第三章

诊断试验评价与疾病诊断

本章要点

- 诊断试验(diagnostic test)是对疾病进行诊断的方法，包括病史和体检所获得的临床资料，各种实验室和辅助检查，以及各种临床公认的诊断标准等。如何合理地选择与安排诊断试验需要对各种试验进行科学的评价以便决策。
- 诊断试验文献科学性的评价可以从金标准的选择、待评价试验是否与金标准进行独立盲法比较、研究人群是否包括临床上应用该试验的各种患者，以及诊断试验方法的描述是否足够详细、具有重复性几个方面来考察。
- 诊断试验的评价指标有很多，包括灵敏度、特异度、阳性预测值、阴性预测值、似然比，和受试者工作曲线下面积等，各有其临床应用价值。
- 联合诊断试验包括平行试验和系列试验，可有效提高灵敏度和特异度，常在临床上根据具体需要而采用。

教学目的

- **掌握**：诊断试验文献科学性评价的方法与标准。
- **熟悉**：诊断试验的评价指标与应用。
- **了解**：提高诊断试验效率的方法。

第一节　诊断试验的范畴与评价意义

临床医师每天接触大量的患者，临床需要解决的首要问题就是明确患者得的是什么疾病，疾病程度范围如何，即诊断问题。在临床诊断中，需要有各种诊断依据，包括症状、体征、既往病史、实验室检查或者特殊检查的结果等。广义地说，这些都属于诊断试验的范畴。

诊断试验(diagnostic test)是对疾病进行诊断的方法，它既包括病史和体检所获得的信息，血液和体液生化、血液细胞学、病原学、免疫学、病理学等实验室检查，以及X线、超声、CT、核磁共振成像(MRI)及放射性核素等影像学检查，还有诸如心电图、内镜、电生理等一些特殊检查，也包括一些复合型的各种临床的诊断标准，如Behcet病及系统性红斑狼疮的诊断标准等。

对诊断试验进行评价，是临床诊治决策的需要。诊断试验首先要有真正的区别疾病的能力，且该能力必须符合临床的需求，还要安全、经济，为人接受。如果不进行评价即随意使用诊断试验，导致其滥用或误用，不但浪费资源，还会误导临床诊疗，给患者带来伤害。

第二节 诊断试验的评价指标及其应用

循证医学是遵循证据的临床医学。在疾病的诊断中，医生也应该通过多种途径，特别是借助现代信息工具获取最新、最全面的关于疾病诊断手段的信息，利用临床流行病学的方法评价各种诊断试验的证据，全面评价后选择合适的诊断性试验提供临床应用，使患者能及早得到正确的诊断，以便接受合理的治疗。

诊断试验可以从很多方面进行评价，如诊断试验对疾病的区分能力，安全性，经济指标等。但作为诊断试验本身而言，其诊断能力，或区分能力（即通过该试验将有病和无病的人区分开来，或是将不同疾病阶段的人区分开来等）是最重要的一个方面。和很多其他医学研究一样，要评价一个诊断试验的表现必须要有一个参考标准，即所谓的“金标准（gold standard）”。金标准是当前临床医学界公认的诊断该病最可靠的诊断方法。常用的金标准有：病理学标准（组织活检和尸体解剖）、外科手术发现、特殊的影像诊断（肺梗死时的肺血管造影）、长期临床随访结果、公认的综合临床诊断标准（如 ARA 诊断标准）等。通过与金标准的同步比较，就能获取一个诊断试验的诊断能力。

表 2-3-1 是经典的诊断试验评价四格表。每一位受试者均接受新诊断试验和金标准检查，诊断试验结果有阳性和阴性两种，而金标准将所有人分为病例组和非病例组。其中 a 为诊断试验结果为阳性的患者，即真阳性；b 为诊断试验结果为阳性、而金标准判定为非病例者，即假阳性（误诊）；c 为诊断试验结果为阴性，但金标准判断为病例者，即假阴性（漏诊）；d 则为诊断试验与金标准均判断为非病例者，即真阴性。在四格表中，a、b、c、d 均为各组的人数，其总人数为 N。通过这个四格表，可以得到诊断试验准确性（区分能力）的众多指标，如灵敏度、特异度、阳性和阴性预测值、似然比等。下文分别介绍一些指标的计算及其意义。

