

数据、方法与解读

Data, method and interpretation

——Application of Statistics in Clinical Research

邓强庭

第三军医大学学报 编辑

国家SPSS数据分析高级工程师

中国医师协会科研与出版专业委员会委员

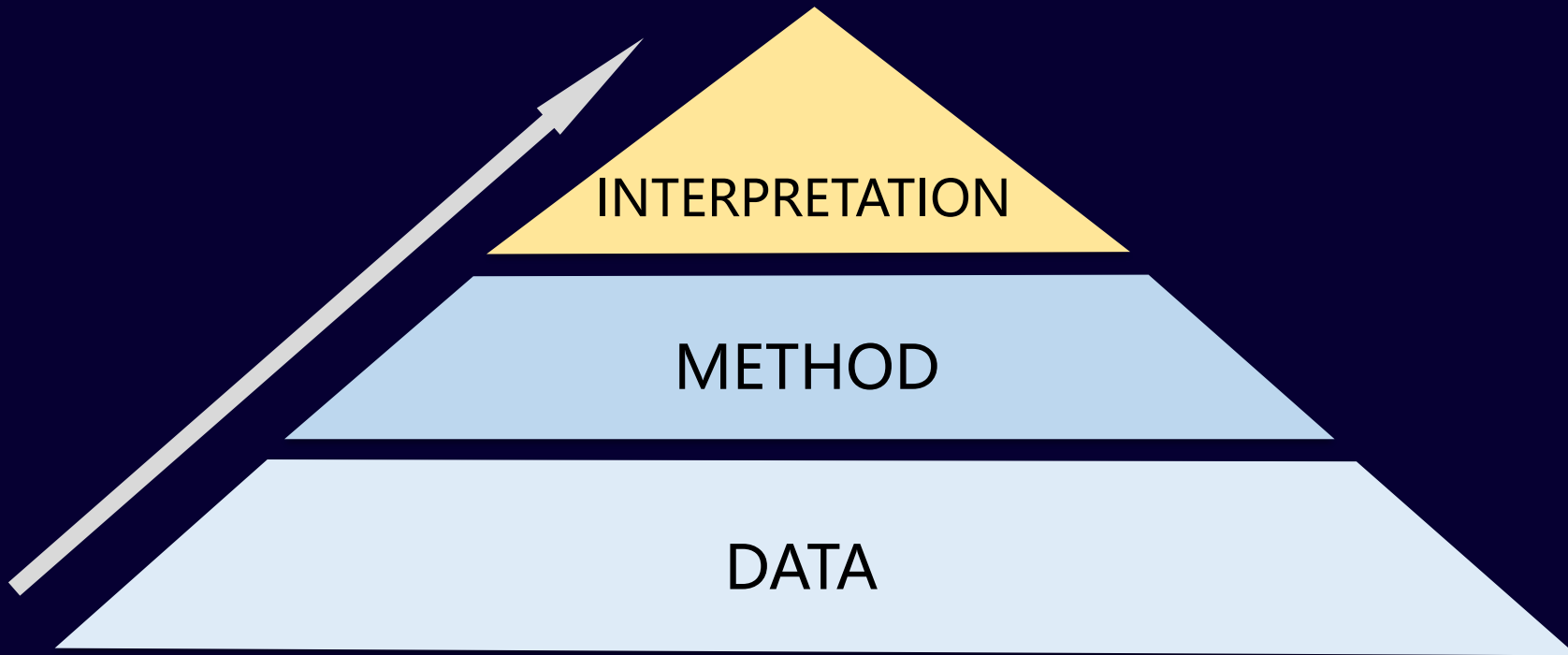
DENG Qiang-ting

Editor of *Journal of Third Military Medical University*

Senior Engineer of National SPSS Data Analysis

Member of the Research and Publishing Committee of CMDA

How to review a manuscript from a statistical perspective ?



Is it appropriate ?

is it right ?

Is it true ?

1 数据 About the Data

We usually identify some data fraud with statistical methods:

- $|t|$ 、 $|Z| > 1.96$, $P < 0.05$, $|t|$ 、 $|Z| > 2.58$, $P < 0.01$;
- 样本量相同时, $|t|$ 、 $|Z| \uparrow$, $P \downarrow$; t 值的正负反映两组均数的大小关系
- 均数差异相同时, $Sd \uparrow$, $P \uparrow$;
- 2×2 contingency table: $\chi^2 > 3.84$, $P < 0.05$; $\chi^2 \uparrow$, $P \downarrow$
- 样本量相同时, $|\text{Proportion}_t - \text{Proportion}_c| \uparrow$, $\chi^2 \uparrow$, $P \downarrow$
- OR 、 $HR = e^\beta$,
 - $\beta > 0$, OR 、 $HR > 1$;
 - $\beta < 0$, OR 、 $HR < 1$

表 2 急诊手术组和 TACE 后二期手术组患者围手术期指标比较

组别	n	术中出血量 ($\bar{x} \pm s$, mL)	术中输血量 ($\bar{x} \pm s$, mL)	围术期死亡 [例 (%)]	肝功能不全 [例 (%)]
急诊手术组	24	980.5±250.31	875.8±384.27	2 (8.3)	2 (8.3)
TACE 后二期手术组	18	492.3±168.05	423.4±173.17	0 (0)	1 (5.5)
t 值		2.547	1.952	—*	—*
P 值		0.028	0.017	0.489	1.000

*. Fisher 确切概率法

心电图检查时,患者保证心态平和环境安宁,保持平卧位并全身松弛。采用 Burdick ECG-9130P 型 12 导联进行连续描记(日本光电工业株式会社生产),增益为 10 mm/mv,纸速为 25 mm/s 以确保记录的 ECG 基线平稳、无干扰且图像清

起止特点、持续时间和伴随症状等并结合心电图检查可明确心律失常诊断,记录各种类型的心律失常,包括激动起源部位、频率、节律、传导时间和途径等一项或多项发生的异常。

冠心病患者可导致心肌梗死者导致心脏性猝死主要通过冠脉上述检查尚未在规心电图和 AEC

1.4 统计方法 采用 SPSS 13.00 软件

表 1 动态与常规心电图的心律失常检出结果

组别	例数	室性			房性			
		期前收缩频发	期前收缩二、三联律	期前收缩成对	期前收缩频发	期前收缩二、三联律	期前收缩成对	
动态	78	44(56.41)	39(50.00)	33(42.31)	47(60.26)	38(48.72)	34(43.59)	
常规	78	43(55.12)	16(20.51)	8(10.26)	46(58.97)	12(15.38)	2(2.56)	
χ^2 值		0.81	0.02598	4.79	12.58	2.48	0.0266267	15.28
P 值		> 0.05	< 0.05	< 0.05	< 0.05	> 0.05	< 0.05	< 0.05

创、费用低廉,对冠心病粗筛诊断非常有益。AECG 所检测的心电信息远大于普通

故结合常规心电图检查更为全面。

综上所述,目前 AECG 在监测心肌

参考文献:

表 1 两组患者治疗后 CRP、TP、ALB 比较 $\bar{x} \pm s$

组别	<i>n</i>	CRP(mg/L)	TP(g/L)	ALB(g/L)
对照组	32	98.1 ± 24.8	53.3 ± 3.2	22.3 ± 3.1
治疗组	32	43.9 ± 25.0	58.8 ± 4.3	26.5 ± 2.8
<i>t</i> 值		<u>5.342</u>	<u>2.474</u>	2.894
<i>P</i> 值		< 0.05 ^{8.707}	< 0.05 ^{-5.805}	< 0.05

表 2 两组患者治疗后血淀粉酶恢复时间、尿淀粉酶恢复时间、住院时间比较 $\bar{x} \pm s$

胸片确诊。纳入标准: 全部患儿均符合美国胸内

表 1 两组患儿相关临床资料比较 [n(%), ($\bar{x} \pm$

组别	例数	性别		抗生素使用		新生儿感染		
		男	女	有	无	有	无	
PNAC 组	85	60(70.59)	25(29.41)	85(100.00)	0(0.00)	69(81.18)	16(18.82)	40
非 PNAC 组	215	129(60.00)	86(40.00)	206(95.81)	9(4.19)	121(56.28)	94(43.72)	89
χ^2/t	-	2.93	1.245		1.352	3.67	16.261	
P	-	0.087	0.141		0.785	0.065	0.000	

组别	例数	新生儿湿肺		新生儿肺透明膜病		机械通气		
		有	无	有	无	有	无	
PNAC 组	85	12(14.12)	73(85.88)	84(98.82)	1(1.18)	65(76.47)	20(23.53)	
非 PNAC 组	215	36(16.74)	179(83.26)	200(93.02)	15(6.98)	103(47.91)	112(52.09)	
χ^2/t	-		0.185		0.152		20.170	
P	-		0.901		0.924		0.000	

Logistic回归分析,以 $P < 0.05$ 为差异有统计学意义。

强直型脑瘫。痉挛型发病有随着年龄增加发病减少的现象(表 1)。年龄段在临床类型分布中差异有统计学意义($\chi^2 = 42.054$, $P = 0.043$)。

2 结果

2.1 年龄、性别与脑瘫临床类型的关系 394例病例中

表 1 年龄与脑瘫临床类型之间的关系

年龄	例数	痉挛型	不随意运动型	共济失调型	肌张力低下型	混合型	χ^2	P
<1岁	144	90	6	0	28	20	42.054	0.043
1岁~	125	95	4	1	20	5		
2岁~	47	39	0	0	3	5		
3岁~	20	19	0	0	1			
4岁~	17	15	0	0	2			
5岁~	14	12	2	0	0			
6岁~	7	6	0	0	0			
>7岁	20	18	0	0	1			

f_x	=CHISQ.DIST.RT(B16, C16)			
	B	C	D	E
Chi2		df	P	
	42.054	35	0.1919162	
	1			

2.2 脑瘫与产前、产时、产后相关因素分析 产前相关因素由胎盘功能不全、孕期不适当用药、孕期接

($P < 0.05$), 见表 2。怀孕年龄、近亲结婚、孕期宫内感染、先兆流产、妊高症、既往流产史、母亲腹部外伤中

接触有毒物质、孕期胆汁淤积、孕期出血、分娩次数、羊水量等因素在脑瘫临床类型之间差异有统计学意义

孕期保胎、孕期抽搐、母亲缺碘、父母癫痫病史、羊膜早破、怀孕次数和羊水浑浊等因素无统计学意义。

表 2 394例脑瘫与产前相关因素分析

相关因素	例数	痉挛型	不随意运动型	共济失调型	肌张力低下型	混合型	χ^2	P
胎盘功能不全	4	0	2	0	2	0	<u>43.072</u>	<u>0.012</u>
孕期不适当用药	12	7	0	1	4	0	<u>36.984</u>	<u>0.004</u>
孕期接触有毒物质	4	1	0	0	1	2	10.546	0.043
孕期出血	6	2	0	0	4	0	14.213	0.039
孕期胆汁淤积	2	0	0	0	0	2	22.47	0.012
分娩次数							23.010	0.003
1产	233	177	6	0	33	17		
2产	142	102	6	0	21	13		
多产	19	15	0	1	1	2		
羊水量							19.777	0.011
羊水多	14	9	0	0	1	4		
羊水少	20	9	2	0	6	3		

