横竖大不同——文章表格中应该提供“率”还是“构成比”

作者：李楠，赵一鸣 来源：临床流行病学和循证医学

本文原地址：https://www.iikx.com/news/statistics/1988.html

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

表 1 新、老两种治疗方法的临床疗效

	无效(%)	有效(%)
新药	25(24.0)	79(76.0)
现有药	22(22.9)	74(77.1)

临床流行病学和循证医学

四格表和各种列联表是论文中最常用的工具之一。从临床研究方法学的习惯出发，我们往往强调表格绘制时的方向性，认为表格的方向应该与从左到右的阅读习惯一致，因果关系/时间关系的顺序也应该保持从左到右的方向。也就是表格的行分类(左侧的分类及对应名称)应该是与原因相关/出现在前的指标，列分类(上边的分类及对应名称)应该是与果相关/出现在后的指标。(表1)

表1似乎看上去没什么技术含量，就是个简单的四格表。其实最大的问题不在例数，而在括号里面给出的百分比。如果您真的核对过的话，您会发现此处每行的百分比之和为100%，而列的百分比则不知。这么做到底对不对呢?

1、如果研究设计类型是RCT、队列/病例系列研究

如果表1中的研究属于这类设计类型，那么表1很大可能是正确的。对这两类研究，我们：

按暴露/干预→结局的顺序收集了每个个体的资料(即便是回顾性的也能确保顺序);

对于每类暴露/干预/具有某些特征的患者，如果没有脱落/失访，我们都能够知道他们的在研究结束后结局如何;

能够知道每个小类研究对象的结局发生率。

表 1.1 新、老两种治疗方法的临床疗效

	无效(%)	有效(%)	总计
新药	25(24.0)	79(76.0)	104(100.0)
现有药	22(22.9)	74(77.1)	96(100.0)

临床流行病学和循证医学

此时我们通过表格描述研究对象的时候往往是用表一的形式。也就是在给出每个格子的例数同时，给出每类患者的结局事件发生率，也就是行百分比。所谓的行百分比，是指以每行的对象总数作为分母、当前格子的对象数量作为分子;也就是每行百分比之和为100%。对应表1，其实可以表示为表1.1这样的情况。

2、如果研究设计类型是病例-对照研究、巢式病例对照研究当然，如果愿意，此时我们还可以计算干预/暴露因素的RR值，也就是新药、现有药物两组有效率的率比。

如果表1的研究出自于这类设计，那么表1就彻底错了。对于这类研究，我们：

从结局出发，按结局→暴露的顺序回溯不同结局患者的暴露情况;

对每类暴露/干预/具有某些特征的患者，这些特征是在选择了病例与非病例之后收集，因此我们并不知道他们的自然分布是多少，只知道病例中各类暴露或特征的构成比、对照中各类暴露或对照特征的构成比;

无法得到某类暴露特征结局对象的发生率。

此时我们不能通过表1的形式展示数据，因为括号中的百分比不再是真实的有效率了，而会随着我们选取的无效/有效对象比例而发生变化，无实际意义。此时我们只能给出每种结局中，不同暴露对象的构成比。当然每个格子中的例数可能还是一模一样的。(表1.2)

表 1.2 新、老两种治疗方法的临床疗效

	无效(%)	有效(%)
新药	25(53.2)	79(51.6)
现有药	22(46.8)	74(48.4)
	47(100.0)	153(100.0)

临床流行病学和循证医学

当然，没有结局的发生率就无法计算RR值，我们只能得到OR值。和1.1相比，可以看到我们给出

的百分比不再是行百分比了，而变成了列百分比。关键原因还是，对于从结局出发的研究(通常是病例对照研究)，行百分比没有任何实际意义(不在代表结局的发生率了)。此时，每类结局中的暴露构成还有一定意义，我们也只能退而求其次的给出列百分比。

3、讨论与观点

我们经常能见到这样的情况，在RCT或者队列研究中，作者给出了列百分比(表1.2这样的类型)。虽然计算出来的百分比有明确的含义，可以代表不同结局中开始的暴露/干预的构成比。但是在队列研究或是RCT中，这一构成比显然没有什么卵用。每一个医生需要知道的当然是：如果患者吃了A要，那他的有效率将会是多少?如果患者具有X特征，那么他在一定时期内死亡的机会有多大?

如果面对的是表1.2中的列百分比，不妨让我们想想一个场景：一脸懵逼的医生对着一个死掉的患者说：'我估计你有20%的机会生前吃的是A药，30%的机会吃的B药，10%的机会吃的是C药……而且我也不知道到底当初吃那个药你活下来的机会更大……'

更多 统计方法 请访问 https://www.iikx.com/news/statistics/

本文版权归原作者所有，请勿用于商业用途，爱科学iikx.com转发