

临床试验精选案例 统计学解读

陈峰 于浩·主编



临床试验精选案例

统计学解读

陈峰 于浩 主编

编者（按姓名拼音顺序排列）：

柏建岭 陈峰 黄丽红
刘丽亚 娄冬华 魏永越
苟鹏程 易洪刚 于浩
曾平 张汝阳 赵杨

人民卫生出版社

图书在版编目 (CIP) 数据

临床试验精选案例统计学解读 / 陈峰, 于浩主编. —北京:
人民卫生出版社, 2014

ISBN 978-7-117-20071-4

I. ①临… II. ①陈… ②于… III. ①临床医学-试验-案例-统计分析 IV. ①R4-33

中国版本图书馆 CIP 数据核字 (2014) 第 280536 号

人卫社官网	www.pmph.com	出版物查询, 在线购书
人卫医学网	www.ipmph.com	医学考试辅导, 医学数据库服务, 医学教育资源, 大众健康资讯

版权所有, 侵权必究!

临床试验精选案例统计学解读

主 编: 陈 峰 于 浩

出版发行: 人民卫生出版社 (中继线 010-59780011)

地 址: 北京市朝阳区潘家园南里 19 号

邮 编: 100021

E - mail: pmph@pmph.com

购书热线: 010-59787592 010-59787584 010-65264830

印 刷: 三河市尚艺印装有限公司

经 销: 新华书店

开 本: 787 × 1092 1/16 印张: 27

字 数: 657 千字

版 次: 2015 年 3 月第 1 版 2015 年 3 月第 1 版第 1 次印刷

标准书号: ISBN 978-7-117-20071-4/R · 20072

定 价: 69.00 元

打击盗版举报电话: 010-59787491 E-mail: WQ@pmph.com

(凡属印装质量问题请与本社市场营销中心联系退换)

前 言

这是一本结合实际案例介绍临床试验中涉及的统计学方法的专著。从国际顶尖杂志精选了 36 个临床试验案例,以 II、III 期临床试验为主。全书按照统计设计方法来安排。案例 1 介绍了史上最早的随机对照试验,案例 2~4、6 分别介绍了单中心、多中心、国际多中心随机双盲安慰剂对照试验,案例 5 是随机开放盲终点对照试验,案例 7 介绍加载试验,案例 8~9 介绍了阳性对照、多剂量平行对照试验,案例 11~14 介绍了临床等效性、非劣效性试验,案例 15~20 介绍了交叉设计、析因设计试验,案例 21~22 介绍了序贯设计,案例 23~24 介绍了多阶段设计,案例 25~29 介绍了几种适应性设计,案例 32~33 介绍了群随机对照和疫苗的临床试验,案例 34~35 介绍了医疗器械的临床试验,最后一个案例介绍了非随机对照试验。此外,案例 4 重点讨论了基线与协变量的处理,案例 5 重点讨论了多重性校正,案例 6 讨论了临床试验的期中分析和独立数据监察委员会,案例 10 介绍了分析人群的不同定义,案例 26 讨论了临床试验中量表的要求,案例 30 讨论了亚组分析,案例 31 讨论了试验的早期终止,案例 33 讨论了替代指标;案例 2 和案例 36 分别介绍了随机对照临床试验报告规范和非随机对照试验报告规范。

在案例选择中,我们尽可能兼顾到临床试验的方方面面,用 180 多个问答介绍了目前临床试验中经常遇到的统计学问题。在试验目的上兼顾了探索性试验和确证性试验;研究设计上涉及了经典设计方法(平行组设计、交叉设计、析因设计、序贯设计)和现代方法(适应性试验设计);比较的形式上兼顾了优效性、等效性和非劣效性;主要终点指标兼顾了连续性变量、分类变量、生存变量;统计推断兼顾了区间估计和假设检验;所用统计分析方法兼顾了经典统计方法(t 检验、 χ^2 检验、确切概率法、秩和检验等非参数方法、方差/协方差分析、多元线性回归、多元 logistic 模型、Poisson/ 负二项模型、Cox 模型、时依 Cox 模型、多重性校正等)和现代统计方法(广义估计方程、多水平模型、随机效应模型、bootstrap 估计、permutation 检验、多重填补等)。并介绍了以连续性变量、二分类变量、生存变量为终点指标的优效性、等效性、非劣效性检验时样本量的估计。

