

Statistical Methods in Diagnostic Medicine

诊断医学中的统计学方法 (第二版)

Xiao-Hua Zhou
Nancy A. Obuchowski
Donna K. McClish 著
侯艳 李康 宇传华 周晓华 译

Statistical Methods in Diagnostic Medicine

诊断医学中的统计学方法 (第二版)

Xiao-Hua Zhou
Nancy A. Obuchowski
Donna K. McClish 著
侯艳 李康
宇传华 周晓华 译

ZHENDUAN YIXUE ZHONG DE
TONGJIXUE FANGFA

图字：01-2014-4798 号

Translation from: *Statistical Methods in Diagnostic Medicine (Second Edition)* by Xiao-Hua Zhou, Nancy A. Obuchowski, Donna K. McClish.

ISBN: 978-0-470-18314-4

Copyright © 2011 by John Wiley & Sons, Inc. All right reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

All Rights Reserved. This Translation published under licence.

图书在版编目 (CIP) 数据

诊断医学中的统计学方法：第2版 / (美) 周晓华，
(美) 奥布乔斯基 (Obuchowski, N. A.)，(美) 麦克林斯
(McClish, D. K.) 著；侯艳等译. — 北京：高等教育出
版社，2016. 2

书名原文：Statistical Methods in Diagnostic Medicine

ISBN 978-7-04-044259-5

I. ①诊… II. ①周… ②奥… ③麦… ④侯… III.
①实验室诊断—医学统计 IV. ①R446-32

中国版本图书馆 CIP 数据核字 (2015) 第 275997 号

策划编辑 王丽萍
责任编辑 李华英
责任印制 毛斯璐

封面设计 姜磊
版式设计 马敬茹

印刷 北京中科印刷有限公司

开本 787mm×1092mm 1/16

出版发行 高等教育出版社
社址 北京市西城区德外大街4号
邮政编码 100120
印 刷 北京中科印刷有限公司
开 本 787mm×1092mm 1/16
印 张 27.75
字 数 540千字
购书热线 010-58581118

咨询电话 400-810-0598
网 址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>
网上订购 <http://www.landaco.com>
<http://www.landaco.com.cn>
版 次 2016年2月第1版
印 次 2016年2月第1次印刷
定 价 89.00元

本书如有缺页、倒页、脱页等质量问题，请到所购图书销售部门联系调换

版权所有 侵权必究

物 料 号 44259-00

前言

诊断试验经常用于确定病人是否需要接受额外的诊断、治疗以及干预,而这将影响病人的最终结局,所以诊断试验研究在医学中起到非常重要的作用。因此,临床医生给出的诊断试验结果对于诊断试验准确度的评价就显得非常重要。为了确保正确地评价诊断试验的准确度,需要临床医生能够较好地设计诊断试验,并理解这些方法如何正确地应用到病人身上进行诊断,本书主要有两个目的:一是提供合适的方法正确地设计诊断试验和分析诊断试验的准确度;二是帮助临床医生更好地理解这些诊断试验的结果,并能够应用到实际。

新版书中根据最新发表的诊断试验评价方法,更新了每章的内容。更新的内容包括:用于检测和定位病变部位的诊断试验评价方法(第二章、第三章和第九章);诊断试验中需要调整的协变量类型(第三章),以及相应的协变量调整方法(第八章);估计和比较诊断试验预测值的方法(第四章和第五章);使用预测值作为诊断试验准确度评价标准的样本量估计方法(第六章);具有预试验结果数据时,多位阅片者研究设计类型的样本量估计方法(第六章);具有连续型结果指标的诊断试验存在证实性偏倚时,ROC曲线估计值新的校正方法(第十章);具有等级型或者连续型结果指标的诊断试验存在非完美金标准情况时,ROC曲线估计值的校正方法(第十一章)。

本版书中还增加了三个实例:正电子发射断层扫描(PET)用于比较检测患病的甲状旁腺准确性的研究;计算机辅助检测结肠息肉准确性的研究;磁共振成像(MRI)用于诊断动脉狭窄准确性的研究(第一章)。这三个实例用于阐述本书中介绍的统计学方法,具体数据详见附录。

本书的整体结构如下:

第二章和第三章分别介绍了有关诊断试验准确度评价的一些基本的衡量指标以及研究设计类型,同时讲解了用于评价诊断试验准确度的统计学方法,并通过实例阐述了如何正确应用这些统计学方法。第四章和第五章分别阐述了准确度的估计方法,并比较了不同设计情况下诊断试验的准确度。第六章介绍了诊断试验样本量的

估算方法. 第七章和第十二章集中介绍了如何进行诊断试验准确度的 meta 分析及相应的统计学方法. 第八章和第九章分别介绍了多种病人情况及多位阅片者设计的诊断试验准确度评价模型. 第十章及第十一章分别介绍了诊断试验研究中存在证实性偏倚及非完美金标准时, 诊断试验准确度估计值的校正方法. 第一章到第三章介绍了诊断试验中基本的统计学概念以及医学术语, 第四章到第七章的内容为生物统计专业人员介绍了基本的数据分析方法. 第八章到第十二章介绍了高级的统计学内容, 同时也提供了更为详细的统计学方法. 本书各章节中的所有例子均面向读者公开, 并且读者可以通过访问提供的网站获得常用的软件 (<http://faculty.washington.edu/azhou/books/diagnostic.html>).

致谢

非常感谢许多同事在本书第一版 (2002) 和第二版 (2011) 的写作和出版过程中给予的支持, 正是因为有了他们对第一版的意见和建议才有了这一版的改进. 特别感谢刘丹萍和王喆宇给予本书的意见, 以及使用计算机程序实现本书中介绍的方法所付出的努力. 感谢 Thomas D. Koepsell 博士为书稿提出宝贵意见.

同时也感谢作者的家人给予的理解和支持. 周晓华博士感谢他的妻子 Yea-Jae 以及他的孩子 Vanessa 和 Joshua; Nancy Obuchowski 博士感谢她的丈夫 Ralph Harvey 博士以及她的孩子 Tucker, Eli 和 Scout; Donna McClish 感谢她的丈夫 Tom 以及他们的女儿 Amanda.

目录

第一部分 基本概念和方法

第一章 绪论	3
1.1 诊断试验准确度研究	3
1.2 实例介绍	5
1.2.1 实例 1: 甲状旁腺疾病	5
1.2.2 实例 2: 结肠癌诊断	7
1.2.3 实例 3: 颈动脉狭窄	8
1.3 软件	9
1.4 本书没有包含的主题	9
第二章 诊断准确度评价指标	10
2.1 灵敏度与特异度	10
2.1.1 衡量诊断试验准确度的基本指标: 实例 2	12
2.1.2 具有连续型检查结果的诊断试验: 人工心脏瓣膜实例	13
2.1.3 具有等级型检查结果的诊断试验: 实例 1	14
2.1.4 疾病患病率与疾病谱对灵敏度和特异度的影响	15
2.1.5 FPR 和 FNR 与 α 和 β 的类比	15
2.2 灵敏度与特异度的综合指标	16
2.2.1 两个及以上诊断试验准确度比较时存在的问题: 实例 1	16
2.2.2 诊断试验结果的正确率	16
2.2.3 优势比和 Youden 指数	17
2.3 受试者工作特征曲线 (ROC)	18
2.3.1 ROC 曲线: 人造心脏瓣膜和实例 1	18

2.3.2 关于 ROC 曲线的假定	19
2.3.3 拟合的光滑 ROC 曲线	19
2.3.4 ROC 曲线的优点	20
2.4 ROC 曲线下面积	21
2.4.1 ROC 曲线下面积的解释	21
2.4.2 ROC 曲线下面积的大小	22
2.4.3 ROC 曲线下面积: 实例 1	23
2.4.4 ROC 曲线下面积的误解	23
2.5 固定 FPR 的灵敏度	25
2.6 部分 ROC 曲线下面积	26
2.7 似然比	27
2.7.1 关于似然比的三个实例	28
2.7.2 似然比的缺点	29
2.7.3 常规与非常规 ROC 曲线	30
2.8 真实诊断结果不是二分类情况的 ROC 分析	31
2.9 比较预测模型的 C 统计量和其他指标	33
2.10 多个病灶的检测与定位	34
2.11 阳性和阴性预测值、贝叶斯定理及实例 2	36
2.11.1 贝叶斯定理	37
2.12 ROC 曲线的最佳决策阈值	39
2.12.1 最大化分类的最优阈值	39
2.12.2 最小化成本的最优阈值	40
2.12.3 最佳决策阈值: 以快速眼动为抑郁症生物标识的实例	41
2.13 多项试验的结果解释	41
2.13.1 平行联合诊断	41
2.13.2 序贯诊断试验	42
第三章 诊断准确度的研究设计	43
3.1 确定研究目标	44
3.2 识别目标患者总体	47
3.3 选择患者抽样计划	47
3.3.1 第一阶段: 探索研究	48
3.3.2 第二阶段: 挑战研究	48
3.3.3 第三阶段: 临床研究	50

