

科普报道中如何理解与使用数据?

江 昀* 杨 雪

(成都理工大学, 成都 610059)

[摘 要] 在科普报道中, 要做到对数据的正确理解与使用, 首先应在充分了解科学规律的基础上, 确定数据是否与科学规律吻合。其次, 警惕某些特殊数据和指标背后的统计谬误。第三, 要重视民意调查数据中隐藏的虚假成分, 不能仅仅考虑公众的态度, 一定要观察大部分公众的实际行为, 并以此作为校正公众态度的基础。

[关键词] 科学 科普报道 数据

[中图分类号] N4 **[文献标识码]** A **[DOI]** 10.19293/j.cnki.1673-8357.2018.02.010

所谓的科普报道, 是指媒体对重大的科学发现、技术发明, 以及与公众生活密切相关的科学方法进行普及性传播的一种形式。由于科普报道是科学普及中最普遍的一种方式, 因此, 在这种普及性的科学传播过程中, 经常出现令公众和媒体十分困惑的问题, 这就是如何理解和使用数据。如果一项科普报道向公众介绍了一种疾病的新疗法, 那么这种新疗法可信吗, 或者说这项研究结论是否在某些存在偏差或疑问的数据中得出的? 假如一位环保人士说工业垃圾会使 30% 的人群诱发癌症, 而一名企业家愤怒地予以否认, 那么谁说的才是正确的呢? 作为一名科学记者, 别一开始就被这些数据所吓到。要想成为一名优秀的科学记者, 其实并不需要高深的科学造诣, 但确实要有某种科学的探索精神, 对可能影响研究结果和其他结论的各种

因素提出质疑。因此, 在科普报道过程中, 记者对科学问题和科学方法的普及性解读, 必须对其中的数据进行科学分析, 学会正确理解和使用数据, 这样才能有效提高公众理解科学的水平。下面, 我们将从三个方面讨论如何正确理解和使用数据。

1 探索数据背后的科学原则

1.1 寻找某些不确定性中的确定性

在卫生保健研究领域, 对于该吃什么才能保持健康, 专家们莫衷一是, 对于生病了该采取什么措施, 他们也在不断改口, 甚至前后矛盾。在新的研究对他们的疗效和安全性提出质疑后, 越来越多的药物和治疗方法已经名声扫地。科学界的这些摇摆不定的说法, 在一定程度上让科学背上了污名。但是, 这也正是科学发展过程中的正常状态, 因为

收稿日期: 2017-09-23

* 通信作者: E-mail: jbs8548@163.com。

科学关注的是“什么是真实”这一统计概率,结论则是基于“强有力的证据”,而不是等待一个难于找到的证据。虽然,自然界和研究过程的复杂性会导致诸多研究结果都带有不确定性,但是,科学之所以能不断向前发展,是因为它是一个持续的过程,总是在不断改进之中,既可以在途中发现,又可以在途中纠错。如果公众能够理解为什么一名科学家只能说:“强大而有力的证据表明某某可能是正确的,为了知道更多,我们还需要做进一步的研究工作,请耐心等待”,那么,一些不确定的事物就不应该成为公众理解科学的障碍。只有当我们深入了解一些科学活动的细节时,才发现并非所有的研究都是客观和平等的^[1]。

1.2 确保数据与结论的可靠性

一般来说,一项医学研究的病例越多,一种新疗法成功的概率就越高,气象学家做的天气观察越多,他们预测下周是否会下雨的准确性就越大。在这些研究过程中,数量关系直接影响着一个个真实事件的统计概率。在通常情况下,这一统计概率P值,用一个公式来表达,即 $P = P\{X < \text{或} > C\}$,X表示检验的统计量,C表示样本数据计算出的统计量值,这个公式考虑了被研究的病例或事件数量。如果在一项研究中,研究对象的P值小于或等于0.05时,研究结论通常被认为具有统计的显著性。这意味着,100次试验中只有5次(或更少)的结论是由于纯粹的巧合得出的。P值越小,巧合的可能性也就越小。换句话说,研究的病例(或其他研究对象)数量越多,P值就越可信。

这里有两个相关的概念,第一个是概率,它是发现某事的可能性。例如,在工作中长期接触某种可疑化学物质的工人,其癌症并发率不断增长。研究中,观察的次数或人数越多,找出结果的可能性就越大。一种新药

所导致的罕见而危险的副作用,只有在投入市场,并被上万甚至上百万的患者服用后才会显现出来。另一个是统计强度,如果一种污染物好像正在使发病率在原有基础上增加10%,这未必就是一种有意义的关联。如果患病几率高出10倍(就像吸烟者和非吸烟者患癌症的几率之比),两者相关的可能性就很大^[2]。

