· 医学检验教育 ·

国内检验医学临床研究常见科研设计缺陷和统计学错误辨析*

胡志德[1]，胡成进[1△]，邓安梅[2▲]
（1. 济南军区总医院实验诊断科，山东济南 250031；2. 第二军医大学长海医院实验诊断科，上海 200433）

DOI：10.3969/j.issn.1673-4130.2013.02.059　　**文献标识码**：B　　**文章编号**：1673-4130(2013)02-0239-03

在开展检验医学临床研究的过程中，严谨认真地对待每一个试验细节，对获取的科学数据进行科学的统计学分析，是保证研究结论正确可靠的基础。笔者注意到，在国内检验专业杂志刊登的临床研究论文中，有很大一部分研究不同程度地存在设计缺陷，滥用统计学方法的现象更是屡见不鲜，严重地削弱了这些研究论文结论的可靠性。在此，笔者拟就目前国内检验专业杂志上刊登的临床研究论文常见的科研设计缺陷和统计学错误作一分析。

1　统计学上存在的一般问题

1.1　定量资料统计中的常见错误　与病理学、影像学检查手段相比，实验室检查最大的优势之一在于其可以用客观的连续变量予以表示。对于两组资料连续变量的比较，首先应分析资料是否服从高斯分布（正态分析），方差是否整齐。只有在满足这两个条件的基础上，才可以采用 t 检验（两组比较）或者方差分析（多组比较）对数据进行统计分析。然而，在临床工作，大部分数据都是呈偏态分布的，对于这些数据之间的比较，一般应采用 Mann-Whitney U 检验（两组比较）和 Kruskal Wallis H 检验（多组比较）。国内的部分检验专业的论文，在进行两组均数比较的时候，并不严格区分两组资料是否符合正态分布，是否为配对设计，几乎千篇一律地采用 t 检验进行两组均数的比较。有的研究在处理两组以上数据间的比较时，甚至反复采用 t 检验进行，增加了Ⅰ类误差的风险。

1.2　分类资料统计中的常见错误　对于分类资料的比较，首先需要明确的两个问题是：(1)目标变量的分类描述之间是否具有等级关系；(2)统计目的是为了明确构成比（率）的差异、目标变量的强度差异、目标变量与分类变量之间的变化趋势、还是目标变量在组间的一致性问题。对于无等级关系的分类资料构成比（率）的比较，一般采用独立样本卡方检验分析构成比（率）的差异是否具有统计学意义，采用配对卡方检验回答两种分类方法是否具有一致性的问题。需要注意的是，对于总体样本量小于 40 或有理论频数小于 1 的方格时，应采用 Fisher 确切概率法分析构成比（率）的差异[1]。对于目标变量为有序等级资料的研究，一般应采用 Ridit 分析比较多组数据之间目标变量的等级的强弱[2]；对于双向有序等级资料的分析，则通常采用趋势性检验分析分组因素与目标变量之间是否存在相同的变化趋势。国内的部分检验专业论文，忽视试验设计的特点，忽视目标变量与分类变量之间是否存在等级关系，忽视专业需求，将卡方检验视为万能检验对分类资料进行分析处理，造成了统计学结论和专业结论的脱节，极大地削弱了研究结论的可靠性。

1.3　未给出可信区间　任何统计学结论必然最终要回归专业结论。在某些情况下，有统计学意义不一定有专业意义，反之亦然。统计学处理结果的 P 值，只能反映这一结论犯Ⅰ类误差的概率，并不能体现实验因素引起效应量的变化幅度。因此，如果仅仅在研究论文中报道 P 值，可能会误导读者，夸大研究的价值。比如，某研究发现了冠心病患者（n=118）血浆载脂蛋白 M 的浓度为（1.375 7±0.149 3）ODu/mm^2，而健康对照人群（n=255）只有（1.350 2±0.128 8）ODu/mm^2，二者的差异具有统计学意义（P<0.05）[3-4]。但众所周知，载脂蛋白 M 的检测误差和生物学变异都远大于 0.02 ODu/mm^2，如此微小的差异虽然有统计学差异，但是并无多大临床价值。如果能列出两组患者载脂蛋白 M 差异的 95%CI，则读者便可以一目了然地判断研究实验因素效应的大小，并根据专业判断该研究是否具有临床价值。