对每个案例,在简单介绍完研究的背景、方案、结果与结论后,通过问答形式,就研究中涉及的主要统计学问题进行解读,包括统计设计方法、样本量估计、统计分析方法和结论表达,并就临床试验中可能遇到的相关问题进行讨论。前面案例已经介绍的方法,后面将不再介绍。需要说明的是,每一个临床试验都有遗憾之处,这里精选的 36 个案例,也并非都是完美的。因此,在介绍完每个案例后,还就每个试验设计、实施或报告中存在的问题进行

了分析和讨论。

除个别文章外,本书所选文章大部分是国际顶尖杂志上发表的、PubMed 上能够免费获得原文的。附录一中列出了所有案例在 PubMed 上的编号,读者在 PubMed 网站上输入相应的编号,即可查到原文。有些研究发表了多篇文章,但本书仅选其中之一进行介绍。

本书不是教科书,在统计学方法的介绍上缺乏系统性。本书按照临床试验中的统计设计方法编排,结合用案例介绍统计方法,每一种设计、每个统计分析方法、统计图表都有其应用的背景,易于理解,便于学习,利于模仿。

苟鹏程、魏永越、柏建岭帮助筛选案例;张汝阳编写了书中所有 SAS 程序;江苏省中医院孙雪梅教授审阅了书稿,纠正了书稿中很多错误的医学表达。在此一并表示感谢。

本书目标读者为临床各科医生、从事临床医学科研的研究人员、高等院校及临床各科学学生、医药企业临床研究相关人员、生物统计工作者,也可以作为生物统计学、医学科研设计与分析课程的教学参考书。

由于作者水平和认识有限,错缪之处在所难免,有些观点不一定正确,敬祈国内外同行及广大读者批评指正。

陈峰 于浩

2014年10月于南京医科大学公共卫生学院

目 录

1 链霉素治疗肺结核——历史上第一个随机对照临床试验	1
1.1 试验方案简介	1
1.2 主要结果与结论	2
1.3 统计学解读	3
Q1.1 临床研究设计的基本原则是什么?	3
Q1.2 临床试验中为什么要随机化?	3
Q1.3 临床试验中为什么要设立对照组?	7
Q1.4 临床试验中为什么要重复?	9
1.4 统计学评价	9
2 硝酸甘油药膏对绝经后女性骨骼密度与强度的作用—— 单中心随机对照试验	11
2.1 试验方案简介	11
2.2 主要结果与结论	13
2.3 统计学解读	16
Q2.1 什么是偏倚?	16
Q2.2 什么是盲法?	16
Q2.3 安慰剂的作用是什么?	17
Q2.4 临床试验中为什么要分主要指标与次要指标?	17
Q2.5 不良事件与不良反应有什么区别?	17
Q2.6 两均数比较时如何估计样本量?	18
Q2.7 如何对两组均数进行假设检验?	18
Q2.8 如何估计两组均数之差的 95% 可信区间?	20
Q2.9 变量变换有什么作用?	20
Q2.10 什么是缺失值的 LOCF 填补法?	20
Q2.11 临床试验统计分析报告应该包含哪些主要内容?	21

Q2.12 在国际期刊发表临床试验报告有哪些要求?	21
2.4 统计学评价	23
3 依法利珠单抗治疗斑块状银屑病——多中心随机对照试验	25
3.1 试验方案简介	25
3.2 主要结果与结论	27
3.3 统计学解读	30
Q3.1 如何用确切概率法估计总体率的可信区间?	30
Q3.2 如何用确切概率法进行组间率的比较?	31
Q3.3 如何用确切概率法估计组间率差的可信区间?	32
Q3.4 用确切概率法比较两个率时如何估计样本量?	32
Q3.5 $k:1$ 分配比例对总样本量有何影响?	33
Q3.6 什么是 Wilcoxon 秩和检验?	34
Q3.7 什么是分层随机化?	35
Q3.8 如何评价和处理中心效应?	35
3.4 统计学评价	38
4 利莫那班治疗动脉粥样硬化——国际多中心随机对照试验	40
4.1 试验方案简介	40
4.2 主要结果与结论	43
4.3 统计学解读	50
Q4.1 什么是方差分析?	51
Q4.2 什么是协变量?	52
Q4.3 什么是协方差分析?	52
Q4.4 如何处理临床试验中的基线?	53
Q4.5 如何进行缺失数据的多重填补?	53
Q4.6 什么是 χ^2 检验?	55
Q4.7 什么是交互式语音应答系统?	56
Q4.8 什么是 MedDRA?	57
4.4 统计学评价	57
5 二十碳五烯酸与冠脉事件——PROBE 设计	59
5.1 试验方案简介	59
5.2 主要结果与结论	61
5.3 统计学解读	64
Q5.1 什么是 PROBE 设计?	64