1.3 在多种理论解释中寻找更科学的解释

事物的关联本身并不能说明二者之间具有因果关系,如同公鸡叫鸣不是导致太阳升起的原因一样。患者体内发现某种病毒,很有可能这种病毒只是无辜的旁观者,而不是病因。所以在做这种因果关系的分析时,实验及其他细节的研究也非常重要。

许多科普报道说,儿童接种疫苗可能引发自闭症。然而,绝大部分专家认为这只是巧合,而不是原因。他们说:“这种‘关联’只表明儿童常在大量接受疫苗的年龄段出现自闭症。”因此,让专家担心的是,由于人们对自闭症的担心,一些家长可能会延误孩子打麻疹和其他危险疾病的预防疫苗时间。然而,对于没有疫苗时,这些疾病所造成的儿童死亡,许多媒体的科普报道却避而不谈。

研究的时间跨度也非常重要。气候研究必须考察多年的数据,才不至于被天气的正常循环弄糊涂。一种治疗也许使癌症病人得到缓解,但只有时间才能告诉我们,它是否有疗效,甚至是否能延长寿命。虽然一些研究人员发现长期接触某种危险化学物质的工人,其平均健康水平比一般人高,但不能因此排除这种化学物质的危害性,因为只有更健康的工人才能得到和做稳这份工作。人是复杂的,即便是同一个人,每天也会有不同的生理变化,一群人中的变化则更多。因此,即使相似的研究也可能会产生不同的结果,甚至有显著的不同。这些都是研究对象的差异性以及研究局限的存在所致。

由此可见,广泛的质疑可使你避免步入歧途。科普报道者经常得问问研究人员,也问问自己:这项研究的数据和结论,还有其他的解释吗?研究持续的时间是否足够长,是否足以支持它的结论?

1.4 注重科学研究的方法与研究等级

出于研究成本和其他原因,并非所有的研究都是在平等的基础上展开的,因此,有些类型的研究更值得你相信。在生物医学研究中,应当特别谨慎对待实验室研究和动物研究,它们能为人类研究提供重要线索。许多流行病学和医学研究采用回溯的办法,回顾那些旧记录、统计数据或记忆。这种方法是必要的,但往往又是不可靠的,因为记忆会消退,而记录又常常是不完整的。前瞻性的研究方法会好得多,这种研究会对选定的研究对象进行长期的跟踪,有时长达数十年。临床病人研究中的“黄金准则”是一种“双盲”研究,病人被随机分配到一个治疗组或控制对照组,控制对照组中的患者只服用安慰剂。“双盲”研究的重要意义在于研究完成前,科学家和病人都不知道谁将被分配到哪一组,从而避免了对研究结果的人为干扰。

虽然不太严谨的研究项目有时可能是重要的,甚至是必要的。但是我们更应该相信严谨的研究。作为科学记者,你得询问各个科学领域的研究人员,为什么用这种方法来设计你的研究,在参考你的研究结论时,人们应当注意什么,而且还要经常问,现在是否还需要开展更确定的研究。

1.5 正确对待同行评审的影响力

在我国科学界,人们普遍认为,凡是已经通过同行专家评审的研究成果,都应该是科学的理论和知识,但这并不能保证它们一定正确,因为评审专家也是人,也有七情六欲。因此,在这些成果还未进行同行评审的科学会议上,甚至是在科学家们刚开始某些

研究时,你也可能会发现好的科学新闻。但对这类研究进行报道时要格外小心,要更多地向其他专家核实。这时,作为科学记者,你应当询问研究人员:还有谁不赞同你的观点?为什么?你的发现和结论与其他科学研究成果有多少吻合?

在长期的科普报道中,实践给了我们这样一条基本原则:聪明的报道者经常用“也许”和“迹象表明”等字眼,而几乎很少用“证实”等字眼,这样才能在你的科普报道中,阐明某种程度的不确定性,增加报道的可信度^[3]。