2　诊断性试验常见的设计缺陷和统计学错误

对目标疾病进行快速且准确的诊断，是制定个体化治疗方案的前提。因为具有客观、微创的优势，实验室检查在疾病的诊疗中占据着十分重要的地位。开展诊断性试验，评价实验室检查手段对目标疾病的诊断能力，是实验室医学家的重要使命之一。国内检验专业杂志上刊登的论文中，有很大一部分属于诊断性试验论文，这些论文都或多或少存在设计缺陷和统计学错误。笔者仅列出以下几点常见设计和统计学错误。

2.1　以健康个体作为对照组　国内开展的部分诊断性试验研究，以健康个体作为对照人群，这其实是一种很不科学的行为。健康个体和疾病患者在症状和体征上已经有了很大的区别，通常无需借助实验室标志物就能进行鉴别诊断。因此采用健康个体作为对照并不足以体现实验室标志物的鉴别诊断效率[5]。正确的对照组应该是在症状和体征上与疾病组高度相似，在临床工作中极易于目标疾病混淆的一类人群。比如，欲评价甲胎蛋白对肝癌的诊断能力，对照组就应该设立为肝硬化、肝囊肿、肝炎等一类与肝癌难以鉴别诊断的疾病。

此外，诊断性试验研究要求研究对象具有较好的临床代表性[5]。因此，最好采用连续招募的方式确定研究对象，以确保诊断性试验的疾病组与对照组的病例分布情况与临床工作一致。而国内的很多诊断性试验研究，未就研究的数据采集方式（前瞻还是回顾）、病例招募方式、纳入和排除标准等进行详细的说明，使得读者无法判断研究结论的可靠性和的临床适用范围。

2.2　采用参考范围上线作为实验室标志物的诊断界值　对于定量分析的实验室标志物，其诊断阈值的确定应该是充分考虑该标志物在疾病人群以及与疾病人群相似的人群中的分布状况，充分权衡漏诊和误诊所带来危害。国内开展的部分诊断性

*　基金项目：国家自然科学基金资助项目（30972370）；上海市科学技术委员会科研计划项目（09JC1405400；11JC1410902）。　△　通讯作者，E-mail：hcj6289@163.com。　▲　通讯作者，E-mail：amdeng70@163.com。

试验，忽视上述原则，错误地以参考范围上限作为诊断界值。众所周知，参考范围上限只是反映实验室标志物在健康个体中的分布状况，并未充分考虑其在疾病患者以及疑似疾病患者中的分布状况，因此不宜作为诊断界值[6]。

对于定性的诊断标志物，可以直接绘制四格表而计算出诊断敏感性、特异性、阴/阳性似然比、阴/阳性预测值等指标。对于连续变量，通常需要采用受试者工作特征曲线(ROC)分析法确定其总体诊断效率。受试者工作特征曲线分将不同诊断界点所对应的敏感性和特异性汇总与同一条曲线上，通过曲线下面积反映目标试验的总体诊断性能[7]。研究人员可以根据专业需要从曲线上选择不同的界点作为推荐的诊断界点。

3 病例对照研究中的常见设计缺陷和统计学错误

开展病例对照研究，旨在分析某种特征与疾病发生的关系，探索疾病发生与发展的原因，为开展队列研究和开发新的治疗手段提供思路[8]。国内检验专业杂志刊登的论文中，有很大一部分属于病例对照研究，然而，这些研究大多不同程度地存在设计缺陷和统计学错误，主要表现在：