3 阴道分娩、产程、出生体质量、阴道分娩中使用催产 (> 0.05)。

表 3 394例脑瘫与产时、产后相关因素分析

相关因素	例数	痉挛型	不随意运动型	共济失调型	肌张力低下型	混合型	χ^2	P
出生胎龄							15.793	0.046
早产儿	120	99	4	0	5	12		
过期儿	7	5	0	0	2	0		
正常	267	190	8	1	48	20		
出生体质量							17.913	0.022
低体质量	122	104	4	0	9	5		
巨大儿	20	13	1	0	6	0		
剖宫产	93	61	1	0	23	8	13.344	0.010
产钳助产	3	1	0	0	0	2	14.295	0.006
窒息	86	56	5	0	19	6	9.770	0.044
缺氧缺血性脑病	29	22	1	1	2	3	<u>13.918</u>	<u>0.008</u>
头部外伤	16	13	0	1	0	2	<u>26.953</u>	<u>0.012</u>
黄疸	42	23	3	0	5	11	24.234	0.000
溶血症	2	0	0	0	0	2	22.740	0.000

2.3 脑瘫临床类型与相关因素的多因素 Logistic回归

物质、羊水多、早产儿、低体质量儿、黄疸等相关因素有

当用药, 混合型中孕期接触有毒 因素差异均没有统计学意义 (表 4)。

表 4 脑瘫痉挛性与其他临床类型的多因素 Logistic 回归分析

相关因素	β	$Sx(\beta)$	Wald	P	OR	OR95% CI
黄疸	1.637	0.81	4.089	0.043	5.142	1.052 ~ 25.14
孕期出血	2.944	1.129	6.803	0.009	19.001	2.079 ~ 173.655
早产儿	-1.823	0.664	7.539	0.006	0.162	0.044 ~ 0.594
产后窒息	0.958	0.375	6.537	0.011	2.608	1.251 ~ 5.436
剖宫产分娩	1.072	0.353	9.234	0.002	2.922	1.463 ~ 5.835
孕期不适当用药	1.679	0.856	3.852	0.049	5.363	1.002 ~ 28.693
孕期接触有毒物质	3.657	1.343	7.411	0.006	38.75	2.785 ~ 539.17
羊水多	1.591	0.756	4.424	0.035	4.909	1.115 ~ 21.624
早产儿	1.256	0.565	4.945	0.026	3.512	1.161 ~ 10.627
低体重儿	-1.782	0.677	6.93	0.008	0.168	0.045 ~ 0.634
黄疸	-2.29	0.521	19.354	0.000	9.875	3.56 ~ 27.392

活质量就越高。早期发现和诊断是早期治疗的关键, 对降低脑瘫发病率, 减轻残疾程度非常重要。通常认

B	Exp (B)
1.637	5.14
2.944	18.99
-1.82	0.162
0.958	2.606
1.072	2.921
1.679	5.36
3.657	38.74
1.591	4.909
1.256	3.511
-1.78	0.168
-2.29	0.101

是 PNAC 的保护因素 ($OR = 0.025$)。

表 3 PNAC 影响因素多因素 Logistic 回归分析结果

相关因素	β	<i>OR</i>	95% <i>CI</i>	<i>P</i>
新生儿感染	1.024	2.352	1.465 ~ 6.785	0.036
PN 持续时间	<u>-1.358</u>	<u>1.854</u>	1.597 ~ 6.574	0.008
禁食时间	<u>-2.867</u>	<u>1.263</u>	1.158 ~ 1.654	0.021
氨基酸热卡比例	<u>8.451</u>	<u>2.865</u>	1.121 ~ 7.85	0.002
脂肪乳热卡比例	<u>1.241</u>	<u>1.854</u>	1.562 ~ 5.563	0.011
奶热卡比例	1.157	0.025	1.012 ~ 6.745	0.032

3 讨论

备
被
酸
危
脂
面
性
力
鹅
强
道

B	Exp (B)
1.024	2.784
-1.358	0.257
-2.867	0.057
8.451	4680
1.241	3.459
1.157	3.18

2 方法 About the method

医学论文统计方法选取和描述方面常见的问题

Common impropriety about methods of medical paper:

- 统计方法描述太简单

Treats the statistical methods too sketchily.

- 统计学方法的选取单纯依据数据类型而忽略研究目的

Choose statistical analysis methods only based on the data characteristics but ignores the purpose of the research.

- 多因素回归纳入的因子不全面

Incomplete factors included in multivariate regression.

选择统计学方法的依据

On what basis did we chose Statistical analysis method ?

研究目的

Purposes

比较

Compare ?

描述

Description ?

观察趋势

Spot the trend ?

分析影响因素

Influencing factors ?

...

设计类型

Design

独立成组

Independent ?

配对设计

Paired ?

正交设计

Orthogonal ?

...

数据类型

Data characteristics

连续型资料

Continuous variables ?

无序分类变量

Nominal ?

有序分类变量

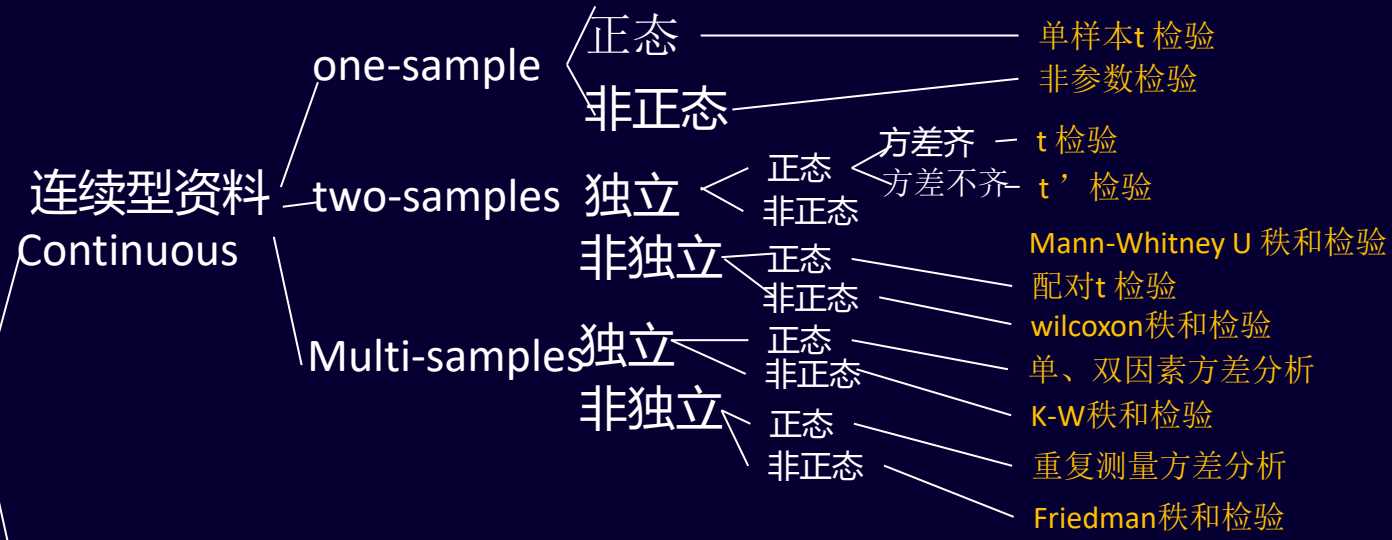
Ordinal ?

时间-结局事件

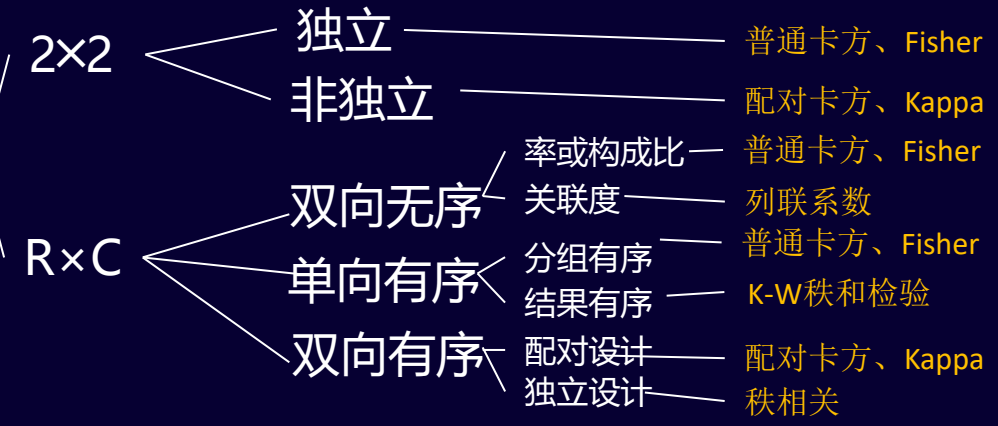
Time-to-Event ?

...

单变量
Single variable



分类资料
Categorical



多变量
Multiple variable



What are your purposes ?

先明确你想做什么

比较各组间某指标的差异
See any difference

Continuous

t test

ANOVA

Categorical

Rank sum test

long-term prognosis
(Time-to-Event)

χ^2 、Fisher、Rank sum test

Survival analysis

评价诊断效能
Diagnostic efficacy

Diagnostic accuracy of
categorical variables

Kappa test

ROC

Diagnostic power of
continuous variables

Pearson correlation

观察两指标的关系
Seek correlation clues

Correlation Analysis

Spearman rank correlation

分析影响因素
/构建预测模型
Find the factors/
Predictive model

Regression

Linear Regression

Logistic Regression

Cox Regression

统计学方法的选取最主要的依据不是数据类型而是研究目的

The main basis for choosing statistical methods is not the type of data but the purpose of the research

“Time-to-Event: Cox Regression” — Is it always the case?

	随访时间 (d)	是否发生 慢性疼痛
1	39.00	0
2	45.00	1
3	33.00	1
4	15.00	0
5	20.00	0
6	28.00	1
7	34.00	0
8	50.00	1
9	67.00	1
10	80.00	1
11	100.00	0
12	109.00	0
13	209.00	1
14	345.00	0
15	78.00	0
16	34.00	0

Data type

—> Cox Regression

“什么样的产妇慢性疼痛来的更晚？”

Which maternals chronic pain occurs later?

Which is better ?

Research purpose

—> Logistic Regression

“什么样的产妇更容易发生慢性疼痛？”

Which maternals is more prone to occurs chronic pain?