2 警惕科普报道中的平均值、比率和风险数字

2.1 平均值的误用

一位名叫赵勇的高一学生到湖边玩,看到湖边牌子上写着平均水深1.3米,不会游泳的赵勇想:我身高1.6米,大于平均水深,因此下水不会被淹死。实际上,据记者调查所知,平均水深1.3米的背后是湖中央,其水深达3米,是非常危险的水域。所以,赵勇身高1.6米,虽然大于平均水深,但是如果下到比他身高还要深的地方,就会有危险。成都市2017年12月12日空气质量检测报告披露:全部九个监测点的PM2.5指数分别是:108、104、102、109、70、74、98、112、94,其中离市区较远的两个监测点是70和74,其余七个点都在市区。按国家空气质量标准,PM2.5指数50~99为良好,100~150为轻度污染,这样,成都市区71%的地区都是轻度污染,而监测报告向公众公布的结果却是求其平均数为96,即使是这样一个貌似代表良好空气质量的平均数,也已经接近了轻度污染的临界值,按美国的标准已经是中度污染。这样一平均,不仅把成都市77%的轻度污染地区的PM2.5指数一下从最高值112降到了96,而且成都市区全部空气质量均被平均值提升在

良好等级上。一项乡村扶贫研究披露,一个仅有三十户人家的村庄,由于政府对其中一农户的养殖野猪项目进行了重点帮扶,使其年收入达到了五百二十八万元,其余二十九户农民年收入均在一万元左右。有记者对此进行了报道,并计算出该村农户的平均年收入为 $(528+29)/30=18.6$ 万。一万和十八万之差,二十九户农民子虚乌有地增加了十七万,这样的平均值能代表什么?这就使我们必须要问一问:在平均值背后是什么?是真理还是谎言?从统计学的角度来讲,在许多情况下,平均值并不能代表总体的平均分布水平,只有在极差(总体中统计样本的最大值与最小值之差)非常小的情况下,平均值才能接近总体的平均水平,才具有实际意义。

2.2 比率的混淆

比率是通过计算有关指标之间的相对数,即两数相比所得的值,也叫比值。它是进行分析评价的一种方法,主要有相关比率分析法、构成比率分析法和趋势比率分析法。科普报道中最容易混淆的比率是构成比率,例如,有一项报道称:据国际民航组织2013年公布的一项调查指出,台湾是全球飞机失事率最高的地区之一,过去10年飞行失事率是美国的10多倍,也比大陆高出许多^[4]。与其他许多误用比率这个词的报道一样,这篇报道根本就没报道比率,只写了死亡人数和坠机总数。后来不得不刊登“更正”,指出每10万次飞机的事故次数(比率)正在逐年上升。在这篇报道中,如果标题更正为“飞机失事次数创10年新高”,既包含了正确的比率关系,也比原句精炼了许多。还有一项报道称,目前中国有2%的年轻人为艾滋病毒的易感人群,这句话的错误就在于比例构成的总体没有圈定,只有在有混乱性交易等情况的高危年轻人群中,才有这样的比例,不属于这种高危人群的年轻人中不存在这种比例。所以,

在许多科学实践中,正确判定比率的总体及相对指标,才是判断比率科学性和正确性表述的重要前提。

2.3 观察风险数字

关于环境风险数字,有人引用每吨排放到空气中的某种物质所引发的死亡人数,或每一万个人中接触到这种物质而死亡的人数。还有一些人引用年死亡数,或用10年死亡总人数来表示风险。到底哪种数据更能准确反映环境风险问题,取决于你得到的数据全面公正的程度。

遇到相对风险和绝对风险时,要注意它们之间的区别。以健康风险为例,相对风险是测量患病或机体失调风险增加程度的一种方法。比如:一项研究总结了接触某种化学物质而患特定癌症的风险是没接触过这种物质的人的两倍,那么相对风险就是2。但就总的受影响人数来说,与罕见疾病的大风险增长相比,常见疾病哪怕是小风险增长也会产生更大的影响。绝对风险概念就把这一点考虑进去了,它计算了每千人每年的发病数。计算相对风险对于发现潜在威胁很重要,而计算绝对风险则对公众健康或临床作用很有用。

此外,还应谨慎观察风险的群发现象。当你根据某些媒体的报道,得知大量癌症病例在一个街道或乡镇集中爆发,在这种情况下,做进一步的研究很有必要,但不用惊慌。我们国家有这么多村社,偶尔有几个村社比其他村社出现更多的癌症病例,这仍然是小概率事件。但是,从这些研究活动中,我们可以找到承担风险的正确对象。有人说:“他们怎么知道这种东西不会造成伤害?”虽然科学自身并不能做反向证明,但受害人有举证义务^[5]。

3 不可轻视的民意调查数据

当代民意调查方法仍然沿袭一般社会调查的基本方法,即以实证主义与现代统计学

作为其基本理论基础,采用问卷调查的方法,直接从一个取自总体的样本那里收集数据资料,并通过对这些资料的统计分析来认识公众的社会态度,从而把握民意的走向。然而,“民意”在“社会态度”与“公共舆论”真实性的获取中并非完全一致。