3.1 未采用多参数的分析方法同时分析患者特征与疾病的关联 病例对照研究从本质上讲属于观察性研究，疾病组与健康对照组的差异可能会同时与多种患者特征有关。因此，如果需要确定待研究的指标与疾病的关联，则需要充分考虑潜在的“混杂因素”的干扰。对于“混杂因素”的排除，一是可以设定严格限制纳入/排除标准或者采用配对的方式进行研究，但这种方法往往增大了研究难度，因此并不常用。目前多采用第二种方法，即将“混杂因素”因素作为一个协变量进行分析，以明确在校正了“混杂因素”的前提下，待研究的患者特征与疾病之间是否还存在关联。因此，在开展病例对照研究时，应尽可能地详细列举疾病潜在的关联因素，以便在进行在统计学分析时能确定各种因素与疾病的关联关系的强弱。

国内检验专业杂志上刊登的部分病例对照研究，对受试对象特征的描述十分简单，有的甚至仅仅提供了性别和年龄等最基本的特征，不仅让读者无从判断研究结论的适用范围，同时也因为未能排除“混杂因素”的干扰，造成研究结论不可靠。正确的处理方式应该是将所有潜在的“混杂因素”作为因变量，采用多参数的数学模型(比如 Logistic 回归模型)，分析在多因素校正的情况下，各个关联因素与疾病的独立关联关系[9]。

3.2 结果的解释与结论脱节 病例对照研究又称“横断面”研究，因为其研究的两个因素：患者“特征”(比如胆固醇增高)与“结局”(发生冠心病)是同时发生的，因此在时序上无法明确因果关系问题。特征的改变与结局的发生之间可能存在三种关联关系：(1)某种特征的改变引发了结局(疾病)；(2)结局(疾病)引起了某种特征的改变；(3)是第 3 个(组)因素同时引起了结局的发生与某种特征的改变，即疾病的发生与特征的改变之间并无直接的因果关系。

国内的部分病例对照研究，受传统观念的影响，忽视研究的“时序性”问题，将“特征的改变”与“疾病”之间的关联解释为某种特征的改变是引起疾病的原因之一，实为不科学和严谨的表现。

4 队列研究中常见的设计缺陷和统计学错误

与病例对照研究不同，队列研究(又称前瞻性研究)是先确定研究人群(队列)，然后对研究人群进行随访，记录结局，即研究的观察终点(通常为疾病的发生或者患者死亡)。然后分析患者进入队列时候的特征(即基线特征)与观察终点的关系[8]。队列研究可以在时序上明确“特征的改变”与“结局”的关系，因此较病例对照研究具有更高的论证强度。有部分国内检验专业杂志上刊登的论文属于队列研究(多以疾病预后研究为主)。但是这些研究都不同程度地存在设计缺陷，主要表现在：

4.1 队列的基线特征、随访方式以及失访人群的介绍不清晰 队列研究的重点在于随访，随访时间的长短、随访频率的高低以及失访率的大小直接决定了研究的质量。因此在进行研究的过程中，有必要浓墨重彩地介绍随访的方式、频率、失访率以及随访时间的长短，以便读者以及循证医学研究者对研究的质量进行评价。高质量的队列研究具在系统综述(system review)中占有更高的权重，是重要的循证医学证据，因此也更容易在疾病指南的制定过程中占有一席之地。国内检验专业同行开展的部分队列研究，在材料与方法中并未详细交代随访的方式与频率，也未交代失访率以及释放数据的处理方式。有的研究为了降低失访率甚至从队列中删除了失访病人数据，是一种极不严谨的科研行为。

4.2 没有采用多参数的数学模型分析各个基线特征与研究对象结局的关系 与病例对照研究一样，队列研究也不可避免地受到一些“混杂因素”的干扰。因此，在纳入研究对象时候，应尽可能地明确患者的“基线特征”，以便再进行统计学分析时能够考虑更多的变量。对于队列研究数据的分析，需要考虑到时间对结局的影响，因此一般以 Kaplan-Meier 生存曲线反映基线特征与受试对象结局的关系，以 Logrank 检验分析某以特征与结局发生的关系，最后以多参数的 Cox 风险比例模型分析基线特征与观察终点的独立关系[10]。

