临床试验的统计学设计与分析

研师：赵杨 南京医科大学

时间：2018-10-25

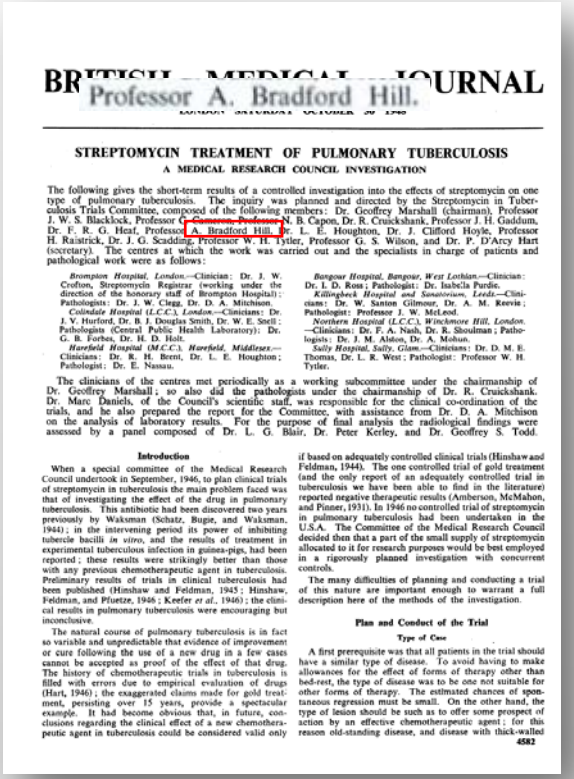
金玉良研
JINYULIANGYAN

目录

统计学在临床试验中的价值

案例1：历史上第一个随机对照试验
链霉素治疗肺结核

- 设计：　多中心、随机、空白对照
- 样本含量：　109例
- 评价指标：
  - 主要：　治疗6个月的生存率
    6个月时基于胸片的明显改善率


BRITISH MEDICAL JOURNAL

Professor A. Bradford Hill.

STREPTOMYCIN TREATMENT OF PULMONARY TUBERCULOSIS

A MEDICAL RESEARCH COUNCIL INVESTIGATION

The following gives the short-term results of a controlled investigation into the effects of streptomycin on one type of pulmonary tuberculosis. The inquiry was planned and directed by the Streptomycin in Tuberculosis Trials Committee, composed of the following members: Dr. Geoffrey Marshall (chairman), Professor J. W. S. Blacklock, Professor G. Cameron, Professor N. B. Capon, Dr. R. Cruickshank, Professor J. H. Gaddum, Dr. F. R. G. Heaf, Professor A. Bradford Hill, Dr. L. E. Houghton, Dr. J. Clifford Hoyle, Professor H. Raistrick, Dr. J. G. Scadding, Professor W. H. Tytler, Professor G. S. Wilson, and Dr. P. D'Arcy Hart (secretary). The centres at which the work was carried out and the specialists in charge of patients and pathological work were as follows:

Brompton Hospital, London.—Clinician: Dr. J. W. Crofton, Streptomycin Registrar (working under the direction of the honorary staff of Brompton Hospital); Pathologists: Dr. J. W. Clegg, Dr. D. A. Mitchison.

Colindale Hospital (L.C.C.), London.—Clinicians: Dr. J. V. Hurford, Dr. B. J. Douglas Smith, Dr. W. E. Snell; Pathologists (Central Public Health Laboratory): Dr. G. B. Forbes, Dr. H. D. Holt.

Harefield Hospital (M.C.C.), Harefield, Middlesex.—Clinicians: Dr. R. H. Brent, Dr. L. E. Houghton; Pathologist: Dr. E. Nassau.

Bangour Hospital, Bangour, West Lothian.—Clinician: Dr. I. D. Ross; Pathologist: Dr. Isabella Purdie.

Killingbeck Hospital and Sanatorium, Leeds.—Clinicians: Dr. W. Santon Gilmour, Dr. A. M. Reevie; Pathologist: Professor J. W. McLeod.

Northern Hospital (L.C.C.), Winchmore Hill, London.—Clinicians: Dr. F. A. Nash, Dr. R. Shoulman; Pathologists: Dr. J. M. Alston, Dr. A. Mobun.

Sully Hospital, Sully, Glam.—Clinicians: Dr. D. M. E. Thomas, Dr. L. R. West; Pathologist: Professor W. H. Tytler.

The clinicians of the centres met periodically as a working subcommittee under the chairmanship of Dr. Geoffrey Marshall; so also did the pathologists under the chairmanship of Dr. R. Cruickshank. Dr. Marc Daniels, of the Council's scientific staff, was responsible for the clinical co-ordination of the trials, and he also prepared the report for the Committee, with assistance from Dr. D. A. Mitchison on the analysis of laboratory results. For the purpose of final analysis the radiological findings were assessed by a panel composed of Dr. L. G. Blair, Dr. Peter Kerley, and Dr. Geoffrey S. Todd.

Introduction

When a special committee of the Medical Research Council undertook in September, 1946, to plan clinical trials of streptomycin in tuberculosis the main problem faced was that of investigating the effect of the drug in pulmonary tuberculosis. This antibiotic had been discovered two years previously by Waksman (Schatz, Bugie, and Waksman, 1944); in the intervening period its power of inhibiting tubercle bacilli *in vitro*, and the results of treatment in experimental tuberculous infection in guinea-pigs, had been reported; these results were strikingly better than those with any previous chemotherapeutic agent in tuberculosis. Preliminary results of trials in clinical tuberculosis had been published (Hinshaw and Feldman, 1945; Hinshaw, Feldman, and Pfuetze, 1946; Keefer *et al.*, 1946); the clinical results in pulmonary tuberculosis were encouraging but inconclusive.

The natural course of pulmonary tuberculosis is in fact so variable and unpredictable that evidence of improvement or cure following the use of a new drug in a few cases cannot be accepted as proof of the effect of that drug. The history of chemotherapeutic trials in tuberculosis is filled with errors due to empirical evaluation of drugs (Hart, 1946); the exaggerated claims made for gold treatment, persisting over 15 years, provide a spectacular example. It had become obvious that, in future, conclusions regarding the clinical effect of a new chemotherapeutic agent in tuberculosis could be considered valid only if based on adequately controlled clinical trials (Hinshaw and Feldman, 1944). The one controlled trial of gold treatment (and the only report of an adequately controlled trial in tuberculosis we have been able to find in the literature) reported negative therapeutic results (Amberson, McMahon, and Pinner, 1931). In 1946 no controlled trial of streptomycin in pulmonary tuberculosis had been undertaken in the U.S.A. The Committee of the Medical Research Council decided then that a part of the small supply of streptomycin allocated to it for research purposes would be best employed in a rigorously planned investigation with concurrent controls.

The many difficulties of planning and conducting a trial of this nature are important enough to warrant a full description here of the methods of the investigation.

