

诊断试验证据评价与临床应用

陈世耀

复旦大学临床流行病学/循证医学中心
复旦大学附属中山医院

1

Focus on

- 诊断试验案例：阅读并评价文献
- 论文报告标准
- 设计案例
- 诊断试验临床应用

2

文献评价内容

- 先进性评估：是否首先报道 (First one)；是否科学评价 (Last one)
- 真实性评估：是否纳入全部患者或代表性如何
- 结果大小：灵敏度/特异度/似然比
- 科学性评估：……

3

临床案例1：

- 男性，35岁
- 既往有乙型肝炎病史，HBsAg(+)、HBeAb(+)、HBcAb(+)，一年前肝肾功能、甲胎蛋白 (AFP)检查结果正常，血常规(包括血红蛋白、血小板、白细胞)正常
- 本次检查，AFP高于正常值高限2倍(定义为阳性)，肝脏B型彩超检未发现占位。这位患者来到门诊就诊，门诊医师该如何处理呢？

4

循证实践的步骤

- ① 依据患者问题提出临床可以回答的问题；
- ② 文献检索或者查找证据
- ③ 文献评价，即证据是否科学
- ④ 应用证据，即证据是否适应我们目前的患者
- ⑤ 后效评价，即临床问题是否解决

5

案例分析：AFP诊断肝癌情况分析

金标准评价最终结果		
	肝癌	非肝癌
AFP(+)	肝细胞肝癌	肝硬化、生殖系统肿瘤、其他肿瘤
AFP(-)	胆管细胞肝癌	非肝癌

6

肝癌的临床诊断依据

- (1) 慢性肝病病史，患者存在乙型肝炎基础，是重要的病史依据；
- (2) 肝病体征，该患者没有；
- (3) 肿瘤标准物，AFP升高是肝癌重要标志物，该患者为阳性，提示肝癌；
- (4) 影像学检查肝内异常占位，该患者B型彩超没有发现异常占位，不符合肝癌诊断。

7

临床问题

- 通过进一步检查明确或排除该患者是否是肝癌。
- 选择CT检查是否能帮助我们确定或者排除该患者肝癌的诊断？
- 或者，选择CT检查诊断或者排除该患者肝癌，其诊断肝癌的灵敏度和特异度是多少？

8

CT诊断肝癌情况分析

金标准评价最终结果		
	肝癌	非肝癌
CT判断(+)	肝癌	肝硬化结节、转移性肿瘤、其他肝占位
CT判断(-)	小肝癌、无强化肝癌、表现类 现脓肿的肝癌、表现为类 化结节的肝癌、表现为类 似多发转移灶肝癌	非肝癌

9

诊断试验的四格表

		黄金标准		
		病例组	非病例组	
诊断 试验	阳性	(真阳性)a	b(假阳性)	a+b
	阴性	(假阴性)c	d(真阴性)	c+d
		a+c	b+d	a+b+c+d
灵敏度=a/a+c		阳性预测值=a/a+b		
特异度=d/b+d		阴性预测值=d/c+d		
准确度=a+d/a+b+c+d				
患病率=a+c/a+b+c+d				

10

评价诊断试验的指标(1)

- 灵敏度(Sensitivity) ↓ 特异度(Specificity)
 - 灵敏度指病人组中经诊断试验查出阳性人数的比例(a/a+c);
 - 病人组中经诊断试验查出阴性人数的比例(c/a+c)即为假阴性率，又称漏诊率；
 - 假阴性率=1-灵敏度。
 - 特异度指非病人组中经诊断试验查出阴性人数的比例(d/b+d);
 - 经诊断试验查出阳性人数的比例(b/b+d)即为假阳性率，又称误诊率；
 - 假阳性率=1-特异度。
- | | | |
|----|---|---|
| | 是 | 否 |
| 阳性 | a | b |
| 阴性 | c | d |

评价诊断试验的指标(2)

- 预测值:
 - 阳性预测值指诊断试验阳性结果中真正患病的比例(a/a+b)
 - 阴性预测值指诊断试验阴性结果中真正未患病的比例(d/c+d)
- | | | |
|----|---|---|
| | 是 | 否 |
| 阳性 | a | b |
| 阴性 | c | d |

12

预测值的特点

- 灵敏度越高的试验，阴性预测值越高；
- 特异度越高的试验，阳性预测值越高；
- 患病率的高低对预测值的影响更大。

13

评价诊断试验的指标(3)

- 似然比(Likelihood ratio, LR)
 - 试验阳性结果的似然比系真阳性率和假阳性率之比
 - 试验阴性结果的似然比系假阴性率和真阴性率之比

14

似然比的计算

- 例. 监护病房CK诊断心梗(CK<80=CK-)