3.4	选择金标准	53
3.5	选择准确度指标	58
3.6	识别目标阅片者总体	61
3.7	选择阅片者抽样计划	61
3.8	数据收集计划	62
3.8.1	试验结果格式	62
3.8.2	阅片者研究的数据收集	63
3.8.3	阅片者培训	69
3.9	数据分析计划	70
3.9.1	统计学假设	70
3.9.2	协变量调整计划	71
3.9.3	报告试验结果	72
3.10	确定样本量	75
第四章	单一样本的参数估计与假设检验	76
4.1	二分类数据	77
4.1.1	灵敏度与特异度	77
4.1.2	阳性或阴性预测值	79
4.1.3	聚类二分类数据的灵敏度、特异度及预测值	81
4.1.4	似然比	83
4.1.5	优势比	85
4.2	有序数据	86
4.2.1	经验 ROC 曲线	87
4.2.2	拟合光滑曲线	87
4.2.3	固定假阳性率的灵敏度估计	91
4.2.4	ROC 曲线下面积与部分面积 (参数模型)	94
4.2.5	置信区间估计	96
4.2.6	ROC 曲线下面积与部分面积的估计 (非参数方法)	98
4.2.7	聚类数据的非参数分析	101
4.2.8	退化数据	102
4.2.9	参数、半参数以及非参数估计方法的选择	104
4.3	连续型数据	105
4.3.1	经验 ROC 曲线	106
4.3.2	拟合光滑 ROC 曲线: 参数、半参数和非参数法	106
4.3.3	估计 ROC 曲线的置信带	111
4.3.4	ROC 曲线下面积和部分面积: 参数、非参数和半参数法	112

4.3.5	ROC 曲线下面积的置信区间	113
4.3.6	固定假阳性率时的灵敏度与决策阈值	115
4.3.7	最佳工作点与决策阈值的选择	118
4.3.8	参数法、半参数法与非参数法的选择	121
4.4	ROC 曲线下整体面积或部分面积为一指定值时的假设检验	122
4.4.1	检验 MRA 是否具有检测显著颈动脉狭窄的能力	123
第五章	两种诊断试验准确度的比较	124
5.1	二分类数据	125
5.1.1	灵敏度与特异度	125
5.1.2	聚类二分类数据的灵敏度与特异度	127
5.1.3	阳性或阴性预测值	129
5.2	有序与连续型数据结果	131
5.2.1	检验两条 ROC 曲线是否相等	132
5.2.2	比较特定点的 ROC 曲线	134
5.2.3	FPR 在一定范围内变化时 TPR 的比较	136
5.2.4	ROC 曲线下整体面积或部分面积的比较	138
5.3	等效性检验	143
5.3.1	实例 3: 检验 ROC 曲线面积是否相等	145
第六章	样本量的估计	146
6.1	单个诊断试验准确度的样本量估计	146
6.1.1	以灵敏度和特异度为评价指标的样本量估计方法: 实例 1	147
6.1.2	以 ROC 曲线下面积为评价指标的样本量估计方法: 实例 2	148
6.1.3	聚类数据的研究	150
6.1.4	ROC 面积等于特定值时的检验假设	151
6.1.5	以固定 FPR 下的灵敏度为评价指标的样本量计算方法: 实例 2	151
6.1.6	以部分 ROC 曲线下面积为评价指标的样本量计算方法: 实例 2	153
6.2	以两种诊断方法准确度差值为评价标准的样本量估计方法	154
6.2.1	样本量计算软件	154
6.2.2	以两种方法灵敏度或特异度比较为评价标准的样本量估计方法: 实例 1	154
6.2.3	以两种方法阳性和阴性预测值为评价标准的样本量估计方法: 实例 1	156

