

· 问题与探讨 ·

数字偏选偏倚及其对医学研究的影响

李丽剑 汪培山

数字偏选(digit preference)是指由于对某些数字的偏好,导致对某种测量结果用近似数字来记录。例如用最接近的整数、偶数、5 或 10 的倍数,或者是 7、14(当时间单位用周时)等^[1],由此可以导致数字偏选偏倚(digit preference bias),这种偏倚可以是观察者变异的一种形式,或者是调查中应答者变异的一种表现。早在 20 世纪 60 年代,人们先是注意到在血压测量的记录中出现数字偏选偏倚,后来人们逐渐注意到,除了血压测量之外,数字偏选还可以出现在多种测量中,如出生体重的测量、自报绝经年龄、药物的剂量、自报妊娠和哺乳时间等^[2]。

1. 血压的测量: 血压测量值末位数字的分布,理论上应是均衡分布。对某些特殊的末位数不成比例的选择、对数字的近似取值,可能会导致血压记录值偏高或偏低,从而影响到患病率的估计。尤其是血压值在有诊断意义的临界值(收缩压 140 mm Hg,舒张压 90 mm Hg; 1 mm Hg = 0.133 kPa)附近时,这种影响就更加明显,同时可能还会影响治疗策略的正确性和可靠性。1999-2000 年在美国进行了一项有关健康与营养状况的研究^[3],在全部的测量结果中,当收缩压 ≤ 90 mm Hg 时,末位数字为 0 的占 44%,为 2 的占 4%;舒张压 ≤ 60 mm Hg 时,末位数字为 0 的占 33%,为 2 的占 11%。当舒张压 ≥ 94 mm Hg 时,末位数字为 0 的降至 15%,而末位数字为 4 的升至 32% ($\chi^2 = 1006.7, P < 0.0001$),发现末位数字偏选在重要的临界值附近又有不同的表现形式。测量结果的准确性可能还受测量人员的警觉性的影响,当血压值较低的时候,其注意力往往不够集中,在临界值附近则较为谨慎。和收缩压的测量相比,测量舒张压可能存在着一一定的技术难度也是一个原因。

数字偏选问题在临床血压测量过程中普遍存在,国外有研究人员通过对缺血性心脏病患者的 85 000 个血压值进行分析后发现^[4]:末位数字 0 偏选占收缩压的 64%,占舒张压的 59%,与预计 10% 比例的差异有非常高的统计学意义 ($P < 0.000 001$)。偶数出现的频率多于奇数,非 0 结尾的收缩压中偶数占 64%,而舒张压中占 65%,预计的比例为 44% ($P < 0.0001$)。同时,在非 0 的偶数中,血压值末位数字以 8 出现的频率最高,占收缩压的 28%,舒张压的 31%,与预计的 25% 的差异有统计学意义 ($P < 0.0001$)。在 5 个非 0 的奇数中,结尾为 5 的最多,收缩压 59%,舒张压 62%,高于预计水平 20% ($P < 0.000 01$),这种偏倚可能对该高危人群治

疗效果的评价有潜在的影响。美国康涅狄格州健康中心回顾性调查了医护人员在 2001 年 9 月份的临床记录,发现不同的测量者之间血压测量结果“0”偏选的情况不同,数字为 0 结尾的收缩压占护士测量结果的 40%,医生的 31%,而对于舒张压分别占 23% 和 36%,护士记录的舒张压中以 2 结尾的占 43%。医生记录的舒张压为 0 结尾的被测者的年龄高于非 0 者(67 岁 \pm 14 岁 vs. 59 岁 \pm 18 岁, $P = 0.008$),舒张压 0 结尾患者的体重指数低于非 0 结尾患者($28 \text{ kg/m}^2 \pm 5 \text{ kg/m}^2$ vs. $32 \text{ kg/m}^2 \pm 6 \text{ kg/m}^2$, $P = 0.02$)^[5]。将高血压的正常临界值由 $< 140/90$ mm Hg 改为 $\leq 140/90$ mm Hg 时,发现高血压患者的比率明显降低,这种下降很可能是由数字偏选引起的^[6],提示我们在比较不同人群高血压的患病率时要考虑到临界值和数字偏选的影响。在评价血管紧张素转化酶抑制剂培那普利疗效的研究中,如果以 $< 140/90$ mm Hg 作为正常界值,疗效为 75.1%。如将正常界值改为 $\leq 140/90$ mm Hg,则疗效提高为 90.7%。其中有 81.3% 的舒张压记录为 5 的倍数,19% 的收缩压记录为 10 的倍数。因此,有人建议通过多次测量或是以舒张压为 5 的倍数、收缩压为 10 的倍数作为正常界值,以期减少数字偏选的影响^[7]。

