

如何撰写高质量的流行病学研究论文

第八讲 非劣效性和等效性随机对照试验的报告规范 ——CONSORT 声明的扩展

朱成斌 詹思延

【导读】 非劣效性和等效性随机对照试验的应用日益广泛,为医疗卫生干预提供了重要的证据。对其结果的准确评价有赖于充分、准确的报告。CONSORT 声明向非劣效性和等效性随机对照试验的扩展是 2006 年提出的又一个报告规范。本文介绍非劣效性和等效性随机对照试验的特殊方法学问题、实施与报告现状以及 CONSORT 扩展声明的清单,并对清单中的部分条目进行了解释与说明。

【关键词】 随机对照试验;非劣效性试验;等效性试验

多数随机对照试验(RCT)的目的是确定一种干预是否优于对照,即优效性试验(superiority trial)。其中,以安慰剂作对照的随机双盲临床试验一直以来被视为药物开发中的金标准,在确认新药疗效方面发挥了重要作用。然而,如果已经有了疗效肯定的药物,仍用安慰剂作对照,会面临伦理学的质疑。随着有效药物的大量出现,有突破性疗效的药物越来越少,临床试验的目的也有所转变,在阳性对照试验中,更多的情形是探索新药与标准药物相比疗效是否不差或相当,并不一定必须超过标准药物,由此提出了非劣效性和等效性试验(noninferiority/equivalence trial)。等效性试验的目的是确定一种处理是否与另外一种处理在疗效上相似;非劣效性试验的目的是确定一种新的处理是否不次于阳性药物或非药物的标准治疗。由于证明完全相同是不可能的,因此只能将处理效果定义为预先指定的一个非劣效性差值(Δ)。等效性被定义为处理效果介于 $-\Delta$ 和 Δ 之间。

1. 非劣效性和等效性试验的方法学问题:在试验设计、实施、分析和结果解读方面,非劣效性和等效性试验有其特点^[1]。

(1) 试验假设:在优效性试验中,无效假设指的是两种处理同样有效,备择假设是它们具有区别。第 I 类错误指的是处理本身无效,却被错误认为有效;第 II 类错误指的是处理本身有效,但其效果未被发现。在非劣效性试验中,无效假设和备择假设正好倒过来;第 I 类错误指的是错误的接受了一种劣效的新处理,第 II 类错误指的是错误地拒绝了一种非劣效的处理。

(2) 试验设计:非劣效性或等效性试验要求参照处理的效果是确定的或已得到广泛应用,因此使用安慰剂或空白对照会被视作不符合伦理学原则。

在非劣效性或等效性试验中,受试者和结局的测量方法

应该与以前的试验相似。利用可信区间方法计算试验所需的样本量。样本量取决于选定的可信水平、第 II 类错误的风险(或想达到的把握度)以及 Δ 。 Δ 可以被理解成两组之间在均数、比例,或比值比、危险比及风险比的对数值方面的差异。经常选择有重要临床意义的最小值作为 Δ 。非劣效性试验通常需要比优效性试验更大的样本量。

(3) 试验实施:试验的实施应该与以前的试验一致。为了避免增加作出错误结论的风险,应尽量避免那些可能会影响不同处理组间真实差异的问题,如依从性较差、失访、募集一些不大可能应答的病例,以及交叉处理。

(4) 试验分析:在设计、分析和报告非劣效性和等效性试验时,可信区间方法则更可取,因为可以得到更多信息。

对于那些未接受分配的处理或违反试验程序的病例,如果分析时将其剔除,那么整个试验可能朝任何方向发生偏倚。在非劣效性和等效性试验中,意向性(ITT)分析往往会增加错误作出非劣效性结论(第 I 类错误)的风险,所以实际处理或符合方案分析这样的非 ITT 分析也许是可取的。当两者结论一致时,结果的可信度会更高。

在中期分析时,非劣效性试验与优效性试验在基本原理上存在一些差异。如果试验结束前已确定为非劣效,从伦理学要求来讲不需要提前终止试验。如果某种处理显然是劣效的,那么终止试验在伦理学上是正当的。

(5) 结果解读:对非劣效性试验的结果进行解读取决于处理效果的可信区间与 Δ 和无效值的相对位置。双侧等效性需要同时考虑 Δ 和 $-\Delta$ 两个差值,而且为了得出等效性结论,需要可信区间位于 $-\Delta$ 和 Δ 之间。