统计学方法的选取最主要的依据不是数据类型而是研究目的

“时间-结局 类型的资料用Cox回归” —— 一定如此吗？

同样是
肿瘤术后的
“时间-结局”
资料
Time-to-Event
after
tumor surgery

中晚期肿瘤术后 intermediate and terminal cancer :

—>Cox Regression

—> “哪些因素影响患者的生存时间”

晚期肿瘤，临床关注的焦点是 “**生命长短**”

terminal, focus: Length of life

早期肿瘤术后：

—>Logistic Regression

—> “哪些因素影响患者复发”

早期肿瘤，临床关注的焦点是 “**是否治愈**”

early-stage, focus: Is it cured?

对统计学方法的描述要详细而规范 Be more detailed.

详细规范的统计学方法描述

设计类型、数据分布类型、数据形态、统计学分析方法、数据转换方法、转换前后的分布类型、缺失值的处理方式、填补策略、变量的赋值方式、检验水准校正方式.....

样本量估算及依据

设计类型、参数 (α 、 β 、 δ 、脱落率.....)、参数确定的依据、估算方式、工具/公式、

恰当的结果展示

结果数据 (均数、标准差、中位数、四分位数.....)

统计参数 (t、 χ^2 、Z、U、AUC、95%CI、OR、HR、.....P)

★只给出 " $P < 0.05$ 、 $P > 0.05$ 、 $P < 0.01$ 、 $P > 0.01$ " 是很不妥的，应该给出 P 值的具体数值以及相应的统计参数的值。

仅给出 $P < 0.05$ 是过去计算机软件没有普及的时代靠手工计算、查表判别统计结果时的做法，早该更正。

Endovascular Therapy after Intravenous t-PA versus t-PA Alone for Stroke

Joseph P. Broderick, M.D., Yuko Y. Palesch, Ph.D., Andrew M. Demchuk, M.D., Sharon D. Yeatts, Ph.D., Pooja Khatri, M.D., Michael D. Hill, M.D., Edward C. Jauch, M.D., Tudor G. Jovin, M.D., Bernard Yan, M.D., Frank L. Silver, M.D., Rüdiger von Kummer, M.D., Carlos A. Molina, M.D., Bart M. Demaerschalk, M.D., Ronald Budzik, M.D., Wayne M. Clark, M.D., Osama O. Zaidat, M.D., Tim W. Malisch, M.D., Mayank Goyal, M.D., Wouter J. Schonewille, M.D., Mikael Mazighi, M.D., Ph.D., Stefan T. Engelker, M.D., Craig Anderson, M.D., Ph.D., Judith Spilker, R.N., B.S.N., Janice Carrozzella, R.N., B.A., R.T.(R), Karla J. Ryckborst, R.N., B.N., L. Scott Janis, Ph.D., Renée H. Martin, Ph.D., Lydia D. Foster, M.S., and Thomas A. Tomsick, M.D., for the Interventional Management of Stroke (IMS) III Investigators

ABSTRACT

BACKGROUND

Endovascular therapy is increasingly used after the administration of intravenous tissue plasminogen activator (t-PA) for patients with moderate-to-severe acute ischemic stroke, but whether a combined approach is more effective than intravenous t-PA alone is uncertain.

METHODS

We randomly assigned eligible patients who had received intravenous t-PA within 3 hours after symptom onset to receive additional endovascular therapy or intravenous t-PA alone, in a 2:1 ratio. The primary outcome measure was a modified Rankin scale score of 2 or less (indicating functional independence) at 90 days (scores range from 0 to 6, with higher scores indicating greater disability).

RESULTS

The study was stopped early because of futility after 656 participants had undergone randomization (434 patients to endovascular therapy and 222 to intravenous t-PA alone). The proportion of participants with a modified Rankin score of 2 or less at 90 days did not differ significantly according to treatment (40.8% with endovascular therapy and 38.7% with intravenous t-PA; absolute adjusted difference, 1.5 percentage points; 95% confidence interval [CI], -6.1 to 9.1, with adjustment for the National Institutes of Health Stroke Scale [NIHSS] score [8–19, indicating moderately severe stroke, or ≥ 20 , indicating severe stroke]), nor were there significant differences for the predefined subgroups of patients with an NIHSS score of 20 or higher (6.8 percentage points; 95% CI, -4.4 to 18.1) and those with a score of 19 or lower (-1.0 percentage point; 95% CI, -10.8 to 8.8). Findings in the endovascular-therapy and intravenous t-PA groups were similar for mortality at 90 days (19.1% and 21.6%, respectively; $P=0.52$) and the proportion of patients with symptomatic intracerebral hemorrhage within 30 hours after initiation of t-PA (6.2% and 5.9%, respectively; $P=0.83$).

CONCLUSIONS

The trial showed similar safety outcomes and no significant difference in functional independence with endovascular therapy after intravenous t-PA, as compared with intravenous t-PA alone. (Funded by the National Institutes of Health and others; ClinicalTrials.gov number, NCT00359424.)

The authors' affiliations are listed in the Appendix. Address reprint requests to Dr. Broderick at the University of Cincinnati Neuroscience Institute, Department of Neurology, 260 Stetson St., Suite 2300, PO Box 670525, Cincinnati, OH, 45267-0525, or at joseph.broderick@uc.edu.

This article was published on February 7, 2013, at NEJM.org.

N Engl J Med 2013;368:893-903.
DOI: 10.1056/NEJMoa1214300
Copyright © 2013 Massachusetts Medical Society.

STATISTICAL ANALYSIS

Participants were randomly assigned in a 2:1 ratio to endovascular therapy or intravenous t-PA alone with the use of an Internet-based, computerized algorithm of minimization and the biased-coin method, which accounted for two factors: clinical center and baseline NIHSS strata (scores of 8 to 19 vs. ≥ 20).²³ We calculated that a sample of 900 patients would provide an effect size of 10 percentage points (the absolute difference between the endovascular-therapy and intravenous t-PA groups in the proportion of participants with a modified Rankin score of ≤ 2 at 90 days), assuming that 40% of the patients had a good outcome in the intravenous t-PA group, as noted in those patients in the NINDS rt-PA Stroke Study who had age and baseline stroke severity similar to the eligibility criteria for the IMS III trial^{1,12}; type 1 and type 2 error probabilities of 0.05 (two-sided) and 0.20, respectively; an inflation factor of 1.03 to account for a noncompliance rate of approximately 2%; and the O'Brien and Fleming-type alpha-spending function²⁴ for three interim efficacy analyses.

The prespecified criterion for futility was based on conditional power of less than 20% under the alternative hypothesis. The primary efficacy hypothesis was assessed with the use of the Cochran–Mantel–Haenszel test, with adjustment for the dichotomized baseline NIHSS score, and the weights of the Cochran–Mantel–Haenszel test were applied in the estimation of the risk difference.²⁵ At both the interim and the final analyses, an unfavorable outcome (defined as a modified Rankin score of >2) was imputed for participants who had missing data for the primary outcome or for whom data on the primary outcome were obtained outside the specified window. For all analyses of predefined secondary outcomes and subgroup and safety analyses, each test was conducted at a two-sided alpha level of 0.01. Prespecified subgroup analyses included NIHSS strata, time from symptom onset to treatment (intravenous t-PA and endovascular therapy), presence or absence of arterial occlusion on CT angiography at baseline, age, sex, and presence or absence of atrial fibrillation. For the analysis of raw modified Rankin scores, we used the van Elteren test.²⁶

衰弱对住院老年冠心病患者短期预后的影响：前瞻性队列研究

张 宁¹, 朱文玲², 刘晓红¹, 陈 伟³, 康军仁³, 朱鸣雷¹, 田 然²

中国医学科学院 北京协和医学院 北京协和医院 ¹老年医学科 ²心内科 ³临床营养科, 北京 100730

通信作者: 朱文玲 电话: 010-69154063, E-mail: zhuwenling@263.net

【摘要】目的 探讨衰弱对住院老年冠心病患者短期预后的影响。**方法** 前瞻性收集并分析 2017 年 12 月至 2018 年 11 月在北京协和医院住院治疗的老年冠心病患者临床资料。根据是否合并衰弱, 将患者分为衰弱组和非衰弱组。对两组患者随访, 终点事件包括非常规就诊、主要不良心脑血管事件 (major adverse cardiac and cerebral events, MACCE)、全因死亡。采用多因素 Cox 回归分析衰弱与冠心病患者预后的关系。绘制两组无 MACCE 的 Kaplan-Meier 生存曲线, 并采用 Log-Rank 检验进行比较。**结果** 共 345 例符合纳入和排除标准的老年冠心病患者入选本研究, 包括稳定性冠心病 250 例, 急性冠状动脉综合征 95 例。衰弱组 74 例 (21.4%), 其中轻度衰弱 38 例、中度衰弱 36 例, 非衰弱组 271 例 (78.6%)。中位随访时间 351 (300, 394) d, 失访 3 例。与非衰弱组比较, 衰弱组非常规就诊发生率 (36.1% 比 21.5%)、全因死亡率 (11.1% 比 4.1%) 均升高 (P 均 <0.05), MACCE 发生率 (9.7% 比 4.8%) 无显著差异 ($P>0.05$)。多因素 Cox 回归分析结果显示, 轻度和中度衰弱是稳定性冠心病患者全因死亡的危险因素 ($HR=4.169$, 95% CI: 1.055~16.474, $P=0.042$), 对其非常规就诊 ($HR=1.704$, 95% CI: 0.947~3.066, $P=0.075$)、MACCE ($HR=1.268$, 95% CI: 0.331~4.863, $P=0.729$) 无显著影响。在急性冠状动脉综合征患者中, 轻度和中度衰弱对其非常规就诊 ($HR=1.159$, 95% CI: 0.342~3.924, $P=0.812$)、MACCE ($HR=0.822$, 95% CI: 0.092~7.369, $P=0.861$) 及全因死亡 ($HR=1.445$, 95% CI: 0.210~9.964, $P=0.708$) 均无显著影响。Kaplan-Meier 生存曲线显示, 衰弱组和非衰弱组患者的无 MACCE 生存曲线无显著差异 ($P>0.05$)。**结论** 住院老年冠心病患者合并轻中度衰弱的比率较高, 其可能增加稳定性冠心病患者近期死亡风险。

【关键词】 老年; 冠心病; 衰弱; 预后

【中图分类号】 R543.3 **【文献标志码】** A **【文章编号】** 1674-9081(2021)01-0059-08

DOI: 10.12290/xhyxzz.20190150

Impact of Frailty on the Short-term Prognosis of Hospitalized Elderly Patients with Coronary Heart Disease: A Prospective Cohort Study

ZHANG Ning¹, ZHU Wen-ling², LIU Xiao-hong¹, CHEN Wei³, KANG Jun-ren³, ZHU Ming-lei¹, TIAN Ran²

¹Department of Geriatrics, ²Department of Cardiology, ³Department of Clinical Nutrition, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100730, China

