验前概率，可通过当地的健康数据获得可参考价值。利用患者的验前概率（基于临床表现或政府数据）和诊断试验文章中的似然比，我们可以通过 Fagan 列线图（似然比列线图）从验前概率估计患者个体化的验后概率。

本文编辑：陈芳