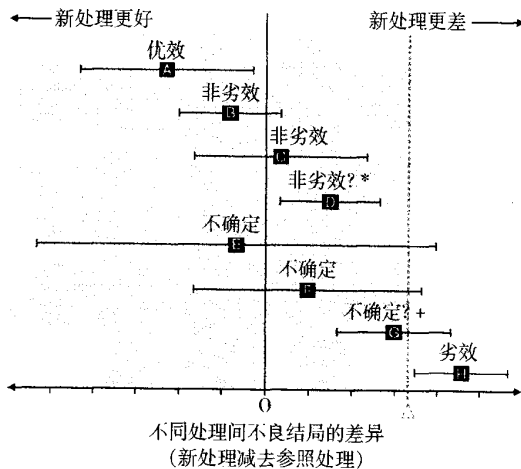
虽然单侧和双侧可信区间都允许进行统计学推断,但在绝大多数非劣效性试验中采用双侧可信区间才是合适的。图 1 显示的是非劣效性试验时若干种可能的双侧可信区间情况。

在优效性试验之后作出非劣效的结论,或者在非劣效性试验后,与安慰剂相比,得出新处理有效的结论,都是不合适的。这种推断的前提是试验在所有方面维持不变,但这不可

作者单位:100083 北京大学医学部公共卫生学院流行病学与卫生统计学系

通讯作者:詹思延,Email:siyan-zhan@bjmu.edu.cn

避免的会受到质疑,所以在进行此类推论时需谨慎。



注:误差条图表示双侧 95% 可信区间,灰色区域表示非劣效性区域。图中除了 D 和 G,其余的结果都很好理解。D 中,由于可信区间没有包括 Δ ,因此可以认为非劣效,但是和标准处理相比,新处理又明显更差,这种罕见的情况往往需要非常大的样本量,因此不大可能发生。另外,非劣效性差值太宽时也容易出现 D 这种情况。G 的不确定性在于,虽然它仍然有可能小于 Δ ,但是新处理又明显比标准处理更差

图1 非劣效性试验中可能观察到的处理之间不良结局(毒副作用)的差异(引自 Piaggio G, 2006)

1980-2000 年间自称对儿童细菌性脑膜炎死亡率进行等效性研究的 25 个 RCT 进行综述的结果表明^[8],只有 2 个(8%)试验被设计成用于等效性检验。对发表于 1988-1998 年间 3 份外科期刊的 90 个无统计学意义或“阴性”RCT 的回顾发现,39% 的试验满足等效性试验的预定标准^[9]。在最近对 188 个阴性结果癌症试验的回顾表明,只有 3 个研究使用了非劣效性或等效性分析^[10]。McAlister 和 Sackett^[11]研究了 4 个高血压方面的大型“阴性”RCT,对等效性试验的方法学进行了评价。只有 2 个试验同时用到了 ITT 和符合方案(或实际处理)分析,只有 1 个试验事先指定了等效性差值,没有一个试验的样本量足以检验等效性假设。

3. CONSORT 声明向非劣效性和等效性试验的扩展:鉴于非劣效性或等效性研究报告中存在的问题,在 McAlister 和 Sackett^[11]的工作基础上,来自世界卫生组织的 Gilda Piaggio 博士领衔一个写作小组对 CONSORT 清单作了修改和扩展^[12],特别是其中的条目 1~7、12、17 和 20。扩展的内容主要集中于非劣效性试验,但也对等效性试验提出了一些参考意见。新的内容在表 1 的清单中以斜体表示。对其他条目的解读请参见本系列讲座第四讲(编者注:本刊 2006 年第 12 期第 1086~1088 页)。

4. CONSORT 扩展声明清单中部分条目的解释与说明:

[条目 1] 标题和摘要 作者应明确说明试验是一个非劣效性或等效性试验。如标题为“口服普那霉素和标准青霉素治疗成人丹毒的比较研究:随机化,非劣效性的非盲试验”,摘要中出现:“设计-多中心,平行组,非盲法,随机化的非劣效性试验”^[13]。

