

证据分级评价系统及其在临床决策分析中的应用

康德英 王家良 洪旗 蒋家林

【摘要】 目的 探讨证据分级评价系统在临床决策分析中应用的可行性。方法 该系统定性分级以及界值点 $NNT_{(T)}$ 的估计过程 结合具体实例 进行临床决策分析。结果 该例 $NNT_{(T)}$ 为 12 与不治疗的效应量 $NNT = 6$ 相比, $NNT < NNT_{(T)}$, 表明利大于弊。结论 证据分级评价系统应用过程中, 正确获取不良事件的发生率以及货币化, 十分关键。

【关键词】 分级评价系统; 临床决策分析; 界值点 $NNT_{(T)}$

Evidence grading system and its usage in clinical decision-making analysis KANG De-ying*, WANG Jia-liang, HONG Qi, JIANG Jia-lin. *Department of Clinical Epidemiology and EBM, West China Hospital, Chengdu 610041, China

【Abstract】 Objective To explore feasibility of applying evidence grading system in clinical decision-making analysis. **Methods** Using grading system as well as a quantitative approach to generate the threshold $NNT_{(T)}$, and conducting clinical decision-making analysis through one study on Asymptomatic carotid atherosclerosis. **Results** $NNT_{(T)}$ was 12, comparing with no-intervention in the ACA study that $NNT = 6$. $NNT < NNT_{(T)}$ indicated the benefits of intervention outweigh the risks and costs. **Conclusion** To obtain the rate of adverse event and the specifying approximate cost were vital in the process of applying grading system in clinical decision-making.

【Key words】 Grading system; Clinical decision making analysis; Threshold $NNT_{(T)}$

在倡导循证医学的今天,大量的研究证据的不断涌现,如何选择高质量证据并将其应用到临床实践与卫生决策中,需要借助证据分级评价系统,以实现证据应用的地区化与个体化。

原理与方法

1. 基本原理:证据分级评价系统是按照论证强度将证据定性分成多个级别,并进一步定量评价证据利弊关系的一系列方法。证据论证强度是指证据的研究质量的高低以及结果真实可靠的程度,针对不同研究类型(病因学研究、诊断学研究、防治研究、预后研究)证据,目前已经有现成的一整套评价方法及原则(具体参见临床流行病学有关内容)。除了考虑证据的论证强度外,还应考察证据的风险性和费用,并计算利弊比,定量确定利弊的平衡点。如果某干预措施尽管有一定的疗效,但风险过大,也应放弃

推荐。

(1)定性评价与分级:综合方法学强度和利弊比,可将证据定性分为七级(1A、1B、1C+, 1C、2A、2B、2C),首先按照设计方案的论证强度分为(A、B、C),再以利弊比是否清楚为依据进行二次分级(利弊比清楚为1,不清楚为2)^[1,2]。例如无重要限制的随机临床试验研究(RCTs)且利/弊清楚的证据为最高级别1A,而观察性研究且利/弊不清楚的证据级别最低,为2C。这里的“利”为获益或疗效;“弊”包括副作用和费用。

(2)界值点量化:确定证据推荐级别,只是用于临床决策分析的第一步。要实现证据的地方化、个体化,还应与当地具体条件及患者的具体特点相结合,定量确立利弊界值点,并通过与效应量的比较,决定最终的取舍。若利大于弊,则推荐应用;反之,则放弃。

2. 方法:

(1)确定效应量及其95%可信区间(CI):效应量(effect size)是有临床意义(实际价值)的差值大

小。以防治性研究为例,效应量表达可以选用相对危险降低率(relative risk reduction, RRR)、绝对危险降低率(absolute risk reduction, ARR)和预防一例不良事件的发生需要治疗的病例数(number needed to treat, NNT)。其中 RRR 值表示与对照组治疗后有关临床事件发生的相对危险度相比,试验组下降的水平, $RRR = [(P - A)/P] \times 100$; ARR 为对照组与试验组事件发生率之间的绝对差值, $ARR = P - A$ (%), 式中 P = 对照组的事件发生率,如病死率, A = 试验组的事件发生率; NNT 是预防一例不良事件的发生,需要治疗的总例数。实际上是 ARR 的倒数。例如在治疗心肌梗死患者时,若使用某药物治疗 300 例才能多减少一例死亡,则 NNT 为 300, NNT 越小,效果越好。NNT 因简便易懂,且可直接由 ARR 计算得到,在临床决策分析时,常常被用来表达效应量。

