

# 流行病学研究数据质量评价方法的探讨

安徽医学院流行病学教研室 薛英杰\* 吴系科 张衍文

流行病学研究的结论正确与否主要取决于原始数据的准确可靠。现代流行病学研究已注意到现场和实验室工作的质量控制(quality control),以保证收集数据的正确性<sup>[1,2]</sup>。

评价原始数据的质量有两种指标,在以往的文献中称之为准确度(accuracy)和精密度(precision)<sup>[1~3]</sup>。准确度是反映观察数据符合变量真实水平程度的指标。精密度亦称可靠度(reliability)是反映若干次重复观察值重现程度的指标。一般情况下,流行病学现场和实验室工作不易预先得到样本变量的实际值,也就不易得到数据的准确度,这种指标常用于某些新的筛检方法的试验性评价<sup>[1]</sup>。对于大规模的流行病学现场和实验室数据,一般仅选用可靠度来评价<sup>[4~7]</sup>。流行病学数据可以根据变量性质分为定性(分类)、定量(测量)以及介于二者之间的半定量(等级)数据。对于不同性质的数据应选用不同的评价方法。

本文根据1982年5月~1983年5月安徽太湖、岳西中学学生血压流行病学研究数据,探讨定量和等级数据质量评价方法。

## 材料和方法

一、材料来源:材料来源于太湖、岳西中学学生血压流行病学研究,主要数据有:

- 1.定量数据:学生血压的三次测量值、身高、体重、血脂以及父母血压。
- 2.等级数据:父母文化水平是按文盲、初小、高小、初中、高中、大学分级;家庭经济收入是按<10元/人·月、10~14元/人·月、15~19元/人·月、20~24元/人·月、>30元/人·月分级;精神行为因素是根据本地中学生的理解

力和其他特征,按照有关文献<sup>[9]</sup>和《MMPI》(Minnesota Multiphasic Personality Inventory)的形式,预先拟定选择式答题,根据学生的答案而分为“焦虑”、“好胜”、“易怒”、“急躁”、“易激动”精神行为等级。

上述数据的获得都进行了必要的现场和实验工作的质量控制。为了评价质量控制措施的效果即数据质量,必须对现场和实验室数据进行盲法复查,我们将精神因素复查时间定为原始调查的半年以后,其它各因素复查均在调查过程中进行,并认为重复观察值的重复性(可靠度)是评价数据质量的恰当指标。

二、定量数据可靠估计方法:通过方差分析可以计算重复测量数据的“个体内”标准差(SDw, intra-person standard deviation),估计测量误差(包括观察者、仪器、环境和宿主状态等因素引起的误差),并可进一步计算重复测量值的“个体内”相关系数( $\rho$ , intra-person correlation coefficient)。

假定方差分析的随机变量模型为:

$$X_{ij} = \pi_i + \lambda_{ij}, \quad i=1, 2, \dots, n$$

$$j=1, 2, \dots, k$$

其中 $X_{ij}$ 为第*i*个受试者第*j*次测量值, $\pi_i$ 为第*i*个受试者全部*k*次测量的期望值, $\lambda_{ij}$ 为第*i*受试者第*j*次测量值的随机测量误差(表1)。

令 $\pi_i$ 总体均数 $\mu$ ,方差 $\sigma_{\pi}^2$ ;  $\lambda_{ij}$ 均数为0方差为 $\sigma_{\lambda}^2$ 。

那末受试者重复测量值的“个体内”标准差为:

$$SDw = \sqrt{MSw} \quad \dots\dots(1)$$

测量数据可靠度指标,“个体内”相关系

\* 研究生

表 1 受试者内测量误差方差分析表

方差来源	自由度 (df)	平方和 (t)	MS	MS的数学期望
受试者间	n-1	t <sub>B</sub>	MS <sub>B</sub>	$\sigma_{\lambda}^2 + k\sigma_{\pi}^2$
受试者内	n(k-1)	t <sub>W</sub> = t <sub>T</sub> - t <sub>B</sub>	MS <sub>W</sub>	$\sigma_{\lambda}^2$
总变异	nk-1	t <sub>T</sub>		

