再谈混杂因素的调整----正确理解混杂效应的连续性

作者：张华，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/1601.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！

在前几天举办的“京津冀临床流行病和循证医学协同发展促进会”上，与会者与临床流行病学专家就混杂因素的调整问题进行了深入讨论，主要争论点在于“单因素分析时两组间某混杂因素差异没有统计学意义时，是否应该将此混杂因素纳入回归方程进行分析”。这个问题曾经困绕很多小编，但经过查阅资料和仔细思考，小编有一些自己的思考，在此与各位分享，不同观点者可以讨论。

首先我们从混杂因素的定义入手：混杂因素指与研究因素和研究结局均有关，若在比较的人群组中分布不匀，可以歪曲(
掩盖或夸大)因素与疾病之间真正联系的因素。**混杂因素存在的三个条件**是：

①必须是所研究疾病的独立危险因子;

②必须与研究因素(暴露因素)有关;

③一定不是研究因素与研究疾病因果链上的中间变量。

以上是流行病教材中的内容，从中可以看出：如果某因素成为混杂因素，必须是比较的组中分布不匀，这是成为混杂的必要条件。但对于“不均匀”却有不同的解释，其是否等同于两个比较组间混杂因素的差异有统计学意义呢?

混杂的实质是暴露因素对疾病的效应和混杂因素的效应混在一起，造成对暴露因素效应的有偏估计。例如研究高血压与心脏病的关系时，年龄也是心脏病的危险因素，假设高血压组和对照组的病例数各100例;

1)高血压组和对照组高龄(>60岁)占30%，即完全平衡，此时虽然年龄是危险因素，但由于在各组分布均匀，所以不会产生混杂效应;

2)高血压组高龄占30%，对照组高龄占29%，两组不全均衡，由于高血压组高龄人多，如果不调整年龄的影响，此时会高估高血压对心脏病的影响，对两组的高龄人比例做卡方检验，p=0.877;

3)高血压组高龄占30%，对照组高龄占20%，两组不全均衡，同上年龄也会产生混杂作用，此时会高估高血压对心脏病的影响，对两组的高龄人比例做卡方检验，p=0.102;

4)高血压组高龄占30%，对照组高龄占18%，两组不全均衡，同上年龄也会产生混杂作用，此时

会高估高血压对心脏病的影响，对两组的高龄人比例做卡方检验，p=0.047;

上面四种情况中，随着两组比例不均衡性增加，混杂效应是增加的，也即混杂效应是一个连续变化的过程，情况3和4相比，对照组高龄比例相差仅2%，如果使用卡方检验差异有统计学意义才调整的话，情况4需要调整，但情况3不用调整，但情况3可能也存在较大的混杂效应。

对于混杂因素的调整，我们不能只简单地认为某因素是(否)混杂因素，混杂效应是一个连续递进的过程(只是我们也没有一个指标去衡量混杂效应大小)，统计学上有无差异与是否产生混杂效应不能等同。在样本量足够且既往文献已经确定某因素是混杂因素时，应将这个因素纳入回归分析。

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