

循证医学和循证保健

第四讲 循证保健中证据的质量评价

陈延 詹思延

循证过程中最重要的一个步骤就是对所收集的证据进行质量评价。循证保健是循证思想在卫生保健系统中的扩展,其目的就是通过循证的过程来指导公共卫生保健领域中干预措施的制订和实施。公共卫生干预与临床上为了防止或治疗个体疾病而采取的干预不同,它主要是指保护和促进社区或整个人群的健康,其特点就是实施过程环节多、情况复杂,并且容易受干预人群所处的社会文化、经济、政治等环境因素的影响^[1]。因此在评价过程中不能完全套用现成的循证医学的评价方法。本讲对此进行概括性介绍。

一、循证保健中证据的等级

许多年来,研究证据的质量评价往往都取决于该研究设计的等级,其中建立在随机对照试验(RCTs)基础上的系统性综述的等级最高(I)队列研究等观察性研究的等级相对较低,最低的是建立在专家或权威机构基础上的证据(VI)^[2]。美国社区指南制定工作组(TASK FORCE)就是根据研究设计来划分证据等级的,并以此作为决定推荐预防保健干预措施先后顺序的依据^[3]。这种等级划分很重要,但对证据的质量评价仅仅依靠研究设计是不够的,所以有学者提出公共卫生领域中的质量评价尽可能不要使用I类或最佳证据(best evidence)的说法,取而代之的是“最佳质量证据”^[1](best quality evidence)。所谓最佳质量证据首先是指证据的针对性好,在周期长、环节多的公共干预措施的实施过程中,每一发展阶段都有具体的证据相匹配,环环相扣;其次,证据要灵敏有效,确实能发现和反映出干预措施产生的重要效果,要作到这一点又要求证据的信息充足,比如措施实施的过程和环境等。最佳质量的证据能帮助评价者判断某项干预措施是否真正有效,区分无效是由于干预措施本身(概念或理论错误)还是由于实施过程导致的^[4]。

二、循证保健中质量的评价

目前,临床用于对治疗措施效果研究证据的质量评价标准已经建立并逐步完善,但这些标准并非完全能适用于公共卫生领域中的质量评价,原因就在于两者干预措施的内容、对象以及所要达到的目的并非完全相同。直到近几年,流行病学关注的热点才由病因学假设转移到干预评价,而对干预措施评价的应用最早是出现在社会科学、教育学和心理学。因而公共卫生领域中的研究质量评价标准是在流行病学和

循证医学的基础上,同时吸收和应用了其他学科的研究方法而建立起来的,但总体上可以通过以下三个方面进行把握。

(一)研究设计的类型:一个好的研究,如RCT不仅可以确定干预措施和结局之间的因果关系,而且可以发现干预措施实施有效或无效的原因,同时易于对实施过程进行质量控制。此外,作为证据的研究采用何种研究方法又直接关系到对干预措施(诊断、治疗、筛查等)和研究结局(干预效果、卫生传递效果、安全性、成本-效益等)评价的效力^[5]。如果要进行完整的评价,往往需要多学科多种研究方法相互结合,取长补短。

(二)研究的结局:①研究结局的测量(包括结局的大小、精确性、是否对干预人群有益等);②除了预期的结局之外,还有哪些是未预期到的结局,有时候未预期到的结果可能是研究者更想要的;③干预措施的效果和效率,对公共卫生干预的评价必然会涉及到对措施效率的评价,包括效率的大小。

(三)研究适用性:适用性的评价是因为公共卫生中的干预措施没有固定的模式,需审时度势,因地制宜,所以不是所有相关的研究都适合作为制定干预措施的参考依据。①研究中干预措施自身的信息(包括措施的实施情况、干预过程中受益或受害人群的特征等);②干预措施实施的大环境(包括政治、经济、文化等,这些都可以影响到证据的适用性);③干预措施和环境因素之间的交互作用,即使最简单的干预措施如健康教育,在实施过程中也会受到效应修饰的影响。