	心梗	无心梗		
CK+	215(93.48%)	16(12.31%)	231	+LR=7.6
CK-	15(6.52%)	114(87.69%)	129	-LR=0.07
Total	230	130	360	

15

似然比的特点和应用

- 比灵敏度和特异度更稳定，不受患病率影响；
- 用于估计疾病概率。

16

诊断试验评价中的常用基本概念

- 疾病概率=病人/(病人+非病人) =a/a+b
- 疾病比率=病人/非病人 =a/b
- 验前比=验前概率/(1-验前概率) =(a/a+b)/[1-a/(a+b)]=a/b
- 验后比=验前比×似然比 =a/b×LR=a1/b1
- 验后概率=验后比/(1+验后比) =(A/B)/(1+A/B)=A/A+B

17

诊断试验似然比应用举例

- 例：女45y, 典型心绞痛就诊, 患冠心病概率多少？
 - 文献资料, 患病率1%
 - 若患者有典型心绞痛, 其似然比=100
 - 验前概率=1%
 - 验前比=0.01/(1-0.01)=0.01
 - 验后比=0.01×100=1
 - 验后概率=1/(1+1)=0.5=50%

18

似然比应用举例

例：女45y, 典型心绞痛, 心电图检查ST段压低2.2mm, 患冠心病概率为多少?

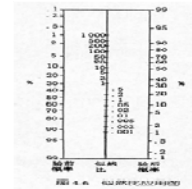
- 文献资料, ST压低2.2, 似然比=11
- 患者有典型心绞痛, 估计患病概率=50%

- 验前概率=50%
- 验前比=0.5/(1-0.5)=1
- 验后比=验前比×似然比=1×11=11
- 验后概率=11/(1+11)=91%

处理综合试验结果

■ 验后比

$$= \text{验前比} \times \text{试验1似然比} \times \text{试验2似然比}$$



文献评价内容

- 先进性评估：是否首先报道 (First one)；是否科学评价 (Last one)
- 真实性评估：是否纳入全部患者或代表性如何
- 结果大小：灵敏度/特异度/似然比
- 科学性评估：……

临床案例2：食管癌术前分期诊断评估

■ 临床问题：

- 如何确定食管癌术前分期 (TNM)
- 三种分期诊断方法如何选择：EUS/增强CT/PET-CT

案例分析：食管癌术前分期诊断评估

■ 研究问题：

- 参考试验 (金标准) 是什么?
- 实验设计?

Surg Endosc
DOI 10.1007/s00461-009-0782-x

Comparison of endoscopic ultrasonography (EUS), positron emission tomography (PET), and computed tomography (CT) in the preoperative locoregional staging of resectable esophageal cancer

Jeongmin Choi · Sang Gyun Kim · Joo Sung Kim · Hyun Chae Jung · In Sung Sung

研究报告摘要

- 目的：
- 方法：
- 结果：
- 结论：

Abstract
Background: Endoscopic ultrasonography (EUS) has been a useful method for the accurate staging of esophageal cancer. This study aimed to compare the diagnostic performance of EUS, positron emission tomography (PET), and computed tomography (CT) in the locoregional staging of resectable esophageal cancers.
Methods: A total of 109 patients with resectable esophageal cancer were prospectively enrolled and retrospectively reviewed for evaluation of preoperative EUS, PET, and CT. The sensitivity, specificity, and accuracy of tumor depth (T) staging and regional lymph node (N) staging for each test were compared with the postoperative histopathologic stage as the gold standard.
Results: The overall accuracy of EUS for T staging was 77%, and it was the only method for delineating the layers of the esophageal wall. The sensitivities for N staging were 42% for EUS, 49% for PET, and 35% for CT, and their specificities were, respectively, 91, 87, and 91%. The accuracy for N staging was 66% for EUS, 68% for PET, and 67% for CT, and it did not differ significantly across the three tests.

诊断试验科学性评价

- 评价试验是否与金标准试验进行了独立、盲法比较？
- 所研究患者样本是否包括临床试验中将使用该诊断试验的各种患者？
- 评价试验结果有无影响参照标准检查的实施；
- 诊断试验的精确性

25

<http://www.stard-statement.org/>



26

题目/摘要/关键词

- 1. 把文章标记为诊断准确性(推荐使用MeSH主题词“灵敏度与特异度”)
- 题目(Title): Comparison of EUS, PET, and CT in the preoperative locoregional staging of resectable esophageal cancer
- 摘要(Results): The sensitivities for N staging were 42% for EUS, 49% for PET, and 35% for CT, and their specificities were, respectively, 91, 87, and 93%.
- 关键词(Keywords): Computed tomography, Endosonography Esophageal cancer, Positron emission tomography, Tumor staging

27

介绍

- 2. 陈述研究问题或目的，如估计诊断准确性或比较不同试验或不同病例群体之间准确性
- This large postoperative study aimed to compare the accuracy of EUS staging with that of PET and CT staging using postoperative pathologic staging as a gold standard.
- It also aimed to determine the impact of EUS in preoperative locoregional staging of resectable esophageal cancer.