Plan and Conduct of the Trial

Type of Case

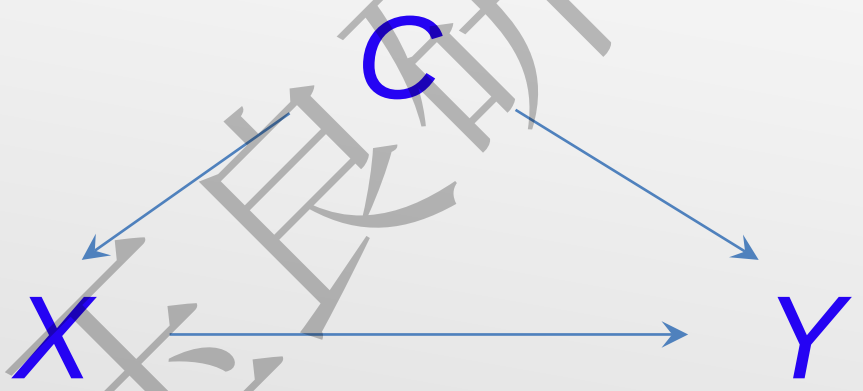
A first prerequisite was that all patients in the trial should have a similar type of disease. To avoid having to make allowances for the effect of forms of therapy other than bed-rest, the type of disease was to be one not suitable for other forms of therapy. The estimated chances of spontaneous regression must be small. On the other hand, the type of lesion should be such as to offer some prospect of action by an effective chemotherapeutic agent; for this reason old-standing disease, and disease with thick-walled

4582

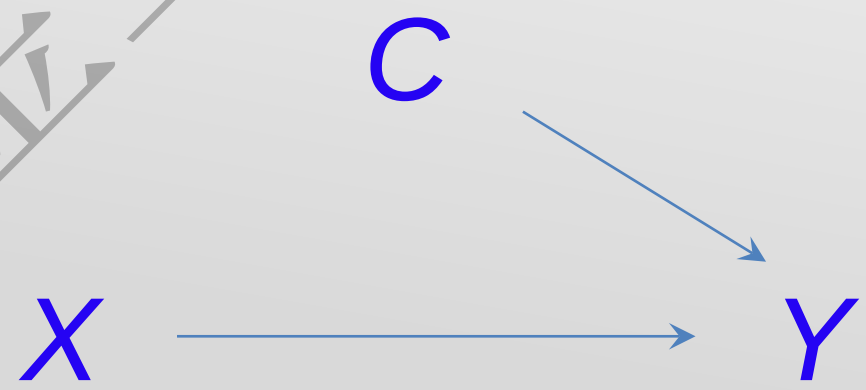
Medical Research Council. Streptomycin treatment of pulmonary tuberculosis. British Medical Journal,1948,2:769822.

随机化的价值：控制了混杂因素的干扰

- 没有随机化时



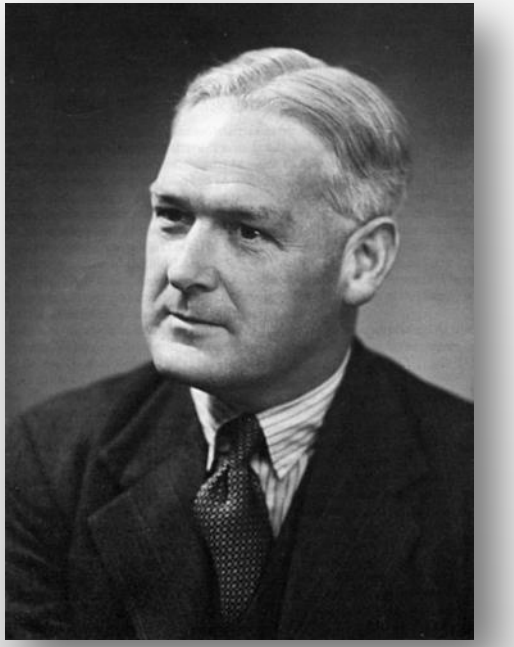
- 随机化后



随机化

1990年，93岁的Hill在回忆录中说：“自1937年我的著作出版后，我一直在寻找机会将随机化应用于临床试验，10年后机会终于来了，而我也早已准备好了”。

I had published my articles in the Lancet, which led to my handbook “Principles of Medical Statistics” just 10 years earlier in1937….I had been thinking about controlled trials for all of those 10 years and hoping for an opportunity that might arise…. Now the occasion arose and I was, therefore, completely ready for it.



Sir Austin Bradford Hill
1897-1991

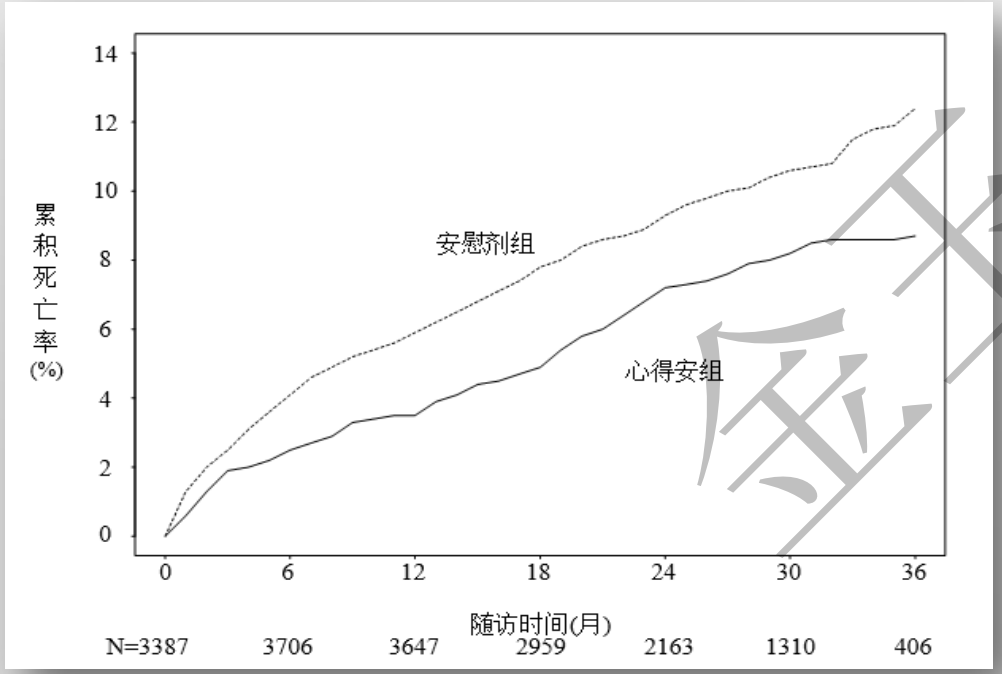
案例2：期中分析
普萘洛尔治疗急性心肌梗死

- 设计：多中心、随机、双盲、安慰剂对照
- 样本含量：4200例(计划)
- 评价指标：
  - 主要：总死亡率
  - 次要：冠心病死亡率，心源性猝死发生率，等等