Q5.2	什么是区组随机化?	64
Q5.3	什么是生存资料?	64
Q5.4	生存资料有哪些常用的分析方法?	65
Q5.5	什么是 Kaplan-Meier 法?	65
Q5.6	什么是 log-rank 检验?	65
Q5.7	什么是 Cox 比例风险模型?	65
Q5.8	生存分析如何估计样本量?	66
Q5.9	什么是重复测量方差分析?	67
Q5.10	什么是多重性问题?	68
Q5.11	什么时候需要校正多重性?	69
5.4	统计学评价	70
6	普萘洛尔治疗急性心肌梗死——临床试验的期中分析	72
6.1	试验方案简介	72
6.2	主要结果与结论	74
6.3	统计学解读	75
Q6.1	什么是期中分析?	75
Q6.2	为什么期中分析要调整 α ?	76
Q6.3	期中分析时如何调整 α ?	76
Q6.4	什么是 Pocock 法?	76
Q6.5	什么是 O'Brien-Fleming 法?	77
Q6.6	什么是 Peto 法?	77
Q6.7	什么是 Lan-DeMets 的 α 消耗函数?	77
Q6.8	什么是数据监察委员会?	78
Q6.9	数据监察委员会章程包含哪些内容?	80
6.4	统计学评价	80
7	介入治疗后早期和持续地口服氯吡格雷——加载试验	82
7.1	试验方案简介	82
7.2	主要结果与结论	84
7.3	统计学解读	89
Q7.1	什么是加载试验?	89
Q7.2	什么是 RRR?	89
Q7.3	两个率比较时如何估计样本量?	90
7.4	统计学评价	91

8 普拉格雷与氯吡格雷治疗急性冠脉综合征——多中心随机阳性对照试验	93
8.1 试验方案简介	93
8.2 主要结果与结论	95
8.3 统计学解读	99
Q8.1 什么是双盲双模拟?	99
Q8.2 什么是 Anderson-Gill 模型?	100
Q8.3 什么是时依 Cox 比例风险模型?	104
8.4 统计学评价	105
9 达格列净治疗 2 型糖尿病——多剂量平行对照试验	107
9.1 试验方案简介	107
9.2 主要结果与结论	110
9.3 统计学解读	113
Q9.1 如何进行临床试验的剂量筛选?	113
Q9.2 如何检验剂量反应关系的总趋势?	114
Q9.3 如何确定剂量反应关系评价中的最优剂量?	116
Q9.4 什么是 logistic 回归?	119
9.4 统计学评价	120
10 扎那米韦治疗 A 型和 B 型流感——分析人群的定义	122
10.1 试验方案简介	122
10.2 主要结果与结论	124
10.3 统计学解读	125
Q10.1 临床试验中有哪些常用的数据集?	125
Q10.2 什么是细菌 / 病毒阳性分析人群?	127
Q10.3 什么是 Hodges-Lehmann 估计?	127
Q10.4 什么是 van Elteren 检验?	128
Q10.5 什么是 bootstrap 估计?	128
10.4 统计学评价	129
11 三联抗 HIV 逆转录病毒治疗——临床等效性试验	131
11.1 试验方案简介	131
11.2 主要结果与结论	133
11.3 统计学解读	135