28

方法：研究对象

- 3. 描述研究人群：纳入和排除标准，数据收集的机构和场所
- 没有具体排除标准，如术前放疗排除等(结果有)

From September 2005 to July 2008, patients with a diagnosis of biopsy-proven esophageal cancer were prospectively enrolled in Seoul National University Hospital to undergo a preoperative staging protocol that included EUS, whole-body PET/CT, and chest CT scan concomitantly. All patients eligible for surgical resection who had no distant metastasis or extensive adjacent organ invasion in the staging protocol were allocated to esophageal resection and extensive regional lymph node dissection. The Insti-

29

方法：研究对象

- 4. 描述研究对象的募集：募集基于表现的症状，还是以前试验的结果，还是研究对象已经接受过目标试验或参考标准的事实？
- 研究对象已经接受过目标试验和参考试验
- 回顾性研究：可接受切除手术的食管癌患者-一段时间、一个单位的总体

30

方法：研究对象

- 5. 描述研究对象的抽样。研究人群是一个根据第3项或第4项定义的选择标准下的连续系列吗?如果不是,说明研究对象是如何选择的
- 对象选择偏移:
- 连续活检证实食管癌-评估可手术(评估的诊断试验影响手术)-患者接受手术-回顾性检查待评估的试验-排除手术未获得标本(如探查不可切除)

31

方法：研究对象

- 6. 描述数据收集:数据收集的计划是在目标试验和参考标准实施之前(前瞻性研究),还是之后(回顾性研究)?
- 数据收集在之后(回顾性研究)-收集信息偏移如缺失、不确定等

and extensive regional lymph node dissection. The Institutional Review Board of the Seoul National University Hospital approved this study, and written informed consents were obtained from all the patients. All information was collected and analyzed retrospectively.

32

方法：试验方法

- 7. 描述参考标准和它的原理

- 参考试验:手术+病理检查

staging with that of PET and CT staging using postoperative pathologic staging as a gold standard. It also aimed to

Pathologic staging was determined using the tumor node metastasis (TNM) classification system according to the American Joint Committee on Cancer (AJCC) esophageal cancer staging [10]. The regional lymph nodes included

33

方法：试验方法

- 8. 描述所使用的材料和方法的技术说明,包括何时、如何进行测量,列出目标试验和参考标准的引用文献

- 描述,非盲法,目标评价试验之间存在互相影响

34

方法：试验方法

- 9. 描述目标试验和参考标准结果单位、截断值和/或分类的定义和原理

The EUS criteria used for T staging were as follows: T1 (tumor invades the mucosa or submucosa but not the muscularis propria), T2 (tumor invades but does not extend

and distant organ metastasis. The lymph nodes with the shortest diameter (>8 mm) were considered positive for metastasis.

35

方法：试验方法

- 10. 描述实施目标试验和参考标准以及阅读结果的人员数量,培训情况和经验

- 测量偏移:未描述

36

方法：试验方法

- 11. 描述目标试验和参考标准里读取结果的人是否对另一个试验的结果设盲，描述任何读取结果可以获得者可以获得的临床信息

- 测量偏移：未设盲
- 诊断试验相互影响，诊断试验本身影响手术
- 手术取材（如淋巴结清扫）依据诊断试验
- 病理取材、切片、读片受手术影响（清楚全部临床信息）

37

方法：统计方法

- 12. 描述计算或比较诊断准确性测量结果的方法，以及对结果不确定性定量的统计方法
- 不确定结果未说明处理

compared. The sensitivity, specificity, positive predictive value, negative predictive value, and accuracy were calculated for each individual T and N stage by using standard definitions. A McNemar test was used to compare the accuracy of EUS, PET, and CT staging [11]. All tests of significance were two-tailed, and p values <0.05 denoted statistical significance. Analyses were performed using the Statistical Package for the Social Sciences, version 12.0 (SPSS, Chicago, IL, USA).