临床实践中,如果观察者将血压 166/96 mm Hg 近似取值为 170/100 mm Hg, 164/94 mm Hg 近似取值为 160/90 mm Hg,那么血压的差异则会由真正的 2/2 mm Hg 扩大到 10/10 mm Hg,这种由于数字偏选导致的偏倚提示我们在诊断和评价疗效时要慎重做出结论^[8]。此外,将血压 88 mm Hg 近似取值为 90 mm Hg,按照某研究舒张压至少为 90 mm Hg 的入选标准就会误选某些个体进入研究队列,这种由于血压测量值精确到 10 mm Hg 而非 2 mm Hg 而导致的结果偏倚,其研究质量值得怀疑^[9]。医生的记录中,如果 42% 的血压值存在末尾数字 0 偏选,在判断患者的降压效果时就会打上一个问号。欧洲开展的一项单纯收缩期高血压临床研究(Syst-Eur)的试验方案规定血压降至 150 mm Hg 以下,且血压是在原有基础上下降了 20 mm Hg 为治疗有效,研究中发现治疗组收缩压 148 mm Hg 的记录值明显多于安慰剂组,提示达到理想血压的患者数有夸大的成分,其真实的血压可能要高于记录值,对于估计最终的结果产生偏倚^[10]。由 Wingfield 等^[11]成立的 GPHSG 小组为深入探讨数字偏选偏倚进行了为期 23 年的前瞻性研究,结果发现,被调查者中血压记录值偏选 88~89 mm Hg 的女性,其校正心血管病死亡率和全死因死亡率均较其他血压组高。可见,如果测量的血压值存在数字偏选偏倚不但会影响临床诊疗的正确性,造

成临床研究质量下降、研究结果丧失真实性和有效性,更为严重的后果是延误治疗、影响患者的生存质量和寿命。

为了控制和减少血压测量过程中出现的数字偏选偏倚,Ataman 等^[12]提出如下的标准来评价血压测量人员对数字“0”偏选的程度:<25%最为理想,<30%为可接受水平,<35%为最低标准。多项研究表明通过培训观测者、规范血压测量过程,使用随机零点血压计(random-zero sphygmomanometer)和自动血压记录仪等手段都能在一定程度上减少偏倚。例如 Fornells 等^[13]报告,培训前结尾数字 0 或 5 分别占舒张压的 57.3% 和 16.6%,收缩压的 66.6% 和 11.8%,经过连续两次培训,分别降至 41.1% 和 7.5%,51.3% 和 8.2%。经压电传感型血压计(Infrasonde)测定的血压结尾 0 偏选的比例(21%)与接受培训的助理医师相近(22%),而低于护士的偏选程度(45%),表明该仪器是一种理想、可靠的设备^[14]。此外,动态血压监测及患者自测血压的结果有着良好的重复性,不受数字偏选和测量者偏倚的影响^[15]。电子血压计的应用减少了末尾数字偏选偏倚,提高了测量精确度。Nuutinen 等^[16]通过测量青少年血压发现,用随机零点血压计测得的收缩压和舒张压低于普通水银血压计,并且两种血压计测得的年龄——血压曲线的形状明显不同,用随机零点血压计测量低的血压值更为准确,既不影响血压值分布,又能消除数字偏选的影响。