首先计算 ARR 的点估计值及其 95% CI, 然后分别取其倒数, 则可得 NNT 值及其 95% CI。例如非症状性颈动脉内膜切除术试验组研究^[3], 颈内动脉出现严重狭窄后有内科和外科两种治疗方法预防脑梗死。颈动脉内膜切除术因其手术技术相对成熟, 国外已成为治疗颈动脉狭窄预防同侧脑梗死的常规方法。在一项国外进行的研究中^[3], 外科治疗 825 例, 内科治疗 834 例, 随访 5 年出现同侧脑梗死的发生率分别为 5.1% 和 11.0% (内科), 则 $ARR = 5.9\%$, $NNT = 17$, 95% CI: 12 ~ 30。与内科治疗相比, 至少外科治疗 12 ~ 30 例颈动脉狭窄患者, 随访 5 年方能多避免一例脑梗死出现。

(2) 利弊平衡点的确定(计算 $NNT_{(T)}$): 由于临床上一项干预措施在实施过程中, 既可能产生临床疗效, 同时也伴随着毒副作用以及费用的产生。因此, 在评价某种干预措施或药物的效果的同时, 还应综合考虑其副作用及产生的费用。所以仅仅凭借 NNT 就决定有无推广价值, 显然是不够的。证据分级评价系统的定量评价的核心问题实际就是建立一个 NNT 界值点 $NNT_{(T)}$, 并以此为标准加以判断比较。

计算界值点 $NNT_{(T)}$ 需要借助卫生经济学的理论与方法, 确定利弊, 并将获益与风险货币化。利弊的货币定量过程, 则是以社会 and 医院角度为出发点, 以使有限的资源能够发挥尽可能大的社会经济效能为原则。该界值点综合考虑了效应量、不良反应以及费用等三方面的关系, 从卫生经济学角度看, 在此界值点上, 利与弊等价, 也就是说干预效果在卫生经济学上等价于副作用与治疗相关费用之和。亦即治疗

投入与治疗产出相等。 $NNT_{(T)}$ 实际上是一个干预措施的利弊平衡点, 在临床决策分析中, 通过与 NNT 比较, 决定证据有无临床推广应用价值。若 $NNT > NNT_{(T)}$ 大于该界值点, 则表明弊大于利, 干预措施得不偿失, 缺乏推广应用价值; 若 $NNT < NNT_{(T)}$ 小于该界值点, 则表明干预利大于弊, 可以推荐应用。

$NNT_{(T)}$ 理论上是指治疗总投入等于治疗总产出时的界值点。也可以解释为: 避免出现一例目标事件而治疗患者所产生的净费用等于治疗(因干预所致)不良反应事件而产生的净费用, 因此建立界值点 $NNT_{(T)}$ 涉及到三大步骤: ① 确定防治的目标事件(target event) 以及不良反应事件(adverse event) 及其发生频率; ② 利用卫生经济学的方法, 将上述两个事件货币化; ③ 估算因干预所产生的一系列费用。

例如, 为预防 ICU 住院患者出现消化道出血, 而预防性服用雷尼替定(ranitidine), 则本例预防目标事件是消化道出血, 不良反应事件为服用雷尼替定的副作用(肝炎及肝功能衰竭以及中枢神经系统毒性), 干预产生的费用为药费及管理费用。

我们以外科治疗颈内动脉严重狭窄者预防同侧脑梗死为例加以阐述^[3]。具体指标与方法:

首先确定目标事件、不良反应事件及其货币化。在此例中, 目标事件为发生同侧脑梗死; 不良反应事件为手术意外, 即术中脑梗死。① $Cost_{(treatment)}$ 为预防目标事件出现, 而采取防治措施治疗单个患者所产生的费用, 本例 $Cost_{(treatment)}$ 为手术费用及其相关费用, 平均 1500 元。② $Cost_{(target)}$ 为出现目标事件时, 治疗每一个目标事件所需的费用。治疗一例脑梗死的费用, 按照中国卫生统计提要(2001 年)提供的数据, 综合医院住院 19 日的平均费用为 12 839 元。③ $Cost_{(AE)}$ 为防治一个不良反应事件的费用。因不良反应事件为术中脑梗死, 因此治疗费用仍被认为 12 839 元。④ $Rate_{(AE)}$ 为治疗患者中发生不良反应事件的比例, 根据国外资料发生术中脑梗死的频率为 0.6%。⑤ $Value_{(target)}$ 为成功预防一个目标事件出现而产生的效益, 相当于我们愿意出多少钱避免一例目标事件, 如何取得这些数值, 具体操作可采取标准博弈法、平衡等价法等方法加以确定, 本例愿意花费 6000 元来预防一例脑梗死的发生, 因此 $Value_{(target)}$ 为 6000 元。⑥ $Value_{(AE)}$ 为治疗该不良事件所需的成本, 相当于我们愿意出多少钱避免一例不良反应事件的发生, 同上本例 $Value_{(AE)}$ 也为 6000 元。

下一步是建立方程,计算 $NNT_{(T)}$ 。 $NNT_{(T)}$ 被定义为治疗总投入等于治疗总产出时的界值点。因此分别计算治疗总投入以及治疗总产出。

治疗总投入的组成^[4]:①治疗 $NNT_{(T)}$ 个患者产生的费用: $Cost_{(treatment)} \times NNT_{(T)}$;②干预(因治疗 $NNT_{(T)}$ 个患者出现的)不良反应事件时需要的成本: $Cost_{(AE)} \times Rate_{(AE)} \times NNT_{(T)}$;③同时减去 $Cost_{(target)}$ 防治一个不良事件的费用。

治疗总产出的组成^[4]:① $Value_{(target)}$ 为预防一个目标事件出现而产生的获益,减去②为预防一个不良反应事件出现而产生的获益, $(Value_{(AE)}) \times Rate_{(AE)} \times NNT_{(T)}$ 。

根据总投入等于总产出,构建等式:

$$Cost_{(treatment)} \times NNT_{(T)} + Cost_{(AE)} \times Rate_{(AE)} \times NNT_{(T)} = Value_{(target)} - (Value_{(AE)}) \times Rate_{(AE)} \times NNT_{(T)}$$

$$NNT_{(T)} = (Cost_{(AE)} + Value_{(target)}) / (Cost_{(treatment)} + Cost_{(AE)} \times Rate_{(AE)} + Value_{(AE)} \times Rate_{(AE)})$$

实例分析

1. 计算界值点 $NNT_{(T)}$:以上述预防性颈内动脉切除术预防同侧脑梗死研究为例,计算 $NNT_{(T)}$ 。

$Cost_{(treatment)} = 1500$ 元[颈动脉内膜切除术手术费用(保守估计)]

$Cost_{(target)} = 12\ 839$ 元(治疗 1 例脑梗死的费用)

$Cost_{(AE)} = 12\ 839$ 元(治疗 1 例手术意外—术中脑梗死的费用)

$Rate_{(AE)} = 0.6\%$ (发生术中脑梗死的比例)

$Value_{(target)} = 6000$ 元(避免 1 例脑梗死发生所产生的获益)

$Value_{(AE)} = 6000$ 元(避免 1 例术中脑梗死发生所产生的获益)

$$NNT_{(T)} = (Cost_{(AE)} + Value_{(target)}) / (Cost_{(treatment)} + Cost_{(AE)} \times Rate_{(AE)} + Value_{(AE)} \times Rate_{(AE)})$$

$$= (12\ 839 + 6000) / (1500 + 12\ 839 \times 0.6\% + 6000 \times 0.6\%) \approx 11 \text{ (例)}$$