38

方法：统计方法

- 13. 如果可能，则描述计算试验可重复性的方法
- 未报告试验重复检查或者审查报告，未采用计算重复性方法，如均数、多数法则、Kpa值等

39

结果：研究对象

- 14. 报告研究完成的时间，包括征集研究对象开始和停止的日期
- 作者在方法学中介绍了时间，可以确定，该研究为回顾性研究；部分患者缺少某项试验或者术前化疗/放疗等影响同样说明回顾性

From September 2005 to July 2008, patients with a diagnosis of biopsy-proven esophageal cancer were prospectively enrolled in Seoul National University Hospital to undergo a preoperative staging protocol that included EUS, whole-body PET/CT, and chest CT scan concomitantly.

40

结果：研究对象

- 15. 报告研究人群的临床和人口学特征(如年龄、性别、症状谱、其他伴随疾病、当前治疗、征集中心)

Table 1 Patient characteristics

Characteristics	n	%
-----------------	---	---

A total of 89 patients underwent transthoracic esophagectomy with two-field (thoracoabdominal) lymph node dissection, whereas 12 patients had three-field (thoracoabdominal and cervical) lymph node dissection, and 8 patients underwent transhiatal esophagectomy with lymph node sampling. Five patients (4.5%) had preoperative neoadjuvant chemotherapy ($n = 1$) or concurrent chemoradiation therapy ($n = 4$) before surgery.

41

结果：研究对象

- 16. 报告满足入选标准进行或未进行目标试验和/或参考标准的研究对象的数量，描述研究对象未能参加试验的原因(强烈推荐流程图)

Examination by EUS was completed for 91 patients (84%). For 18 patients (16%), obstructing tumors did not permit

Because CT and PET cannot differentiate the individual layers of the esophageal wall, no attempt was made to compare T staging among the three methods. Findings

42

结果：试验结果

- 17. 报告从目标试验到参考标准的时间间隔，以及期间采取的任何治疗措施
- 测量偏移：缺少盲法比较，诊断试验相互影响，放化疗影响手术分期
- 评价试验在先，参考标准在后，但没有具体间隔时间说明
- 需要比较的试验之间的先后顺序以及间隔时间也没有说明

43

结果：试验结果

- 18. 有目标状态的研究对象，报告疾病严重性的分布程度；对没有目标状态的，描述其他的诊断
- 疾病轻重程度的分析，本例中分期诊断。在一种分期状态下，仍然有不同程度会有影响。如对远处转移的不同部位描述、淋巴结转移的不同部位描述
- 诊断试验影响处理（部分阳性不手术，剔除-选择偏移）

44

结果：试验结果

- 19. 报告根据参考标准结果的目标试验结果(包括不确定和缺失的结果)的交叉表；对连续型结果，报告根据参考标准结果的目标试验结果的分布

Table 2 Accuracy of endoscopic ultrasonography (EUS) tumor depth (T) staging

	Pathologic stage				Sensitivity (%)	Specificity (%)
	T1	T2	T3	T4		
EUS T stage						
T1	39	0	0	0	80	100
T2	8	8	9	0	53	82
T3	2	7	32	4	78	80
T4	0	0	0	0		

PPV positive predictive value, NPV negative predictive value

45

结果：试验结果

- 20. 报告实施目标试验或参考标准期间的任何不良事件
- 真实性评估：
 - 回顾性收集资料，未报告，如CT过敏、麻醉意外、内镜操作损伤或者出血等

46

结果：结果估计

- 21. 报告诊断准确性估计结果和统计学不确定性的测量结果(如95%CI)
- 缺少估计结果（95%CI）

47

结果：结果估计

- 22. 报告目标试验里不确定结果，缺失结果和异常结果是如何处理的
- 没有具体交待，按照缺失处理(EUS病例数不同)

	Pathologic stage				Sensitivity (%)	Specificity (%)
	T1	T2	T3	T4		
EUS T stage						
T1	39	0	0	0	80	100
T2	8	8	9	0	53	82
T3	2	7	32	4	78	80
T4	0	0	0	0		

	Pathologic stage	
	N0	N1
EUS		
N0	49	32
N1	5	23

48

结果：结果估计

- 23. 报告诊断准确性在不同亚组，不同读取结果者或不同中心之间差异的估计
- 测量偏移：单中心，回顾性研究，缺少估计，直接使用临床报告结果
- 不同读取结果者之间的差异如何估计（重新检查评估-读片）

49

结果：结果估计

- 24. 如果可能，则报告试验可重复性的估计结果
- 理论上可以，未报告
- 对每一试验计算Kpa值，包括参考试验

50

讨论

- 25. 讨论研究结果的临床适用性
- 人群特征与代表性（不可切除，狭窄不能超声）

Another factor that may affect accuracy in staging of esophageal cancer is the presence of strictures that limit passage of an echoendoscope. Up to one-third of patients

Our data are limited by the inclusion of patients with a somewhat narrowed clinical spectrum of disease because only surgically resectable patients were enrolled in this study. Patients with advanced disease stages were excluded, which may be related to underestimation of both sensitivity and accuracy in the detection of metastasis to lymph nodes.

