

如何理解流行病学调查中校正的“率”

谷鸿秋, 李卫

关键词 校正; 标准化; 率

在阅读大样本的流行病学调查文章时, 我们经常会看到“经 XX 校正的 XX 率”, 如“经性别校正的死亡率”、“经年龄校正的患病率”等诸如此类的术语。那么校正究竟是怎么一回事? 什么是校正? 为什么要校正? 如何做校正? 本文将对这些问题作一简单介绍。

校正是一种统计学的数据处理过程, 常用于性别、年龄等人口结构不同的人群间的健康结局的比较。通常由于人们在生理、心理、社会等各个方面的变化会综合反应于性别和年龄, 因而健康结局在不同的性别、年龄中都会存在差异^[1]。例如, 乳腺癌多见于女性, 在男性中则很少见。大部分的慢性病(高血压、糖尿病、肿瘤)常见于老年人群, 而各种伤害(溺水、烫伤、交通事故)则多见于儿童。因此, 如果两个人群的性别、年龄结构存在差异, 就有可能“歪曲”健康结局在这两个人群中的真实反应。这也就是流行病学中常常被提及的混杂因素。例如, 表 1 中, 在各个年龄段基本都是甲县的肺癌死亡率低于乙县, 但是总体上却是甲县的高于乙县^[2], 出现这种不一致结果的原因就是: ①肺癌在各年龄层中的死亡率不同(大体上是年龄越高, 肺癌死亡率越高); ②两县的人口结构不同(甲县的人口结构偏老龄化)。甲县因有更多的老年人从而使整体的死亡率被拉高。因此, 直接比较两县的肺癌死亡率是不恰当的, 我们需要“平衡”两县的年龄结构, 然后再做比较其死亡率。由于两县的人口结构不同, 我们可以用一个统一的标准(常用两县的整个人口分布作为标准), 对其人口结构进行调整, 然后按此标准在各个年龄层对粗死亡率进行调整, 最后合计调整的率就可以获得经年龄标化的死亡率。如此, 可见甲县的肺癌标化死亡率是低于乙县。具体过程如表 2 所示。

此外, 还有一种情况是我们想获得某地某病的患病率, 于是做抽样调查, 但是抽样获得的人群年龄结构与总体的人群年龄结构有差异, 此时我们也经常用总体的人群结构去调整我们调查所得的率, 以此来更好的估计总体人群的患病率。例如, 2002 年中国居民营养与健康状况调查时, 便对调查结果用 2000 年全国的人口数据进行了年龄和地区的调整^[3]。

表 1 甲乙两县各年龄组人口及肺癌死亡率(1/10 万)

| 年龄组 | 甲县 | | 乙县 | |
|-----|--------|--------|--------|--------|
| | 人口构成 | 肺癌死亡率 | 人口构成 | 肺癌死亡率 |
| 0~ | 0.5031 | 3.26 | 0.6227 | 4.29 |
| 30~ | 0.1522 | 9.23 | 0.0976 | 19.20 |
| 40~ | 0.1272 | 44.18 | 0.1110 | 45.81 |
| 50~ | 0.1025 | 144.23 | 0.0919 | 122.27 |
| 60~ | 0.0664 | 328.16 | 0.0499 | 332.32 |
| 70~ | 0.0486 | 298.54 | 0.0268 | 319.17 |
| 合计 | 1.0000 | 56.67 | 1.0000 | 46.03 |

表 2 甲乙两县各年龄组肺癌粗死亡率及标化死亡率(1/10 万)

| 年龄组 (1) | 标准人口 分布(2) | 甲县 | | 乙县 | |
|------------|---------------|-------------|------------------------------------|-------------|------------------------------------|
| | | 粗死亡率 (3) | 校正后死亡率 (4)=(2) ⁺ (3) | 粗死亡率 (5) | 校正后死亡率 (6)=(2) ⁺ (5) |
| 0~ | 0.5589 | 3.26 | 1.82 | 4.29 | 2.40 |
| 30~ | 0.1267 | 9.23 | 1.17 | 19.20 | 2.43 |
| 40~ | 0.1196 | 44.18 | 5.29 | 45.81 | 5.48 |
| 50~ | 0.0976 | 144.23 | 11.15 | 122.27 | 11.93 |
| 60~ | 0.0587 | 328.16 | 19.26 | 332.32 | 19.51 |
| 70~ | 0.0385 | 298.54 | 11.48 | 319.17 | 12.28 |
| 合计 | 1.0000 | 56.67 | 51.17 | 46.03 | 54.02 |

以上“率的校正”的方法都是“标准化法”, 需要以一个“标准人群”对所研究人群的率在各年龄层进行调整。另一种方法便是“模型调整法”, 利用 Logistic 回归模型、广义线性模型, 混合效应模型等统计模型调整多个因素(甚至包括随机效应)后, 依据模型估算我们需要的率, 从而得到经多因素校正的率。这种方法可以控制多个“混杂”因素, 可以在“平衡”多个因素后来比较率的大小。限于篇幅, 具体方法在此不做详细介绍, 感兴趣的读者可以查阅相关资料来学习。

简而言之, 校正的率使“率”在同一个平台上, 增加了可比性, 也使得样本信息更好的反映了总体的情况。在阅读流行病学调查报告时, 在比较不同流行病学调查结果时, 我们应注意校正率背后的隐含信息。

参考文献

- [1] Kleinbaum D G, Kupper L L, Morgenstern H. Epidemiologic Research: Principles and Quantitative Methods[M]. Wiley, 1982.
- [2] 方积乾. 卫生统计学[M]. 第 5 版. 北京: 人民卫生出版社, 2006.
- [3] 杨晓光, 孔灵芝, 翟凤英, 等. 中国居民营养与健康状况调查的总体方案[J]. 中华流行病学杂志, 2005, 7:471-474.

(收稿日期: 2013-02-15)
(编辑: 梅平)