2. 临床决策分析(NNT 与 $NNT_{(T)}$ 的比较与分析): $NNT_{(T)}$ 为临床决策分析提供了一个定量的判断标准。仍以颈动脉内膜切除术预防脑梗死研究为例,分不同情况加以分析。

在国外进行的颈动脉内膜切除术多中心研究中^[3],内科组干预措施为研究对象每天服用 325 mg 阿司匹林预防脑梗死,外科治疗组则在每天服用等剂量阿司匹林的基础上行内膜切除术。颈动脉内膜

切除术的获益是减少脑梗死的发生率,弊端则是手术意外与手术费用。

(1) 总体分析:在外科手术组 825 例中,随访 5 年有 42 例发生脑梗死,而内科治疗组 834 例中发生 92 例,效应量 NNT 点估计及 95% CI 为 17(12~30),均大于 $NNT_{(T)} = 12$,说明结合当地患者的实际情况,手术后服用阿司匹林与单纯服用阿司匹林相比,风险和费用要高于获益。

(2) 亚组分析:患者的病情严重程度不同,目标事件与不良反应事件发生几率也不相同,根据患者的具体特点加以取舍判断。由于效应量受性别、年龄的影响,以男性及年龄 < 68 岁的颈动脉狭窄两个亚组患者的长期疗效最佳,即在外科手术组 544 例男性患者中,随访 5 年发生脑梗死 22 例;内科治疗组 547 例中,发生 66 例,则效应量 NNT 为 13,95% CI 9~21;年龄 < 68 岁男性患者中,随访 5 年发生 19 例;内科治疗者发生 47 例,则效应量 NNT 为 14,95% CI :10~29。两组亚组分析结果表明,针对男性及年龄 < 68 岁的颈动脉狭窄特定患者,其效应量 95% CI 包含平衡界值点 $NNT_{(T)}$ 。颈动脉内膜切除术的利弊比较接近。手术虽然能取得一些效果,但综合考虑副作用与费用后,其最终的纯效应与服用阿司匹林非常接近。

(3) 与空白对照比较:国外研究资料表明,若颈动脉狭窄者若不预防性地服用阿司匹林,则同侧脑梗死的发生率为 26%,而手术方法则可降至 9%。此时的 NNT 为 6,小于 $NNT_{(T)}$,利大于弊,与空白对照相比,功效显著。

当然上述界值点是根据我国的一般情况与平均经济水平确定的,当临床证据用于临床实践时,针对具体病例,可根据当地的价格水平,重新赋值并相应调整界值点。

讨 论

在临床决策分析过程中,首先应利用证据分级评价系统,综合考察利弊比与证据方法学强度,分级评价并筛选出高质量的证据。为便于临床应用,最直观易懂的效应量表达为 NNT ,可以利用 ARR 计算其点估计值及 95% CI 。

NNT 多大才有推广价值呢?应结合当地的技术条件与价格水平,建立界值点 $NNT_{(T)}$,然后与之比较,进行决策分析。鉴于临床上一项干预措施在实施过程中,既可能产生临床疗效,同时也伴随着毒副

作用以及费用的产生,界值点实际上就是疗效、毒副作用与产生费用三者关系的平衡点。在此点上获益等价于风险。

若 $NNT > NNT_{(T)}$, 大于该界值点, 则表明干预弊大于利, 干预措施得不偿失, 缺乏推广应用价值; 若 $NNT < NNT_{(T)}$, 小于该界值点, 则表明干预利大于弊, 干预可以接受。

计算 $NNT_{(T)}$ 需要确定目标事件、不良反应事件的发生率、防治费用, 以及预期获益。正确获取不良事件的发生率以及货币化过程(如对成功避免一例目标事件与不良反应事件所产生成本与收益的认定)对于利弊界值点确定十分关键, 需要借助卫生经济学的衡量与评价方法(如 trade off 法)等。