以上三个方面贯穿于循证保健中证据质量评价的全过程,目前在对以社区为基础的研究质量的评价中已经被广泛的接受和使用。在循证保健中,无论采用的是哪种研究证据,对其质量评价基本上都是围绕这三个方面展开的。当然在具体实际应用时,需结合不同研究证据的侧重点有的放矢。现按不同设计类型获取证据的力度对下列研究的评价重点进行介绍。

1. 系统性综述(systematic review):在各种研究方法中,系统性综述在评价干预措施和研究结局方面的效果是最佳的^[5],原因在于系统性综述属于二次研究(secondary research),它可以是一种类型的研究,也可以是不同类型研究方法的综合。与一次研究(primary research)相比,其关注的不再是个体或者人群,而是在一次研究的基础上对其结论进行复习、分析和整理。在系统性综述的过程中有3个明显的特征,即收集检索文献的范围清楚明确,对所收集到文献质量

作者单位:100083 北京大学公共卫生学院流行病学与卫生统计学系

评价的标准具体客观,对使用的一次研究结论的分析科学有效。与之相比,传统综述在这些方面显得相形见绌。另外,系统性综述使得得出结论的过程更加透明化,而且结论与单一的研究证据相比更加可信,但是对其进行仔细谨慎的质量评价仍然不可缺少。评价过程中的重点是^[6]①综述所采用的研究文献讨论的主题和研究结果一致吗?②综述覆盖的范围广而全吗?③综述所采用的研究文献是否有统一的选择标准?④选择的文献经过评价后真实性如何?评价是否具有可重复性?⑤研究的结果是什么?精确性如何?对人群有益吗?

2. 随机对照试验(RCTs) :RCTs是检验临床干预措施效果的金标准,但不是所有开展的RCT都能令人信服。国外学者通过对250个RCT的检查发现,大约40%的研究由于未能严格做到随机分组,从而夸大了试验产生的效果^[7]。在公共卫生领域,由于开展大规模人群干预措施的实施过程中,很难做到以个体为单位的随机化,而且还可能会遇到伦理学问题,所以较少使用传统的RCT^[1],而是以整群随机,即社区干预项目(CIP)多见。无论是哪种类型都需要严格的质量评价:①研究过程是否采用了随机分组和盲法,是否对此有详细的描述?②研究的统计方法(比如使用意向性分析)?③对干预对象的随访情况?④解释结果时是否考虑了所有研究对象?对CIP的评价还需要在上述基础上注意以下两个问题:①选择的社区间是否有可比性,要求在社区干预之前评价基线资料;②由于CIP经常使用中间结局的指标来评估干预的效果,那么试验所选用的中间结局的相关性如何(与主要结局的相关性,能否最大限度的反映干预的效果?),可行性如何(是否有条件完成所要测量的结局),接受性如何(测量方法人群能接受吗?)。

3. 观察性研究(observational studies):与RCT相比,观察性研究在评价公共卫生干预措施方面有着明显的优势,相对简单、可行性好,易于接受(避免了伦理学问题)^[8]。但在研究的过程中容易产生偏倚,而且难以估计和控制,所以对观察性研究的评价主要是侧重于偏倚的评价。

(1) 病例对照研究:①选择的对照是否来自产生病例的人群?是否不患有研究的疾病,而且除了具有和病例一致的某些特征之外是否和病例具有可比性?②对对照的测量是否能避免产生各种偏倚?

(2) 队列研究:①研究对象选取的方法是否有明确的说明?②是否采取的某些措施可能会选取或排除了一些重症患者?③如果采用死亡率作为标准,有哪些步骤可以确认死亡状态?④如果选用其他的指标作为标准,测量使用的方法是否有效?⑤在分析中是否考虑了疾病的严重程度?⑥在分析中是否考虑到了并发症的出现?

(3) 现况研究^[5]:①数据是通过何种调查方式获得的,普查还是抽样调查,如果是抽样调查,如何抽取样本?②在数据收集过程中使用的调查表效果如何?研究者是否考虑到了不同的访问者可能有不同的结果?③数据收集的质量控

制,如何来确保数据的真实无误?④根据对数据分析的结果能否得出结论,或者在此基础上还要进行推导?