数为:

$$\rho = \frac{\sigma_{\pi}^2}{\sigma_{\pi}^2 + \sigma_{\lambda}^2}$$

$$\hat{\rho} = \frac{MS_B - MS_W}{MS_B + (k-1)MS_W} \dots\dots(2)$$

三、等级数据可靠度评价：等级数据重复观察值的等级相关系数表明等级数据重复性，可以作为可靠度的评价指标。两次重复观察值可以列成方形  $\gamma \times \gamma$  列联表计算 Kendall 相关系数，现以  $3 \times 3$  表为例说明计算方法(表2)。

表 2 两次重复测量值的频数分布 ( $3 \times 3$ )

第一次测定等级	第二次测定等级		
	1	2	3
1	$n_{11}$	$n_{12}$	$n_{13}$
2	$n_{21}$	$n_{22}$	$n_{23}$
3	$n_{31}$	$n_{32}$	$n_{33}$

$$P = n_{11}(n_{22} + n_{23} + n_{32} + n_{33}) + n_{12}(n_{23} + n_{33}) + n_{21}(n_{32} + n_{33}) + n_{22}n_{33}$$

$$Q = n_{13}(n_{21} + n_{22} + n_{31} + n_{32}) + n_{12}(n_{21} + n_{31}) + n_{23}(n_{31} + n_{32}) + n_{22}n_{31}$$

$$S = P - Q$$

$$T_1 = n_{11}n_{12} + n_{11}n_{13} + n_{12}n_{13} + n_{21}n_{22} + n_{21}n_{23} + n_{22}n_{23} + n_{31}n_{32} + n_{31}n_{33} + n_{32}n_{33}$$

$$T_2 = n_{11}n_{21} + n_{11}n_{31} + n_{21}n_{31} + n_{12}n_{22} + n_{12}n_{32} + n_{22}n_{32} + n_{13}n_{32} + n_{13}n_{33} + n_{23}n_{33}$$

则 Kendall 相关系数为:

$$\tau = \frac{P - Q}{\sqrt{(P + Q + T_1)(P + Q + T_2)}} \dots\dots(3)$$

如果各受试者观察值之间较少存在重复的等级可以应用 Spearman 等级相关系数作为估计。

### 结 果

#### 一、测量数据可靠度估计的结果:

1. 对 4 名观察者训练结果的测量误差和可靠度的分析：选择 12 名 12~18 岁的青少年作为测量对象(n=12)，采用双联单头听诊器，每次由两名观察者同时测量，随机决定观察者的组合和次序。每名受试者测压两次，间隔 15~20 分钟，合计有 4 次血压记录值(K=4)。根据有关文献<sup>[4]</sup>和以往的经验，我们确定可靠度  $\rho > 0.7$  作为预试合格的标准。列方差分析表(表 3)。

表 3 4 名观察者训练测量 SBP 的方差分析 (n=12, k=4, 岳西、太湖中学 1982.5.)

方差来源	df	t	MS	SD
受试者间	11	3201	291 (MS <sub>B</sub> )	—
受试者内	36	732	20 (MS <sub>W</sub> )	4.51 (SD <sub>W</sub> )
总变异	47	3933	—	9.15

根据表 3 进一步计算 SBP 的可靠度:

$$\hat{\rho} = \frac{MS_B - MS_W}{MS_B + 3MS_W} = \frac{291 - 20}{291 + 3 \times 20} = 0.77$$

可靠度已超过规定标准，说明预试过程对观察和其他因素的综合控制已达设计的要求。

同样方法可以计算 4 名观察者测量 12 名受试者的 Korotkoff 第 IV 相舒张压 (DBP<sub>IV</sub>) 和第 V 相舒张压 (DBP<sub>V</sub>) 的测量误差和可靠度，结果见表 4。

表 4 4 名观察者训练测量 SBP、DBP<sub>IV</sub>、DBP<sub>V</sub> 的测量误差和可靠度 (n=12, k=4)

变量	n	k	MS <sub>B</sub>	测量误差		可靠度 $\rho$
				MS <sub>W</sub>	SD <sub>W</sub>	
SBP	12	4	291	20	4.51	0.77
DBP <sub>IV</sub>	12	4	362	35	5.91	0.70
DBP <sub>V</sub>	12	4	309	82	9.08	0.40

