doi: 10.3978/j.issn.2095-6959.2015.02.010
View this article at: http://dx.doi.org/10.3978/j.issn.2095-6959.2015.02.010

·AME 科研时间专栏·

专栏导读：AME Groups 旗下出版了 *Journal of Thoracic Disease*（《胸部疾病杂志》）、*Annals of Cardiothoracic Surgery*（《心胸外科年鉴》）、*Chinese Journal of Cancer Research*（《中国癌症研究》）和 *Annals of Translational Medicine*（《转化医学年鉴》）等近 20 本英文医学学术期刊。2014 年，AME Groups 中文平台——“科研时间”的诞生，为广大从事临床和基础研究的科研工作者带来了福音，提供了更多科研交流和学习分享的机会。欢迎广大读者关注我们“AME 科研时间专栏”，订阅我们的公众微信号（AME 科研时间：amegroups），给我们提出宝贵的建议和意见，以便于将这个专栏建设得更好，成为读者喜闻乐见的一个栏目。

AME统计学专栏 | 如何正确认识“*P*值”

胡志德

(济南军区总医院实验诊断科，济南 250031)

记得上研究生期间学习医学统计学，统计学老师十分幽默地对我们说：统计学就是个*P*！因*P*与虽然“屁”同音，但意思大相径庭，所以在课堂上引发了哄堂大笑。的确，在医学统计学的所有名词术语中，大家最熟悉的莫过于“*P*值”了。然而，*P*值到底是什么意思？笔者曾试探性地接触了一些同行，能准确说出*P*值含义的人不敢说是凤毛麟角，但是个人经验应该不超过30%。因此，笔者在此撰写短文一篇，浅谈自己的*P*值的理解。

1 如果没有了统计学，这个世界会咋样？

我们假定在一个没有统计学的世界里，路人甲做了一个关于帕洛西汀治疗抑郁症的研究。他起早贪黑、废寝忘食地收集了100个抑郁者患者，其中50个接受帕洛西汀的治疗，另外50个病人接受安慰剂的治疗(为便于说明问题，此处暂且不考虑医学伦理学问题)。在治疗了1年以后，路人甲发现接受帕洛西汀治疗的50名抑郁症患者中，有40名患者完全康复了，治愈率达80%。而接受安慰剂治疗的患者中，仅有5例康复，治愈率仅为10%。这个结果看起来十分令人兴奋，于是路人甲赶紧将这些结果写成论文，投给了某本学术杂志。杂志邀请了路人乙作为审稿人，路人乙看了文章后就提了一个意见：无法排除治愈率之间的差异可能是抽样误差造成的！帕洛西汀和安慰剂的疗效可能是一样的，都是10%。作者之所以得到80%的治愈率，完全是运气，不信您再做一遍！做上十万个病人试试？

这样一个审稿意见恐怕会令这个星球上任何有血有肉的作者欲哭无泪。虽然大多数读者都会认为这不可能是抽样误差，但是问题在于，科学不能靠直觉，必须拿客观证据说事！这就是没有统计学的世界，人们总没有办法排除抽样误差的可能性。

如果有了统计学，结果就不一样了。路人甲完全可以说，我经过统计学分析(Fisher确切概率法)，发现*P*值是小于0.001的。如果路人乙也懂统计学，他就不会再发出任何不和谐的声音了。

2 *P*值的含义

*P*值的含义，简单地说就是差异的来源是有抽样误差(随机误差)的可能性。在上述案例中，*P*值小于0.0001的意思就是：帕洛西汀治疗组和安慰剂治疗组在有效率(80%和10%)上的差异当然可

收稿日期 (Date of reception)：2015–01–04
通信作者 (Corresponding author)：胡志德，Email: hzdlj81@163.com

能是由抽样误差造成的，但是这种事情的可能性不到万分之一(0.0001)。我们通常定义*P*小于0.05表示有统计学意义，实际上就是说：当差异可以用抽样误差来解释的可能性不足5%时，我们就会认为这事跟抽样误差无关，而是由试验因素造成的了。

再用一句通俗的话来描述这个问题：假定有人闲着无聊用相同的研究方案(研究对象，研究方法，样本量等都相同)去重复了路人甲的研究，理论上讲，每次得出的结果不可能与路人甲完全相同，但是研究结果显示安慰剂的疗效好于帕洛西汀的可能性不足万分之一。

3 应用*P*值时需要注意的几个问题

部分学者可能会错误地认为*P*<0.05表示差异很显著；*P*<0.01表示差异十分显著。实际上，*P*值只是表示结果可以由抽样误差解释的可能性，与差异大小无关，更与差异是否有专业价值无关。打个极端的比方，为研究某降糖药的降糖效果，研究人员招募了10例患者进行研究。这些患者服药前的血糖都是7 mmo/L，服药后都变成了6.9 mmol/L，如果做统计学分析，服药前后血糖值之间的差异是有统计学意义的(*P*<0.001)，但是问题在于：服药前后血糖值的差异仅有0.1 mmo/L，显然谈不上十分显著。更重要的是，在临床上，如果只能把血糖降低0.1 mmo/L的话，这种药可以说是一无是处。因此，有统计学意义不见得有专业意义。