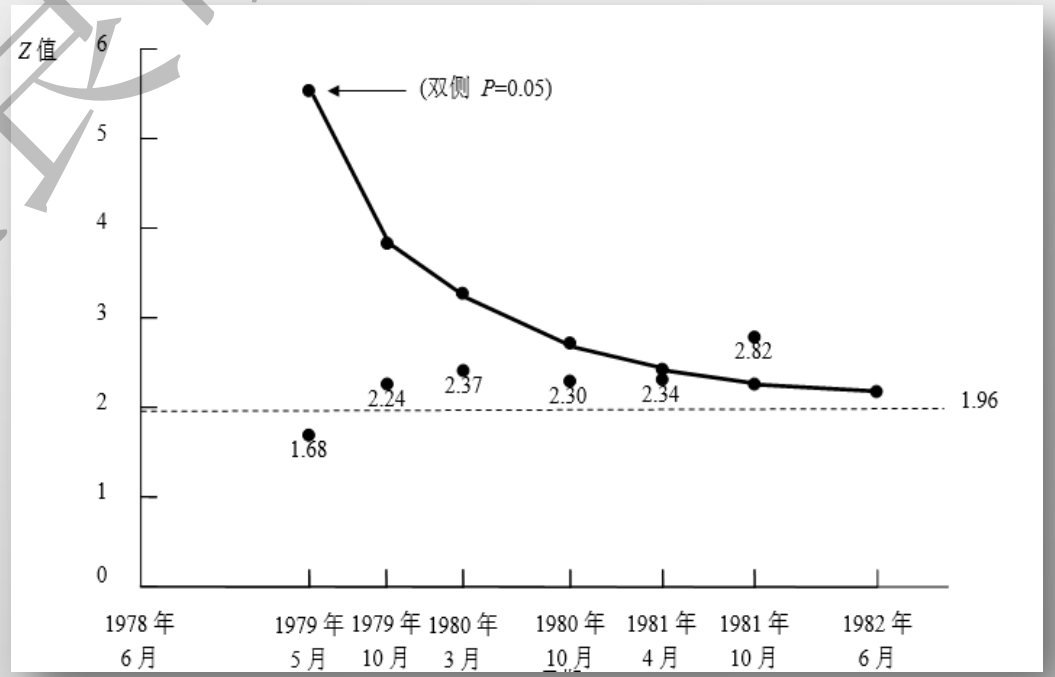
Beta-blocker heart attack trial research group. A randomized trial of propranolol in patients with acute myocardial infarction: mortality results. Journal of American Medical Association,1982,247:1707-14.

期中分析的价值：早期终止

- 独立政策和数据监察委员会(PDMB)



普萘洛尔 死亡率 7.2%
安慰剂 死亡率 9.8%



根据1981/10数据提前终止试验

临床试验中需要统计学

- 临床试验管理规范(ICH E6)

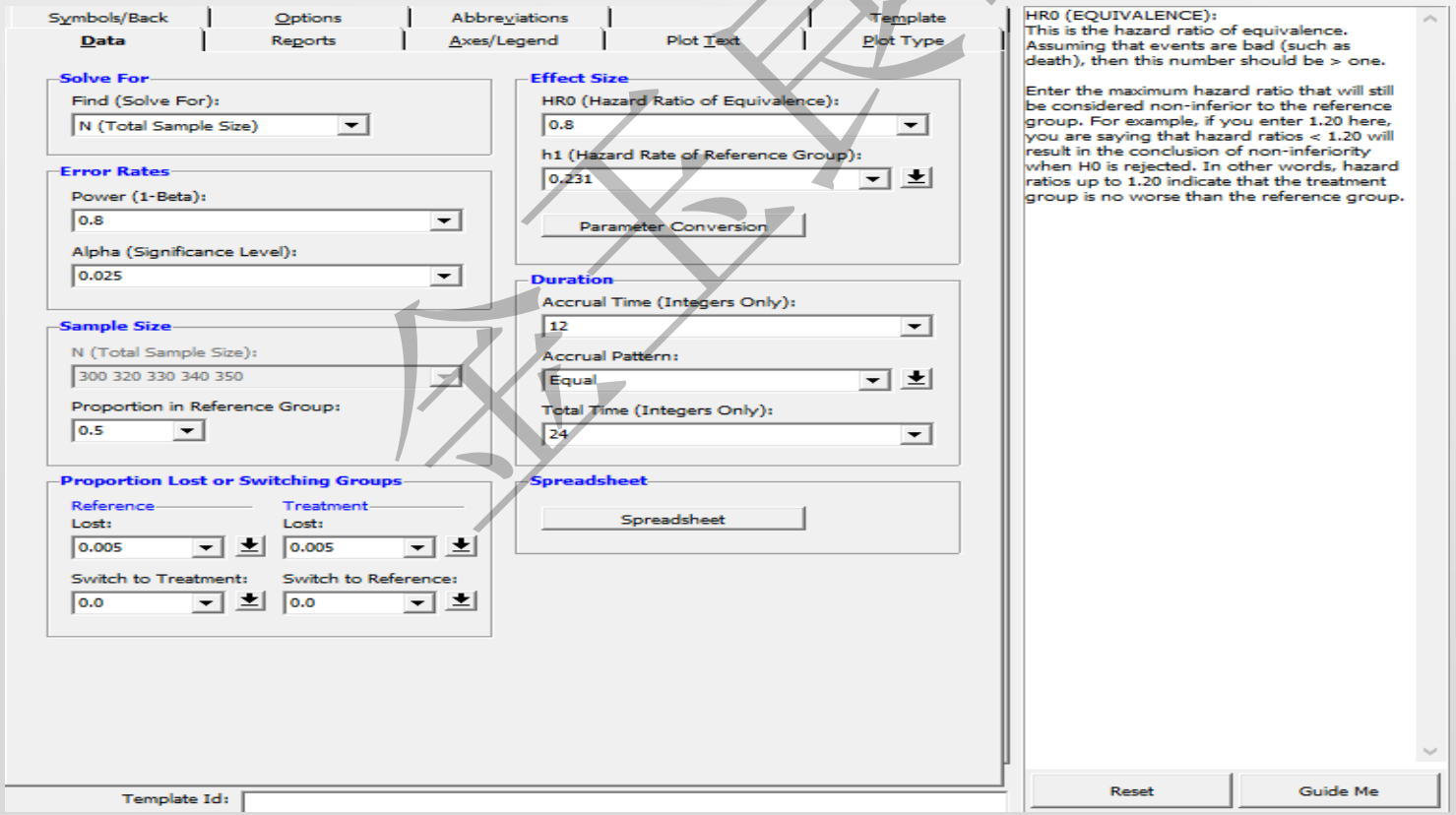
  所有与临床有关的统计工作需由有资质且有经验的统计专家负责；

- 临床试验指导原则 (GCP)

  在临床试验的统计结果的表达及分析过程中都必须采用规范的统计学分析方法，并应贯彻于临床试验始终。各阶段均需有熟悉生物统计学的人员参与。

沟通：这个研究做多少例？能不能省一点？

- 样本含量问题



02

临床试验中统计设计的价值

临床试验中常见的统计设计

- 按检验的假设分
  - 差异性研究
  - 等效性研究
  - 优效性研究
  - 非劣效性研究
- 按受试者分组和处理的方法分
  - 平行组设计
  - 交叉设计
  - 析因设计
  - 单组设计
  - ……

不同假设的价值

差异性研究　　差异性检验(inequality test)

试验药是否和对照药疗效不等?

H_0: $|T - S| = 0$　　H_1: $|T - S| <> 0$

不拒绝H0，无法回答“是否等效”的问题!

II类错误大小?
样本含量不够?