4. 决策分析(decision analysis)循证保健中的决策是在对最佳质量的研究证据仔细评价后才制定出来的。如何将许多研究证据结合起来形成卫生决策呢?一种比较系统的方法就是“决策分析”,即对那些必须考虑的研究证据加以描述,然后对各种可能的卫生保健措施在实施后产生的效果进行估计。Lilford, Royston^[9]认为决策分析是理论到实践的桥梁。事实上,决策分析具有双向性,既可以根据理论指导实践,也可以从实践中获得新的理论知识。从证据到决策的产生,这个过程至关重要,亦需严格的质量评价^[5]:①在决策分析的树状图中,所评价的方案是否包括了所有重要的健康结局和决策?进行比较的卫生决策是否都是常用的?其中建立在较高质量基础上的决策方案占多大的比例?②开展或不开展某些卫生保健措施的依据是什么?③是否有明确的方法来确定决策分析中所使用的参数?在决策分析中使用的参数大多来源于已发表的研究结果,为了对原始参数进行评价,需要了解这些研究在哪些范围内对哪些因素做了敏感性分析,如:关于副作用的分析或者关于经济成本方面的分析。④是不是已经包括了所有需要考虑的成本?

5. 定性研究(qualitative studies):以上提到的研究都属于定量研究,但在收集研究证据的过程中我们有时会碰到定性研究。其研究的基本方法是社会人类学和社会学^[5]。通过提出问题为定量研究建立假说,比如:为什么人们不依照处方用药?为什么临床医生和护士不能很好地协调工作?最后通过定量研究或者两者相结合来验证假说。对定性研究的质量评价主要考虑该研究的理论基础、数据收集的范围和方法以及对数据的描述,还有就是该研究的代表性^[10]。

三、循证保健质量评价中的问题

在具体的评价过程中还要注意两个方面的问题:第一,区分质量和效益;在考虑评价证据的质量和预防措施产生的效益时,往往会让研究者感到困惑。比如说,某个干预措施的实施对健康结局有较大的影响时,就认为证据为高质量。事实上,尽管预防措施所产生的效益和危害的程度对于决策和建议有非常大的影响,但是这和该研究证据的质量无关;第二,区分证据评价和政策制定:虽然在循证保健中对证据评价的目的是为了指导公共卫生干预措施的制订和实施,但是证据的评价和政策的制定是相互独立的,例如:美国的TASK FORCE在起草报告时,由循证实践中心(EPCs)的成员组成的课题组负责书写证据的系统性综述,包括证据的缺点及优点,这与决策之间有着独立性,时间上也有明显的分界线。本文讨论的都是以研究为基础的证据,真正在制定卫生决策时还有许多因素都需要加以考虑。

总之,目前在公共卫生领域,围绕证据的构成和有效性^[11]、各种研究设计类型的论证强度和适用性等^[1],还存在争议,但通过这样的讨论,有助于提高卫生保健措施评价的严格性,同时也督促研究实施者重视措施的合理性和有效

性^[12],使证据质量评价的总体水平得到提高。

参 考 文 献

- 1 Rychetnik L, Frommer M, Hawe P, et al. Criteria for evaluating evidence on public health interventions. *J Epidemiol & Community Health*, 2002, 56:119-127.
- 2 NHS Research and Development Center for EBM. <http://cebml.jr2.ox.ac.uk/docs/levels.html>
- 3 Russell PH, Helfand M, Steven HW, et al. Current method of the U.S. preventive services task force, a review of the process. *Am J Prev Med*, 2001, 20:21-35.
- 4 Hawe P. How much trial and error should we tolerate in community trial? (letter). *BMJ*, 2000, 320:119.
- 5 Gray JAM. Evidence-based healthcare. How to make health policy and management decisions. The second edition. London: Churchill Livingstone, 2001. 122-130.
- 6 Hunt DL, McKibbin KA. Locating and appraising systematic review.

Ann Intern Med, 1997, 126:532-538.