[条目 2] 背景 基本原理的科学背景和解释中要包括非劣效性或等效性设计的理论基础。如果以前的试验或其系统综述表明参照处理优于安慰剂,那么应该列出其效果大小和可信区间。如果没有这样的试验结果,那么应该提供其他证据证明参照处理有效。同时,应该列出新处理优于参照处理的其他优点,以证明新处理的应用是合理的。

[条目 3] 受试者 受试者的人选标准中要详细说明非劣效性或等效性试验的受试者是否与确立参照处理有效性的试验的受试者相似,以及数据的收集方法和地点。和以前的试验相比,如果受试者的特征发生了变化,应该报告并加以说明。

[条目 4] 干预措施 详细说明非劣效性或等效性试验中用到的参照处理是否与以前确立有效性的试验中用到的处理相同(或非常相似),以及实际上它们是什么时候如何接受的干预。当前试验中给予的干预措施和以前试验中明确有效的参照处理之间的任何区别,都应该报告并加以解释。

[条目 5] 目的 给出特定的研究目的和研究假设,包括有关非劣效性或等效性的假设。如“一种快速灌注剂替奈普酶的效果和安全性与阿替普酶疗法等同。在这个双盲,随机化对照研究中,对这个假设进行检验。”^[14]作者应该说明非劣效性或等效性假设应用于哪个结局,优效性假设又是应

2. 非劣效性和等效性试验报告现状:2004 年 10 月在 Cochrane 对照试验登记系统内搜索关键词“等效”或“非劣效”,结果在 415 918 个试验中仅发现 1021 个(占 0.2%)对应的试验。对 1997-2007 年 7 月我国 CNKI 数据库的检索发现,自称等效性研究的文献有 800 篇,其中生物等效性研究就占了 766 篇,而非劣效性研究仅 9 篇。上述检索结果可能会产生误导,因为不是所有的此类试验都用这些术语,而且在报告优效性试验的阴性结果时,等效性这个术语经常被不合适的使用。此外,非劣效性试验经常被冠以等效性试验的标题。

由此看来,似乎非劣效性和等效性试验相对很少,实际情形并非如此。由于这类设计主要用于仿制药疗效的评价,虽然客观上使用很多,但作为学术论文发表的较少。近几年,国内多名学者对这两种试验的概念、基本原理和方法学做了介绍和阐述^[2,5]。药品注册管理机构也提出,对国内尚未上市的药品,无论是创新药还是仿制药,如选国内已上市同一治疗领域的药物作为阳性对照药,则应至少验证其具有非劣效性。与已上市药物具有相同活性成分的药品,应进行生物等效性或临床等效性验证^[6]。

近年来,国外几个系统综述关注了此类研究报告的质量。Greene 等^[7]对发表于 1992-1996 年间的 88 篇自称等效性的研究作了系统综述,发现一些方法学上的缺陷。67% 的文献根据优效性的非显著性检验不恰当的直接得出等效性结论。只有 22% 的研究采用了合适的程序。对发表于

用于哪个。通常,非劣效性或等效性假设指的是主要终点,而对新处理的期望是能带来其他优点,如给药方便,耐受性较好,毒性较低或价格便宜等。

[条目 6] 结局 要详细说明非劣效性或等效性试验中用到的结局,是否与以前确立参照处理有效性试验的结局相同(或非常相似),同时说明提高测量质量的方法。新试验和以前试验中在结局测量上的任何区别都应该指出并加以解释,特别是评价标准的改变。理想中,结局应该维持不变,但是随着对某种疾病的理解、治疗和预防水平的提高,经常导致某些观念发生变化。如早期的艾滋病试验使用死亡结局,后来研究者将指标转向临床事件,现在又转向新的替代指标。

[条目 7a] 样本量 需要详细说明样本量的计算是否

按照非劣效性或等效性标准,而且需要指明选定等效性差值的理论基础。如“参考以前的研究,估计每年两组处理病例的主要事件发生率约为 3.1%。为了在单侧 α 为 0.025 时得到 90% 的统计学把握度,每个处理组大约需要 1600 人年的暴露,才能对 ximelagatran(一种凝血酶抑制剂)与华法令进行非劣效性比较。假定平均随访 16 个月,那么大约需要 2400 例患者。”^[15]应该指明非劣效性或等效性的差值,且最好是依据临床意义加以确定。