国内检验专业杂志上刊登的部分队列研究论文，对于队列的基线特征介绍不够详细，让读者无法判断可能存在的混杂因素。在描述观察终点的发生状况时，仅仅简单地以“一年生存率”、“一年发生率”等文字进行简单的描述，而未采用 Kaplan-Meier 生存曲线来展示结果，更没有以多参数的 Cox 风险比例模型校正潜在的混杂因素，导致研究结果可靠性大打折扣。

5 方法学对比研究中常见的设计缺陷和统计学错误

受经济、地域、观念等因素的限制，对于同一个检验项目，往往会有不同的检测方法。因此，有必要开展检验方法学之间的对比实验，评价针对同一检验项目多种检测方法的可比性，为这些检验方法的临床解释提供参考，为不同医疗单位检验结果的“共享”提供依据，最终达到节约医疗资源的目的。目前在国内检验专业杂志上刊登的论文中，有一部分内容属于检验方法学比对的研究。我们以定量资料的方法学比对为例，浅析国内检验方法学对比研究存在的设计缺陷和统计学错误，探讨正确的统计学处理方法。

5.1 对两种准确性均欠佳的方法进行比对 一种新的检验方法之所以能应用于临床实践，检测结果的准确性(与真实值的差异)是基础。换而言之，这种新的检测方法应该具有“溯源性”。如果待评价的两种方法本身“无源可溯”，那么即使两种方法有良好的相关性和一致性，也不能说明两种方法具有临床应用价值。比如，有 A～E 五个浓度不等样本，但已知其中某种物质的真实的浓度分别为 2、3、1、5、4(单位略)，分别用甲乙两方法进行对该物质的浓度进行检测，两种方法的检测结果均为 1、2、3、4、5(单位略)，虽然两种方法具有较好的一致性，但是这种一致性并无多大临床价值，因为两种检测方法的检测结果均准确性欠佳。

5.2 以 t 检验进行比对 有部分检验方法学对比研究，以独立样本 t 检验或者配对样本 t 检验比较两种方法的检测结果，试图以“两种检测方法的结果均数之间无差异”这一统计学结

论来说明两种方法具有良好的一致性。这种统计学处理方式是完全错误的。t 检验回答的是两种检测方法所得出的检验结果均数之间无差异，并未回答两种检测方法的一致性问题。比如，有 A～E 5 个浓度不等样本，分别与甲乙两法进行检测，甲法的检测结果为 1、2、3、4、5(单位略)，而乙法的检测结果为 5、2、3、1、4。若采用 t 检验对数据进行分析，虽然两组检测结果均数之间无差异($P=1.00$)，但两种方法并无一致性可言。

5.3 只评价了两种方法的相关性，而未评价一致性　有部分检验方法学对比研究，采用 Pearson 法对两种检验方法进行比对，试图通过相关系数来反映两种方法的可比性。这种统计学处理方式也是不严谨的，因为相关性分析回答的是“相关性”问题，而非“一致性”问题。当存在系统误差时，两种检测方法完全可以具有良好的相关性，而无一致性。比如，有 A～E 五个浓度不等样本，分别与甲乙两法进行检测，甲法的检测结果为 1、2、3、4、5(单位略)，而乙法的检测结果为 1.5、2.5、3.5、4.5、5.5。若采用 Pearson 法对数据进行分析，两组检测具有良好的相关性($r^2=1$，$P<0.01$)。但是乙法的检测结果较甲法高出了 0.5 个单位，因此，两种方法也并无一致性可言。

5.4 推荐的统计学方法　对于检验方法比对实验，应该分别从统计学上和专业解释上证实两种方法是否具有可比性。正确的统计学方法为：首先以配对 t 检验分析两种方法检验结果之间的差异是否具有统计学意义和专业意义。同时，进一步采用 Bland-Altman 法绘制 Bland-Altman 图，计算两种检测方法的一致性限度，并且分析两种方法的一致性限度是否符合专业要求[11]。若两种方法的一致性限度符合行业标准(比如 CLIA′88)或者一些行业共识。若两种方法的一致性限度已经符合了专业要求，则进一步采用 Pearson 法或者 Spearman 法分析两种方法的相关性，并对相关方程的截距和斜率进行假设检验，分析截距与 0 之间的差异，斜率与 1 之间的差异是否具有统计学意义。这样就可以从统计学上和专业上同时回答“两种检验方法是否具有可比性”的问题。