因此当全球性证据转化为当地临床实践时, 或者将证据个体化时, 界值点需要调整。如果认为 $NNT_{(T)}$ 中主要成分的价值定义不合适, 或者具体条件不合适, 可以根据自己的实际情况重新计算一个 $NNT_{(T)}$, 利用现有的证据做出相应的临床决策。

在已知的影响界值点的因素中, 首先应注意地域特点与患者的个体特征, 如不同地点、不同患者的界值点有所不同; 同时应该考虑所有可能对患者造成的损害与不便, 额外增加的卫生服务系统管理负担和其他治疗成本, 以及不同费用测评方法等方面因素, 必要时还应加入临床医生及患者的主观期望和评价内容。

参 考 文 献

- 1 Gordon G, Holger S, Deborah C, et al. Grades of recommendation for antithrombotic agents. Chest 2001, 119:3s-7s.
- 2 Materials of KM+ team B Meeting, INCLIN knowledge management plus+ program. INCLIN XIX Global Meeting, Kunming 2003.
- 3 Executive committee for asymptomatic carotid artery stenosis. Asymptomatic carotid atherosclerosis study. JAMA, 1995, 273:1421-1428.
- 4 Gordon G, Jack S, Deborah C, et al. Moving from evidence to action-grading recommendations, a quantitative approach. User's guides to the medical literature: a manual for evidence-based clinical practice. Chicago, Ill, AMA Press 2002. 609-619.

(收稿日期 2003-07-18)
(本文编辑:张林东)

· 疾病控制 ·

医院职工乙型肝炎病毒感染的现状分析

钟珍 赵学军 陈文璞

医院是乙型肝炎(乙肝)传播的主要场所,为评估医院职工乙肝病毒(HBV)感染状况,2002年我们对全院1723名职工进行健康体检,现将其中乙肝表面抗原(HBsAg)、乙肝表面抗体(抗-HBs)的检测资料分析报告如下。

1. 资料与方法:1723名职工(含离退休人员),其中男性588人,女性1135人,按年龄分为20~、30~、40~、>50岁四组;按工作性质的不同分为医生、护士和其他人员三个组。检查项目:HBV血清标志物(HBsAg(>8.0为阳性)、抗-HBs。采集受检者早晨空腹静脉血3ml,分离血清,采用酶联免疫吸附试验(ELISA)检测HBV五项。试剂盒为深圳月亮湾生物工程技术有限公司生产,按说明书要求统一操作步骤和结果判定。

2. 结果:

(1)HBV检测结果与性别的关系:1723名职工检出HBsAg阳性123例,感染率7.1%,其中男性52例(8.8%),与女性71例(6.3%)比较,差异无显著性($P > 0.05$);而抗-HBs阳性726例,感染率为44.2%,其中女性547例(48.2%)与男性215例(36.6%)比较,差异有高度显著性

($P < 0.01$)。

(2)HBV检测结果与职业的关系:护士组HBsAg阳性率升高(8.1%),与医生组(5.9%)和其他组(7.8%)相比较,差异无显著性意义($P > 0.05$);护士组抗-HBs阳性率较高(52.6%),与医生组(36.6%)和其他组(43.9%)比较,差异有显著性($P < 0.05$)。

(3)HBV感染检测结果与年龄的关系:HBsAg在不同年龄组间差异无显著性($P > 0.05$),而青年组抗-HBs阳性率(50.9%)高于其他年龄组(33.7%~46.3%),差异有高度显著性意义($P < 0.01$)。

3. 讨论:HBsAg是诊断或筛检乙型肝炎的重要指标。本文通过对1723名医院职工体检调查发现,HBsAg感染率为7.1%,低于全国平均水平(9.8%)。HBsAg阳性率男性高于女性,可能与男性在外应酬就餐的机会多,自体免疫差异等因素有关。抗-HBs阳性率,女性48.2%高于男性36.6%,表明女性是HBV易感的高危人群,对HBV的易染人群的预防健康教育工作不容忽视。通过本次的体检结果发现,护士及青年人为HBV感染高危人群。

作者单位 524001 湛江,广东医学院附属医院预防保健科

(收稿日期 2003-09-29)
(本文编辑:张林东)