Q11.1	什么是等效性?	135
Q11.2	临床等效与生物等效有何区别?	135
Q11.3	常用的等效性检验方法有哪些?	136
Q11.4	如何确定临床等效性界值?	137
Q11.5	如何确定等效性试验所需样本量?	138
Q11.6	什么是 AUC,如何确定 AUC 的 95%CI?	139
11.4	统计学评价	139
12	HIV 抗病毒治疗中两种监测方案的比较——临床非劣效性试验	141
12.1	试验方案简介	141
12.2	主要结果与结论	142
12.3	统计学解读	144
Q12.1	什么是非劣效性临床试验?	144
Q12.2	如何确定非劣效性界值?	144
Q12.3	常用的非劣效性分析方法有哪些?	145
Q12.4	如何确定非劣效性研究所需样本量?	146
Q12.5	期中分析的 I 类错误控制方法有哪些?	147
12.4	统计学评价	148
13	超长效德谷胰岛素与甘精胰岛素治疗 1 型糖尿病—— 临床非劣效性试验	149
13.1	试验方案简介	149
13.2	主要结果与结论	151
13.3	统计学解读	152
Q13.1	什么是负二项回归?	152
Q13.2	什么是 post-hoc 检验?	153
Q13.3	什么是多个终点指标的顺序检验?	154
13.4	统计学评价	154
14	奈韦拉平联合齐多夫定预防 HIV-1 母婴传播—— 三臂自适应非劣效试验	156
14.1	试验方案简介	156
14.2	主要结果与结论	158
14.3	统计学解读	161
Q14.1	非劣效是指试验药不比阳性对照药差吗?	161
Q14.2	非劣效设计中, α 应取 0.05 还是 0.025?	161

Q14.3	三臂非劣效试验是不是最优的非劣效试验设计?	161
Q14.4	非劣效试验的分析应针对 ITT 人群还是 PP 人群进行?	161
14.4	统计学评价	161
15	HD203 与依那西普的生物相似性——2×2 交叉试验	163
15.1	试验方案简介	163
15.2	主要结果与结论	165
15.3	统计学解读	166
Q15.1	什么是交叉设计?	166
Q15.2	什么是洗脱期?	167
Q15.3	什么是滞后效应?	167
Q15.4	交叉设计与平行设计、配对设计的区别?	168
Q15.5	交叉设计有哪些条件和优缺点?	169
Q15.6	如何分析 2×2 交叉试验设计的资料?	169
Q15.7	交叉设计的样本量如何估计?	171
Q15.8	生物制品的生物相似性与化学药品的生物等效性有何区别?	174
15.4	统计学评价	175
16	氧吸入与丛集性头痛——多阶段重复交叉设计	178
16.1	试验方案简介	178
16.2	主要结果与结论	180
16.3	统计学解读	183
Q16.1	什么是多阶段重复交叉设计?	183
Q16.2	为什么要采用高阶交叉设计?	184
Q16.3	FDA 为什么推荐只有两种用药顺序组的高阶交叉设计?	186
Q16.4	什么是多水平模型?	187
Q16.5	如何分析滞后效应?	190
16.4	统计学评价	191
17	山楂提取物治疗高血压——4×4 交叉设计	194
17.1	试验方案简介	194
17.2	主要结果与结论	196
17.3	统计学解读	198
Q17.1	什么是 Williams 设计?	198
Q17.2	4×4 交叉设计的资料如何进行统计分析?	199
Q17.3	建立剂量反应关系时如何估计样本量?	200

17.4	统计学评价	201
18	维生素 E 与前列腺癌风险——2×2 析因试验	203
18.1	试验方案简介	203
18.2	主要结果与结论	205
18.3	统计学解读	208
Q18.1	什么是析因设计?	208
Q18.2	什么是主效应? 单独效应? 联合效应?	208
Q18.3	什么是交互作用?	208
Q18.4	什么是相加和相乘交互作用?	209
Q18.5	什么是交互作用的阶数?	209
Q18.6	析因设计的生存分析资料如何分析?	209
Q18.7	如何估计析因设计的生存分析资料的样本含量?	210
18.4	统计学评价	210
19	利格列汀、二甲双胍与 2 型糖尿病——2×3 析因试验	213
19.1	试验方案简介	213
19.2	主要结果与结论	215
19.3	统计学解读	218
Q19.1	什么是不完全析因设计?	218
Q19.2	如何分析不完全析因设计资料?	218
Q19.3	什么是多重比较的顺序检验程序?	219
Q19.4	什么是 FWER?	219
Q19.5	临床试验中有哪些常用的控制 FWER 的方法?	219
19.4	统计学评价	220
20	氨氯地平、阿托伐他汀与高血压及血脂异常——3×5 析因试验	222
20.1	试验方案简介	222
20.2	主要结果与结论	224
20.3	统计学解读	227
Q20.1	如何对析因设计的资料进行协方差分析?	227
Q20.2	进行协方差分析需要具备哪些条件?	227
Q20.3	什么是多重比较的逐步向下程序?	227
20.4	统计学评价	228

