

· 基础理论与方法 ·

需治疗人数——衡量临床治疗试验效果的重要指标

徐文体 许能锋

为了将临床研究结果应用于临床实践,使随机对照试验结果易于接受, Laupacis 等^[1]提出了一个新的临床流行病学术语——需治疗人数(number needed to treat, NNT)。NNT 把抽象的率转变为一个具体的频数,使临床试验结果转化为临床实践应用的指标,具有表达统计学意义及临床意义的双重作用,是一个衡量临床治疗效果、指导临床决策的有用工具^[2,3]。

1. 定义:①如果试验组同对照组相比,接受治疗的试验组患者发生不良事件的概率减少了, NNT 可以定义为:某种治疗措施实施一段时间后,需要治疗多少例患者可以预防一例发生不良事件(the number of patients who need to be treated to prevent one adverse event)^[4]。②如果试验组同对照组相比,接受治疗的试验组患者出现好的结局事件的概率增加了, NNT 可以定义为:某种治疗措施实施一段时间后,需要治疗多少患者可以使一例出现好的结局(the number of patients who need to be treated to achieve one additional good outcome)^[4]。

2. 适用条件: NNT 用于系统评价某个治疗措施的效果,主要应用于临床随机对照试验。假设临床试验是预防不良结局事件的发生,结局变量为二分类变量(发生事件/不发生事件),试验组事件发生率用 EER(experimental event rate)、对照组事件发生率用 CER(control event rate)表示。

3. NNT 的可信区间及其计算:由于精确的危险度降低率是未知的,存在于某个特定范围内,因此 NNT 本身是个点估计^[5],其可信区间为绝对危险度降低率(ARR)/绝对获益增加率(ABI)可信区间的倒数。如果 $CER > EER$, 指积极治疗措施可以预防不良事件发生,则

$$NNT = 1 / (CER - EER) = 1 / ARR \quad (1)$$

如果 $EER > CER$, 指积极治疗措施可以使患者获

益,则

$$NNT = 1 / (EER - CER) = 1 / ABI \quad (2)$$

此外, NNT 还可以表示为(式中 RRR 为相对危险度降低率)

$$NNT = 1 / (CER \times RRR) \quad (3)$$

即 NNT 表达了没有接受治疗的危险度和接受治疗后减少了的危险度,而且, NNT 能让医生和患者了解他们必须花费多少努力去预防一例不良事件发生(或者出现一例好的结局),能够比较一种治疗方案和其他治疗方案,预防一例相同事件发生,而需要努力的事件数。

4. NNT 的列线图算法: Chatellier 等^[6]提出在 NNT 的应用前提下,用列线图来表示 NNT 的大小,而无需计算。NNT 的应用前提是:①RRR 和基线事件发生率不相关;②不在特定临床试验的时间点上也可以外推试验结果。列线图有左、中、右 3 个竖线轴,左手轴代表 CER,中轴代表 RRR,右手轴上代表 NNT。从左手轴的 CER 处引一条直线经过 RRR 点,到达右手轴上的截点就是 NNT。通过 RRR 的可信区间,能得到 NNT 的可信限。这样就可以直接获得 NNT,进行临床决策。

该计算用 RRR 代替 ARR 是基于试验得出的 NNT,不是对每一例接受治疗患者的估计,而是一个均值,用 RRR 代替 ARR 可以估计未接受治疗患者的基线事件发生率。

5. 连续性变量如何计算 NNT:对于连续性变量来说,一个重要问题是个体间的差异,如何用连续性变量来计算 NNT,2002 年 Shirer 等^[5]提出用最小临床重要差异(MCID)来解决这个问题。MCID 是指临床上认为有意义的最小效果,以 MCID 为界线,将患者接受治疗前后效果的差值分为 3 组($< MCID$, $MCID$, $> MCID$),试验组同对照组相比较,根据频数计算每一组的事件发生率,就可以得到 ARR 和 NNT。

6. 评价:

(1) NNT 的优点:相对危险度(RR)、RRR、

作者单位:300011 天津市卫生防病中心消毒与病媒生物控制学部(徐文体);福建医科大学公共卫生学院流行病学与卫生统计学教研室(许能锋)

ARR、ABI 都是常用的表达临床试验结果的评价指标,但是这些指标在临床实践应用中存在缺陷。如 RR 在临床随机对照试验中

$$RR = EER/CER \tag{4}$$

表达相对指标的还有 RRR,当 $EER < CER$,指积极治疗措施预防不利事件发生的概率,用 RRR 指标

$$RRR = (CER - EER)/CER \tag{5}$$

RR、RRR 受基线事件发生率的影响^[7],可以保证大概率事件在统计学上有显著性意义及临床上对患者有意义,而对于小概率事件,考虑到治疗的意义及成本花费,治疗措施可能不值得。而且,RR、RRR 受时间的影响,仅仅在特定时间获得的资料不可能外推到其他时间,否则 RR、RRR 可能会误导医生的治疗决策。