一般来说,被调查主体的社会态度具有相对稳定性。但是,在特定情境下,公众的具体行为会出现权宜性、伪装性和暂时性,并与自身所持的真实社会态度相背离。简单说来,就是在特定事件的问卷调查中,民众的个体认知与公共认知在认识层面上,具有正向一致性,但是,基于“成本—收益”的价值权衡,公众的实际行为选择恰恰走向反面。例如,发生于2011年3月17日的“中国大陆抢盐事件”^[6],大多数民众的行为选择,其实都是基于个体理性在特定环境下对利益和风险的权衡,最后使个体的行为选择走向了公共认知的反面。例如,面对日本核辐射可能影响中国的所谓“谣言”,在个人理性上,基本都认定它是一种“谣言”,即与社会认知的公共理性相一致,但在实际行动中,却采用了“宁信其有,不信其无”的社会态度。因为,相信“谣言”的交易成本很低,只需支付几十元“抢盐”代价,就可规避不可预期的社会风险,而不相信“谣言”的交易成本很高,万一真的发生核辐射扩散,损失无穷。因此,我们的科学记者,在此情况下,如果完全相信调查的民意数据,即公众不相信谣言,完全保持公共理性,闻风不动,那就大错特错了。所以,参与抢盐的民众的行为选择与社会认知的公共理性相矛盾。但是,他们并不是因为谣言而盲动,大多数的抢盐者,实际上都清楚自己为什么要这么做。

当然,民意调查数据也同时让我们知道公众是否希望我们进行更多的科学探索或其他科学研究,但是,民意调查数据必须经过

科学的分析之后才可相信。例如,被采访的人必须是总人口中随机抽取的样本,而不是我们指定抽取的,因为这样做不符合统计学上的“大数定律”。有些电视科教节目经常通过电话来进行民意调查,但这种调查的对象,只是该节目的观众,在这些观众中只有那些有强烈意见要表达的人才可能打电话,因此致电者不算随机样本,这种民意调查也不科学。通常来说,一项民意调查,被访问者越多,抽样误差就越小,民意调查的正确性就越大。此外,要精心设计民意调查中的问题,以排除任何可能诱导受访者以某种特定方式回答问题的因素。在这种情况下,作为科学记者,我们要问一问:这种民意调查的结果到底是如何对受访者提问的?谁为该项民意调查出资?而且,民意调查只是人们在某个特定时刻的说法而已,人们的说法有可能发生变化,这是民意调查和所有科学研究的基本属性。与此同时,我们还要查看具体数字,问问自己,对民意调查或研究的结论是否还有别的解释?要考虑到任何可能存在的有意无意的偏差,并且牢记不确定事物中的确定性。只有这样,才能使我们获取的数据资料基本或完全符合客观世界的规律性,我们才能将科学普及真正做到准确无误。

4 结语

综上所述,科普报道中,真正做到对数据的正确理解与使用,应该从以下几个方面去考虑。首先,必须尊重数据背后的科学规律。科学数据只是某些科学规律和知识的外部表征,它本身既不是知识也不是规律。只有在充分了解科学规律的基础上,才能确定数据是否与科学规律吻合,是否具有科学意义,从而判断数据是否准确而有效地解释现实问题^[7]。其次,警惕某些特殊数据和指标背后的谬误。平均值的科学意义在于解释调

查对象中的大部分样本分布规律，其先决条件是样本之间的性状差异很小，否则，它不但不具有任何科学意义，而且还会产生荒谬的结果。比率和风险数字误用的主要根源是调查对象所在总体边界的不确定性，只要总体边界科学准确地确定了，各种比率及风险指标就能客观正确地反映研究对象的真实状况和风险的严重程度。再次，要重视民意调查数据中隐藏的虚假成分。面对一场社会危

机或灾难，公众在社会心理机制中的“从众心理”作用下，他们的态度容易取得一致，从而表现出貌似统一且经过实证的态度指标。但是，个体的行为会受到个人利益的驱使，其外在表现则是另外一回事，甚至与被测态度完全相反。因此，分析与使用民意调查数据时，一定要观察大部分公众的实际行为，并以此作为校正公众态度的基础，只有这样，我们的科普报道才能反映真正的民意。