表 2-3-1 诊断试验评价四格表

		金标准		
		病例组	非病例组	合 计
新诊断试验	+	a 真阳性	b 假阳性	a+b
	−	c 假阴性	d 真阴性	c+d
合 计		a+c	b+d	N

注：a：真阳性，为病例组内试验阳性的例数；b：假阳性，为对照组内试验阳性的例数；c：假阴性，为病例组内试验阴性的例数 d：真阴性，为对照组内试验阴性的例数；N 总人数。敏感度＝a/(a+c)；特异度＝d/(b+d)；阳性预测值＝a/(a+b)；阴性预测值＝d/(c+d)；准确度＝(a+d)/N；诊断比值比＝ad/bc；患病率＝(a+c)/N；阳性结果似然比＝[a/(a+c)]/[b/(b+d)]＝敏感度/(1−特异度)；阴性结果似然比＝[c/(a+c)]/[d/(b+d)]＝(1−敏感度)/特异度。

一、灵敏度和特异度

灵敏度（sensitvity）是指由金标准诊断方法确诊有病的人群（病例组）中经诊断试验查

出阳性结果人数的比例[a/(a+c)]，而病例组中诊断试验未查出即结果阴性的人数比例[c/(a+c)]称假阴性率，又称漏诊率，等于1－灵敏度。

特异度(specificity)是指由金标准诊断方法确诊无病的人群(对照组)中经诊断试验检出阴性结果人数的比例[d/(b+d)]，而对照组中查出阳性的结果人数的比例[b/(b+d)]称假阳性率，又称误诊率，等于1－特异度。

灵敏度和特异度是诊断试验诊断特性的两方面。高灵敏度的试验不易漏诊，有利于排除诊断(可记作 SnNout，即高灵敏度试验阴性结果可排除诊断，highly SeNsitive test when Negative rules OUT)，而高特异度的试验不易误诊，有利于肯定诊断(可记作 SpPin，即高特异度试验阳性结果可确立诊断 highly SPecific when Positive rules IN)。在医生需要排除或肯定诊断时，可以分别选用高灵敏度或高特异度试验。很多时候，人们希望有同时具备高灵敏度和高特异度的诊断试验，然而这样"好"的诊断试验不多。通常情况下鱼与熊掌不可兼得，高灵敏度时常常有更多的假阳性，高特异度时常常有更多的假阴性。临床具体应用时，要根据实际情况权衡假阳性和假阴性可能造成的后果，考虑高灵敏度(如果疾病漏诊可能带来很大的危害)，或者高特异度(如果疾病误诊可能给患者或者家庭带来很大的经济或者精神负担)可能带来的好处。

当试验方法和阳性结果标准固定时每个诊断试验的灵敏度和特异度是恒定的。区分诊断试验正常和异常的临界点会影响灵敏度和特异度。临界点变动时，灵敏度和特异度往往呈反向变化(图 2－3－1，表 2－3－2)。

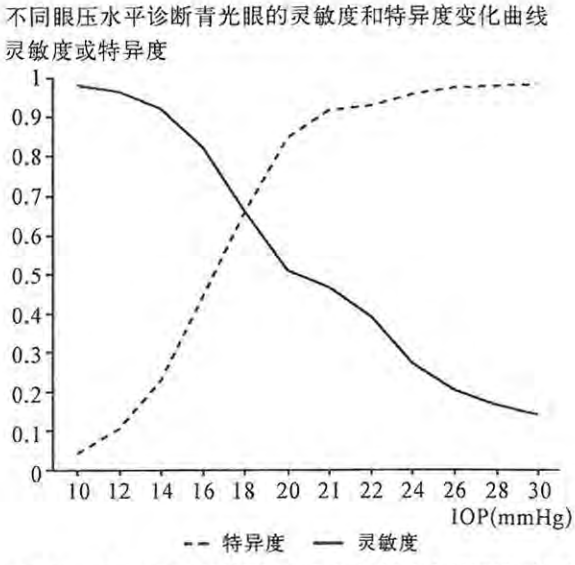

图 2－3－1 不同眼压(IOP)水平作为青光眼诊断阈值时灵敏度、特异度的变化(Baltimore Eye Survey，1991)