而绝对危险度能够克服 RRR 的缺陷^[7],区分受益或受害的高低,表达没有接受治疗的结果。当 $EER < CER$,用 ARR 表示

$$ARR = CER - ERR \tag{6}$$

当 $EER > CER$,用 ABI 表示

$$ABI = ERR - CER \tag{7}$$

但是 $ARR/ABI < 1$,并以小数的形式来表达,很难把小数形式的表达方式同临床结合起来。在临床上无法做出合理的比较,而 ARR/ABI 的倒数 NNT 可以克服 ARR、ABI 的缺陷,在临床上比两者更实用。例如^[7],对同一种病有两种治疗方案(A、B),A 方案治疗 250 例患者才能防止一例出现不利事件,B 方案治疗 2.5 例患者就能防止一例出现不利事件,基于治疗方案和高成本考虑,尽管两者有相似的 RR 和 RRR 值,人们选择治疗方案 B,因此 NNT 比 RRR、ARR 更易于解释临床事件。

此外,NNT 可以推广到群体。Heller^[8] 提出从公共卫生角度而不单纯从临床决策来考虑 NNT,因为 NNT 仅仅关注接受治疗的患者,并没有给出患病者中有多少人或整个人群中的患病者能从治疗中受益。于是他提出两个统计学名词:疾病影响数(disease impact number)和人口影响数(population impact number)。

(2)NNT 的局限性:不同的治疗措施^[7],甚至是同一种治疗措施治疗同一种疾病,由于基线率和治疗期的改变,NNT 也会随着改变。只有同一种疾病、同样的预后和同样治疗期的条件下,不同干预措施的李 NNT 才能进行比较。NNT 有多种局限:①

RRR 存在可变性:NNT 成立的前提条件之一是 RRR 在某些疾病中临床试验是不变的。实际上,RRR 可能改变,对于一些疾病来说,RRR 是随着疾病严重程度的差异而改变。②治疗时间的影响:大多数临床情况下,某病治疗时间越长,接受治疗的病例数会越来越多,对于同样的 RRR 5%,对照组发病率为 20% 的 NNT 比 2% 的 NNT 会更低,同时 RRR 或多或少会随着时间而改变^[6],因此 NNT 只可能有效估计短期效果,并不能预测长期的获益。否则,可能会低估或高估获益,从而导致决策失误。③主观概率和数值本身的意义:高估 ARR,导致 NNT 降低,高估了干预效果,数值本身可能影响临床决策。对一些临床医生来说可能认为某个 NNT 值代表一个重要的卫生获益,而对另一些医生可能持相反的观点。NNT 数值仅仅描述治疗性效果的概要性指标,它对临床决策影响的大小是未知的,应该在临床实践中融合医生和患者双方的意见。④基线率的影响:NNT 显著受患者基线发生率的影响,对照组有更高的疾病发生率^[1],NNT 将更小,在将已发表的已知试验的典型病例外推到不同基线发生率的单个病例时,需要进行校正。⑤可信区间的影响:假如 NNT 95% 的可信区间估计是无限的^[7],试验治疗措施也是无益或者引起副作用;假如 95% NNT 的点估计包括在可信区间内,证明 NNT 点估计是一个有益结果。

总之,尽管 NNT 本身有局限性和需要解决的问题,比如时间、基线率,但是 NNT 在其正确应用前提下,仍不失是一个好的评价临床治疗效果的指标,因为它是一个比较直观的指标,易于为临床医生接受和理解。国内将其应用于临床证据分级评价系统、慢性病防治效果的评价及药物疗效及不良反应的评价中^[9-11],但如果不正确应用,NNT 也可能误导临床决策。

参 考 文 献

- 1 Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *N Eng J Med*, 1988, 318: 1728-1733.
- 2 Szatmari P. Some useful concepts and terms used in articles about treatment. *Br Med J*, 1998, 314: 39-40.
- 3 Rembold CM. Number needed to screen: development of a statistic for disease screening. *BMJ*, 1998, 317: 307-312.
- 4 Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. *Br Med J*, 1995, 310: 452-454.
- 5 Shrier, Ian MD, Dip SM, et al. Clinical relevant? *Clin J Sport Med*, 2002, 12: 328-330.
- 6 Chatellier G, Zapletak E, Lemaitre D. The number needed to treat:

a clinically useful nomogram in its proper context. Br Med J, 1996, 312:426-429.