样本量的计算通常是基于这样一个假设:处理组之间差异的点估计值为 0。如果观察到的点估计值为 0,那么图 1 中的例子 F 和 G 能够满足非劣效性标准。如果样本量足够大,即使点估计值在 0 和 Δ 之间,也可以得出非劣效性的结论。如果真实的效果大于 0,就需要增大样本量。

表1 非劣效性或等效性试验报告的条目清单(增补或修改的内容以斜体表示)

文献内容和主题	条目编号	条目描述(根据非劣效性或等效性试验作了调整)
标题和摘要	1*	受试者是被如何分配到干预组的,明确说明试验是一个非劣效性或等效性试验
介绍背景	2*	基本原理的科学背景和解释,包括非劣效性或等效性设计的理论基础
方法		
受试者	3*	受试者的人选标准(详细说明非劣效性或等效性试验的受试者是否与确立参照处理有效性的试验的受试者相似)以及数据收集的设置和地点
干预措施	4*	每个试验组要接受的干预措施的细节,详细说明非劣效性或等效性试验中用到的参照处理是否与以前确立有效性的试验中用到的处理相同(或非常相似),以及实际上他们是什么时候如何接受的干预
目的	5*	特定的研究目的和研究假设,包括有关非劣效性或等效性的假设
结局	6*	对主要结局和次要结局的测量作出清楚的定义,详细说明非劣效性或等效性试验中用到的结局是否与以前确立参照处理有效性的试验中用到的结局相同(或非常相似),同时,说明用于提高测量质量的方法
样本量	7*	样本量是如何确定的,需要详细说明样本量的计算是否按照非劣效性或等效性标准,而且需要指明选定等效性差值的理论基础
随机化序列的产生	8	产生随机分配序列号的方法,包括任何限定情况(如分组、分层)
分配隐匿	9	按照产生的序列实施随机分配的方法(如封筒法或电话总机),指明在给予干预前是否隐匿序列号
实施	10	谁产生分配序列号,谁征集受试者,谁将受试者分配给各处理组
盲法	11	受试者,干预措施的执行者以及结局的评价人员是否知道各处理组的分配情况。如果相关,如何评价盲法成功与否
统计方法	12*	用于对主要结局进行组间比较的统计方法,指明是否用了单侧或双侧可信区间的方法。进一步分析的方法,如亚组分析和调整分析
结果		
受试者流程	13	受试者经历每个试验阶段的流程(强烈建议绘成图表)。对于每一个处理组,明确报告受试者有多少人得到随机分配,接受了预定的处理,完成了试验方案并依据主要结局进行了分析。对试验中与原计划有出入的方案进行描述并给出理由
征集受试者	14	规定征集和随访的周期和日期
基线数据	15	每个处理组的人口统计学基线数据和临床特征
分析的数目	16	分析中每组受试者(分母)的人数,以及是否进行了“ITT”和/或替代分析。可行的话,将结果以绝对值标出(如 10/20,而不只是 50%)
结局和估计	17*	对于每一个主要和次要结局,列出每组的结果和估计的效果大小及其精度。对于非劣效性或等效性假设的结局,可以利用图形表示可信区间和等效性差值
辅助分析	18	为了应对多样性的问题,需要对任何其他的分析作出报告,包括亚组分析和调整分析,揭示哪些是预期的,哪些是探索性的
不良反应事件	19	每个干预组中发生的所有重要的不良反应或副反应
讨论		
解读	20*	对结果进行解读,考虑非劣效性或等效性假设以及其他试验假设,发生潜在偏倚或偏差的来源,以及与分析和结局的多样性产生关联的危险
可推广性	21	试验结果的可推广性(外部有效性)
总体证据	22	结合现有的证据,对结果进行全面解读

[条目 7b] 终止原则 如果可行的话,应该对所有的中期分析和终止原则(以及是否与非劣效性或等效性假设相关)作出说明。如“调查了所有妇女人数的 40% 至 70% 后,需要对中期安全性分析作出计划。如果第一次中期分析时任何 P 值小于 0.007,并且第二次 P 值小于 0.012,那么可以认为和长期疗法相比,短程疗法的 HIV 传染率增加……”^[16],依据 P 值作为试验中期的终止标准已经成为惯例,这些调整过的 P 值与加宽的可信区间是一致的。