51

总体评价

- 选择偏移：
 - 可切除患者
 - 食管狭窄患者
- 测量偏移
 - 金标准客观，但未实现盲法评估
 - 诊断试验之间影响/放化疗对金标准影响
 - 时间间隔影响/报告人之间影响
- 处理：全部试验重新盲法双评价，计算Kpa

52

如何评估诊断试验结果测量重复性

- 评价方法：Kappa值计算
- Kappa值是排除了机遇一致率后的观察一致率

53

Kappa值计算

甲、乙医生阅读同样100张胸部X线片的诊断一致性

		甲医生		
		肺门淋巴结	正常	
乙医生	肺门淋巴结	46 a	b 10	56(r ₁)
	正常	12 c	d 32	44(r ₂)
		58(c ₁)	42(c ₂)	

观察一致率(po)= a+d/N

po= 46+32/100=78%

机遇一致率(pc)= ((r₁ × c₁)/N+ (r₂ × c₂)/N)/N pc=((56×58)/100+(44×42)/100)/100=51%

非机遇一致率=1-pc

非机遇一致率=100%-51%=49%

实际一致率=po-pc

实际一致率= 78%-51%=27%

Kappa值=po-pc/1-pc

Kappa值 =27%/49%=0.55

问题：如何进行诊断研究设计

- 胃癌是一种常见恶性肿瘤，通过胃镜检查+活检确诊。处理包括早期胃癌的内镜下治疗、进展期胃癌的根治性手术治疗和晚期胃癌的化疗。采用何种处理方式，除了根据患者意愿、病人状态，最主要的依据是胃癌的分期。目前临床上用于胃癌分期(TNM分期)的手段主要包括超声内镜(EUS)、增强CT。
- 请设计一项临床研究，比较EUS、增强CT检查用于胃癌侵犯深度和是否周围淋巴结转移的分期诊断价值。

55

设计要求

- 研究目的：
- 研究设计类型：
- 人群与病例选择：
- 诊断试验金标准：
- 金标准与诊断试验的实施方法：
- 评价指标：

56

设计要点-1

- 研究目的：
- 比较超声内镜、增强CT检查用于胃癌侵犯深度和是否周围淋巴结转移的分期诊断价值

57

设计要点-2

- 研究设计类型：
 - 诊断试验评价(病例-非病例对照)：2个评价试验之间无相互影响，2个试验联合应用时相互作用大小
 - 比较(RCT研究)：增加样本，减少患者损害，或者2个诊断试验之间有明确相互影响

58

RCT设计-比较诊断试验之间相互影响

The **NEW ENGLAND**
JOURNAL *of* **MEDICINE**

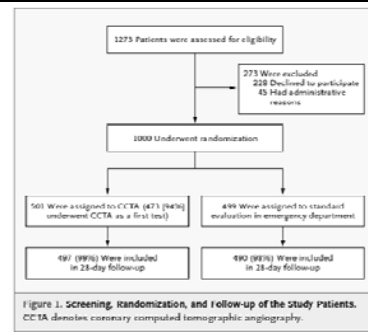
ESTABLISHED IN 1912 JULY 26, 2012 VOL. 367 NO. 4

**Coronary CT Angiography versus Standard Evaluation
in Acute Chest Pain**

Udo Hoffmann, M.D., M.P.H., Quynh A. Truong, M.D., M.P.H., David A. Schoenfeld, Ph.D., Eric T. Chou, M.D., Pamela K. Woodard, M.D., John T. Nagurny, M.D., M.P.H., J. Hector Pope, M.D., Thomas H. Hauser, M.D., M.P.H., Charles S. White, M.D., Scott G. Weiner, M.D., M.P.H., Shant Kalanjian, M.D., Michael E. Mullins, M.D., Issam Mikati, M.D., W. Frank Peacock, M.D., Pavel Zakrzkovsky, B.A., Douglas Hayden, Ph.D., Alexander Goehler, M.D., Ph.D., Hang Lee, Ph.D., G. Scott Gaziano, M.D., M.P.H., Ph.D., Stephen D. Wiviott, M.D., Jerome L. Fleig, M.D., and James E. Udelson, M.D., for the ROMICAT-II Investigators

59

Randomized controlled trial



Outcome

METHODS

In this multicenter trial, we randomly assigned patients 40 to 74 years of age with symptoms suggestive of acute coronary syndromes but without ischemic electrocardiographic changes or an initial positive troponin test to early CCTA or to standard evaluation in the emergency department on weekdays during daylight hours between April 2010 and January 2012. The primary end point was length of stay in the hospital. Secondary end points included rates of discharge from the emergency department, major adverse cardiovascular events at 28 days, and cumulative costs. Safety end points were undetected acute coronary syndromes.