2. 自报年龄:医学研究中人们常常分析年龄的分布及其与其他因素的关系。误报年龄会造成年龄分组的误差。例如乳腺癌筛查建议的年龄是从 50 岁开始,如果年龄报告不准确,可以造成延误检查时间,或者造成不必要的资源浪费。有人指出^[17],发展中国家,特别是文化程度低和老年人中末尾数字 0 或 5 偏选的问题比较严重,年龄资料的报告质量偏低,而患者误报年龄则会掩盖不同年龄组之间的差异,降低研究的检验效能,影响结果的外推。

在 1945 年、1955 年和 1970 年的土耳其人口普查中,对年龄的报告都存在着数字偏选,20 岁以上各年龄组均有不同程度的误报现象,其中以 40~44 岁组最多,表现为以 0 或 5 结尾的数据占多数,但随着时间的推移,数字偏选的程度有所下降,特别是在女性当中,可能与女性的受教育水平提高有关^[18]。回顾性调查妇女在育龄期的怀孕时间间隔,受数字偏选的影响,多数回答间隔 6 或 12 个月经期,部分妇女回答为 3 个月经期^[19]。以末次月经时间估计怀孕时间,存在数字偏选,产妇回忆末次月经的时间多集中在每月的 15 日,而报告的时间中很少有 27-31 日以及 9、11 和 19 日这些日期^[20]。

绝经期激素水平的改变会引发某些疾病,人群绝经年龄的分布状况能够解释患病率随之所发生的变化,如估计骨质疏松的患病率等。自报绝经年龄不准确可能会影响绝经年龄与健康的关系的一些研究。尽管横断面研究不能评价报告的有效性和可靠性,但可以通过测量数字偏选来达到这一目的。Crawford 等^[21,22]通过横断面调查访问 2151 名年龄在

40~55 岁自然绝经妇女,比较其自报绝经年龄的末位数字的频数分布与 10 个数字(0~9)等概率分布(各占 10%)的差别,研究发现报告中频繁出现末位数字 4 或 5($\chi^2 = 841.60$, $P < 0.001$),提示绝经年龄末位数字的概率分布偏离等概率分布。然而,这种分布形式与前瞻性的研究结果(无数字偏选和回忆偏倚)相一致,能够反映被调查者的绝经年龄分布,可能更为准确,而非由数字偏选引起。进一步以调查的年份分层,考察各年的分布是否相同,与 1996 年相比,1997 年的年龄分布变化了一位数字,提示仍无数字偏选或系统性变动。该项研究指出当研究对象年龄范围较小,年龄分布受某种潜在因素的限制时,评价是否存在数字偏选应该考虑等概率分布以外的其他分布形式。

3. 出生体重及孕期体重:国外有报道指出^[2],出生体重的数字偏选与测量单位有关,如以克(g)为单位受数字偏选的影响要小于以盎司为单位。但是即使以克为测量单位,数字 0 或 5,特别是 0 结尾的测量值在各出生体重组出现的频率均较高($\chi^2 = 2187.0$, $df = 8$, $P < 0.001$)。而其他数字在 ≥ 1000 g 的各组中尚不足 1%,在 < 1000 g 组为 20%。对体重的最后两位数字进行分析得出,以 10 和 90 结尾的数值明显少于 10 的其他倍数结尾值,提示观察者考虑到出生体重是以 100 g 划分等级可能会采取近似取值。与其他各组相比, < 1000 g 组最后两位数字为 10 的倍数的体重值较少($\chi^2 = 194.5$, $df = 10$, $P < 0.001$)。以“00”结尾的数值占全部体重值的 16.1%,且随体重的增加其比例也不断升高:9.1% (< 1000 g)、12.6% (1000 g~)、14.3% (2000 g~)、16.2% (3000 g~) 和 18.7% (4000~4999 g),而预计末尾数字为“00”的体重值在以 00~99 结尾的数值中的比例为 1%,这无疑会影响低出生体重发生率的计算。在 < 1000 g 组数字偏选出现较少,可能是因为体重在这个范围内的新生儿需要特殊的临床用药和护理,观察者在测量过程中会比较谨慎。对低体重儿发生率的低估可能会在地区间的比较以及时间趋势分析中产生影响。由此可见,数字偏选对出生体重的临床测量记录的影响是很重要的,第一,可以引起出生体重分布曲线的改变,在探索危险因素的研究中,降低检验统计量值;第二,当用切点值作诊断时,容易出现分类错误。