Corresponding author: ZHU Wen-ling Tel: 86-10-69154063, E-mail: zhuwenling@263.net

1.3 样本量计算

一项前瞻性临床研究显示合并衰弱的住院老年患者出院 180 d 全因再住院率为 55%, 而不合并衰弱者为 32%^[11]。应用临床研究样本量计算器 (CRESS V1.3) 软件^[12], 估算衰弱患者样本量为 53 例, 非衰弱患者为 162 例, 共 215 例。

1.4 质量控制

参与本研究的评估者均为接受过 CGA 培训的老年科医生, 并在研究前进行测量方式和评估方法的标准化培训。

1.6 统计学处理

采用 SPSS 24.0 统计分析软件进行统计分析。BMI、PA 为正态分布计量资料, 以均数±标准差表示, 组间比较采用 t 检验; 年龄、MNA-SF 评分、CCI 等为偏态分布计量资料, 以中位数 (四分位数) 表示, 组间比较采用 Wilcoxon 秩和检验; 性别、有/无跌倒史、有/无尿失禁等计数资料以频数 (百分比) 表示, 组间比较采用 χ^2 检验或 Fisher 精确概率法。采用 Cox 回归分析衰弱对老年冠心病患者发生各终点事件的影响; 无 MACCE 生存曲线的绘制采用 Kaplan-Meier 法, 比较采用 Log-Rank 检验。双侧检验, 以 $P<0.05$ 为差异有统计学意义。

1.2 研究方案

1.2.1 样本量估算过程 使用 Bland-Altman 方法进行一致性评价目前尚无公认的样本量计算方法, Bland-Altman 方法的创始人 BLAND 和 ALTMAN 本人也只是经验性地推荐认为其样本量不应少于 100 例^[13], 本文采用基于 Monte-Carlo 模拟的方法计算样本量^[14], 公式如下: $n = (2 + Z_{1-\alpha/2}^2) (t_{\alpha/2, n-1} + t_{\beta/2, n-1})^2 S^2 d / 2 (\bar{d} + Z_{1-\alpha/2} Sd - \delta)^2$, 公式基于的 H_0 是两种方法的检测差值为 0, 公式中 t 分位数与样本量有关, 需要采用迭代法, 第一次迭代时使用标准正态分布替代 t 分布获得初始 n 值, 然后代入这个 n 值计算相应的 t 分位数计算相应的 n 值, 依次迭代下去, 直到迭代前后两次结果无明显差异时结束, 以此时的结果作为最终 n 值, 即样本量。关于不同无创脑氧饱和度检测设备对比研究的相关文献甚少, 通过预实验我们发现, 两种仪器脑氧饱和度监测读数差值均数为 -8.77, 通常临床上可接受的脑氧饱和度指数的差异不超过 5, 故 δ 取 5, 取 -8.77。在检验水准 (α) 0.05, 检验效能 ($1-\beta$) 0.9 的条件下带入公式计算, 在第 4 次迭代时可得到稳定的样本量 $n = 147.0494$, 向上取整为 148, 再考虑共 15% 的脱落和退出率, 可得 $148/0.85 \approx 174.12$, 故本研究招募 174 名受试者。

1.3.3 一致性比较

1.3.3.1 rSO_2 测量值的一致性 由于每种仪器的测量范围不同, 无法直接进行测量值的一致性比较, 需要将数据标准化后进行数据分析。公式为 $z = (X - \mu) / \sigma$, 其中 $X - \mu$ 为离均差, σ 表示标准差。采用 Bland-Altman 方法, 以美敦力 INVOS 5100C 的 rSO_2 作为基准, 分别进行平原和模拟 4 500m 高海拔时 rSO_2 测量值的一致性分析。

1.3.3.2 rSO_2 值下降幅度的一致性 为对比研究 4 种仪器检测受试者从平原至模拟高原环境时 rSO_2 值下降幅度的一致性, 以 INVOS 5100C 的 rSO_2 下降百分比作为标准, 即 $\Delta rSO_2 (\%) = (\text{平原 } rSO_2 - \text{高原 } rSO_2) / \text{平原 } rSO_2 \times 100\%$, 采用 Bland-Altman 方法^[17], 将其他 3 种国产脑氧仪的 $\Delta rSO_2 (\%)$ 与 INVOS 5100C 的 $\Delta rSO_2 (\%)$ 做一致性分析。

1.4 统计学方法

采用 SPSS 25、Medcalc 19.0.7 软件进行统计学分析。计量资料以 $\bar{x} \pm s$ 表示, 正态分布连续变量对比采用 t 检验、方差分析, 非正态分布数据采用非参数检验; 相关性采用 Pearson 检验; 率的比较采用 χ^2 检验, 当数据不符合 χ^2 检验要求时, 采用基于 Monte Carlo 的 Fisher's 确切概率法。使用配对 χ^2 (McNemar's test) 检验^[15] 评估设备之间在检测 rSO_2 的显著差异。一致性比较采用 Bland-Altman 方法。 $P < 0.05$ 认为有统计学意义。

数据的分布类型及转换方式的交代:

血浆 mtDNA 水平与体外循环心脏术后炎症反应的关联性研究

范景秀¹, 胡佳¹, 郭应强¹, 康焯² 610041 成都, 四川大学华西医院: 心脏大血管外科¹, 重症医学科²

[摘要] **目的** 探究体外循环心脏术后 ICU 患者血浆 mtDNA 水平与炎症细胞计数和炎症因子表达之间的潜在联系。**方法** 2015 年 9 月 1 日到 2016 年 12 月 1 日, 华西医院心外科行体外循环心脏手术患者, 返回 ICU 半小时内采集血常规、肝功和肾功、NT-proBNP、IL-6、CRP、PCT 和 mtDNA。规范完成检测和结果判定及录入, SPSS21.0 进行统计分析。**结果** ①纳入患者 962 名, 男性 459 名, 女性 503 名。②血浆 mtDNA 水平对数值与体外循环心脏术后血小板活化程度对数值之间存在直线相关 ($r=0.1027, P=0.0028$)。③血浆 mtDNA 水平平方根数值与体外循环心脏术后中性粒细胞绝对值平方根数值之间存在直线相关 ($r=0.1069, P=0.0040$)。④血浆 mtDNA 水平对数值与体外循环心脏术后 IL-6 对数值之间存在直线相关 ($r=0.0798, P=0.0453$)。**结论** 血浆 mtDNA 水平与体外循环心脏术后早期患者的全身炎症反应密切相关。体外循环过程中活化炎症细胞可能是血浆 mtDNA 的释放来源。

[关键词] 血浆线粒体 DNA 水平; 体外循环; 炎症反应

[中图分类号] R364.5; R446.11; R654.1 **[文献标志码]** A

Association of plasma mtDNA level with cardiopulmonary bypass-related inflammation in ICU patients

FAN Jingxiu¹, HU Jia¹, GUO Yingqiang¹, KANG Yan² ¹Department of Cardiovascular Surgery, ²Department of Critical Care Medicine, West China Hospital, Sichuan University, Chengdu, Sichuan Province, 610041, China

[Abstract] **Objective** To explore the potential association of plasma mtDNA level with inflammatory cell counts and inflammatory factor levels in ICU patients after cardiac surgery under cardiopulmonary bypass (CPB). **Methods** From September 1, 2015 to December 1, 2016, all the patients who underwent cardiac surgery under CPB at our department of cardiovascular surgery were recruited in this study. Blood routine, liver and renal functions, and the levels of N-terminal brain natriuretic propeptide (NT-proBNP), IL-6, C-reactive protein (CRP), procalcitonin (PCT), and mtDNA were tested and measured within half an hour after cardiac surgery according to the respective requirements. Statistical software SPSS 21.0 was used for statistical analysis. **Results** ① A total of 962 patients were included in this study, including 459 males and

他炎症细胞研究, 并最
机制和致
梗死导致
升高提示
[7]。在脂
浆 mtDNA
异常免疫
大量细胞
症反应关
心脏术后
术后 24 h
积极参与
A 与体外
关系, 有助
关联, 以及

在华西医
患者, 年龄
成人心脏
委员会的批

VPCR×1/VEXT [c 为血浆 mtDNA 浓度; Q 为荧光定量 PCR 检测的样本中 mtDNA 的拷贝数; VDNA 为提取 DNA 过程中最后一步 AE 缓冲液的体积 (200 μL); VPCR 为荧光定量 PCR 配置体系中加入样本的体积 (1 μL); VEXT 为提取 DNA 过程中第一步使用的血浆总量 (50 μL)]。

1.3 统计学方法

经正态性检验, 血浆 mtDNA 水平、血小板活化水平、中性粒细胞绝对值、IL-6 水平、CRP 水平均不服从正态分布。血浆 mtDNA 水平、血小板活化程度、IL-6 水平、经自然对数转换后服从正态分布; 中性粒细胞绝对值经开平方根转换后服从正态分布; CRP 水平经取倒数转换后服从正态分布。经过数据转换后相关性分析采用直线相关, 以皮尔逊积矩相关系数 (Pearson product-moment correlation coefficient, PPMCC 或 PCCs) r 来表示相关程度; 当数据缺失率 < 10% 时, 采用多重填补 (multiple imputation) 方法处理缺失数据; 当数据缺失率 > 10% 时, 该指标不纳入研究。

所有的统计学处理采用 SAS 9.4, R 3.2.4, SPSS 21.0 统计软件。检验水准 $\alpha=0.05$ 。

2 结果

2.1 病例资料基本情况

本研究共纳入 962 名患者, 其中男性 459 名, 女性 503 名, 年龄 18~75.8 (51.6±12.5) 岁; BMI 为 14.8~34.9 (22.9±3.29) kg/m²; 机械通气时间 2~232.5 (29.07±36.09) h; 住 ICU 时间 1~12.15 (2.89±1.79) d。

多因素回归 Multivariate regression :

纳入哪些因子, 应该在统计学分析的基础上结合专业讨论确定

Which factors should be included, more consideration should be given to clinical significance.

仅把单因素分析 $P < 0.05$ 的因子纳入多因素回归不妥

It is not appropriate to include only factors of $P < 0.05$ in univariate analysis into multivariate regression.

2.3 脑瘫临床类型与相关因素的多因素 Logistic 回归分析 将患者产前、产时、产后有统计学意义的 15 个相关因素 ($P < 0.05$) 作为自变量引入 Logistic 回归模型中, 并以脑瘫临床类型为因变量进行多因素分析, 譬如以痉挛性脑瘫为参照, 结果显示, 不随意运动型中黄疸, 肌张力低下型中孕期出血、早产儿、产后窒息、剖宫产分娩及孕期不适当用药, 混合型中孕期接触有毒

2.3 影响宫内妊娠的多因素 logistic 分析

对单因素分析中有统计学意义的变量进行多因素分析, 考虑到近端阻塞和远端阻塞患者治疗方式不同, 分别对近端阻塞和远端阻塞的患者进行分析。

Why?