*P*值固然重要，可以直观地告诉我们结果属于抽样误差的可能性，但是也不要忽视科研设计的作用。科研设计已经错了，再高明的统计学手段也是无济于事。可能有的同行已经看出来了，血糖本身有一定的波动性，所以不能采用前后对照的方式进行研究，应该同时设立平行对照。我们假定再设立一组安慰剂对照组，让10名患者接受安慰剂治疗，然后检测这10名患者服药前后的血糖值，发现这些患者服用安慰剂前的血糖浓度都是是7 mmo/L，服用安慰剂后却都变成了6 mmo/L。安慰剂的降糖作用还可能大于降糖药，那岂不是说这种降糖药不仅不能降血糖，反而还会升高血糖？

*P*值大于0.05的解释不是试验因素没有作用，而是目前还不能认定试验因素有作用。这句话听起来很拗口，不妨举个例子来说明。

路人甲研究帕洛西汀治疗抑郁症，他收集了80例男性抑郁症患者，其中40例接受了帕洛西汀治疗，另外40例接受安慰剂治疗。得到如下结果(表1)：

表1 帕罗西汀治疗男性抑郁症疗效观察

	有效/例	总体/例
帕洛西汀	35	40
安慰剂	20	40

经卡方检验后作者发现*P*=0.08，因此路人甲认为：帕洛西汀对治疗男性抑郁症无效。

路人乙也在研究这个课题，不过他研究的是女性抑郁症患者，他的研究结果与路人甲出奇地相似(如表1)。统计结果也表明*P*=0.08，因此路人乙认为：帕洛西汀对治疗女性抑郁症无效。路人丙闲着无聊把路人甲和路人乙的结果进行了汇总分析，得出表2：

表2 帕罗西汀治疗抑郁症疗效观察

	有效/例	总体/例
帕洛西汀	70	80
安慰剂	40	80

经卡方检验后，*P*=0.01，因此路人丙认为：帕洛西汀对治疗人类抑郁症有效。

问题来了：这药咋回事？对男性无效，对女性也无效，对人类却有效？这不是滑天下之大稽吗?

出现这种自相矛盾的结果，其根源就在于路人甲和路人乙没有正确理解*P*值的含义，给出了错误的结论。在男性患者和女性患者中开展的研究均得到*P*=0.08，其确切的含义是：目前还不能认为帕洛西汀对治疗男/女性抑郁症有效，而不是帕洛西汀对治疗男/女性抑郁症无效。实际上，路人甲和路人乙之所以没有在统计学上观察到帕洛西汀有效，主要是因为样本量太小，统计效率低下。这里面牵涉到一个I类误差和II类误差的概念，以及样本量估计等问题，限于篇幅所限，在此不再赘述。

*P*值多见于两组或者多组数据的比较，但某些比较很隐秘，不容易被人看出来。笔者一位朋友曾投给国内某杂志一篇论文，其中有一个统计学分析步骤是计算受试者工作特征(receiver operating

characteristic，ROC)曲线的面积。作者的表述是：曲线下面积为0.88(P<0.01)。审稿专家立即质疑：一个曲线下面积，也没有和其他数据相比，何来P值？审稿专家之所以犯这种错误，实际上是没有理解ROC分析以及P值的含义。实际上，这个P值也是经过“比较”得出来的，只不过是这种比较很隐秘，不容易被发现而已。ROC的曲线下面积(area under curve，AUC)是介于0.5~1.0之间的，AUC越大，表示诊断效率越高，若AUC为0.50，则表示该诊断手段无任何诊断价值。AUC本身有一定的抽样误差，本研究得出AUC为0.88，但无法排除这是抽样误差所致，真实的AUC可能是0.50(即毫无诊断价值)。因此需要比较0.88和0.5之间的差异有无统计学意义。P<0.01的含义其实是：AUC等于0.50的可能性不足1%，因此我们认为该指标是有诊断价值的。

本文引用：胡志德. AME统计学专栏 | 如何正确认识“P值”[J]. 临床与病理杂志, 2015, 35(2): 206-208. doi: 10.3978/j.issn.2095-6959.2015.02.010

4 总结

本文主要阐述了统计学P值的含义，简而言之，P值表示结果可以由抽样误差解释的可能性。在医学统计分析中，我们不仅需要看P值的大小，更需要关注差异是否足够大，是否有专业意义。任何脱离了专业解释的统计学结果都是没有价值的。

声明：作者宣称没有利益冲突。

作者：胡志德，*Journal of Thoracic Disease*学术沙龙委员、Section Editor (Systematic Review and Meta-analysis)，工作于济南军区总医院实验诊断科，现为第二军医大学临床检验诊断学博士研究生，以第一作者或通讯作者身份发表SCI论文十余篇，并主持国家青年科学基金一项。。

本文首先以中文发表于【科研时间】(doi: 10.3978/kysj.2014.1.387). 本文已获科研时间和作者同意将该文内容以中文在本刊发表。