21 长春瑞滨和氟尿嘧啶联合治疗晚期乳腺癌——成组序贯临床试验	230
21.1 试验方案简介	230
21.2 主要结果与结论	231
21.3 统计学解读	232
Q21.1 什么是成组序贯试验?	232
Q21.2 如何进行无对照成组序贯试验设计?	233
21.4 统计学评价	234
22 除颤器与冠状动脉病变患者的生存期——序贯试验	236
22.1 试验方案简介	236
22.2 主要结果与结论	238
22.3 统计学解读	239
Q22.1 序贯试验有哪些设计类型?	239
Q22.2 序贯设计后如何进行数据分析?	239
22.4 统计学评价	240
23 阿西替尼治疗晚期非小细胞肺癌——无对照二阶段试验	241
23.1 试验方案简介	242
23.2 主要结果与结论	243
23.3 统计学解读	245
Q23.1 什么是无对照二阶段设计?	245
Q23.2 二阶段设计如何估计样本量?	247
Q23.3 二阶段设计适用于什么条件?	252
Q23.4 二阶段设计需要考虑 α 校正吗?	252
Q23.5 按照二阶段设计的临床试验是否可被批准上市?	252
23.4 统计学评价	253
24 贝伐单抗联合卡培他滨治疗晚期肝癌——无对照三阶段试验	254
24.1 试验方案简介	254
24.2 主要结果与结论	256
24.3 统计学解读	258
Q24.1 什么是三阶段设计?	258
Q24.2 Ensign 三阶段设计的基本思想是什么?	259
Q24.3 如何选择单阶段、二阶段和三阶段设计?	265
Q24.4 如何选择 Simon 的最优设计和 Ensign 设计?	265

24.4	统计学评价	266
25	马立马司他治疗转移性乳腺癌——适应性设计	267
25.1	试验方案简介	267
25.2	主要结果与结论	269
25.3	统计学解读	272
Q25.1	什么是适应性设计?	272
Q25.2	什么时候需要改变分组概率?	273
Q25.3	临床试验中有哪些适应性分组方法?	274
Q25.4	如何进行样本量再估计?	275
Q25.5	什么是两阶段适应性设计?	276
Q25.6	什么是分层 Cox 模型?	277
Q25.7	什么是分层 log-rank 检验?	278
25.4	统计学评价	279
26	氟维司群与阿那曲唑治疗晚期乳腺癌——剂量筛选的适应性试验	282
26.1	试验方案简介	282
26.2	主要结果与结论	284
26.3	统计学解读	287
Q26.1	什么是剂量筛选的适应性设计?	288
Q26.2	如何进行 II / III 期无缝设计剂量筛选?	288
Q26.3	什么是肿瘤病人生存质量“FACT”量表?	289
Q26.4	如何考评一个量表使其用于临床试验结果的评价?	290
26.4	统计学评价	292
27	卡培他滨辅助治疗早期乳腺癌——Bayes 适应性设计	294
27.1	试验方案简介	294
27.2	主要结果与结论	296
27.3	统计学解读	298
Q27.1	什么是条件把握度?	298
Q27.2	什么是预测概率?	300
Q27.3	什么是 Bayes 预测概率?	302
Q27.4	何时进行期中分析?	302
Q27.5	什么是 Wald 检验?	302
27.4	统计学评价	303