表 1 不同群体医护人员 SAS、SDS、PSS 得分比较 ($\bar{x} \pm s$)

群体	n	SAS	SDS	PSS
男性	65	39.90 ± 9.82	42.48 ± 11.84	41.77 ± 5.15
女性	186	43.60 ± 11.31	46.13 ± 12.25	43.17 ± 5.78
t/p		2.35/0.02	2.09/0.04	1.73/0.09
已婚	148	42.66 ± 11.03	44.62 ± 11.99	43.11 ± 5.64
未婚	103	42.62 ± 11.11	46.00 ± 12.57	42.59 ± 5.66
t/p		0.03/0.95	0.87/0.41	0.71/0.53
医生	61	39.86 ± 8.65	42.60 ± 11.04	41.84 ± 5.14
护士	179	44.08 ± 11.66	46.67 ± 12.37	43.44 ± 5.70
t/p		2.59/0.01	2.27/0.02	2.04/0.04
本地	145	43.46 ± 10.93	46.23 ± 12.44	43.37 ± 5.59
外援	106	41.53 ± 11.15	43.75 ± 11.84	42.02 ± 5.65
t/p		1.36/0.17	1.59/0.11	1.87/0.06
相关专业	58	42.80 ± 12.06	46.03 ± 12.65	42.64 ± 6.27
非相关专业	193	42.60 ± 10.75	44.93 ± 12.12	42.85 ± 5.46
t/p		0.12/0.90	0.60/0.55	0.26/0.80

注：SAS：焦虑自评量表；SDS：抑郁自评量表；PSS：知觉压力量表；

单因素分析有差异的
未必真有差异

(可能是别的因素间接作用的结果)

univariate analysis: $P < 0.05$
May be an indirect effect of other factors

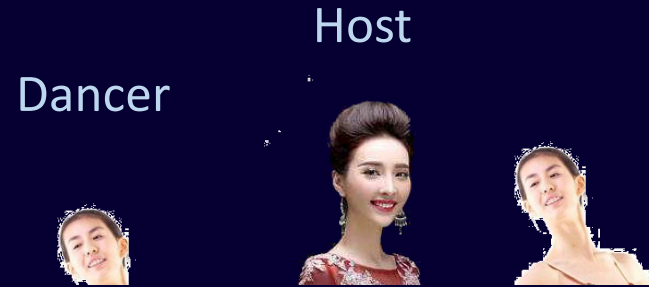
单因素分析没差异的

未必真的没差异

(可能被另一个因素的特征抵消了)

univariate analysis: $P > 0.05$
May be offset by the inverse effect of other factors.

An image description of Multivariate regression:



An image description of Multivariate regression:



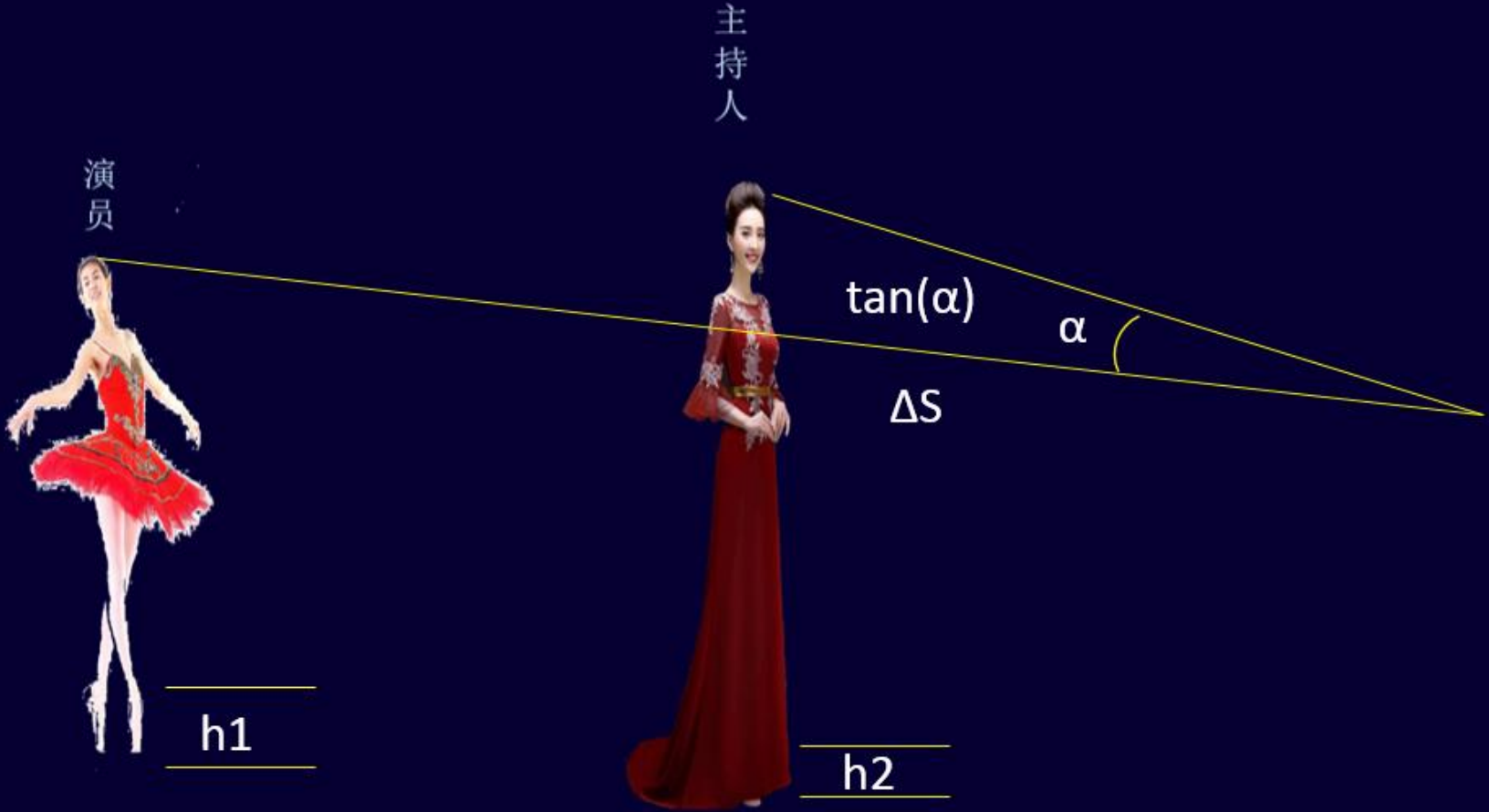
An image description of Multivariate regression:



An image description of Multivariate regression:



An image description of Multivariate regression:



回归与预测模型 Regression and prediction models

注意事项:

- 需要有足够大的样本量 (尤其是因素多的时候)
通常为纳入分析的因子数量的20倍以上 (经验要求)
More than 20 times the number of factors.
- 对连续型自变量最好转换为分类变量再进行分析
Categorical → Categorical
- 理论驱动的数据分析。回归分析也不一定说明因果关系,
因果关系更多是通过专业判断。
Theory driven, not data driven.

3 解读 About the Interpretation

医学论文对于统计学结果解读中应该注意的问题：
Statistical significance vs. Clinical significance

- 有统计学差异未必有专业意义上的差异
- 回归分析的自变量与因变量，在专业上要有影响的可能性 $X \rightarrow Y$
- cut-off值的选取要结合具体的临床问题
(Cost of missed diagnosis vs. Cost of misdiagnosis)
- 注意定性诊断敏感度/特异度 与 阳性预测值/阴性预测值 代表的含义
- 大数据、人工智能、机器学习等新方法应用时一定要紧密结合医学

有统计学差异未必有专业意义上的差异

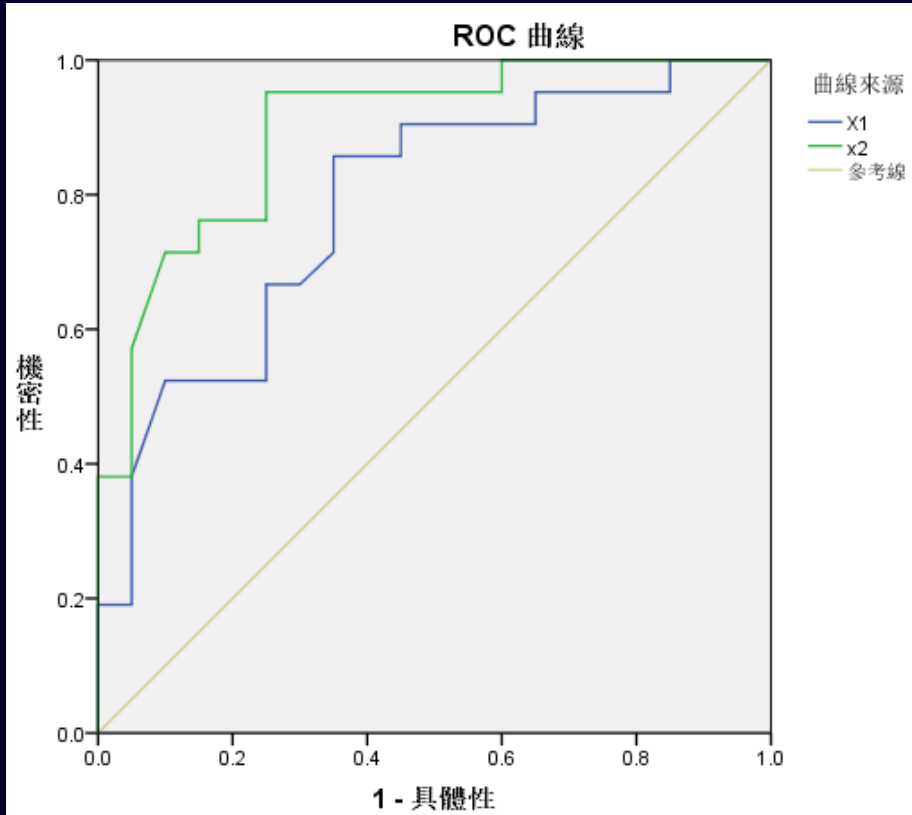
statistical significance vs. clinical significance

	grp	h
37032	0	175.27
37033	0	174.43
37034	0	174.30
37035	0	175.44
37036	0	175.71
37037	0	174.08
37038	0	174.66
37039	0	175.36
37040	0	174.97
37041	0	174.98
37042	0	174.73
37043	1	175.27
37044	1	174.63
37045	1	174.97
37046	1	176.37
37047	1	175.64
37048	1	175.56
37049	1	175.65
37050	1	176.02
37051	1	175.43
37052	1	176.03
37053	1	175.59
37054	1	175.12
37055	1	175.00
37056	1	175.38
37057	1	175.15

grp	N	平均數	標準偏差	標準錯誤平均值
h 0	37042	174.9958	.57755	.00300
1	37042	175.5006	.57633	.00299