育龄妇女在孕期体重增加的多少不但与胎儿的生长发育密切相关,还会对分娩和产后恢复造成影响,但这一数据多数要在产后通过产妇自己报告,因此报告的准确与否直接影响着公共卫生人员了解育龄妇女的健康状况和制定相应的指导原则。有研究表明,自报的体重增加值多以 0 或 5 结尾,且存在此类数字偏选的妇女通常会低估自己的体重^[23]。

4. 研究结果中的统计量值和 P 值:近年来,随着医学研究的蓬勃发展,人们越来越关注科研文献的质量。Emili 等曾经抽取 2001 年 Nature 上发表的论文 181 篇,共收集 610 个检验统计量及相应的 P 值,发现统计量的末位数字偏离等概率分布(Kolmogorov-Smirnovtest, $Z = 2.7$, $P < 0.0005$),提示科研人员可能经常错误地随意近似取值、凑整,特别是在

末位数字较大或是以 4、6 或 9 结尾的情况下。同样,论文中报告的 P 值也存在类似问题(Kolmogorov-Smirnovtest, $Z = 1.4, P = 0.043$),很少见到末位数字为 0、4 或 9 的 P 值,部分研究结果的检验统计量与相应的 P 值不符^[24]。上述情况表明,即使是发表在著名的期刊杂志上的文章其研究过程中统计学处理的质量也存在着一定的问题。因此,研究人员有必要在注重提高科研能力的同时,改进统计学方法应用的准确性,确保研究结果的真实性和可信性,以提高论文质量。

5. 其他:

(1) 药物有效成分的含量:要达到理想的药物疗效,患者用药的剂量应依赖于药物的生物活性和临床试验的结果。为了计算更方便,多数片剂和胶囊这类剂型药物中的含量,大多是取容易计算的 2.5、10 等的倍数。药物含量的确定可能受到数字偏选的影响,如最早生产的阿替洛尔片剂为 100 mg,然而制备这一剂量是因为更高剂量的疗效并没有明显的提高,这样看来,或许当初选择制备 80 mg、70 mg 或 60 mg 也会同样有效。地高辛片剂从最早的 250 μg 到 125 μg ,再到 62.5 μg ,如此减半的药剂剂量只是近似取值,并无严格的规定。一项对 41 种药物含量的调查结果显示,各种剂型的药剂剂量偏选数字 2 和 5 的(不包括 0)分别有 22% 和 41%。不必要的剂量过大势必在部分患者中导致本来可以避免的不良反应。因此,作者提出需要应用严格的方法来制定推荐药剂剂量及药物片剂或者其他剂型中的含量^[25]。

(2) 自报吸烟量:美国第二次健康和营养调查发现^[26],在 4275 名吸烟的成人中,绝大多数被调查者自报每天吸烟的支数为 10 的倍数,这与暴露于吸烟的客观指标碳氧血红蛋白测量结果的分布明显不同,而这种差异并非由碳氧血红蛋白的波动造成。研究得出,经常吸烟、白种人或受教育水平低者多存在数字偏选问题,他们的回答更倾向于每天吸烟 20 支。

(3) 职业暴露时间:Palmer 等^[27]报告,在调查职业接触生产性振动工人自我报告接触振动持续时间的时候,发现与接触全身振动者相比较,接触手传振动的工人明显地高估了自己的接触时间,并且暴露时间相同的工人之间报告的变异性也较大。但工人们报告的时间都存在着近似取值到最为接近的 5 或 10 的整数倍分钟数的现象。

(4) 结核菌素试验硬结直径的测量:在结核菌素试验中,分析硬结直径的频数分布也会遇到末位数字偏选 5 的倍数的问题,引起人们对测量的准确性产生怀疑^[28]。Eilers 等^[29]通过建立统计学模型的方法可以定量描述数字偏选的大小,并对其加以校正。