[条目 12] 统计方法 要指明是否用了单侧或双侧可信区间的方法。以二分变量的结局为例,“两组 ITT 受试者中每年发生主要事件的比例,以及与差异有关的单侧 97.5% 可信区间,将利用出现首发事件的时间进行估计……主要分析中的 Δ 是基于绝对事件的率差……如果对主要事件率差估计值的 97.5% 可信区间上限低于 Δ ,那么 ximelagatran 与华法令相比的非劣效性将被接受”^[15]。为了证明在显著性水平 α 上存在非劣效性,处理效果的单侧 $(1-\alpha)$ 可信区间上限[或者相对应的双侧 $(1-\alpha/2)$ 可信区间]必须低于差值 Δ 。

[条目 17] 结局和估计 对于非劣效性或等效性假设的结局,可以利用图形表示可信区间和等效性差值。如一项多中心随机双盲临床试验,比较在第三产程口服米索前列醇是否与催产素等效,就用图 2 表示了 2 组发生严重产后出血(失血 ≥ 1000 ml)的相对危险($RR = 1.39, 95\% CI: 1.19 \sim 1.63$)^[17]。图 2 中垂直的虚线代表事先确定的临床等效性差值,实线代表无效假设。可见,米索前列醇与催产素并不等效。

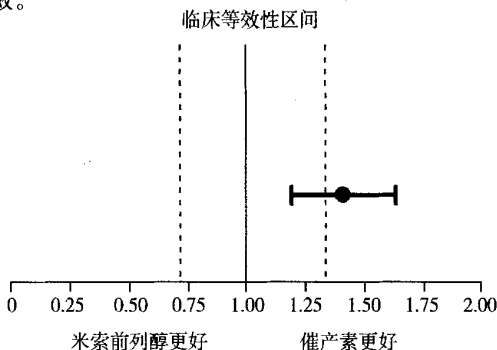


图2 比较米索前列醇和催产素发生严重产后出血的相对危险(引自:Gülmezoglu AM, 2001)

[条目 20] 对结果的解读 考虑非劣效性或等效性假设以及其他试验假设,发生潜在偏倚的来源,以及与分析和结局的多样性产生关联的危险。如对非劣效性下结论。“根据等效性的定义,从统计学上来说长期-短程疗法与长期-长期疗法的效果是等价的……两组之间差异的 95% 可信区间为 5.3% (接近 6.0% 的界限)”^[16]对新药的劣效性(或参照药物的有效性)下结论。“虽然试验旨在评价替罗非班与阿昔单抗比较时的非劣效性,但是结果发现,与阿昔单抗相比,替罗非班对主要缺血性事件提供的保护更少……为了满足对等效性的定义,两者比较时风险比的 95% 可信区间上限必须

小于 1.47……有 2398 例患者的替罗非班组与 2411 例患者的阿昔单抗组相比,前者主要终点事件的发生率更高(7.6% vs. 6.0%; 风险比 1.26; ……双侧 95% 可信区间是 1.01 vs. 1.57,表明阿昔单抗优于替罗非班; $P = 0.038$)”^[18]在一个被设计用于优效性评价的试验中,对新药的非劣效性下结论。“SYNERGY 规范规定,如果依诺肝素未能显示出优于普通肝素,那么就应该进行非劣效性分析……对于非 ST 段抬高 ACS 高危患者的治疗,依诺肝素不优于普通肝素,但却是非劣效的。”^[19]

5. 结论:在试验设计、实施、分析和结果解读方面,非劣效性和等效性试验的报告应当做到明确无误,从而使读者能够全面准确地理解报告的结果。希望本扩展能对非劣效性和等效性试验的报告质量产生促进作用。最新版本的所有 CONSORT 指南可以在以下网站查询: <http://www.consort-statement.org>。