		Levene 的變異數相等測試		針對平均值是否相等的 t 測試				
		F	顯著性	T	df	顯著性 (雙尾)	平均差異	標準誤差
h	採用相等變異數	.305	.581	-119.077	74082	.000	-.50481	.00424
	不採用相等變異數			-119.077	74081.667	.000	-.50481	.00424

注意正确解读ROC分析的结果



曲線下的區域

測試結果變數	區域圖	標準錯誤 ^a	漸進顯著性 ^b	漸進 95% 信賴區間	
				下限	上限
X1	.786	.072	.002	.645	.926
x2	.896	.049	.000	.801	.992

測試結果變數: X1, x2 在正數實際狀態與負數實際狀態群組之間至少有一個連結空間。統計資料可能有偏差。

a. 在非參數式假設下

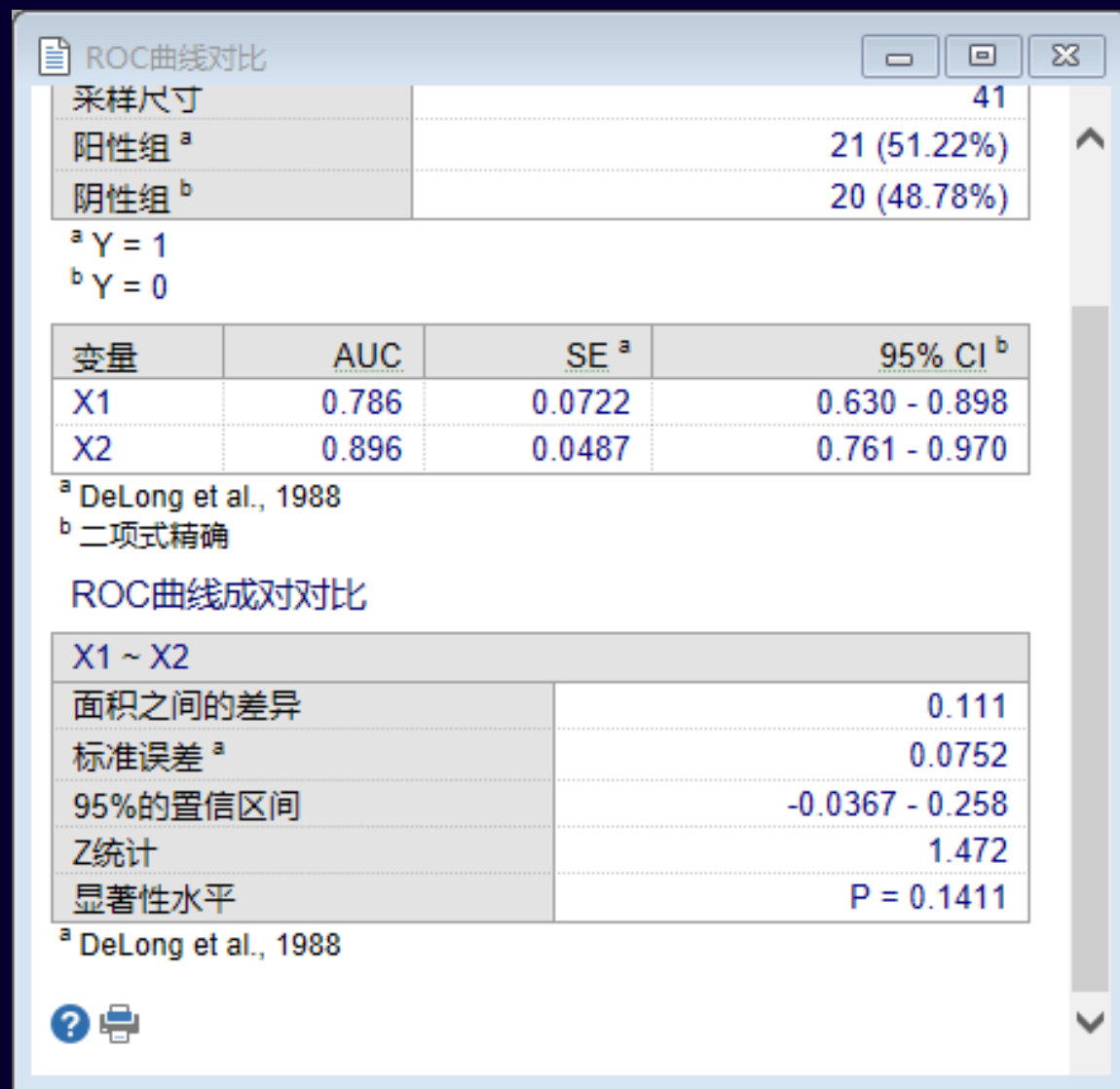
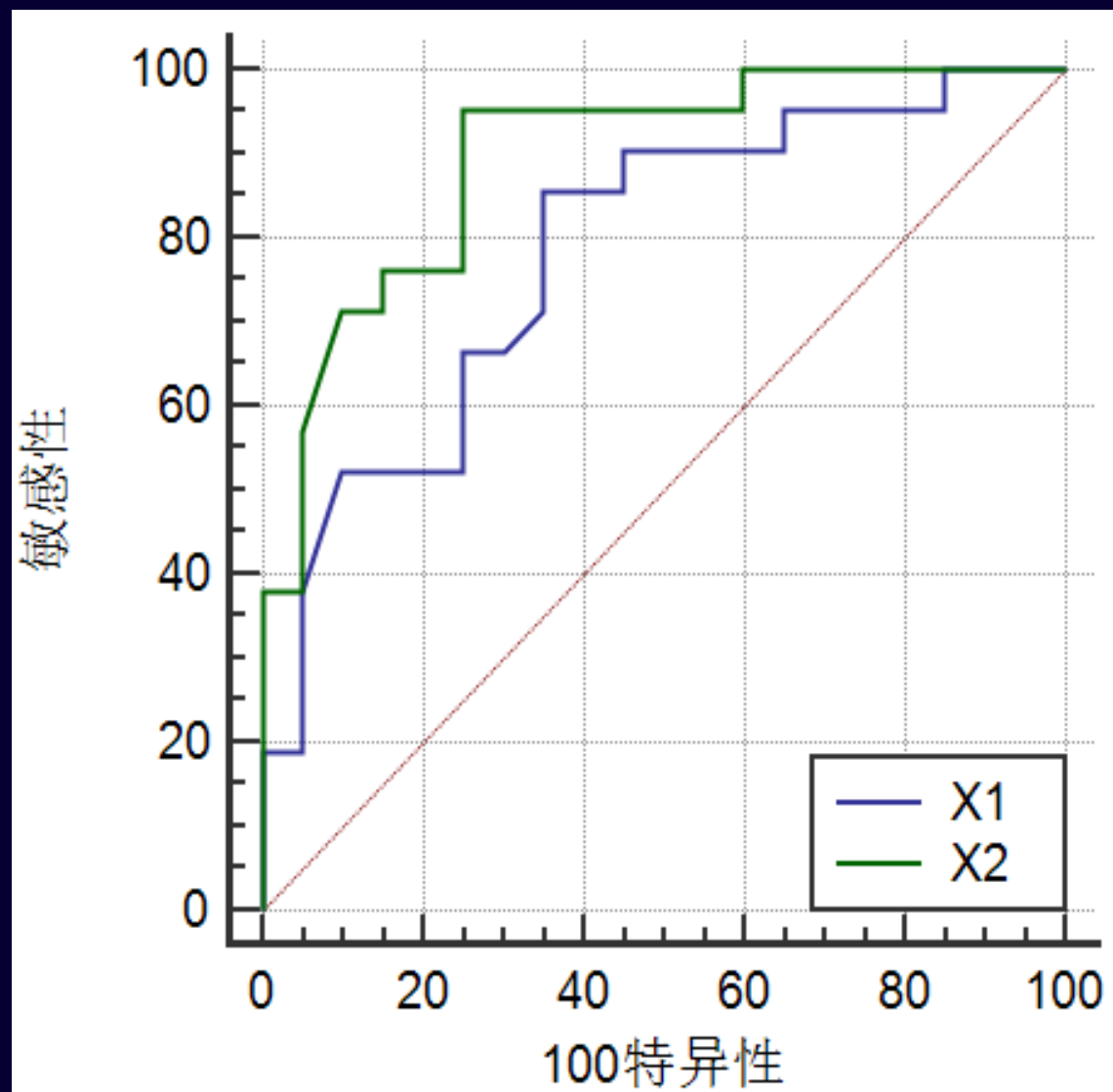
b. 空值假設: true 區域 = 0.5