7 Osiri M, Suarez-Almazor ME, Wells GA, et al. Number needed to treat (NNT); implication in rheumatology clinical practice. the Annals of the Rheumatic Disease, 2003, 62:316-321.

8 Heller D. Disease impact number and population impact number; population perspectives to measures of risk and benefit. Br Med J, 2000, 321:950-952.

9 康德英,王家良,洪旗,等.证据分级评价系统及其在临床决策分析中的应用.中华流行病学杂志,2004,25:351-354.

10 詹思延,秦颖,李立明. NNT 在常见慢性病防治效果评价中的应用.中国慢性病预防与控制,2000,8:99-102.

11 詹思延,秦颖. NNT:综合评价药物疗效及不良反应新指标.药物不良反应杂志,2001,1:62-65.

(收稿日期:2005-03-07)
(本文编辑:张林东)

优势比在诊断试验评价中的应用

童峰 陈坤 何寒青

诊断试验是辅助临床医师进行疾病诊断的试验方法,评价诊断试验准确度的常用指标,包括灵敏度(真阳性率, TPR)、特异度(真阴性率, TNR)、假阴性率(FNR)、假阳性率(FPR)、约登指数、似然比(LR)、阳性预测值(PPV)、阴性预测值(NPV)等。这些指标各有优缺点。其实,可以利用流行病学中的优势比(OR)概念,将其应用到诊断试验的评价中,与传统的指标相比,OR 在某些方面有其独特的优势和用途,为此拟对诊断试验 OR 的含义及应用做一介绍。

1. 基本原理:评价一项诊断试验的基本方法是对金标准已诊断为患者和非患者的研究对象,采用某种试验方法进行检测,根据测定结果,整理成四格表(表 1),诊断试验 OR 即为病例组中试验阳性的比值与对照组中试验阳性的比值之比^[1]。

$$OR = \frac{a/c}{b/d} = \frac{ad}{bc}$$

表1 单个诊断试验的四格表

诊断试验	金标准		合计
	有病	无病	
阳性	a	b	a + b
阴性	c	d	c + d
合计	a + c	b + d	N

当有某一单元格的频数为 0 时,可以在每格的数值上各加 0.5,再进行计算。OR 表示诊断试验的结果与疾病的联系强度。OR 数值越大表示该诊断试验的判别效果越好;OR = 1,表示该试验无法判别患者与非患者;若 OR < 1,则意味着对照组比病

例组更有可能得到诊断试验的阳性结果。

OR 值的 95% 可信区间(CI)估计,可以采用 Woolf 自然对数转换法。

OR 值的概念还可以应用于诊断试验的 Meta 分析。对诊断试验进行 Meta 分析时,如果只是简单地合并各个研究的灵敏度和特异度作为衡量某一诊断试验准确性的合并效应值并不恰当。因为各研究在报道某项试验的阳性和阴性结果时,采用的截断阈值可能不尽相同,所以研究间灵敏度和特异度的变异,部分可能是由于阈值的变动所致^[2]。而采用 OR 作为合并效应值,就可以克服这个问题。要计算合并 OR 采用 SROC 曲线的方法^[3],首先建立一个线性回归方程: $D = \alpha + \beta S$, 方程中 $D = \log(OR)$, $S = \logit(TPR) + \logit(FPR)$, S 与诊断界点的选择有关,可视为诊断阈值;β 为回归系数,反映 $\log(OR)$ 依赖于阈值(S)的程度;α 为方程的截距项,是 $S = 0$ 时的 $\log(OR)$ 值。如果 β 与 0 的差异无统计学意义,则说明 OR 值不随着诊断阈值的变动而改变,那么各个原始研究的合并 $OR = e^\alpha$ 。要比较两种诊断试验的准确性有无差异,可以在线性回归模型中加入一个哑变量(x),表示不同的诊断试验进行回归分析,那么对自变量 x 的回归系数进行反自然对数转换后,就得到了诊断试验的相对优势比(ROR)。此外,也可以将其他可能导致研究间异质性的协变量(如研究设计类型、患者临床特征、文献发表年份等)加入到模型中,最终得到调整的合并 OR 值。

2. 实例分析:为了系统地评价核磁共振血管造影术(MRA)中的二维模式(2-D MRA)和三维模式(3-D MRA)诊断下肢动脉闭塞的准确性,以传统动脉造影术和动脉内数字减影血管造影术作为参照金标准,利用 Koelmay 等^[4]报道的其中 40 个研究资

作者单位:310031 杭州,浙江大学医学院公共卫生系流行病学与卫生统计学教研室

通讯作者:陈坤, E-mail:ck@zju.edu.cn