表 2－3－2 老年人中不同的眼压水平作为青光眼诊断的临界值时的灵敏度和特异度(引自 Baltimore Eye Survey，Tielsch JM，et al.，1991)

眼压临界值(mmHg)	灵敏度(%)	特异度(%)
≥16	83	40
≥18	67	63

（续表）

眼压临界值(mmHg)	灵敏度(%)	特异度(%)
≥20	52	81
≥21	47	91
≥22	40	92
≥24	26	96

二、阳性预测值与阴性预测值

灵敏度和特异度是诊断试验本身的特性，只考虑到病例组或非病例组诊断试验结果的阳性率与阴性率的情况。然而在临床实践中，医师更关心的是当某诊断试验是阳性结果时，就诊者患病的可能性有多少，阴性结果时不患有该病的可能性有多少，这就是预测值(predictive value，PV)。阳性预测值(positive PV，+PV)是指试验阳性结果中真正患病的比例[a/(a+b)]，阴性预测值(negative PV，−PV)是指试验阴性结果中真正未患病的比例[d/(c+d)]。

预测值与试验的灵敏度、特异度和患病率均有关。通过公式推导不难得出阳性预测值与灵敏度、特异度和患病率的关系。一般说来，越是灵敏的试验(高灵敏度)，阴性预测值越高；反之特异性越高的试验，阳性预测值越高。值得注意的是患病率[(a+c)/N]对预测值的影响要比灵敏度和特异度更为重要。例如，磷酸肌酸激酶(CPK)测定作为心肌梗死诊断试验，灵敏度 93%，特异度 88%。如果将该试验用于监护病房(心肌梗死的患病率为 64%)，CPK 阳性(80 IU 以上)者患有心肌梗死的概率为 93%。但如果用于普通病房(心肌梗死的患病率 10%)，试验的灵敏度与特异度不变，CPK 阳性者患有心肌梗死的概率只有 56%，也就是猜测 CPK 阳性者是否为心肌梗死患者就跟抛硬币决定差不多(表 2-3-3)。

表 2-3-3 CPK 诊断心肌梗死在不同患病率下的预测值

	监护病房(患病率 64%)			普通病房(患病率 10%)		
	心肌梗死	无心肌梗死	总 计	心肌梗死	无心肌梗死	总 计
试验(+)	215(a)	16(b)	231(a+b)	34(a)	27(b)	61(a+b)
试验(−)	15(c)	114(d)	129(c+d)	2(c)	197(d)	199(c+d)
总 计	230(a+c)	130(b+d)	360(N)	36(a+c)	224(b+d)	360(N)
+PV		93%			56%	
−PV		88%			99%	

注：CPK 诊断临界点 80 IU，灵敏度 93%，特异度 88%。+PV：阳性预测值，−PV：阴性预测值。

在本例中，CPK 诊断心肌梗死的灵敏度和特异度没有改变，但患病率下降后，阳性预测值明显下降，而阴性预测值有所上升。由此可见，患病率对于预测值有非常重要的影响。因此在评价文献报道的诊断试验时，应考虑研究中受试人群的患病率是否与本单位情况相同；一项在三级医院阳性预测值很高的试验，在一级医院或者在社区筛查时可能就很低。患病率在不同临床情况相差甚大。一般人群普查时，即使试验的特异度很高，当用于患病率很低的人群时，仍会出现大量假阳性结果；同样，一种灵敏度非常高的试验，当用于患病率很高的人群，仍会出现较多假阴性结果。所以应用一项诊断试验时，必须要考虑其目标应用人群的患病率。

三、阳性结果似然比和阴性结果似然比

诊断试验的灵敏度与特异度分别从两个方面反映了患患者群和不患该病的对照人群试验结果的信息，不能仅根据一个指标来评价诊断试验及估计疾病概率；而且诊断试验结果为计量资料时，诊断临界点的划分会影响灵敏度与特异度。预测值尽管为临床诊断提供了很好的信息，但受患病率影响明显，因而不能用于评价诊断试验。