SPSS中ROC曲线分析的P值是指的是该AUC

与0.5的参考线比较的P值，不一定说明该诊断方法的优劣

若要做多条ROC曲线的AUC的比较，需要使用 MedCalc

注意正确解读ROC分析的结果



注意正确解读ROC分析的结果

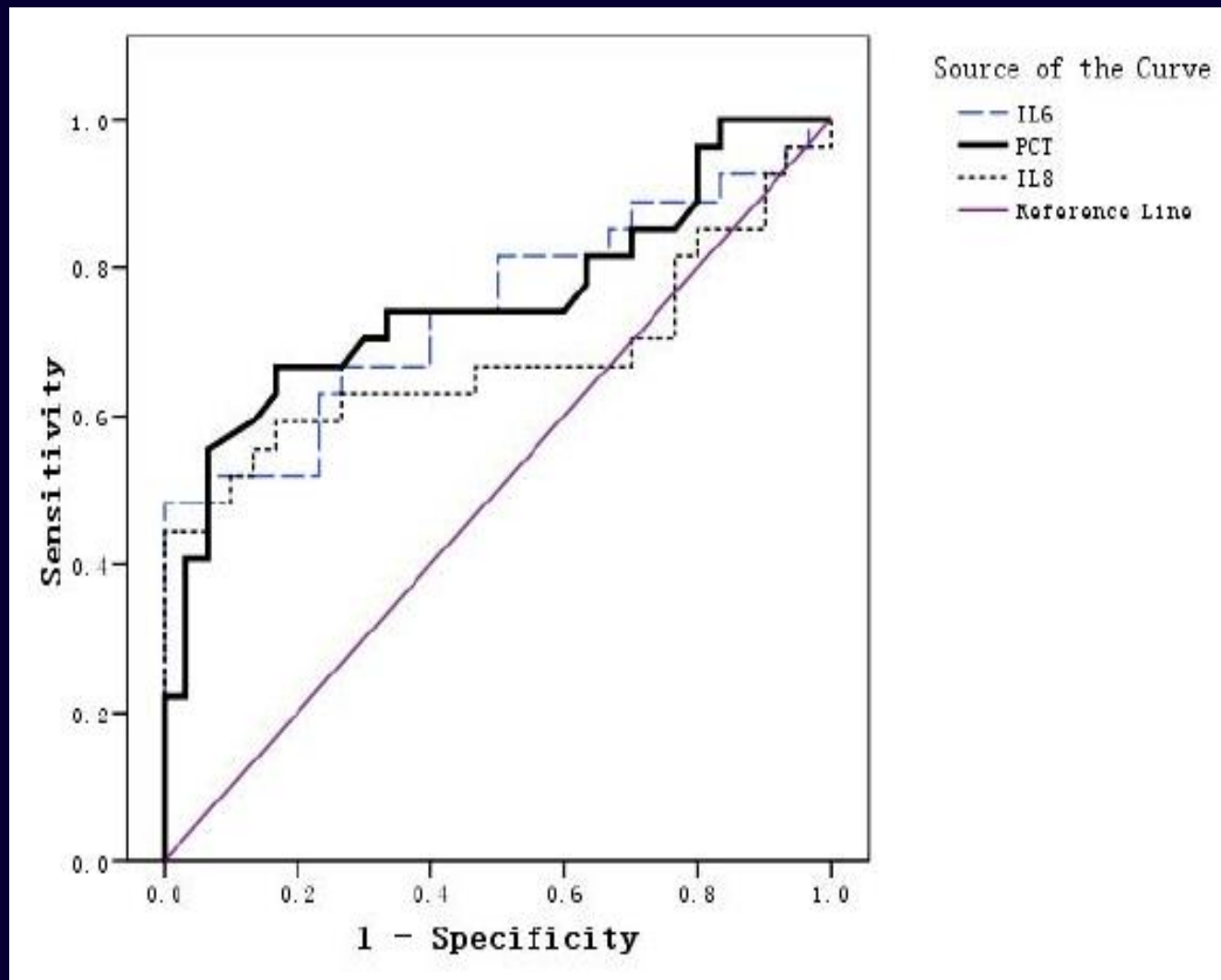
【例】应用不当的例子：

SPSS软件做ROC分析给出的P值是与0.5参考线相比的P值，通常没有什么临床上的实际意义

2.2 两组患者血液中 PCT 阳性检出率比较

ROC 曲线面积下两组患者中的 IL-6 及 IL-18 与 PCT 含量对比显示其 3 种含量对急性重症胰腺炎继发感染检出率较高，且灵敏度、阳性预测值和阴性预测值准确率均达到最高值 (AUC=0.632、0.571、0.689, P=0.001)。

见图 1。



注意正确解读ROC分析的结果

【例】应用不当的例子：

不同的ROC曲线的AUC之间没有做假设检验，无法说明AUC大小之间的差别是否有统计学意义

表5 APACHE II评分、SOFA评分、血清CC-16、MCP-1浓度对ARDS患者转归的推断价值

项目	最佳截断值	AUC	灵敏度	特异度	约登指数
APACHE II评分	19.60 (分)	0.808	78.57	68.97	47.54
SOFA评分	14.79 (分)	0.684	66.67	67.81	34.48
血清CC-16浓度	20.26($\mu\text{g/L}$)	0.911	90.48	80.46	70.94
血清MCP-1浓度	32.12(ng/L)	0.889	80.95	90.80	71.76

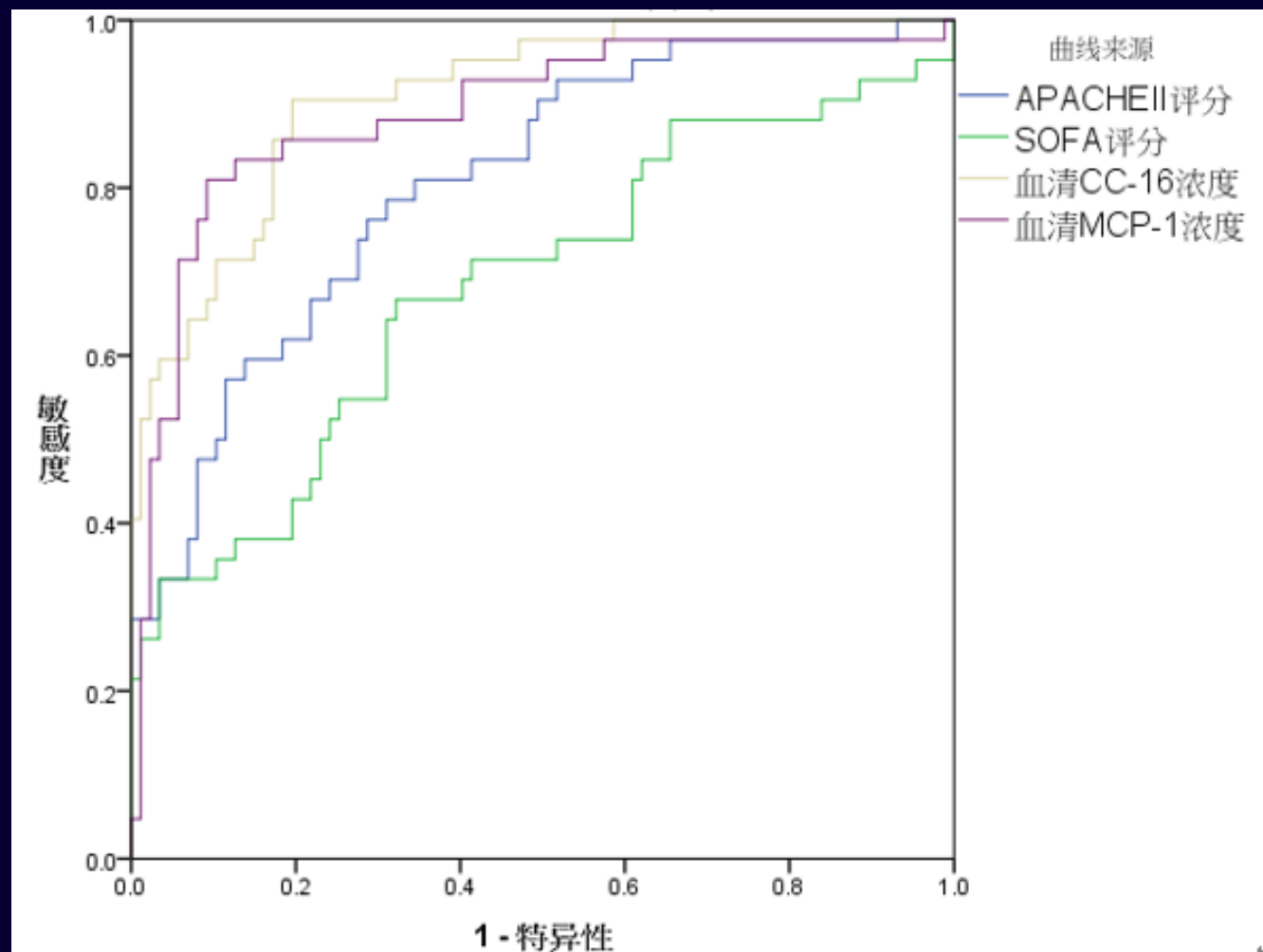


图1 APACHE II评分、SOFA评分、血清CC-16、MCP-1浓度对ARDS患者转归的ROC曲线

诊断研究几个参数的意义

【例】评价CT诊断原发性肝癌的准确性

Kappa检验
及一系列衡量诊
断效能的指标:

CT	活检（金标准）		
	+（患者）	-（非患者）	
+	29 a	6 b	a+b
-	2 c	48 d	c+d
	a+c	b+d	

敏感性 (SEN) = $a/a+c$

特异性 (SPE) = $d/b+d$

阳性预测值 (PPV) = $a/a+b$

阴性预测值 (NPV) = $d/c+d$

阳性似然比 (+LR) = $(a/a+c)/(b/b+d)$

阴性似然比 (-LR) = $(d/b+d)/(c/a+c)$

“病人里你能正确检出多少”

“非病人里你能正确检出多少”

“查出了+结果，患病的概率多少？”

“查出了-结果，不患病的概率多少？”

患者得到+结果的概率是健康人得到+结果的多少倍？

健康人得到-结果的概率是患者得到-结果的多少倍？

应用思想：用频率来推论概率

回归与预测模型

【例】Logistic回归示例：
小儿哮喘的危险因素

方程中的变量							
		B	S.E.	Wals	df	Sig.	Exp (B)
步骤 1 ^a	母亲过敏史	.384	2.182	.031	1	.860	1.468
	是否母乳喂养	-4.483	1.488	9.072	1	.003	.011
	常里	2.241	1.092	4.210	1	.040	9.405

a. 在步骤 1 中输入的变量: 母亲过敏史, 是否母乳喂养.

结果解读： e^B 即 OR (比值比/优势比) (发病率较低时，OR可以近似RR)

1、母亲过敏史、是否母乳喂养是小儿患哮喘的独立影响因素

2、母亲有过敏史的小儿患哮喘的风险

是母亲没有过敏史的小儿的 $e^{0.384}$ 倍(约1.5倍)；

母乳喂养的小儿患哮喘的风险

是非母乳喂养的小儿的 $e^{-4.483}$ 倍(约1%)

cut-off值的选取要结合具体的临床问题（漏诊还是误诊后果更严重？）

A Laparoscopy-Based Score To Predict Surgical Outcome in Patients With Advanced Ovarian Carcinoma: A Pilot Study

Anna Fagotti, MD,¹ Gabriella Ferrandina, MD,² Francesco Fanfani, MD,¹ Alfredo Ercoli, MD,² Domenica Lorusso, MD,³ Marco Rossi, MD,³ and Giovanni Scambia, MD¹

¹Division of Gynaecologic Oncology, Catholic University of the Sacred Heart, Largo A. Gemelli 1, 86100 Campobasso, Italy

²Division of Gynaecologic Oncology, Catholic University of the Sacred Heart, Largo A. Gemelli 8, 00168 Rome, Italy

³Department of Anaesthesiology and Intensive Care, Catholic University of the Sacred Heart, Largo A. Gemelli 1, 86100 Campobasso, Italy

Background: Our objective was to set up a more objective quantitative laparoscopy-based model in predicting the chances of optimal cytoreductive surgery in advanced ovarian cancer patients.

Methods: Sixty-four advanced ovarian cancer patients were submitted to both laparoscopy and standard longitudinal laparotomy sequentially, to define the chances of optimal debulking surgery (residual disease ≤ 1 cm). Three patients could not be evaluated by laparoscopy because of the presence of multiple and tenacious adhesions. Sensitivity, specificity, positive predictive value, negative predictive value, and overall accuracy were calculated for each laparoscopic parameter. On the basis of the statistical probability of each factor to predict surgical outcome, seven laparoscopic features were selected for inclusion in the final model. Each parameter was assigned a numerical score based on the strength of statistical association, and a total predictive index value was tabulated for each patient. Receiver operating characteristic curve analysis was used to assess the ability of the model to predict surgical outcome.