不同假设的价值

等效性研究　　等效性检验(equivalence test)

试验药是否和对照药等效?

$H_0: |T - S| \geq \delta$　　$H_1: |T - S| < \delta$

优效性研究　　差异性检验(superiority test)

试验药是否能获得预期甚至更好的收益?

非劣效性研究　非劣效性检验(inferiority test)

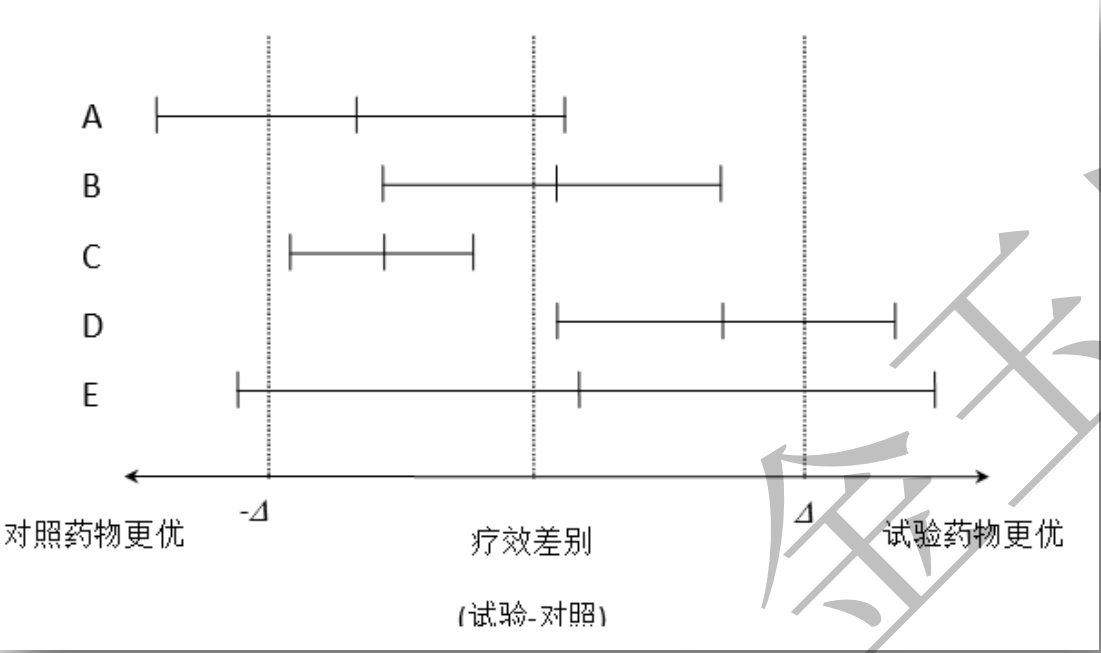
试验药的疗效是否更优或等效?

案例3：临床等效性
三联抗HIV逆转录病毒治疗

- 设计：　国际多中心、双盲、阳性药物平行对照
- 样本含量：　550例(计划)
- 评价指标：
  - 主要：　治疗48W周后病毒学抑制率(HIV RNA≤400 拷贝/ml)
  - 次要：　治疗48W周后病毒学显著抑制率，等等

Staszewski et al. Abacavir-lamivudine-zidovudine vs indinavir-lamivudine-zidovudine in antiretroviral-naive HIV-infected adults: A randomized equivalence trial. JAMA. 2001

利用可信区间法检验等效性



- ITT分析显示：

试验组抑制率 40%，对照组46%，率差的95% CI为 -15%~2%

试验药(三联疫苗)与对照药(逆转录治疗)在相同的试验条件下，显示出相似的有效性及安全性，差别在临床可接受范围内，即可认为两者疗效相等，或治疗等效。

临床试验中常见的统计设计

- 平行组设计
- 交叉设计
- 析因设计
- 单组设计

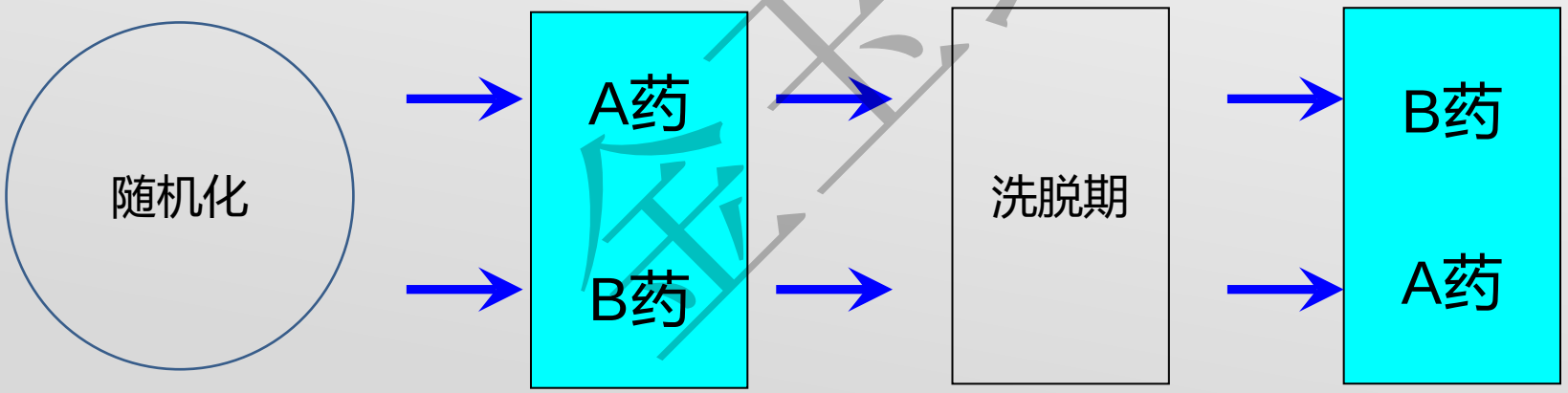
平行组设计的价值

- 最简单，便于实施

- 平行组+标准治疗 加载研究 (add-on study)
- 平行组+整群抽样和分组 群随机研究 (cRCT)
- 导入期给药+平行组+随机 随机撤药(withdrawal)研究
- 平行组+转组/无缝。。。 适应性设计(adaptive design)
- 平行组+开放标签+盲态评估 PROBE设计

交叉设计 （crossover design）

- 设有试验药A和对照药B，2×2交叉设计为



案例 4：交叉设计
生物仿制药HD203与依那西普的生物相似性

- 设计：　随机、双盲、单剂量、2*2交叉
- 样本含量：　37例
- 评价指标：
  - 主要：　药代动力学参数(T_{max},C_{max},AUC_{0-t},AUC_{0-inf})

Yi et al. Comparative pharmacokinetics of HD203, a biosimilar of etanercept, with marketed etanercept (Enbrel®): a double-blind, single-dose, crossover study in healthy volunteers. BioDrugs, 2012.