似然比(likelihood ratio，LR)是可以同时反应灵敏度和特异度的复合指标，是有病者得出某一试验结果的概率与无病者得出这一结果的概率的比值。当试验结果只有阴性和阳性两种结果时，似然比分为阳性试验结果似然比和阴性试验结果似然比：

$$\text{阳性似然比}(+\text{LR})=\frac{\text{真阳性率}[a/(a+c)]}{\text{假阳性率}[b/(b+d)]}=\frac{\text{灵敏度}}{1-\text{特异度}}$$

$$\text{阴性似然比}(-\text{LR})=\frac{\text{假阴性率}[c/(a+c)]}{\text{真阴性率}[d/(b+d)]}=\frac{1-\text{灵敏度}}{\text{特异度}}$$

与灵敏度和特异度不一样，应用似然比可以避免将计量试验结果简单地划分为正常和异常，从而可以全面反映诊断试验的诊断价值。此外，似然比灵敏度和特异度更稳定，更不受患病率的影响。根据表 2-3-3，可以计算 CPK 诊断心肌梗死的似然比。阳性似然比为(215/230)/(16/130)=7.6，也就是说 CPK 阳性则该患者心肌梗死的可能性为非心肌梗死的 7.6 倍；CPK 阴性结果的似然比为(15/230)/(114/130)=0.07，也就是说 CPK 阴性时该患者心肌梗死的可能性不到非心肌梗死的 1/10。

似然比不仅能更好地评价诊断试验，更重要的用途在于估计疾病的患病概率。

疾病诊断与鉴别诊断的过程实质是肯定疾病与排除疾病的过程，也是对患病可能性大小的判断。人群患病的基础概率是人群的患病率(prevalence)，与地区、年龄、性别等一般资料有关，可以通过查阅特定疾病的人群患病率确定。通过病史询问与体格检查，临床医师通常对患者的患病概率有重新判断，然后决定选择进一步检查即诊断试验。

验前概率(pre-test probability)是指患者在做某项诊断试验前的可能的患病概率。验前概率多根据流行病学资料、其他人的报告，或根据患者的病史、体格检查和医生在临床实践中遇到此类患者的概率来估计的。

似然比是评价诊断试验价值的有效指标，其重要应用就是试验的结果使验前概率提高或降低的多少，根据试验前患者的患病率(验前概率)和做了某项试验后得出的结果的似然比(根据结果为阳性或者阴性分别选择阳性似然比或者阴性似然比)，可以得出验后概率(post-test probability)。验后概率是患者得到某个诊断试验结果后可能的患病概率。计算中，概率必须先化成比数(odds)后才能与似然比相乘，而相乘后得出的验后比，再转变为概率，即验后概率。

