
为什么以数据为导向的论文缺乏科学价值?

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/1917.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！

在临床研究咨询过程中，经常会有一些人拿了一堆数据，让看看这堆数据能否写篇文章。这种数据写文章的难度极大，而且往往写出来的文章没有科学价值，可能成为科研垃圾，是什么原因呢？

首先由于没有事先设计，重要的因素可能没有收集

，导致这类数据在分析时十分困难，重要的混杂因素没有收集或者缺失太多，混杂因素无法调整，得到的结论可靠性比较低。

其次，收集的数据往往不真实

。这类研究往往是观察性研究，数据是为临床医疗或者其它工作而服务的，因此在数据收集时，不重要的数据往往是根据经验填写，并未实际测量。如门诊病历中有些人是为开药而来，在病历中诊断是为了开某个药而写的，诊断往往不真实。住院病历中血压、体温、呼吸频率等并未测量或者测量并不精确。科研数据要求需要相对精确，如血压测量，要有规范培训，并对测量员做测量的一致性检查，在测量时往往测量2-3次，才可以确定血压值或者诊断高血压。

再者，即使数据完整准确，但以数据为导向的分析往往没有研究假设

，往往采取“数据挖掘”，将差异有统计意义的结果写成文章，而没有统计学意义的结果忽略。这种分析造成一类错误的概率极速增大，很可能文章的结果是错误的。例如从同一群人随机抽取两组病例，理论上两组病例都没有差异，但如果测量两组病例的500个指标，做两组差异的统计学分析，可能会出现几个指标的差异是有统计学意义的，这就是犯了一类错误，这类结果也是科研垃圾。

当然小编并不反对数据挖掘，但数据挖掘的结果一定要经外部数据验证，否则结果也是无意义的。

总结一下，要做好科研，无论是前瞻性的还是回顾性的研究，都要从科研设计做起，有一个可以证伪的研究假设，设计一个合适的研究类型，确定好主要研究指标，严格地进行记录和测量，统计方法合理，让科研结果经得起检验。

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](https://www.iikx.com)转发