案例 5：交叉设计
托烷司琼与昂丹司琼对化疗致呕吐的疗效

- 设计：　　多中心、随机、盲法、2*2交叉
- 样本含量：　120例(计划)
- 评价指标：
  - 主要：　　呕吐次数

交叉设计的价值

- BE研究的标准设计
- 提高估计精度， 降低成本
- 一定程度上降低入组难度

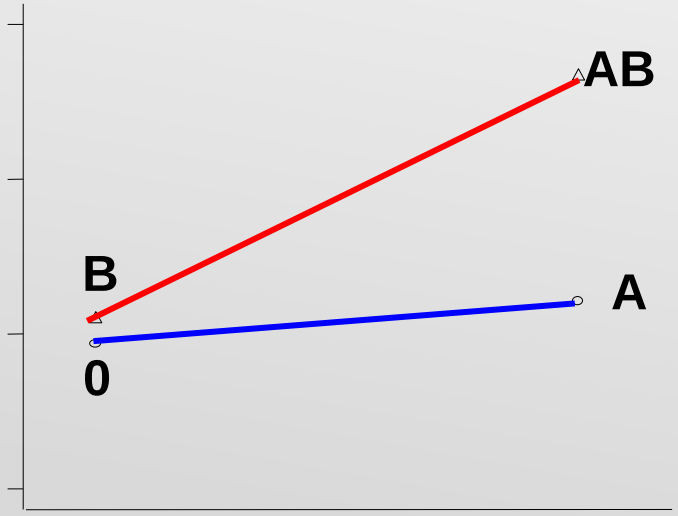
交叉设计的限制

- 统计分析复杂
- 往往用于症状缓解治疗
- 研究过程中难度增加

析因设计

是指包括两个或多个研究因素，且对各因素各水平的所有组合进行试验的一种研究设计方法。

	B0	B1
A0	第一组：（0） A药0　剂量 B药0　剂量	第二组：（b） A药0　剂量 B药1mg剂量
A1	第三组：（a） A药0.1mg剂量 B药0　剂量	第四组：（ab） A药0.1mg剂量 B药　1mg剂量



案例6：析因设计
维生素E与前列腺癌的风险

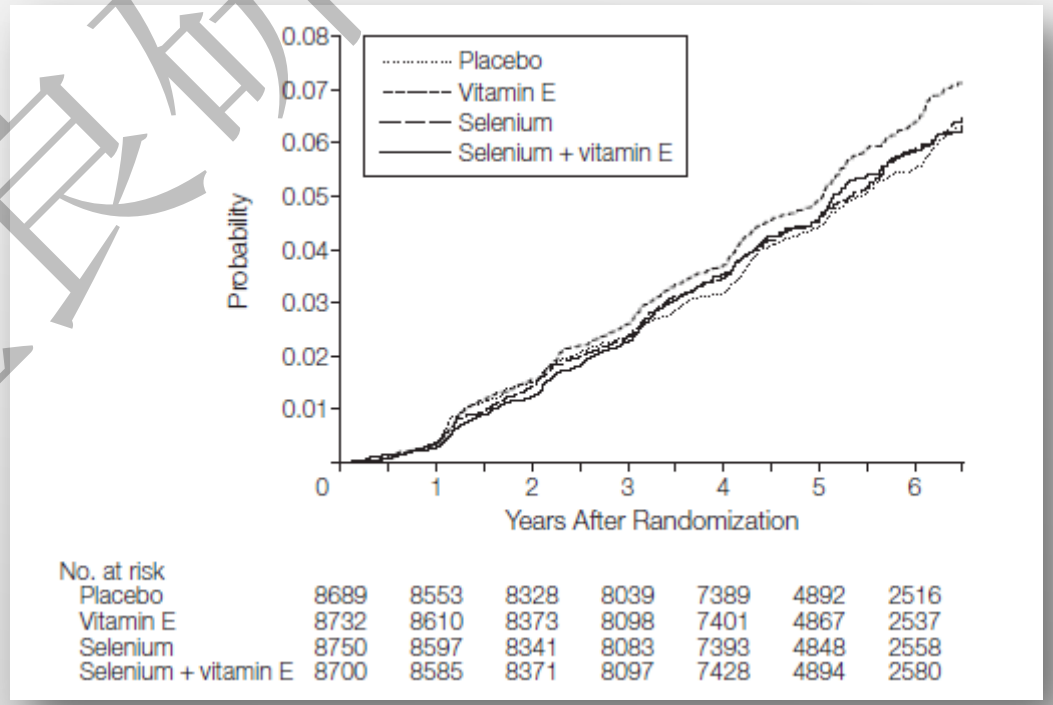
- 设计：多中心、随机、双盲、安慰剂对照、2*2析因、基于人群的临床试验
- 样本含量：35533例
- 评价指标：
  - 主要：前列腺癌的临床发病率
  - 次要：肺癌、结肠癌及其他肿瘤发生率……

Lippman et al. Effect of selenium and vitamin E on risk of prostate cancer and other cancers: the Selenium and Vitamin E Cancer Prevention Trial (SELECT). JAMA, 2009.

未能发现硒和维生素E的交互作用

四组间，被诊断患有前列腺癌的绝对人数（或5年发病率）之间无统计学差异：安慰剂组，416例（5年发病率为4.43%）；硒组，432例（4.56%）；维生素E组，473例（4.93%）；硒+维生素E组，437例（4.56%）。

在本试验所用剂量及配方下，硒和维生素E无论单独使用还是联合使用都没有预防健康成年男性人群前列腺癌发生的作用。



析因设计的价值

- 效率上具有优势
- 可以分析交互作用
- 可以寻找最佳组合

析因设计的限制

- 交互作用检测需较多样本
- 交互作用的定义影响了分析
- 天花板效应

单组设计的价值

- 便于实施
- 往往与多阶段设计整合
- 一般用于探索性研究
- 可用于罕见病、严重疾病

单组设计的限制

- 判断和解释需慎重

03

统计分析方法的价值

一般临床试验统计表格的结构

- 入组、数据集定义情况
- 脱落与方案违背情况
- 人口学和基线情况

研究基本轮廓

- 疗效分析
- 安全性分析
- 合并用药分析

安全性和有效性

- 亚组分析
- 敏感性分析

支持性数据

临床试验中常用的分析方法

- t 检验与 F 检验
- 卡方检验与Fisher确切概率法
- 基于秩次的非参数检验
- 生存时间数据的logrank检验

- 多因素分析：分层分析、回归模型

定量资料均数的比较：t检验

- 成组 t 检验用于两组均数的组间比较

  两组资料的总体均数是否相等?

- 配对 t 检验用于同组资料的前后比较

  前后差值的总体均数是否为0？

指标	时间点		后-前		
			A	B	组间比较
体温 ℃	第 2 周期	N	95	41	t=1.03，P=0.3031
		Mean ± Std	-0.02 ± 0.44	0.06 ± 0.30	
		Median(P25,P75)	0.00(-0.20,0.20)	0.00(-0.10,0.10)	
		Min~Max	-1.70-1.00	-0.50-1.00	
		组内比较	t=-0.44，P=0.6620	t=1.25，P=0.2175	

定量资料均数的组间比较：F检验

- F 检验用于两组或多组均数的组间比较

  多组资料的总体均数是否都相等?

- 往往有多重比较的问题(multiple comparisons)

等级资料或定量数据分布的组间比较

- Wilcoxon成组秩和检验

两组资料所来自总体的分布是否相同?

- Wilcoxon符号秩检验

差值的中位数是否为0?