$$\text{验前比}=\text{验前概率}/(1-\text{验前概率})$$

$$\text{验后比}=\text{验前比}\times\text{似然比}$$

$$\text{验后概率}=\text{验后比}/(1+\text{验后比})$$

例如医生判断某名就诊者的患病可能性(验前概率)为 50%，验前比=50%/(1-

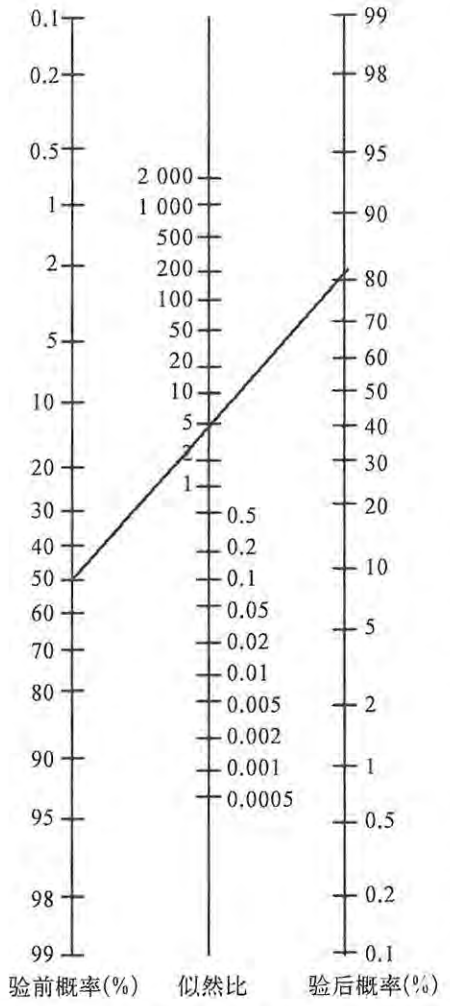

图 2-3-2 验前概率、似然比和验后概率的列线图(引自 Fagan TJ, 1975)

50%)=1;做某项阳性似然比为 5 的诊断试验后得出阳性结果,则验后比为 1×5=5,验后概率为 5/(1+5)=83%。该诊断试验的阳性结果将该名就诊者可能患病的概率提高了 33%。

已知验前概率和似然比,除了通过公式计算之外,还可以根据一些参考书上提供的图表直接估算验后概率。如图 2-3-2 所示,在做某项诊断试验之前,医生认为就诊者的患病只有 50%,如果阳性似然比为 5 的诊断试验结果为阳性,将验前概率和相应诊断试验的似然比连线,其右侧延长线与右侧直线相交处即可得到约 83%的验后概率。这样在临床上快速得到验后概率,有助于床旁循证决策。有些网站,如 http://araw.mede.uic.edu/cgi-bin/testcalc.pl,也提供一些快速的交互式计算模块,有兴趣的读者可以试试看。

四、ROC 曲线与曲线下面积

如前文所述,当诊断试验的结果呈连续性数据时,区分正常、异常(即阳性、阴性)的临界点(cut of points)划在哪里,将会影响灵敏度和特异度。灵敏度和特异度一般呈反比关系。临床上如何合理划分临界点?如何比较两种或者两种以上诊断试验的临床价值?此时可以采用受试者工作特性曲线(receiver operator characteristic curve,简称 ROC 曲线)。

ROC 曲线是用真阳性率(灵敏度)和假阳性率(1-特异度)作图所得出的曲线,它可表示灵敏度和特异度之间的相互关系。假如利用血清某物质 x′诊断心肌梗死,可以得到表2-3-4数据,则依据表 2-3-4,以假阳性率作为横坐标,灵敏度为纵坐标,可以得出图 2-3-3 的 ROC 曲线。ROC 曲线中最接近左上角的一个点(A)点,其灵敏度和特异度之和相对来说最高,其对应的诊断试验结果往往作为最佳的诊断临界点。

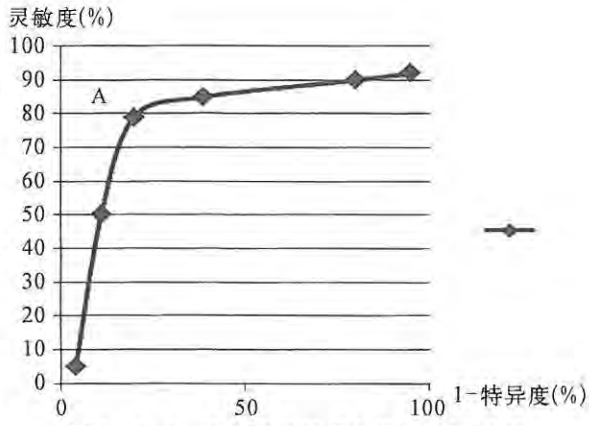

图 2-3-3 血 x′浓度诊断心梗的 ROC 曲线

表 2-3-4 血 x′浓度诊断心肌梗死的灵敏度和特异度

血 x′浓度(ng/mL)	灵敏度(%)	特异度(%)	假阳性(%)
12	92	5	95
24	90	20	80
48	85	50	39
96	79	60	20
192	50	80	11
384	5	96	4