6.2.4	两条 ROC 曲线下面积比较的样本量计算: 实例 2	157
6.2.5	聚类数据检验的样本量计算方法	159
6.2.6	以固定 FPR 时两个灵敏度差值为评价标准的样本量估计方法: 实例 2	160
6.2.7	以部分 ROC 曲线下面积比较为评价标准的样本量估计方法: 实例 2	161
6.3	评价两种诊断方法非劣效性或优效性的样本量估计方法	162
6.4	确定合适诊断阈值的样本量	165
6.5	多位阅片者研究的样本量估计方法	166
6.5.1	MRMC 样本量估计软件	167
6.5.2	无预试验数据的 MRMC 样本量估计方法	167
6.5.3	有预试验数据的 MRMC 样本量估计方法	172
6.6	其他样本量计算公式	174
第七章	诊断准确度研究中的 meta 分析	175
7.1	目的	176
7.2	文献检索	176
7.2.1	文献检索: 超声诊断周围动脉疾病的 meta 分析	180
7.3	纳入/剔除标准	180
7.3.1	纳入与剔除标准: 超声诊断周围动脉疾病的 meta 分析	183
7.4	提取文献信息	183
7.4.1	数据提取: 超声诊断周围动脉疾病的 meta 分析	185
7.5	统计分析	185
7.5.1	二分类数据	185
7.5.2	有序或者连续型数据	185
7.5.3	ROC 曲线下的面积	195
7.5.4	其他方法	197
7.6	公开发表	197
7.6.1	结果展示: 超声诊断周围动脉疾病的 meta 分析	198

第二部分 高级方法

第八章	独立数据的 ROC 回归分析	201
8.1	四项临床研究	202
8.1.1	MRA 诊断颈动脉血管病变准确性的实例	202
8.1.2	胰腺癌生物标志物诊断准确性的实例	202

8.1.3 畸变产物耳声发射诊断准确性的实例	203
8.1.4 影像学诊断前列腺癌分期准确性的实例	203
8.2 连续型检查结果的回归模型	204
8.2.1 ROC 曲线的间接回归模型	205
8.2.2 ROC 曲线的直接回归模型	208
8.3 有序型检查结果的回归模型	219
8.3.1 潜在光滑 ROC 曲线的间接回归模型	220
8.3.2 潜在光滑 ROC 曲线的直接回归模型	222
8.3.3 实例分析: 超声检查前列腺癌浸润情况	223
8.4 连续型检查结果调整协变量的 ROC 曲线	226
第九章 多位阅片者多项诊断试验的分析与评价研究	228
9.1 具有协变量的多项诊断试验效果比较研究	229
9.1.1 两项临床研究	229
9.1.2 有序型检查结果的间接回归模型	229
9.1.3 连续型检查结果的直接回归模型	234
9.2 多位阅片者多项诊断方法 (MRMC) 的评价研究	237
9.2.1 三种 MRMC 研究	238
9.2.2 MRMC 研究的分析方法	238
9.2.3 间质疾病案例分析	248
9.2.4 MRMC 研究分析方法的比较	249
9.3 多个诊断方法定位、诊断病灶的分析方法	250
9.3.1 LROC 分析方法	250
9.3.2 FROC 曲线分析方法	250
9.3.3 ROI 分析方法	251
第十章 校正证实性偏倚的方法	253
10.1 实例	254
10.1.1 肝闪烁造影术	254
10.1.2 阿尔茨海默氏症筛检试验	254
10.1.3 实例: 不确定性来源的发热	255
10.1.4 实例: 胰腺癌分期的 CT 和 MRI 检查	255
10.1.5 实例: 阿尔茨海默氏症 (AD) 国家协调中心的最小数据集	255
10.2 证实性偏倚的影响	256
10.3 具有二分类检查结果指标的单一诊断试验	257
10.3.1 MAR 假定下证实性偏倚校正方法	257