参 考 文 献

- [1] D'Agostino RB, Massaro JM, Sullivan LM. Noninferiority trials: design concepts and issues—the encounters of academic consultants in statistics. *Stat Med*, 2003, 22:169-186.
- [2] 刘玉秀,姚晨,陈峰,等. 非劣效性/等效性试验中的统计学分析. *中国临床药理学杂志*, 2000, 16(6):448-452.
- [3] 夏结来,于莉莉,蒋红卫. 新药疗效评价的区间检验. *第四军医大学学报*, 2003, 24(24):2209-2212.
- [4] 闫宇翔,王洪源,王仁安. 等效性评价方法研究现状. *药物流行病学杂志*, 2003, 12(5):253-256.
- [5] 于莉莉,夏结来,陈启光,等. 显著性检验与等效性检验的区别与联系. *中国卫生统计*, 2005, 1(22):37-38.
- [6] 黄钦,赵明. 对临床试验统计学假设检验中非劣效、等效和优效性设计的认识. *中国临床药理学杂志*, 2007, 23(1):63-67.
- [7] Greene WL, Concato J, Feinstein AR. Claims of equivalence in medical research: are they supported by the evidence? *Ann Intern Med*, 2000, 132:715-722.
- [8] Krysan DJ, Kemper AR. Claims of equivalence in randomized controlled trials of the treatment of bacterial meningitis in children. *Pediatr Infect Dis J*, 2002, 21:753-758.
- [9] Dimick JB, Diener-West M, Lipsett PA. Negative results of randomized clinical trials published in the surgical literature: equivalency or error? *Arch Surg*, 2001, 136:796-800.
- [10] Costa LJ, Xavier ACG, del Giglio A. Negative results in cancer clinical trials—equivalence or poor accrual? *Control Clin Trials*, 2004, 25:525-533.
- [11] McAlister FA, Sackett DL. Active-control equivalence trials and antihypertensive agents. *Am J Med*, 2001, 111:553-558.
- [12] Piaggio G, Elbourne DR, Altman DG, et al. Reporting of noninferiority and equivalence randomized trials an extension of the CONSORT statement. *JAMA*, 2006, 295(10):1152-1160.
- [13] Bernard P, Chosidow O, Vaillant L. French erysipelas study group. Oral pristinamycin versus standard penicillin regimen to treat erysipelas in adults: randomised, non-inferiority, open trial. *BMJ*, 2002, 325(7369):864.
- [14] Assessment of the Safety and Efficacy of a New Thrombolytic (ASSENT-2) Investigators. Single-bolus tenecteplase compared with front-loaded alteplase in acute myocardial infarction: the ASSENT-2 double-blind randomised trial. *Lancet*, 1999, 354:716-722.
- [15] Halperin JL. Executive committee, sportif III and V study investigators. Ximelataragan compared with warfarin for prevention of thromboembolism in patients with nonvalvular atrial fibrillation: rationale, objectives, and design of a pair of clinical studies and baseline patient characteristics (SPORTIF III and V).

- Am Heart J, 2003, 146:431-438.
- [16] Lallemand M, Jourdain G, Le Coeur S, et al. A trial of shortened zidovudine regimens to prevent mother-to-child transmission of human immunodeficiency virus type 1. *N Engl J Med*, 2000, 343: 982-991.
- [17] Gülmezoglu AM, Villar J, Ngoc NTN, et al. WHO Collaborative Group to Evaluate Misoprostol in the Management of the Third Stage of Labour. WHO multicentre randomised trial of misoprostol in the management of the third stage of labour. *Lancet*, 2001, 358:689-695.
- [18] Topol EJ, Moliterno DJ, Herrmann HC, et al. Comparison of two platelet glycoprotein II b/III a inhibitors, tirofiban and abciximab, for the prevention of ischemic events with percutaneous coronary revascularization. *N Engl J Med*, 2001, 344:1888-1894.
- [19] SYNERGY Trial Investigators. Enoxaparin vs unfractionated

heparin in high-risk patients with non-ST-segment elevation acute coronary syndromes managed with an intended early invasive strategy: primary results of the SYNERGY randomized trial. *JAMA*, 2004, 292:45-54.

(收稿日期:2007-07-18)

(本文编辑:张林东)

[编后语:至此由北京大学公共卫生学院詹思延教授负责组织编写的系列讲座 1~8 讲已全部刊出,为此本刊深表由衷感谢。我们清楚知道,对这样一个崭新的内容和涉及面广泛的选题,是很难在有限的章节中讲解透彻,颇有意犹未尽之感,如此请那些关注本讲座内容的读者与詹思延教授联系: Email: siyan-zhan@bjmu.edu.cn]

