



谎言、可恶的谎言和统计

Ira Salkin, 道德编辑

【编辑注：原文发表于 Editorial Office News 2012 年 4 月刊】

Mark Twain 有一句很有名的话，不过这句话很有可能是他错误地推崇英国首相 Benjamin Disraeli 的一句话，这句话是：

「有三种谎言：谎言、该死的谎言和统计。」最近有篇文章让我想起这句话，这篇文章描述有很多中作者可能操控他们的数据设置的方法，让统计分析支持他们研究中获得的结果的意义。一篇发表在 Psychological Sciences 的文章 False-Positive Psychology 中，作者 Joseph Simmons 和他的同事们报道，数据采集方法中的灵活性允许作者提出几乎任何东西都是统计显著的¹。后者可以大致定义为结果可以通过偶然产生，而不是由研究人员研究不大于百分之五的因素产生。

Simmons 和同事指出，研究样本的大小是研究者（作者）决定的，他们也决定何时停止收集数据，并且将那些他们研究中发现的异常值从分析中排除。通常，这些研究人员在调查过程中才定义他们的假设，而不是在研究开始时就定义，换句话说，作者会根据他们研究过程中发现的结果决定他们的目标。通过各个方面操纵他

们的方法，作者可以扭曲数据，比如即使结果不被科学支持，他们还是可以声称有统计意义。这其中的一个原因可能是：大多数编辑喜欢出版正面的而不是负面的结果。

为了示范调查人员可以操纵数据的方式，Simmons 等人制定了两个大小差不多的试验人群。每组听不同的流行歌曲，之后要求他们完成一个调查，调查包含出生日期等许多变量。出人意料的是，听了曲子 A 的成员被发现比听过歌曲 B 的人年轻一岁多。要特别注意，开始是并没有试图用年龄来分组。

这要怎么样能描述成一个统计显著结果？嗯，当 Simmons 等人回顾了他们的调查中的众多变量，他们报告了一个可以明显支持结果的统计分析。如果你把这个跟研究人员在选择时可以自由地停止数据收集的事实结合的时候，就可以增强了显著结果发现的现象。Simmons 审查了数据块中十个意见或对调查的内容，并在共同作者跟他方便的时候停止分析。非常偶然地，出现了两组之间年龄差距的最高点。

编辑们太希望他们的期刊能发表前所未料的“风头”调查或研究，尤其是这些研究有统计分析支持时。如果有统计的支持，一般保证了研究的结果可以被大众媒体迅速传播。有多少次你读过、听过或者

1 Simmons JP, Nelson LD, and Simonsohn U. False-positive psychology. Undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychological Science 2011;22(11):1359-66.

看过临床调查的结果、新的治疗方法或慢性疾病的结果，被报告在统计上有不寻常的变量联系，例如参与者的年龄。每当我看到特定比例的长者或白种人或意大利裔美国人更容易罹患特定的疾病，或者可以由特定药物得到帮助，我都会记得，群组里有 98% 的成员都刷牙的这个事实。因此，统计分析清楚地证明，刷牙一定是获取该疾病或可以用新药物成功治疗的诱发因素。请记住，是作者负责选择要进行统计分析的变量，但统计数据的误用不会被其他同行发现吗？现实情况是，许多编辑都不愿意发表先前已出现在文献中的复制研究。但是，有一个对这个广泛编辑方向的概括的警告，如果后续确认研究比原始调查研究产生一样或更多的关注，编辑更有可能发表该报告。

Simmons 博士作出一些限制作者滥用统计分析的建议，这些作者可能投稿他们的工作到你们的期刊：

- 作者必须为数据收集和描述研究工作制定具体的截止日期，以我最熟悉的领域举例：“Bacteria were grown for 72 hours on the new medium, after which the colonies were

observed for morphologic variations.” 看到这个之后，读者知道所有数据是在特定的时间内收集；

- 作者必须描述他们在调查过程中收集到的所有变量。例如，作者应该表明“只有测量细菌细胞的长度和宽度并进行进一步分析”；
- 作者必须描述所有的实验情况，即使它们在调查期间中止。此外，作者应该报告从研究中排除任何实验情况的理由。另外，如果没有终止任何条件，作者应该讨论统计结果。如此以来，作者就不能够只选择那些由技术和程序中得到能够支持其假设的结果。

Simmons 给作者以及审稿人的建议的完整讨论，读者可以参考 Psychological Science 网站，可以免费下载完整文章。

大多数可能的统计操纵处理的实例是起源于良性的想法，作者不是有意报告有问题的数据。然而，忽略合适的方法会危害结果的完整性。因此，作者、编辑和主编应该在此议题上接受教育和警惕，确保批准发表的论文达到有效的统计标准。