- Kruskal-Wallis检验

多组资料所来自总体的分布是否都相同?

例如：生存质量评分的比较

表 1 两组病人第 2 周期末生活质量情况分析（FAS）

指标		对照组(n=140)	试验组(n=284)	组间比较
躯体功能	N	67	220	z=-1.11，P=0.2661
	Mean±Std	71.36±22.68	75.51±19.72	
	Median±QRange	80.00±26.67	80.00±20.00	
	Min~Max	13.23-100.00	6.67-100.00	

组间比较

表 2 两组病人第 2 周期末生活质量变化(后-前)变化情况分析(FAS)

指标		差值（后-前）		组间比较
		对照组(n=140)	试验组(n=284)	
躯体功能	N	67	220	z=-1.99，P=0.0469
	Mean±Std	-9.31±22.13	-3.33±17.84	
	Median±QRange	0.00±20.00	0.00±20.00	
	Min~Max	-66.37-52.33	-83.67-60.00	
	组内比较	S=-328.50,P=0.0002	S=-1202.50,P=0.0586	

改变值组内和组间比较

(无序)定性资料的组间比较：

- Pearson χ^2 检验

两组资料的率/构成比是否相等?
多组资料的率/构成比是否都相等?

- Fisher 确切概率计算法

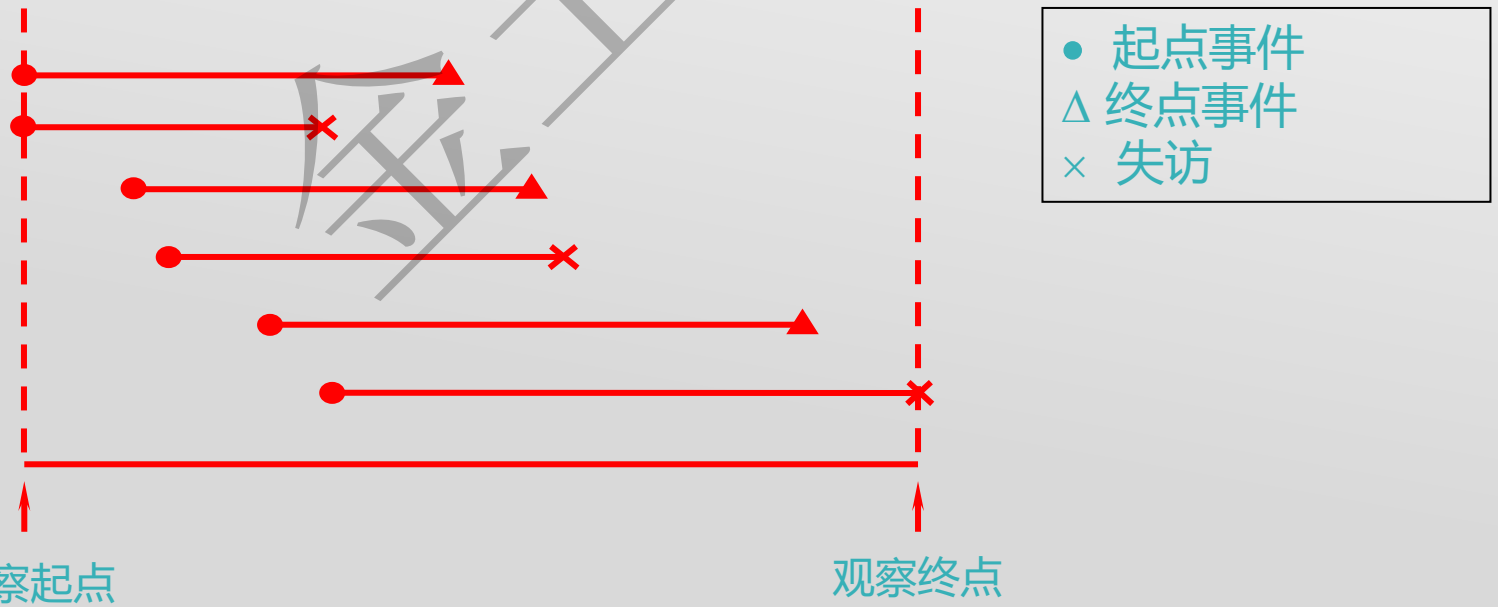
两组资料的率/构成比是否相等?