综上所述,数字偏选在医学研究中广泛存在,由此导致的数字偏选偏倚将对医学科研质量和临床诊疗决策带来一定的影响。这些数字偏选造成的偏倚,有些是可以通过制定有效的控制措施进行控制的。笔者在一项心血管病随访的研究中,对最初 2 个月的血压测量资料进行初步分析,发现这种数字偏选的存在,立即制定措施,重新对调查员进行培

训,并加强了对数据的监测,这种测量偏倚在此后的测量中得以控制。在医学科学研究和临床实践中,有必要加强对医务工作者和研究人员的培训和管理,制定有效的质量控制措施,以减少数字偏选造成的误差,提高医学科研质量。

参 考 文 献

- 1 Last JM, ed. A dictionary of epidemiology. 4th ed. New York: Oxford University Press, 1994. 51.
- 2 Edouard L, Senthilselvan A. Observer error and birthweight: digit preference in recording. *Public Health*, 1997, 111:77-79.
- 3 Ostchega Y, Prineas RJ, Paulose-Ram R, et al. National Health and Nutrition Examination Survey 1999-2000: effect of observer training and protocol standardization on reducing blood pressure measurement error. *J Clin Epidemiol*, 2003, 56:768-774.
- 4 Lusignea S, Belsey J, Hague N, et al. End-digit preference in blood pressure recordings of patients with ischaemic heart disease in primary care. *J Hum Hypertens*, 2004, 18:261-265.
- 5 Thavareajah S, White WB, Mansoor GA. Terminal digit bias in a specialty hypertension faculty practice. *J Hum Hypertens*, 2003, 17:819-822.
- 6 Green BB, Kaplan RC, Psaty BM. How do minor changes in the definition of blood pressure control affect the reported success of hypertension treatment? *Am J Manag Care*, 2003, 9:219-224.
- 7 Alcocer L, Novoa G, Sotres D. Digit preferences observed in the measurement of blood pressure: repercussions on the success criteria in current treatment of hypertension. *Am J Ther*, 1997, 4:311-314.
- 8 Johnston GD. Orlistat associated with hypertension: digit preference lays conclusions about orlistat open to doubt. *BMJ*, 2001, 322:110.
- 9 Kuznetsova T, Staessen JA, Kawecka-Jaszcz K, et al. Quality control of the blood pressure phenotype in the European Project on Genes in Hypertension. *Blood Press Monit*, 2002, 7:215-224.
- 10 Wingfield D, Cooke J, Thijs L, et al. Terminal digit preference and single-number preference in the Syst-Eur trial: influence of quality control. *Blood Press Monit*, 2002, 7:169-177.
- 11 Wingfield D, Freeman GK, Bulpitt CJ, et al. Selective recording in blood pressure readings may increase subsequent mortality. *QJM*, 2002, 95:571-577.
- 12 Ataman SL, Cooper R, Rotimi C, et al. Standardization of blood pressure measurement in an international comparative study. *J Clin Epidemiol*, 1996, 49:869-877.
- 13 Fornells JM, Borrás JM, Balaguer-Vintro I. The preference for the final digits in blood pressure determination: a continuing education experience. *Aten Primaria*, 1990, 7:102-105.
- 14 Hia KM, Vokaty KA, Feussner JR. Observer error in systolic blood pressure measurement in the elderly. a case for automatic recorders? *Arch Intern Med*, 1986, 146:2373-2376.
- 15 Staessen JA, O'Brien ET, Thijs L, et al. Modern approaches to blood pressure measurement. *Occup Environ Med*, 2000, 57:510-520.
- 16 Nuutinen M, Turtinen J, Uhari M. Random-zero