- 7 Schulz KF, Chalmers L, Hayes RJ, et al. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA*, 1995, 273:408-412.
- 8 Black N. Why we need observational studies to evaluate the effectiveness of healthcare. *BMJ*, 1995, 312: 1215-1218.
- 9 Lilford R, Royston G. Decision analysis in the selection, design and application of clinical and health services research. *J Health Services Research & Policy*, 1998, 3:159-166.
- 10 Popay J, Rogers A, Williams G. Rationale and standards for the systematic review of qualitative literature in health services research. *Qualitative Health Research*, 1998, 8:341-351.
- 11 McQueen DV. The evidence debate. *J Epidemiol & Community Health*, 2002, 56:83-84.
- 12 McQueen DV. Strengthening the evidence base for health promotion. *Health Promotion International*, 2001, 11:261-268.

(收稿日期 2002-07-19)

(本文编辑 段江娟)

· 短篇报道 ·

乌鲁木齐市肺结核患者中艾滋病病毒抗体检测结果分析

周梅 杨彩云 凌敏

为了解艾滋病病毒(HIV)与结核杆菌协同感染的情况,1998年1月至2000年12月,对新疆维吾尔自治区人民医院的结核病人进行HIV抗体检测。随机采集医院呼吸科门诊及住院部肺结核患者200例(男123例,女77例),年龄17~78岁,汉族86例、维吾尔族99例、回族5例。诊断标准参照中华医学会结核病学分会制定的肺结核诊断治疗指南。实验室检测用ELISA初筛HIV抗体(上海科华试剂),微粒子酶免法,美国Abbott试剂,在AXSMY自动免疫分析仪上进行复检,阳性血清送新疆HIV/AIDS监测中心用WB法进行确认。免疫球蛋白(IgG、IgA、IgM、C3、C4)检测在BEKMAN-ARRAY-360自动免疫分析仪上进行,梅毒初筛用快速血凝纸片法,阳性及可疑血清用梅毒螺旋血凝法确认,痰涂片抗酸染色确定抗酸杆菌为阳性。AIDS诊断参照我国1996年新修订的HIV/AIDS诊断标准。

实验室检测结果:200例结核患者中HIV抗体阳性9例,阳性率4.5%,其中HIV感染者占2%、AIDS患者占3.5%;男性占6.5%、女性占1.3%。9例HIV抗体阳性中20岁组7例(77.78%),30岁组2例(22.22%),均为维吾尔族;无业者6例(66.66%),工人2例(22.22%),外出经商回国人员1例(11.11%);文化程度小学2例(22.23%)、初中4

例(44.44%)、高中3例(33.33%);静注毒品者4例(44.44%)、静注毒品加多性伴者5例(55.56%)。胸水渗出液、痰结核杆菌阳性各4例,梅毒3例。

临床表现:2例HIV感染者为门诊就诊者,7例AIDS患者中均持续发热(>1个月),进行性体重下降(>10%)6例,咳嗽、气短、纳差各4例,盗汗、腹泻(4~5次/d)、呼吸困难各3例,胸痛2例,神经系统症状3例,肺内结核症状7例,肺内、肺外结核症状2例,死于呼吸衰竭2例。X线胸片:肺门阴影增大3例,双上肺呈斑片状阴影2例,双肺纹理增重1例,胸腔积液单侧4例,双侧1例,从胸片上看表现为各种各样,可见肺泡和肺间质浸润,肺门淋巴结肿大。

通过临床表现及检测结果显示,我区从结核人群中查出的HIV抗体阳性率(4.5%)明显高于广东省(0.308%),多数为青年及男性吸毒者,HIV感染者(1%)与AIDS患者(3.5%)阳性率之更高说明结核感染是AIDS引起机会性感染常见病之一,其临床表现复杂,X线胸片多种多样,结核菌涂片阳性、胸水渗出液检测均说明艾滋病与结核虽有不同的致病因子和传播方式,但它们的致病作用是相互影响、相互促进的。乌鲁木齐市是结核病与HIV/AIDS高发区,我院又地处吸毒高发区。因此,建议医院对结核患者进行HIV抗体检测,以便对症治疗。

(收稿日期 2001-11-06)

(本文编辑:段江娟)

作者单位:830001 乌鲁木齐 新疆维吾尔自治区人民医院检验科
(周梅、杨彩云),呼吸科(凌敏)