· 疾病控制 ·

水痘疫苗保护效果配对病例对照研究

傅传喜 梁建华 王大虎 王鸣

采用 1:1 配对病例对照研究。水痘病例为临床诊断病例,来源于国家传染病报告网;病例需居住在广州市且能在儿童免疫接种管理系统找到其记录。病例均按如下配比条件选择 3 个对照作为备选:①与病例住址接近(同一居委会或村内);②与病例性别相同;③与病例出生日期最近者(信息用计算机搜集);④户口性质相同;⑤从未出现过水痘样皮疹症状(通过电话、走访等方式确认)。最后随机选择 1 名作为对照。信息采集:在两家社区卫生服务中心进行预调查。每张问卷包括 1 名病例和对照,其中既往免疫史信息以儿童免疫接种管理系统的记录为准。发病 1 月之前注射水痘疫苗记为有免疫;发病后又注射水痘疫苗者记为未免疫。疫苗效力(EV) = (1 - OR) × 100%。由设计者对本次调查员分区进行培训,并进行市及区的两级问卷质量控制,抽取 10% 的问卷进行双录入核对,对录入不合格高于 10% 的问卷全部重新输入。全部问卷以 Epi Data 软件建库录入,采用 SPSS 11.5 软件进行分析。

结果与分析:①样本代表性:预计调查病例 1020 例,实际调查 856 例(占 83.9%);总体中男女性别比为 1.61:1,样本男女性别比为 1.53:1;总体中平均年龄 6.21 岁,样本平均为 6.30 岁,其性别和年龄间差异均无统计学意义($\chi^2 = 2.650, P = 0.071$; $t = 2.534, P = 0.060$)。②病例与对照的可比性:856 对病例和对照中平均年龄分别为(6.30 ± 2.61)岁及(6.27 ± 2.62)岁,两者年龄无差异($t = 1.202, P = 0.230$)。③免疫情况:病例中有免疫史者 104 例(占病例总数 12.1%),年龄中位数为 4.74 岁,自免疫至发病间隔中位数为 803 d,自免疫至调查时间为 1323 d。④疫苗效果:总体保护效果:病例和对照均已免疫的有 76 对,均未免疫的 472 对,病例免疫对照未免疫 28 对,病例未免疫对照免疫 280

对。则 $OR = 28/280 = 0.10$ (95% $CI: 0.073 \sim 0.137$), $\chi^2 = 206.18, \nu = 1, P < 0.01$; $EV: 1 - 0.10 = 90.0\%$ (95% $CI: 86.3 \sim 92.7$)。国产水痘疫苗保护效果: $OR = 11/168 = 0.065$ (95% $CI: 0.041 \sim 0.103$), $\chi^2 = 137.70, \nu = 1, P < 0.01$; $EV: 1 - 0.065 = 93.5\%$ (95% $CI: 89.7 \sim 95.9$)。进口水痘疫苗保护效果: $OR = 3/29 = 0.103$ (95% $CI: 0.039 \sim 0.271$), $\chi^2 = 21.13, \nu = 1, P < 0.01$; $EV: 1 - 0.103 = 89.7\%$ (95% $CI: 72.9 \sim 96.1$)。有效的疫苗应继续进行上市后大规模人群应用后的研究,这样能得到真实的保护效果。本研究提示,1 剂水痘国产疫苗的免疫效果和进口的类似,这与相关的研究相同^[1,2]。随着时间的延长,儿童水痘抗体的滴度衰退,保护效果也随之下降,达不到初免时高水平的保护效能。本研究为病例对照研究设计,根据儿童免疫接种管理系统的接种记录而不是根据监护人的口述来确定免疫史,原因是根据主观记忆而得到的信息不够准确,而接种和发病日期等记录是有明确前后关系的,如果病例发病后再接种则不记为有免疫史,大大减少了传统病例对照研究中的回忆偏倚。本研究显示,接种 1 剂水痘疫苗有 3.6 年的保护效果,但限于对各个品牌疫苗保护效果的研究样本量较小,还不能进一步确定其疫苗的保护效果,应继续深入探讨不同品牌水痘疫苗的保护作用。

参 考 文 献

- [1] 孙海燕,袁家麟,王树巧.人群中水痘-带状疱疹病毒流行率及其疫苗免疫效果. *中华流行病学杂志*, 2001, 22(1): 60.
- [2] Arnedo-Pena A, Puig-Barbera J, Aznar-Orenga MA. Varicella vaccine effectiveness during an outbreak in a partially vaccinated population in Spain. *Pediatr Infect Dis J*, 2006, 25(9): 774-778.

(收稿日期:2007-01-18)

(本文编辑:尹廉)