ROC 曲线下的面积(area under the ROC curve, AUC),反映了诊断试验的价值,曲线下面积越接近 1.0,其诊断的真实度越高,区分能力越强;越接近对角线,面积约接近 0.5,则诊断的真实度越差,区分能力越弱。ROC 曲线可以据此比较几种诊断试验的诊断效率:在图 2-3-4 中,Stratus OCT 和 GDx VCC 在青光眼的诊断区分能力相似,但均优于 HRT II。

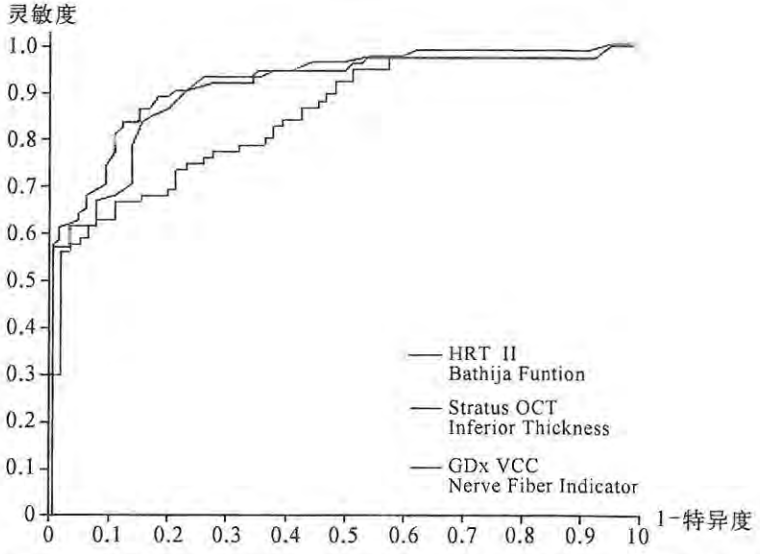

图 2-3-4 几种青光眼诊断试验的 ROC 曲线比较(引自 Medeiros FA, et al., 2004)

五、诊断能力的其他评价指标

诊断试验的准确性还有其他指标,如诊断比值比(diagnostic odds ratio)、诊断准确度(diagnostic accuracy)等,有兴趣的读者可以翻阅相关参考书籍进一步了解。

六、联合试验及其作用

疾病漏诊可能会给患者的健康带来影响,伤害患者及其家庭,而疾病误诊又可能给他们带来很大的经济或者精神负担,引起进一步不必要的、过度的,甚至是有害的检查和治疗,因此临床上常常需要有高灵敏度和高特异度的诊断试验以减少疾病的漏诊或误诊,来明确或者排除某些疾病。然而很多情况下,现有的诊断试验却未必有足够高的灵敏度和特异度,即使有也可能费时费力,或者较为昂贵,或者是有创的。

对于诊断试验结果呈连续分布的计量数据而言,通过改变诊断临界点是单方面提升

灵敏度或特异度的一种有效的办法。联合试验方法是提高灵敏度或者特异度的另一种有效办法。联合试验包括平行试验(parallel tests)和系列试验(serial tests)。选择平行试验或者系列试验依据临床对灵敏度或者特异度需要。运用得当的话,联合试验不失为合理利用现有诊断试验,提高诊断效率的一个方法。

平行试验是同时做几个试验,只要有一个阳性,即判定为阳性,认为就诊者有患病的证据;平行试验增加了诊断的灵敏度和阴性预测值。

系列试验系依次相继的试验,要所有试验皆阳性才能作出就诊者患病的判定;系列试验提高了特异度和阳性预测值。

第三节 诊断试验证据的评价

上一节内容介绍了关于诊断试验的诊断能力方面的评价指标。在循证临床实践中,医师通过文献检索获取关于诊断试验的一些证据,不能仅仅看这些指标的大小,更要对证据进行全面的评价,包括证据的真实性、证据的重要性和适用性三大方面。