- sphygmomanometer, Rose's tape, and the accuracy of the blood pressure measurements in children. *Pediatr Res*, 1992, 32: 243-247.
- 17 Denic S, Khatib F, Saadi H. Quality of age data in patients from developing countries. *J Public Health (Oxf)*, 2004, 26:168-171.
- 18 Mukherjee BN, Mukhopadhyay BK. A study of digit preference and quality of age data in Turkish censuses. *Genus*, 1988, 44:201-227.
- 19 Ridout MS, Morgan BJ. Modelling digit preference in fecundability studies. *Biometrics*, 1991, 47:1423-1433.
- 20 Savitz DA, Terry JW, Dole N, et al. Comparison of pregnancy dating by last menstrual period, ultrasound scanning, and their combination. *Am J Obstet Gynecol*, 2002, 187:1660-1666.
- 21 Crawford SL, Johannes CB, Stellato RK. Assessment of digit preference in self-reported year at menopause: choice of an appropriate reference distribution. *Am J Epidemiol*, 2002, 156:676-683.
- 22 Crawford S, Johannes C, Bradsher J, et al. Digit preference in year at menopause. Data from the study of women's health across the nation. *Ann Epidemiol*, 2000, 10:457.
- 23 Schieve LA, Perry GS, Cogswell ME, et al. Validity of self-reported pregnancy delivery weight: an analysis of the 1988 National Maternal and Infant Health Survey. *Am J Epidemiol*, 1999, 150: 947-956.
- 24 Garcia-Berthou E, Alcaraz C. Incongruence between test statistics and *P* values in medical papers. *BMC Med Res Methodol*, 2004, 4: 13.
- 25 Herxheimer A. How much drug in the tablet? *Lancet*, 1991, 337: 346-348.
- 26 Klesges RC, Debon M, Ray JW. Are self-reports of smoking rate biased? evidence from the Second National Health and Nutrition Examination Survey. *J Clin Epidemiol*, 1995, 48:1225-1233.
- 27 Palmer KT, Howard B, Griffin MJ, et al. Validity of self reported occupational exposures to hand transmitted and whole body vibration. *Occup Environ Med*, 2000, 57:237-241.
- 28 Enarson DA. Use of the tuberculin skin test in children. *Paediatr Respir Rev*, 2004, 5 suppl 1:135-137.
- 29 Eilers PH, Borgdorff MW. Modeling and correction of digit preference in tuberculin surveys. *Int J Tuberc Lung Dis*, 2004, 8: 232-239.

(收稿日期:2004-12-16)

(本文编辑:张林东)

· 征稿通知 ·

第九届全国临床流行病学学术会议第一轮通知

第九届全国临床流行病学学术会议将于 2006 年 6 月 15-18 日在上海市召开,会期 3 天。大会主题为“创造一流知识,应用最佳证据”。会议由中华医学会临床流行病学分会/中国临床流行病学网主办,复旦大学临床流行病学资源和培训中心与复旦大学循证医学中心承办。会议选题包括:(一)重大疾病的临床诊治和流行病学研究进展;(二)加强卫生研究能力及临床科研方法学进展;(三)循证医学临床实践的最新进展;(四)中医药现代化及其研究方法学;(五)合理用药。会议期间同时举办循证医学、临床统计学进展以及知识转化与管理学习班。

此次会议现征集与上述选题相关的学术论文。论文内容可涉及临床病因诊断和治疗方法、疾病预后、流行病学研究及新进展等。入选论文将被收入会议的论文汇编。征文要求来稿必须是未在其他学术期刊或会议发表过的中文学术论文,并附 800 字的中文摘要及相应的英文题目。论文摘要应包括研究目的、方法、结果和结论四个部分。

投稿截止日期为 2006 年 2 月 20 日。

论文以电子版形式发送至: chinaclen@zshospital.net 或 jianglindi@medmail.com.cn 或 jin@zshospital.net

拟参会人员请将姓名、性别、年龄、职称、专业、单位及通讯地址、电话号码等信息发送至上述电子邮箱。

联系人:姜林娣、金雪娟、王志军

通讯地址:200032 上海市枫林路 180 号中山医院内科

联系电话:021-64041990-2940,021-27992051,13120660799

第九届全国临床流行病学学术会议组委会