参考文献

- [1] 贾鹤鹏. 科学打假的逻辑 [J]. 青年记者, 2010, 10(30): 65-66.
- [2] 徐坤, 蔚晓慧, 毕强. 基于数据本体的科学数据语义化组织研究 [J]. 图书情报工作, 2015, 9(5): 120-121.
- [3] 屈宝强, 王凯. 科学数据引用现状和研究进展 [J]. 情报理论与实践, 2016, 39(5): 135-136.
- [4] 许巧娜. 人为失误、维修失当, 多次酿成台湾重大空难 [EB/OL]. [2014-03-11]. http://taihai.fjsen.com/2014-03/11/content_13662974.htm.
- [5] Will'f L, Il D. Citation Analysis of Data Files Use Trends [J]. Journal of Scientific News, 1982, 31 (3): 467-477.
- [6] 张玉. 民意调查中真实性“公共民意”获取的方法论路径 [J]. 社会科学, 2011(11): 22-30.
- [7] 叶法丞, 冯伟民, 郭震宇. 中美互联网科学新闻传播对比 [J]. 传媒观察, 2015, 8(4): 21-22.

(编辑 涂珂)

(上接第 26 页)

- [4] 李睿. 大学英语教材辅助移动学习资源系统建设[J]. 科技与出版, 2014(8): 107-110.
- [5] 王晨. 高校教学素材库建设探讨[J]. 教育与职业, 2014(35): 191-192.
- [6] 孙宁, 卢春艳, 孙晨. 关于优质教学资源建设的思考[J]. 中国电化教育, 2013(11): 91-94.
- [7] 孙曙辉, 刘邦奇, 李新义. 大数据时代智慧课堂的构建与应用[J]. 中国信息技术教育, 2015(7): 112-114.
- [8] Liarokapis F, Anderson E. Using Augmented Reality as a Medium to Assist Teaching in Higher Education[C]//Eurographics 2010—Education Papers, 2010: 9-16.
- [9] 隗静秋. 场景革命背景下数字出版用户需求满足探析[J]. 中国出版, 2017(4): 9-12.
- [10] Schlueter C, Shaw M J. A Strategic Framework for Developing Electronic Commerce[J]. IEEE Internet Computing, 1997, 1(6): 20-28.
- [11] 钟岑岑. 国内数字教材研究现状文献综述[J]. 数字教育, 2016, 2(5): 12-18.
- [12] 王光玲, 宋廷山. 基于建构主义的网络教学浅论[J]. 山东师范大学学报(人文社会科学版), 2003(4): 136-138.
- [13] 孙曙辉, 刘邦奇. 智慧课堂[M]. 北京: 北京师范大学出版社, 2016: 7-9.
- [14] 王志刚, 余宏亮. 基础教育数字教学资源体系架构研究[J]. 数字教育, 2016, 2(3): 1-7.

(编辑 袁博)

(College of Political Science, Nanjing Agricultural University, Nanjing 210095)¹

(Institute for Science and Technology Development, Nanjing Agricultural University, Nanjing 210095)²

Abstract: American Science Festival has established its own unique science festival model in the continuous innovation and development. The study found that American science festival had some characteristics of various forms, the level of sponsors, the combination of tourism, the new form of science popularization activities and unique content. Compared with the case of the United States, there are some problems and shortcomings in the current domestic science festival. It is proposed to promote the establishment of the Science Festival Alliance, strengthen the cooperation between the science festival and the local tourism department, establish the cooperation model of the science festival with the sponsors, as well as improve science popularization activities with active participation and interaction, the aim is to comprehensively improve scientific literacy of the citizens.

Keywords: America; science festival; Science Festival Alliance; Cambridge Science Festival

CLC Numbers: N4 **Document Code:** A **DOI:** 10.19293/j.cnki.1673-8357.2018.02.009

On How to Understand and Use Data in Popular Science Reports

Jiang Yun Yang Xue

(Chengdu University of Technology, Chengdu 610059)

Abstract: In order to gain an accurate understanding and the use of data in popular science report, following steps shall be taking into consideration, firstly it should be on the basis of fully understanding the scientific laws so as to determine whether the data is in accordance with the scientific laws. Secondly, we should be on guard against the statistical errors concerning some special data and index. Thirdly, we should attach great importance to false underlying information of the public opinion survey data, in this regard we shall consider the public's attitude and delve into the actual behavior of the public, so as to make it the basis which helps to correct the public attitudes.

Keywords: science; science report; data

CLC Numbers: N4 **Document Code:** A **DOI:** 10.19293/j.cnki.1673-8357.2018.02.010

On Scientific Method and Enlightenment of Modernity: An Analysis of Science Articles by Liang Qichao and Lu Xun

Geng Xian

(China Science and Technology Museum, Beijing 100012)