一、证据的真实性评价

1. *金标准选择是否得当并在每个受试者中应用* 首先要检查文章中使用的金标准是否得当。金标准的选择应结合临床具体情况决定,例如肿瘤诊断应选用病理诊断,胆石症应以手术发现为标准。如果要判断肌酸磷酸激酶(CPK)诊断心肌梗死的价值,以心电图为金标准就不甚妥当,而应选用冠状动脉造影作为金标准。诊断试验中金标准的选择要有足够的依据,要能被大家接受;疾病的定义非常重要,需要清晰明确不含糊。

其次,要判断研究证据中是否对每一位受试者都采用了合适的金标准诊断。有些时候,比如金标准检查费用昂贵或为有创性,不能保证所有的患者都做了金标准检查。此时,不少研究者常常将被考核试验结果阳性者,送去做金标准试验,而阴性者只抽一部分人去做金标准试验,这样就可能会带来所谓的确认偏倚(verification bias)。因为试验阴性者也可能是患者,这样的研究结果必然夸大了灵敏度,造成偏倚。

2. *评价试验是否同金标准进行独立的盲法比较* 新诊断试验应与金标准检查作同期独立盲法对比,即要求评价试验结果的人不能预先知道该病例使用金标准诊断为"有病"还是"无病",金标准检查者亦不知新诊断试验的结果,两种检查要互相独立进行评价。了解金标准试验的结果往往会影响对被考核试验结果的解释;有时候,金标准检查结果模棱两可时,如果检查评定者知道新诊断试验的结果就可能会有倾向性结果,从而引起偏倚。

3. *研究人群是否具有代表性,能包括临床上应用该试验的各种患者* 研究人群应包括两组:一组是用金标准确诊"有病"的病例组,另一组是用金标准证实为"无病"的对照组。病例组应包括各型病例:如典型和不典型,早、中与晚期病例,有无并发症,经治和未治及复发等,以便使诊断试验的结果更具有临床应用价值。在临床上,诊断试验最有价值的是能将相对早期的疾病与易与该病混淆(症状、体征相近)的其他疾病鉴别开来。故而,诊断试验的研究应该纳入那些临床实践中可能遇到的、将会使用这种试验的各种患

者。这个意义上来讲，对照组可选用金标准证实没有目标疾病的其他病例，特别是与该病易混淆的病例，以明确鉴别诊断的价值；而完全正常的健康人群一般不宜作为对照组，否则会夸大其灵敏度和特异度。

4. 试验设计及受试者入组方法是否合理　诊断试验的研究一般是属于横断面研究(cross-sectional study)，同期比较新诊断试验和金标准检查。研究最好采用前瞻性设计，这样能更好地减少选择性偏倚，减少数据缺失的情况；对于疾病、诊断试验的具体条件、方法和阳性值等也可事先清楚定义。受试者的入选亦最好采用在一定时间内、一个或数个研究中心符合入选标准的受试者连续招募入组的方法，以避免选择偏倚。

5. 诊断试验的方法描述是否详细，能否重复　诊断试验一定要有明确的实验方法，清晰的实验程序和正确的科学依据。试验报告应该对诊断试验方法有详细的描述，以便别人重复和印证。对于诊断阈值(阳性值)的确定及其依据亦应详细说明。

6. 是否报告了所有的试验结果　例如有的试验结果中除了有阳性和阴性，还有一些是无法判断或可疑的病例，如果试验报告和分析中弃用了这些数据，仅报告阳性或阴性结果，则可能会带来偏倚。

二、诊断性试验证据的重要性

在判断诊断试验证据的真实性之后，还需要判断该证据的重要性。

首先要看研究报告是否给出了诊断试验的重要指标，如上文介绍的灵敏度、特异度、似然比、ROC曲线下面积，以及预测值等。医生可以根据这些指标来观察该诊断试验的表现，了解其区分有病和无病的能力。即使一个诊断试验的证据真实性很好，但诊断试验本身对疾病的区分能力不佳，其临床应用也大打折扣。诊断试验的这些重要指标的数值，如灵敏度、特异度及似然比等的大小，为临床医生在各种不同的诊断试验中选取合适的方案提供了重要的参考依据。以上指标的可信区间有助于判断结果的精确性。