生存时间的组间比较：log-rank检验

- 肿瘤临床试验中，生存时间是非常关心的指标
- 以时间为指标时，存在着“截尾/删失”现象

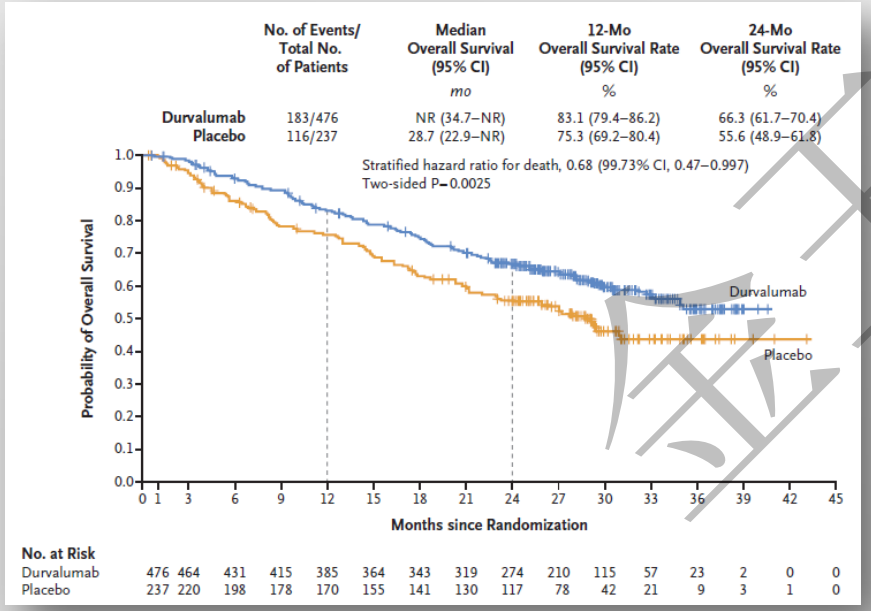
总生存期 overall survival, OS

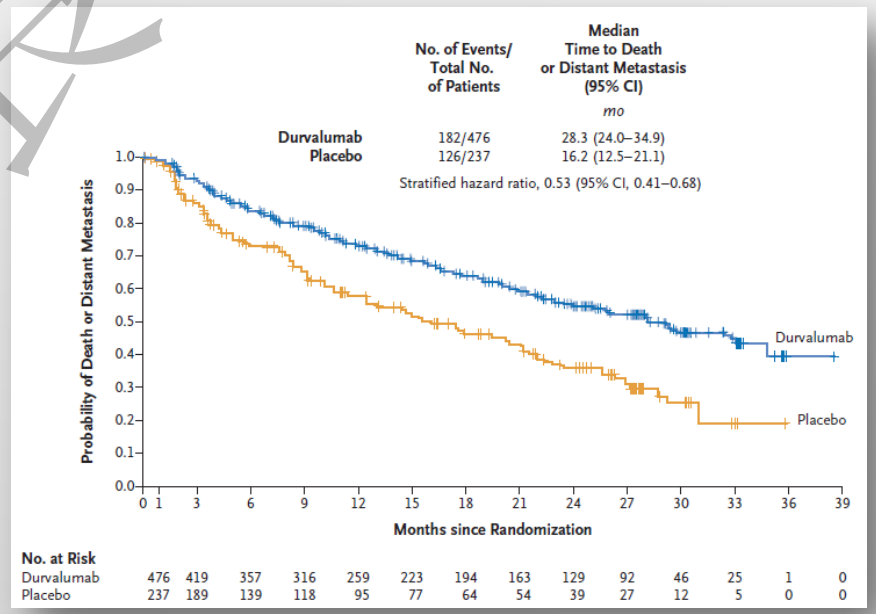
无进展生存期 progression free survival, PFS



生存过程的描述：Kaplain-Meier法
生存过程的比较：log-rank检验

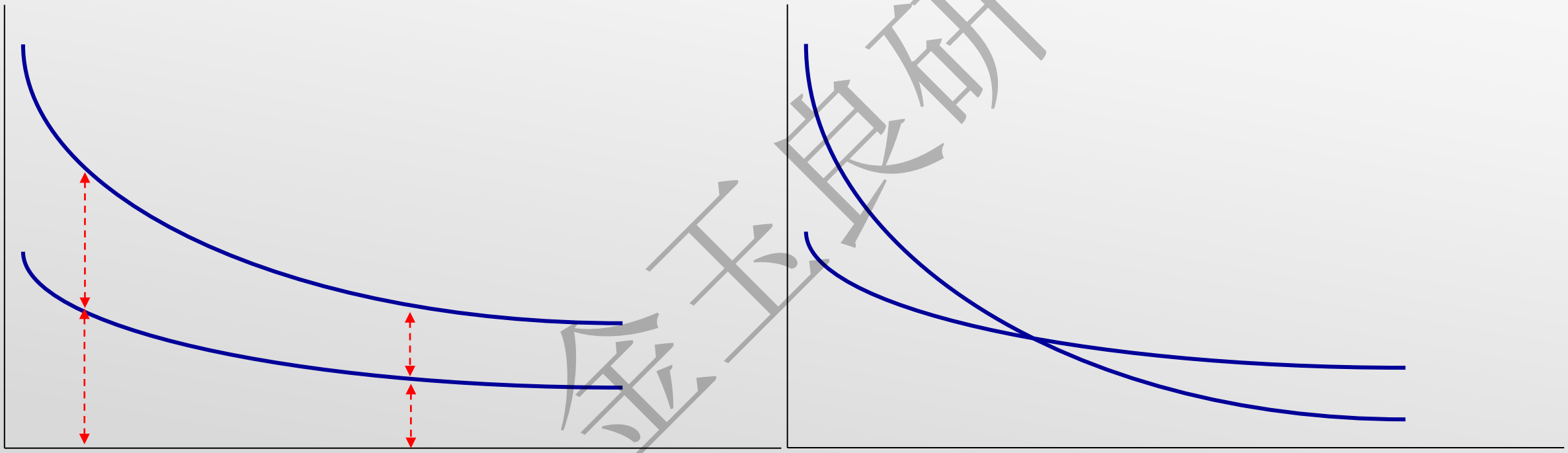
- 利用乘积极限法估计各时间点的累积生存率，并绘图





Antonia et al. Overall Survival with Durvalumab after Chemoradiotherapy in Stage III NSCLC. NEJM. 2018.

Logrank检验的要求：等比例风险假设



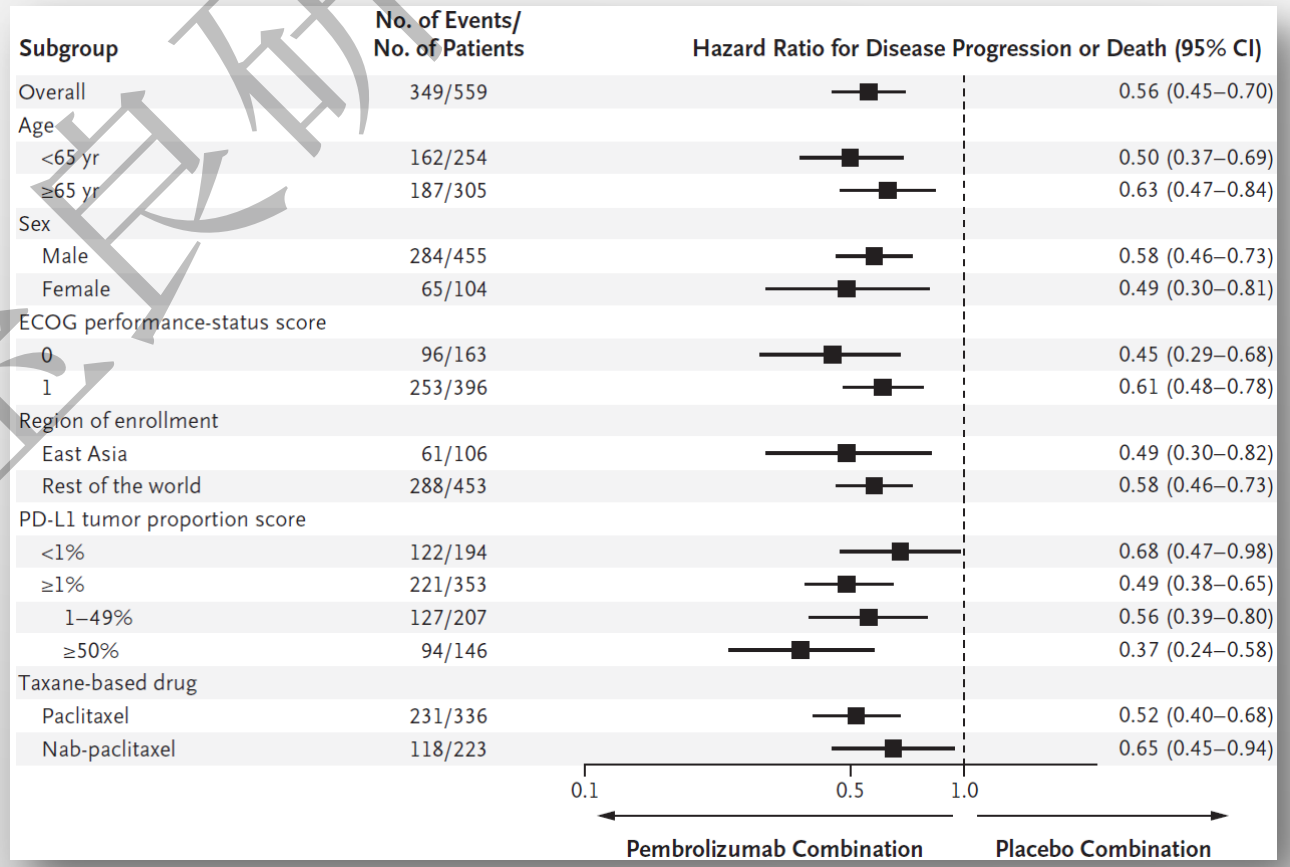
等比例风险　　不等比例风险

分层分析的价值

按照协变量分层

排除分层因素混杂(confounding)后，估计各层效应；

了解层间异质性(heterogeneity)