RESULTS

The rate of acute coronary syndromes among 1000 patients with a mean (±SD) age of 54±8 years (47% women) was 8%. After early CCTA, as compared with standard evaluation, the mean length of stay in the hospital was reduced by .6 hours (P<0.001) and more patients were discharged directly from the emergency department (47% vs. 12%, P<0.001). There were no undetected acute coronary syndromes and no significant differences in major adverse cardiovascular events at 28 days. After CCTA, there was more downstream testing and higher radiation exposure. The cumulative mean cost of care was similar in the CCTA group and the standard-evaluation group (\$4,289 and \$4,060, respectively; P=0.05).

设计要点-3

- 人群选择:
- 胃癌或者其他检查确诊的胃癌患者, 包括早期胃癌、进展期胃癌, 是否纳入晚期胃癌根据金标准实施确定, 但推广应用与选择人群需要一致。
- 样本量估计

62

设计要点-4

- 金标准:
- 采用手术探查或者切除结合病理分析, 将胃癌按局部侵犯和淋巴结转移分别进行分期(T分期和N分期)。
- 手术探查的指证不依据本次增强CT检查结果或者超声内镜检查结果 (如可以采用MR或者PET等检查结果)

63

设计要点-5

- 诊断试验实施:
- 超声胃镜检查检查和CT检查, 两种检查在手术治疗前完成(如2周内, 术前放化疗剔除), 两种检查独立进行, 互相不知道对方结果, 检查结果的判定均需要至少两名有经验的专业医师做出, 有争议时采用第三方观点。
- 如果RCT设计, 一定做到随机化分组 (分别接受其中一种诊断试验) 及随机化隐藏。
- 针对两名判断者, 计算Kpa值

64

设计要点-6

- 评价指标:
- 采用灵敏度、特异度、似然比等指标评价, 结论为两种手段的比较, 何种更好, 或者相似。

65

临床应用-1: 诊断试验结果的代表性

- 研究人群
- 入选与排除标准
- 关注剔除人群

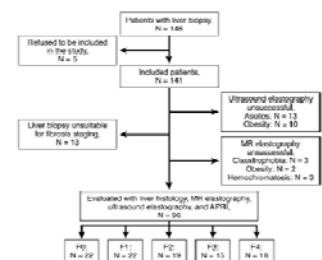


Figure 1. Flow diagram of patients who underwent liver biopsy for chronic liver disease in the Department of Gastroenterology during the 16-month study period.

临床应用-2: 诊断试验结果大小

- 结果是否适用于并可提供给我自己的患者?
 - 阳性预测值很高, 明确吗?
- 诊断试验结果是否改变了对患病率的估计?
 - 患病可能性多大?
- 诊断试验结果是否改变了对患者的处理?
 - 如何进行科学决策

67

诊断试验应用理论—治疗理论

- 不要问: 这个试验有多好?
- 而要问: 这个治疗有多好?
- 治疗原则: 如果好处多于坏处, 进行治疗
- 试验原则: 如果有改变处理的潜在可能, 试验!
- 情况:
 - 不确诊病例而只有一种试验可用。
 - 不确诊病例, 而有2种或2种以上的试验可用

68

确诊疾病(1)

- 结果

Disease疾病	Treat治疗	活	死亡
Present存在	Yes	80%	20%
Present存在	No	50%	50%
- Benefit收益:
 - =80% - 50% = 30%

69

不确诊疾病(2)

- 结果

Disease疾病	Treat治疗	活	死亡
Present存在	Yes	80%	20%
Present存在	No	50%	50%
Absent无	No	100%	0%
Absent无	Yes	90%	10%
- 收益Benefit for D+
 - =80% - 50% = 30%
- 危害harm for D-
 - =100% - 90% = 10%

70

患病率较高时

- 收益Benefit for D+
 - =80% - 50% = 30%
- 危害harm for D-
 - =100% - 90% = 10%
- 如果病人患病可能性为90%, 你治疗吗? (假设无试验可用)
 - Yes!
 - Benefit = 90% * 30% = 27%
 - Harm = 10% * 10% = 1%