诊断试验的可重复性(repeatability)，又称精确性(precision)，是指诊断试验在相同条件下，进行重复操作获得相同结果的稳定程度。所有的研究都存在一些测量变异(measurement variation)。对于一些结果判断相对主观的诊断试验，应该描述其观察者间及观察者内变异，以了解其可重复性，方便对试验结果的稳定性和变异性进行判断。除了单个研究报道之外，大家还要关注是否有其他验证研究，是否提供类似的研究结果。

有的诊断试验可能存在一定的不良事件，特别是一些有创的诊断试验，应该在研究报告中描述这些事件及其发生频率，便于临床医生权衡利弊、进行决策。

三、诊断性试验证据的适用性

在确定诊断证据的真实性、重要性之后，接下来要考虑的就是该证据是否能适用于本地区、本医疗单位或具体到某个患者。证据的适用性可以从以下几个方面进行考虑：

(1) 该诊断试验是否能在本单位本部门合理、正确地开展：不同医疗单位的技术力量和医疗资源不一，所接待患者的疾病谱也存在很大的差别。这些都会影响证据的适用性。另外，诊断试验的费用也是一个需要考虑的问题。即使是一个非常好的诊断试验，如果费用非常昂贵，也常常会因此而使得其适用性大大受到限制。

(2) 是否能合理地估计患者的验前概率，以考虑诊断试验的适用性：验前概率常常根

据医生自己或同事的临床经验、地区或国家的流行病学调查结果、特定的数据库信息，或是经过评价的一些文献报道等来估计。同样的一个诊断试验，用于患病率不同的人群，其结果会存在很大的不一样。在应用一个诊断试验的时候，医生考虑得更多的是拿到一个检查结果后判断患者的患病概率(验后概率)。如前文所述，即使试验的特异度很高，当用于患病率很低的人群时，仍会出现大量假阳性结果；同样，一种敏感度非常高的试验，当用于患病率很高的人群，仍会出现较多假阴性结果。因此，合理地估计患者的验前概率，是选择诊断试验、确定验后概率的一个重要条件。

(3) 验后概率是否能改变医生后续的诊疗方案：这个问题虽然放在最后，但非常重要。因为无论一个诊断试验有多好，假如做和不做该试验最终给患者的处理是一样的，这个检查就没有意义了。验后概率取决于验前概率和诊断试验的似然比。在面对有一定验前概率的患者时，那些能进一步明确或排除疾病、为下一步诊疗计划提供方向的才是合适的诊断试验。换句话说，只有那些可能会使得验前概率发生足够程度改变的试验才有意义。

临床上，假如患者不太可能患有某病(即患病概率 p_1 低于某个数值)，则医生一般让他观察无须进一步处理；或者患者极有可能患有某病(即患病概率 p_2 高于某个数值)，医生也无须进一步诊断即可直接治疗。根据患者的验前概率和诊断试验的阳性似然比可以估计患者的验后概率 p_1'，如果 $p_1' \leqslant p_1$，则即使是阳性的诊断结果，患者验后概率亦非常低，无须进一步处理，只需观察即可；另一方面，根据患者的验前概率和诊断试验的阴性似然比可以估计患者的验后概率 p_2'，如果 $p_2' \geqslant p_2$，则即使是阴性的诊断结果，患者的验后概率亦非常高，还是需要直接治疗，故而诊断试验是多余的。只有当患者的验后概率介于 p_1 和 p_2 之间时，诊断才可能带来诊疗措施的改变。因此要根据患者验前概率和诊断试验的特性判断该诊断试验的应用价值与意义。

关于诊断性试验的证据(研究报告)有个国际通用的标准和声明，即 STARD(诊断试验研究报告标准，STAndards for the Reporting of Diagnostic accuracy studies)。该标准不但可用于指导研究报告写作，亦可用作研究设计和文献评价的重要参考。详细内容可以在其官网获得，网址 http://www.stard-statement.org/，有兴趣的读者可以进一步研究和学习。

(陈世耀　袁源智)

【思考题】

1. 在阅读诊断试验研究文献时，如何科学性地评价诊断试验？
2. 诊断试验评价指标有哪些？各自基本概念和特点是什么？有何应用价值？
3. 如何提高诊断试验的效率？