10.3.2	无 MAR 假定的证实性偏倚校正方法	260
10.3.3	实例分析: 继续肝闪烁造影术的分析评价	261
10.4	相关的二分类结果诊断试验准确度评价	263
10.4.1	无协变量的 ML 方法	263
10.4.2	实例分析: 两种阿尔茨海默氏症筛检试验准确度比较	265
10.4.3	具有协变量的 ML 估计方法	265
10.4.4	实例分析: 两个筛检痴呆的诊断试验效果评价	268
10.5	单一等级结果指标的诊断试验准确度评价	268
10.5.1	无协变量的 ML 方法	269
10.5.2	不确定性来源发热数据的分析	272
10.5.3	具有协变量的 ML 方法	273
10.5.4	实例分析: 阿尔茨海默氏症筛检试验的分析	277
10.6	相关的等级结果指标诊断试验准确度评价	278
10.6.1	潜在光滑 ROC 曲线的加权估计方程方法	279
10.6.2	基于似然方法的 ROC 面积	285
10.6.3	实例分析: 胰腺癌分期的 CT 和 MRI 诊断效果评价	287
10.7	连续型结果指标诊断试验准确度评价	288
10.7.1	MAR 假定下 ROC 曲线和曲线下面积的估计	289
10.7.2	非 MAR 假定下 ROC 曲线和曲线下面积的估计	295
第十一章	校正非完美金标准偏倚的方法	301
11.1	实例	302
11.1.1	粪便化验诊断粪类原虫感染病	302
11.1.2	结核菌素试验诊断结核菌病	302
11.1.3	实例: 二分类检查结果的 X 射线诊断胸膜增厚病	303
11.1.4	实例: 生物检测诊断 HIV	303
11.1.5	病理学家用有序型检查结果的试验诊断宫颈原位癌	303
11.1.6	等级和连续型检查结果的磁共振诊断颈内动脉狭窄	303
11.2	非完美金标准偏倚的影响	304
11.3	单一总体二分类结果的诊断试验	305
11.3.1	模型识别的条件	306
11.3.2	基于频率最大似然法估计可识别模型的参数	307
11.3.3	基于贝叶斯方法估计不可识别模型的参数	307
11.3.4	实例: 类原虫感染病例的分析	310

11.4	G 个总体单个二分类结果的诊断试验	311
11.4.1	估计方法	312
11.4.2	实例: 肺结核实例	314
11.5	单一总体多个二分类终点指标的诊断试验	316
11.5.1	检验模型的可识别性	316
11.5.2	条件独立假设下的最大似然估计	317
11.5.3	实例: 胸膜增厚病例的分析	318
11.5.4	可识别条件相关模型的最大似然法	319
11.5.5	HIV 生物检测实例	323
11.5.6	条件相关模型的贝叶斯方法	327
11.5.7	颈动脉狭窄病例 MRA 诊断的实例分析	327
11.6	G 个总体多个二分类终点指标的诊断试验准确度评价	328
11.6.1	CIA 条件下的 ML 估计方法	328
11.6.2	无 CIA 条件下的 ML 估计方法	329
11.7	单一总体多个等级终点结果指标的诊断试验准确度评价	330
11.7.1	CIA 条件下非参数 ROC 曲线估计方法	330
11.7.2	一些条件相关模型的 ROC 曲线估计方法	332
11.7.3	具有等级变量终点结果指标的宫颈原位癌诊断准确性评价 实例分析	333
11.8	单一总体多种类型终点结果的诊断试验准确性评价	334
11.8.1	具有连续型终点结果指标的 MRA 诊断严重颈动脉狭窄病 准确性分析	336
第十二章	Meta 分析中的统计学分析方法	338
12.1	二分类结果指标	338
12.1.1	随机效应模型: 超声诊断 PAD 的 meta 分析实例	339
12.2	等级或者连续型结果指标	340
12.2.1	随机效应模型	340
12.2.2	双变量方法	341
12.2.3	二分类回归模型	343
12.2.4	分层 SROC 曲线	344
12.2.5	其他方法	346
12.3	ROC 曲线下面积	346
12.3.1	实例分析: 经验贝叶斯法用于 DST 的 meta 分析	348

附录 A 实例介绍以及第八章数据	349
附录 B 估计方差和置信区间的刀切法及 bootstrap 方法	380
参考文献	384
索引	414

第一部分 基本概念和方法