71

患病率较低时

- 收益Benefit for D+
 - =80% - 50% = 30%
- 危害harm for D-
 - =100% - 90% = 10%
- 如果病人患病可能性为10%, 你治疗吗? (假设无试验可用)
 - No!
 - Benefit = 10% * 30% = 3%
 - Harm = 90% * 10% = 9%

72

治疗或行动点(AT_{ODDS})

- Prob*Benefit=(1-prob)*Harm
- Prop/(1-prob)=Harm/benefit=AT_{ODDS}

$$\text{Prop}/(1-\text{Prop})=10\%/30\%=\text{AT}_{\text{ODDS}}$$

$$\text{Prop}=25\%$$

73

不确诊疾病，但有试验可用(3)

■ 原则:

- 只有在有潜在可能会改变管理的情况下才试验

74

患病率较低时

- 假设已知试验灵敏度为90%，特异度为70%
- 问题1

- 如果病人患病可能性为1%，你试验吗？

Disease疾病	Treat治疗	活	死亡
Present存在	Yes	80%	20%
Present存在	No	50%	50%
Absent无	No	100%	0%
Absent无	Yes	90%	10%

75

患病率较低时-No

- 假设已知试验灵敏度为90%，特异度为70%
- 问题1

- 如果病人患病可能性为1%，你试验吗？

■ No!(No test, No treat)

- 试验阳性似然比LR+=90%/(1-70%)=3
- 验前比Pre-Odds=1/99
- 验后比Post-Odds=(1/99)*3=3/99
- 验后概率Post-Prob=3/(3+99)%=3%

76

患病率较高时

- 假设已知试验灵敏度为90%，特异度为70%
- 问题2

- 如果病人患病可能性为90%，你试验吗？

Disease疾病	Treat治疗	活	死亡
Present存在	Yes	80%	20%
Present存在	No	50%	50%
Absent无	No	100%	0%
Absent无	Yes	90%	10%

77

患病率较高时-No

- 假设已知试验灵敏度为90%，特异度为70%
- 问题2

- 如果病人患病可能性为90%，你试验吗？

■ No!(No test, But treat)

- 试验阴性结果的似然比=(1-90%)/70%=1/7
- 验前比Pre-Odds=90/10
- 验后比Post-Odds=(90/10)*(1/7)=9/7
- 验后概率Post-Prob=9/(7+9)%=56%

78

诊断域值

- 在什么可能之下对疾病仅观察而不试验，也不治疗？
 - 诊断域值 $ODDS * LR+ = AT_{ODDS}$
 - 诊断域值 $ODDS = AT_{ODDS} / LR+ = (1/3) / [(90\%) / (1-70\%)] = 1/9$
 - 诊断域值 $prob = 1 / (1+9) = 10\%$

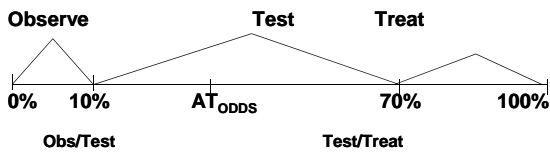
79

治疗域值

- 在什么可能之上对疾病凭经验作治疗，甚至不作试验？
 - 治疗域值 $ODDS * LR- = AT_{ODDS}$
 - 治疗域值 $ODDS = AT_{ODDS} / LR- = (1/3) / [(1-90\%) / 70\%] = 7/3$
 - 治疗域值 $prob = 7 / (7+3) = 70\%$

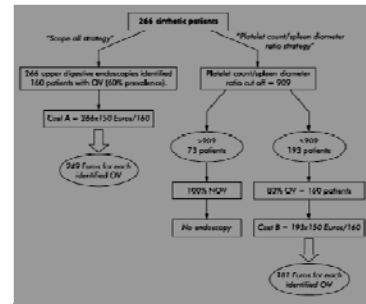
80

试验—治疗域值示意图



81

经济学结果帮助决策：成本效果



82

临床应用-3：提高诊断试验效率

- 在高危人群中应用诊断试验
- 应用联合试验
 - 平行试验：提高灵敏度和阴性预测值(漏诊减少，但误诊增加)。
 - 系列试验：提高特异度和阳性预测值(误诊减少但漏诊增加)。

83

例、患病率对预测值的影响(1)

- AFP诊断肝癌，灵敏度=80%，特异度=90%
- 一般人群

	肝癌	非肝癌		
AFP+	8	9999	10007	+PV=0.08%
AFP-	2	89991	89993	
Total	10	99990	100000	Pre=10/10万

84

例、患病率对预测值的影响(2)