6 结　语

严谨的科研设计与科学的统计学处理是开展高质量检验医学临床研究的基础。只有注重试验设计的每一个细节，根据分析目的正确地运用统计学方法对数据进行分析，并合理解释统计学结论，研究结果才具有较高的论证强度，才会引起同行特别是循证医学工作者的注意，并最终实现科学研究成果向临床应用的转化。

参考文献

[1] 陆运清. 用 Pearson′s 卡方统计量进行统计检验时应注意的问题[J]. 统计与决策，2009，19(15)：32-33.

[2] 刘明华，张晋昕. Ridit 分析与秩和检验在等级资料处理时的关系[J]. 循证医学，2010，10(5)：282-285.

[3] Arora BM, Singh MK. Evaluation of ApoM as a biomarker of coronary artery disease[J]. Clin Biochem, 2009, 43(10/11): 932.

[4] Su W, Jiao G, Yang C, et al. Evaluation of apolipoprotein M as a biomarker of coronary artery disease[J]. Clin Biochem, 2009, 42(4/5): 365-370.

[5] Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies[J]. Ann Intern Med, 2011, 155(8): 529-536.

[6] Geffre A, Friedrichs K, Harr K, et al. Reference values: a review[J]. Vet Clin Pathol, 2009, 38(3): 288-298.

[7] Cook NR. Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve[J]. Clin Chem, 2008, 54(1): 17-23.

[8] Song JW, Chung KC. Observational studies: cohort and case-control studies[J]. Plast Reconstr Surg, 2010, 126(6): 2234-2242.

[9] LaValley MP. Logistic regression[J]. Circulation, 2008, 117(18): 2395-2399.

[10] Benitez-Parejo N, Rodriguez del Aguila MM, Perez-Vicente S. Survival analysis and Cox regression[J]. Allergol Immunopathol (Madr), 2011, 39(6): 362-373.

[11] 萨建，刘桂芬. 定量测量结果的一致性评价及 Bland-Altman 法的应用[J]. 中国卫生统计，2011，28(4)：409-413.

(收稿日期：2012-09-28)

· 医学检验教育 ·

中国与美国医学检验专科教育课程学时比较*

唐　宜，尹　红，甘晓玲

(重庆医药高等专科学校医学技术系，重庆 401331)

DOI：10.3969/j.issn.1673-4130.2013.02.060　**文献标识码**：B　**文章编号**：1673-4130(2013)02-0241-03

我国医学检验专科教育是高等医学检验教育的重要组成部分，为基层医疗单位培养高素质技能型实用人才。课程设置和教学内容是人才培养计划的核心，直接反映了人才培养计划的科学性、合理性和实用性，而课程的学时长短和相互间学时比例是保证人才培养计划完成的重要保证。我国医学检验专科教育与美国社区大学两年制医学检验教育在学生来源、人才培养层次、就业去向和从业岗位等有相似性和可比性。本文拟通过中美两国医学检验高职基础课程、专业基础课程和专业课程学时比例以及专业课程间学时比例的比较，找出其共性、差异和差距，分析其成因，取长补短，相互借鉴。

我国医学检验教育招收国家统考高中毕业生，学制 3 年，毕业生授予大专文凭。美国的医学检验职业教育招收国家统考高中毕业生，学制 2 年，毕业生授予副学士文凭。我国的大专文凭和美国的副学士文凭都是本科以下文凭，具有等同性。本研究通过书信、网络查询和访问等形式，获得 5 所国内医学检验专科教育人才培养计划，5 所美国两年制社区大学的医学检验专科教育人才培养计划。对所得资料进行了分析、整理和归类。将所有必修课程按照基础课程、专业基础课程和专业课

* 基金项目：重庆市教委立项资助项目(103441)。