28 索拉非尼治疗转移性肾细胞癌——随机撤药试验	305
28.1 试验方案简介	305
28.2 主要结果与结论	306
28.3 统计学解读	308
Q28.1 什么是随机撤药试验?	308
Q28.2 如何分析随机撤药试验资料?	309
Q28.3 如何评价随机撤药试验?	309
28.4 统计学评价	309
29 阿立哌唑辅助抗抑郁治疗——随机序贯安慰剂平行对照试验	311
29.1 试验方案简介	311
29.2 主要结果与结论	314
29.3 统计学解读	315
Q29.1 什么是序贯安慰剂平行对照设计?	315
Q29.2 如何分析序贯安慰剂平行对照试验资料?	317
Q29.3 如何评价序贯安慰剂平行对照设计?	318
29.4 统计学评价	318
30 替卡格雷减少急性冠脉综合征患者的心血管事件——亚组分析	320
30.1 试验方案简介	320
30.2 主要结果与结论	322
30.3 统计学解读	328
Q30.1 什么是亚组分析?	328
Q30.2 如何正确看待亚组分析的结果?	329
Q30.3 进行亚组分析有哪些统计学考虑?	330
Q30.4 如何规范报告亚组分析的结果?	331
30.4 统计学评价	332
31 静注钙/镁减少奥沙利铂诱发的感觉神经毒性——试验的早期终止	334
31.1 试验方案简介	334
31.2 主要结果与结论	336
31.3 统计学解读	337
Q31.1 什么时候可以早期终止临床试验?	337
Q31.2 早期终止临床试验有哪些优缺点?	337
Q31.3 早期终止临床试验有哪些统计学考虑?	338

31.4	统计学评价	339
32	2型糖尿病患者的健康教育和自我管理程序的有效性—— 群随机对照试验	341
32.1	试验方案简介	341
32.2	主要结果与结论	342
32.3	统计学解读	345
Q32.1	什么是群随机对照试验?	345
Q32.2	基于群随机样本的研究如何估计样本量?	346
Q32.3	什么是广义估计方程?	346
32.4	统计学评价	348
33	Vi 伤寒疫苗的有效性和安全性——群随机对照试验	350
33.1	试验方案简介	350
33.2	主要结果与结论	352
33.3	统计学解读	354
Q33.1	疫苗临床试验的特点是什么?	354
Q33.2	疫苗人群试验的研究指标是什么?	355
Q33.3	什么样的指标可以作为替代指标?	356
Q33.4	基于群随机样本的研究如何估计样本量?	357
Q33.5	流行病学调查在疫苗群随机试验中有何作用?	357
33.4	统计学评价	358
34	氢氟烷抛射剂加压计量吸入器治疗哮喘——器械的临床试验	359
34.1	试验方案简介	359
34.2	主要结果与结论	361
34.3	统计学解读	363
Q34.1	什么是医疗器械?	363
Q34.2	医疗器械上市前也要进行临床试验吗?	364
Q34.3	医疗器械临床试验中有哪些统计学考虑?	364
34.4	统计学评价	364
35	颈动脉支架术联合栓塞保护性治疗的上市后研究——单组目标值法	366
35.1	试验方案简介	366
35.2	主要结果与结论	367

35.3	统计学解读	370
Q35.1	什么是靶值? 什么是单组目标值?	370
Q35.2	单组目标值试验中如何估计样本量?	371
Q35.3	什么时候采用单组设计?	372
Q35.4	如何对 OPC 进行统计推断?	373
Q35.5	单组目标设计方法需要注意哪些问题?	373
35.4	统计学评价	373
36	体外膜肺氧合对严重 H1N1 甲型流感患者的疗效——非随机对照试验	375
36.1	试验方案简介	375
36.2	主要结果与结论	376
36.3	统计学解读	378
Q36.1	什么是个体匹配?	378
Q36.2	什么是倾向指数匹配?	379
Q36.3	什么是多元匹配法?	379
Q36.4	什么是倾向指数分层法?	380
Q36.5	什么是倾向指数校正法?	380
Q36.6	非随机对照临床试验有哪些统计学考虑?	381
Q36.7	发表非随机对照研究有哪些要求?	381
36.4	统计学评价	383
 附录		
附录一	案例原文 PubMed 文献编号一览表	385
附录二	样本量估计方法索引	386
附录三	ICH E9: 临床试验统计学指导原则	387