- AFP诊断肝癌，灵敏度=80%，特异度=90%
- HBsAg(+)人群

	肝癌	非肝癌		
AFP+	80	9990	10070	+PV=0.8%
AFP-	20	89910	89930	
Total	100	99900	100000	Pre=100/10万

85

例、患病率对预测值的影响(3)

- AFP诊断肝癌，灵敏度=80%，特异度=90%
- HBsAg(+) + 年龄>40岁人群

	肝癌	非肝癌		
AFP+	800	9900	10700	+PV=7.48%
AFP-	200	89100	89300	
Total	1000	99000	100000	Pre=1000/10万

86

例、患病率对预测值的影响(4)

- AFP诊断肝癌，灵敏度=80%，特异度=90%
- HBsAg(+) + 年龄>40岁 + 肝硬化人群

	肝癌	非肝癌		
AFP+	8000	9000	17000	+PV=47.06%
AFP-	2000	81000	83000	
Total	10000	90000	100000	Pre=10000/10万

87

例、阳性预测值的计算

- AFP诊断肝癌，灵敏度=80%，特异度=90%
- 患病率=10/10万人

	肝癌	非肝癌		
AFP+	8	9999	10007	+PV=0.08%
AFP-	2	89991	89993	
Total	10	99990	100000	Pre=10/10万

$$+PV = \frac{\text{患病率} \times \text{灵敏度}}{\text{患病率} \times \text{灵敏度} + (1 - \text{患病率}) \times (1 - \text{特异度})}$$

88

临床应用-4: 提高诊断试验准确性

- 应用似然比

89

似然比的计算

	疾病	非疾病	
试验+	a(a/a+c)	b(b/b+d)	+LR=Sen/(1-Spe)
试验-	c(c/a+c)	d(d/b+d)	-LR=(1-Sen)/Spe
Total	a+c	b+d	a+b+c+d

Sen=灵敏度; Spe=特异度, LR=似然比

90

似然比的计算

	疾病	非疾病	
试验+	a(a/n1)	b(b/n2)	+LR=(a/n1)/(b/n2)
试验+	c(c/n1)	d(d/n2)	+LR=(c/n1)/(d/n2)
试验-	e(e/n1)	f(f/n2)	-LR=(e/n1)/(f/n2)
Total	n1=a+c+e	n2=b+d+f	n1+n2=a+b+c+d+e+f

91

疾病分期诊断试验评价

	T1	T2	T3	
试验+	a(a/n1)	d(d/n2)	g(g/n3)	+LR=(a/n1)/(d+g/n2+n3)
试验+	b(b/n1)	e(e/n2)	h(h/n3)	+LR=(b/n1)/(e+h/n2+n3)
试验-	c(c/n1)	f(f/n2)	i(i/n3)	-LR=(c/n1)/(f+i/n2+n3)
Total	n1=a+b+c	n2=d+e+f	n3=g+h+i	

92

联合试验似然比应用

- 验后比
= 验前比 × 试验1似然比 × 试验2似然比
- 应用条件
诊断试验独立性

93

如何估计联合试验中的相互作用

- 直接应用联合试验似然比计算验后概率
- 校正：计算实际联合试验似然比与无相互作用状态下两个独立试验似然比乘积
- 定量数据应用回归方程回答临界点，同时解决相互作用问题。

94

多种肿瘤标志物诊断肺癌

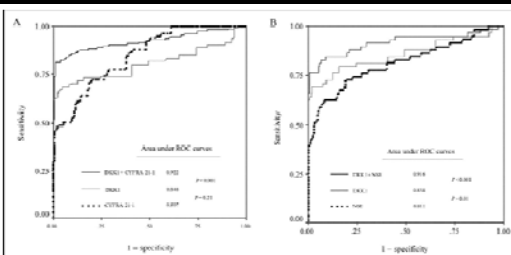


Fig. 3. (A), ROC curve analysis of DKK1, CYFRA 21-1, and the combination of 2 markers in patients with NSCLC (n = 154). (B), ROC curve analysis of DKK1, NSE, and the combination of 2 markers in patients with SCLC (n = 38).

95

诊断试验循证实践

- 先进性评估：是否首先报道 (First one)；是否科学评价 (Last one)
- 真实性评估：是否纳入全部患者
- 科学性评估：.....
- 应用价值的大小不仅取决于其效率，还包括可接受性、安全性、费用等

96

小结

- 诊断试验研究设计要点
- 诊断试验国际报告规范(STRAD标准)
 - 选择偏移/测量偏移/混杂偏移
- 诊断试验临床应用注意点

97

Thank you

syaochen@163.com

98