Subgroup	No. of Events/ No. of Patients	Hazard Ratio for Disease Progression or Death (95% CI)
Overall	349/559	0.56 (0.45–0.70)
Age		
<65 yr	162/254	0.50 (0.37–0.69)
≥65 yr	187/305	0.63 (0.47–0.84)
Sex		
Male	284/455	0.58 (0.46–0.73)
Female	65/104	0.49 (0.30–0.81)
ECOG performance-status score		
0	96/163	0.45 (0.29–0.68)
1	253/396	0.61 (0.48–0.78)
Region of enrollment		
East Asia	61/106	0.49 (0.30–0.82)
Rest of the world	288/453	0.58 (0.46–0.73)
PD-L1 tumor proportion score		
<1%	122/194	0.68 (0.47–0.98)
≥1%	221/353	0.49 (0.38–0.65)
1–49%	127/207	0.56 (0.39–0.80)
≥50%	94/146	0.37 (0.24–0.58)
Taxane-based drug		
Paclitaxel	231/336	0.52 (0.40–0.68)
Nab-paclitaxel	118/223	0.65 (0.45–0.94)

分层分析的森林图

多因素回归模型的价值

- 在调整了…某因素后，研究因素是否有意义

分析指标	常用模型
血糖变化值	协方差分析模型(一般线性模型)
呕吐次数	Poisson回归模型
改善率	Logistic回归模型
OS,PFS	Cox比例风险模型

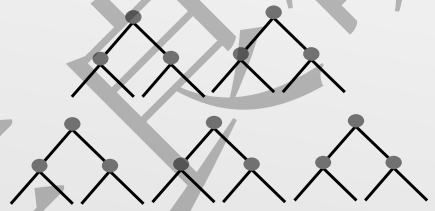
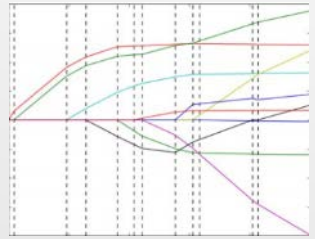
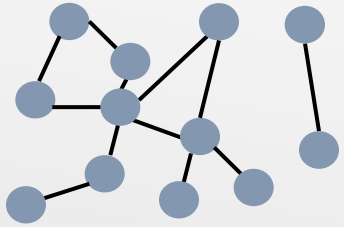
沟通：多因素分析很好很强大！多做一点？

- 一般作为探索性分析、影响因素分析或敏感性分析
- 主要指标分析中往往不去调整太多协变量
- 亚组分析，往往不对多个因素的组合形成亚组

- 随机分层因素作为调整变量，纳入主要分析模型

沟通：很多很fancy的方法，拿来用用？

- 支持向量机？
- 随机森林？
- 人工神经网络？
- 深度学习？

确证性试验

是一种事先提出假设，并用有对照组的试验对其进行确证性检验。
常用于证明有效性和安全性。
确证性试验的结果必须稳健。

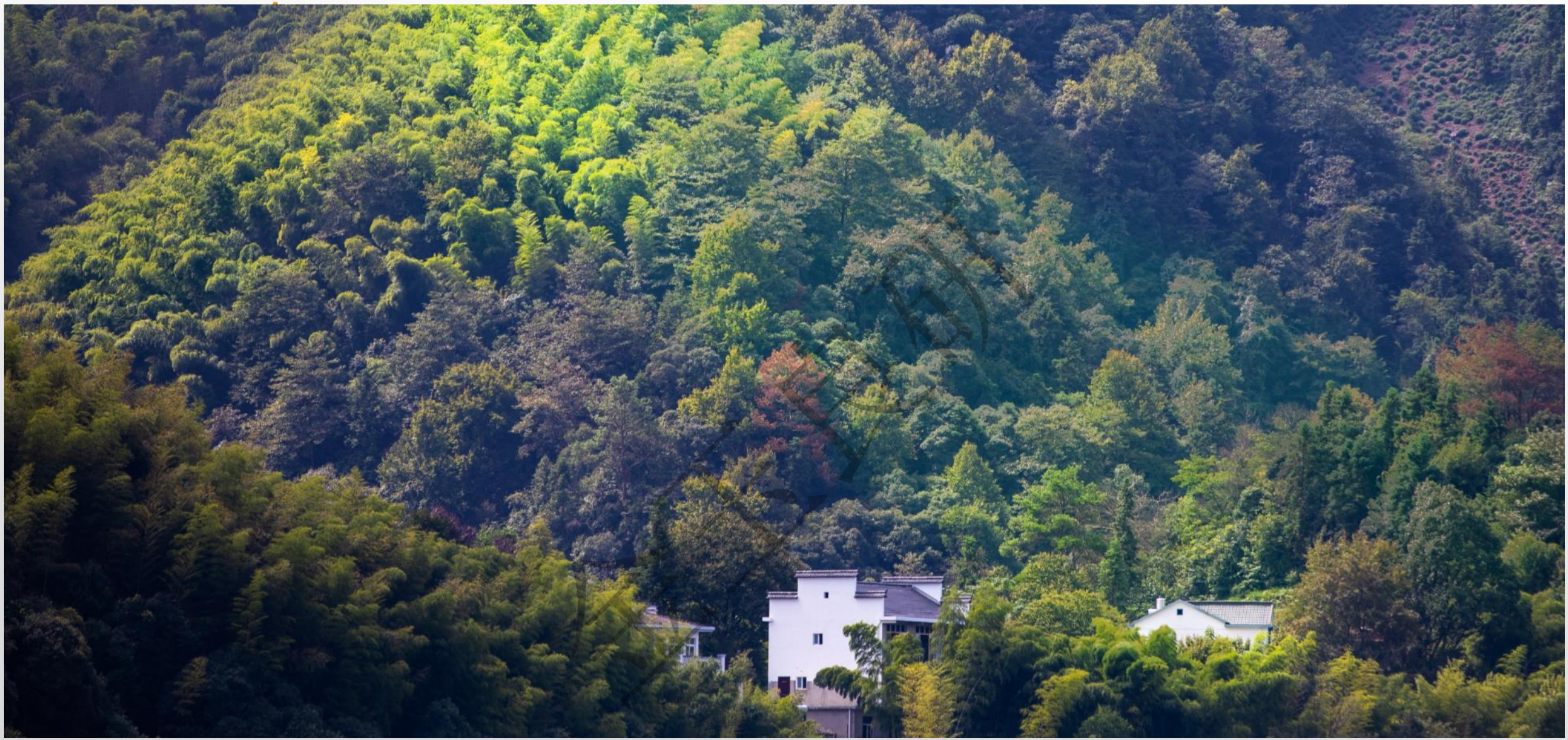
沟通：多做一些分析行不行？

在与其他临床试验专家的合作中，统计专家的作用和职责是确保新药临床试验中统计学原理的正确应用；

保证试验方案及修订方案中所涉及的统计学问题均描述得清晰、准确，并使用专业术语；

保证统计学报告和表格中结果的正确性和表达的合理性；

区别：临床试验的统计分析？临床试验数据的统计分析？



谢谢关注