Results: After debulking surgery, 41 (67.2%) of 61 patients were left with optimal residual disease. The presence of omental cake, peritoneal carcinosis, diaphragmatic carcinosis, mesenteric retraction, bowel and/or stomach infiltration, and liver metastases satisfied the basic inclusion criteria and were assigned a final predictive index value of 2. In the final model, a predictive index score ≥ 8 identified patients undergoing suboptimal surgery with a specificity of 100%. The positive predictive value was 100%, and the negative predictive value was 70%.

Conclusions: The reliability of laparoscopy in assessing the chance of optimal cytoreduction can be improved by using a simple scoring system.

Key Words: Predictive index score—Laparoscopy—Ovarian cancer—Cytoreduction.

presence or absence of some specific features and ranged from 0 to 12 points (Fig. 1). Individual PIVs demonstrated the largest concentration of cases in the group with a predictive index of 0 ($n = 30$). The sensitivity, specificity, PPV, and NPV and the accuracy of each PIV of 0 through 10 are shown in Table 2. The discriminating performance of the PIV was confirmed by a receiver operating characteristic curve analysis in which the area under the curve was .797 (data not shown).

The overall accuracy rate of the model in predicting surgical outcome was approximately 75% (range, 74%–77%), corresponding to a wide range of PIV (from ≥ 2 to ≥ 8). In this situation, sensitivity and specificity ranged from 74% to 30% and from 76% to 100%, respectively. In other words, the percentage of cases correctly identified as suboptimal surgery (sensitivity) increased with the PIV cutoff level, and the likelihood that a patient would have a suboptimal surgical result (PPV) increased to 100% in the case of a PIV ≥ 8 ($P = .00056$). However, the highest NPV, that is, the likelihood of having an optimal cytoreduction, was 83% and corresponded to a PIV of 0 or 2.

cut-off值的选取要结合具体的临床问题（漏诊还是误诊后果更严重？）

TABLE 2. Predictive index model

PIV	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (%)	Unnecessarily explored (%)	Inappropriately unexplored (%)
0	78	66	58	83	70	17	42
≥2	74	76	65	83	75	17	35
≥4	61	87	74	79	77	21	26
≥6	43	95	83	73	75	27	17
≥8	30	100	100	70	74	30	0
≥10	17	100	100	66	69	34	0

PIV, predictive index value; PPV, positive predictive value; NPV, negative predictive value.

在这个评分系统中，发生“inappropriately unexplored”（本来能在手术中获益而没有进行手术）的后果显然比“unnecessarily explored”（不必要的手术）严重的多，所以，此处就选择了PPV刚好能达到100%处（评分 ≥ 8 分）作为诊断界值，而没有选取约登指数最大处（ ≥ 2 ）作为诊断界值

大数据、机器学习等新方法应用中的问题

例：某论文研究肝硬化并发肝性脑病的预测模型

表1 肝性脑病组和对照组基线特征

Tab1 The baseline characteristics of hepatic encephalopathy group and control group

指标	肝性脑病组 N=285	对照组 N=1213	统计值	p 值
年龄(岁)	63.00(54.00,69.00)	56.00(49.00,66.00)	-7.061	<0.001
性别(男)	201(70.53%)	817(67.35%)	0.926	0.336
吸烟史	149(52.28%)	530(43.69%)	6.525	0.011
饮酒史	151(52.98%)	509(41.96%)	10.929	<0.001
住院天数	11.00(6.00,18.00)	10.00(6.00,16.00)	-1.944	0.052
消化道出血	61(21.40%)	232(19.13%)	0.622	0.430
丙肝	16(5.61%)	59(4.86%)	0.138	0.710
酒精肝	84(29.47%)	121(9.98%)	72.637	<0.001
自身免疫性肝病	27(9.48%)	101(8.33%)	5.438	0.020
Hb(g/L)	95.00	107.00	-5.843	<0.001
	(76.00,114.00)	(85.00,128.00)		
总蛋白(g/L)	64.60(58.70,70.40)	65.70(59.20,72.20)	-2.325	0.020
总胆红素(umol/L)	43.60(25.50,77.70)	25.00(15.90,49.00)	-8.496	<0.001
中性粒细胞百分比(%)	73.00(62.90,80.40)	66.30(57.40,75.20)	-6.008	<0.001
直接胆红素(umol/L)	21.60(12.30,48.00)	12.10(7.00,26.80)	-8.090	<0.001
血小板计数($\times 10^9/L$)	65.00(43.00,112.00)	71.00(49.00,109.00)	-1.575	0.115
血红蛋白总量(g/L)	95.00(76.00,114.00)	107.00(85.00,128.00)	-5.843	<0.001
天门冬氨酸氨基转移酶(TU/L)	48.00(33.00,70.00)	50.00(31.00,94.00)	-1.066	0.286
全血钠(mmol/L)	137.60(133.30,141.60)	140.00(137.00,142.00)	-5.888	<0.001
全血钾(mmol/L)	3.91(3.57,4.37)	3.87(3.58,4.18)	-1.837	0.066
白蛋白(g/L)	28.80(26.00,31.40)	32.30(28.10,36.60)	-9.277	<0.001
凝血酶原活性(%)	51.00(41.00,62.00)	69.00(58.00,81.80)	-14.914	<0.001
凝血酶原标准化值	1.54(1.36,1.85)	1.30(1.17,1.48)	-12.896	<0.001
肌酐(mmol/L)	6.36(4.83,10.40)	5.32(4.17,9.27)	-7.504	<0.001

表3 回归分析

指标	95%CI	P 值
酒精肝	(2.405,5.065)	<0.001
年龄	(1.040,1.069)	<0.001
凝血酶原活动度	(0.919,0.941)	<0.001
尿素氮	(1.070,1.139)	<0.001
常数	-	0.011

消化道出血是临床上公认的肝性脑病的危险因素，不能简单地因为P值大，就不考虑进去

消化道出血差异不显著的原因值得分析：

资料中数据可能未反映实际情况（消化道出血隐匿性强）血红蛋白的下降很可能反映了这一点，但模型也未纳入血红蛋白

- 本研究先做Logistic回归→再用LASSO筛选→再用Logistic建模，这种做法其实值得商榷
- 预测模型的研究应该考虑临床应用实际

大数据、机器学习等新方法应用中的问题

基于机器学习算法鉴别儿童发热待查病因的分类模型研究

1. 2. 3. ...

400000 重庆, 医学信息学院¹; 400065 重庆, 大学数据工程与可视计算重点实验室²; 400014 重庆, 大数据智能应用重庆市高校工程研究中心³

[摘要] 目的 筛选鉴别儿童发热待查病因为感染性疾病或非感染性疾病的独立危险因素并建立机器学习算法模型, 为病因鉴别提供计算机辅助诊断方法。方法 通过医院临床大数据平台, 收集重庆医科大学附属儿童医院 2015 年 1 月至 2019 年 12 月符合纳排标准的 873 例发热待查住院患儿的临床资料。采用 SPSS 26.0 软件进行单因素分析和 Logistic 回归分析并筛选出独立危险因素, 以此分别构建 Logistic 回归、BP 神经网络(back propagation neural network, BPNN)、支持向量机(support vector machine, SVM)、随机森林(random forest, RF)和 LightGBM 5 种机器学习算法模型。计算并比较 5 种模型的准确率、ROC 曲线下面积(AUROC)和 F1 值 3 个指标。结果 研究的有效样本共 873 例, 单因素分析结果显示 41 项指标差异有统计学意义($P < 0.05$), 采用 Logistic 回归分析筛选出年龄、钠离子、氯离子、乳酸脱氢酶、球蛋白、红细胞压积、C 反应蛋白、白细胞酯酶共 8 项独立危险因素, 基于以上因素建立 5 种机器学习算法模型。Logistic 回归、BPNN、SVM、RF、LightGBM 的准确率依次为 0.748、0.769、0.803、0.800、0.784, ROC 曲线下面积(AUROC)为 0.751、0.784、0.806、0.776、0.796, F1 值为 0.856、0.864、0.877、0.873、0.871, 相比之下, SVM 的各评价指标均略高于其他算法。结论 年龄、钠离子、氯离子等是鉴别儿童发热待查病因为感染性疾病或非感染性疾病的独立危险因素, 基于独立危险因素建立了 Logistic 回归、BPNN、SVM、RF 和 LightGBM 这 5 种机器学习算法模型, 其中 SVM 的分类预测效果更准确。获得患儿入院时常用的各项相关指标, 即可较准确地预测病因类型, 为临床进一步明确诊断提供决策依据。

表 1 儿童 FUO 相关指标的单因素分析

指标	感染性疾病组 (653)	非感染性疾病组 (220)	Z/ χ^2	P
男性[例(%)]	380(58.19)	139(63.18)	1.70	0.192
年龄 (岁)	2.91 (1.18,6.77)	5.76 (2.38,9.46)	-5.64	<0.001
生化指标				
总钙 (mmol/L)	2.37 (2.26,2.47)	2.27 (2.15,2.40)	-6.57	<0.001
镁 (mmol/L)	0.91 (0.86,0.97)	0.90 (0.84,0.96)	-1.32	0.185
钠离子 (mmol/L)	139.1 (137.20,141.00)	138.20 (135.80,140.30)	-3.43	0.001
钾离子 (mmol/L)	4.33 (4.02,4.62)	4.24 (3.93,4.57)	-2.19	0.029
氯离子 (mmol/L)	102.55 (100.20,104.80)	101.20 (98.70,103.60)	-4.92	<0.001
磷 (mmol/L)	1.55 (1.36,1.71)	1.51 (1.29,1.67)	-1.98	0.048
乳酸脱氢酶 (U/L)	275.70 (228.50,338.00)	316.50 (241.50,502.58)	-4.74	<0.001
谷草转氨酶 (U/L)	32.00 (24.00,44.80)	36.95 (24.90,57.83)	-2.76	0.006
尿酸 ($\mu\text{mol/L}$)	266.70 (171.00,216.00)	219.35 (168.63,279.75)	-0.18	0.860
尿素 (mmol/L)	3.20 (2.41,4.06)	3.25 (2.40,4.10)	-0.21	0.830
肌酐 ($\mu\text{mol/L}$)	27.00 (22.00,35.00)	28.50 (22.98,37.58)	-1.41	0.158
总胆红素 ($\mu\text{mol/L}$)	4.80 (3.10,7.20)	4.85 (3.38,7.13)	-0.48	0.633
球蛋白 (g/L)	24.80 (20.63,29.10)	28.80 (24.25,34.23)	-7.64	<0.001

研究工作中应用数据统计分析应注意什么

- 1、往往是理论驱动而非数据驱动
Theory driven, not data driven.
- 2、从研究目的入手看待统计学分析方法的应用
Focus on research purposes
- 3、在论文中报告方法和数据宜尽量详细（体现专业性和真实性）
Be more detailed.
- 4、数据分析结果一定要结合专业临床意义探讨
Closely integrated with clinical significance

真实的数据
规范的方法
准确的解读

《第三军医大学学报》欢迎赐稿：

<http://aammt.tmmu.edu.cn/>