2. 太湖、岳西中学学生 1982 年血压流行病学现场和实验室测量数据的可靠度：表 5 是太

湖、岳西中学学生1982年血压断面调查SBP与抽样复查测量值的方差分析表。复查人数n=40, 复查一次连同原始调查共2次, K=2, 按公式(2)计算 $\hat{\rho}$ :

$$\hat{\rho} = \frac{MS_B - MS_W}{MS_B + MS_W} = \frac{210 - 27}{210 + 27} = 0.77$$

表5 断面调查SBP测量误差的方差分析 (太湖、岳西中学1982, n=40, k=2)

方差来源	df	SS	MS	SD
受试者间	39	8177	209.66	—
受试者内	40	1097	27.42	5.24(SD <sub>w</sub> )
总变异	79	9274	—	10.82

身高、体重、两次随访测量的血压、血脂以及父母血压等测量值也都按此种方法计算测量误差和可靠度。

二、等级数据可靠度评价结果: 表6是学生“焦虑”行为等级重复测定值的频数分布, 第一次调查405人, 半年后随机抽查111人, 根据111人的两次测定结果, 可进行如下计算:

Kendall相关系数:

$$\tau = \frac{P - Q}{\sqrt{(P + Q + T_1)(P + Q + T_2)}} = \frac{1952}{\sqrt{4290 \times 4563}} = 0.44$$

表6 学生“焦虑”行为两次测定等级的频数分布 (n=111, k=2, 太湖、岳西中学, 1982)

第一次测定	第二次测定				合计
	≤ -5	< 0	≤ 5	> 5	
≤ -5	8	4	1	0	13
≤ 0	6	12	2	3	23
≤ 5	2	13	20	14	49
> 5	3	2	10	11	26
合计	19	31	33	28	111

同样方法可以评价学生的“急躁”、“易怒”、“易激动”、“好胜”等4种精神行为等级以及父母文化水平、家庭经济收入的可靠度。

## 讨 论

一、可靠度评价方法的正确选择: 流行病学资料偏差主要来源于两方面, 其一来源于抽样偏差, 其二来源于现场和实验室的观察误差, 包括观察者、仪器以及环境、宿主状态所引起的误差。防止第一类偏差主要依赖于正确的实验设计, 减少第二类偏差主要依赖于现场和实验工作的质量控制。

由于观察偏差不如抽样误差那样为人们所重视, 经常在质量评价方面出现各种类型的错误, 主要有如下两方面:

1. 将预试训练结果作为评价依据。预试工作一般在较小的人群中进行, 容易进行质量控制, 但是在较大的人群、较长时间的现场和实验室工作的质量控制较为困难。如果没有严格的质量控制, 观察数据的质量将会有较大的下降。

2. 选择不恰当的评价指标。不少研究应用成组均数或配对差别均数的t检验来断定观察之间差别的大小, 用以评价数据的质量。实际上t检验只能推断两次观察值之间可能存在的系统改变而产生的偏高或偏低估计的测量误差的显著性, 不能确定个体内随机测量误差的大小。

要对数据质量进行正确的评价, 必须对现场和实验数据进行复查, 选择能表达重复测量值的重复性(可靠度)的指标。“个体内”相关系数和Kendall等级相关系数则能恰当地反映定量数据和测量数据的重复性。

通过方差分析可以计算定量数据的“个体内”标准差和“个体内”相关系数。“个体内”标准差和“个体内”相关系数呈相反的关系, 它们评价问题的实质是一致的, 但SD<sub>w</sub>具有单位的限制, 而 $\rho$ 没有单位, 因此 $\rho$ 可以用来比较不同变量、不同研究的可靠度。

等级数据不符合方差分析的条件, 不能应用上述两种指标来评价, 并且观察值之间存在较多的重复等级, 因此选择Kendall相关系数

评价两次观察值的重复性是合理的。 $\tau$ 有近似于线性相关系数的性质,如果 $\tau$ 增大趋于1,则每个个体内两次观察值的等级越相似,反之差别增大。

二、太湖、岳西中学血压流行病学研究数据的质量评价:定量数据质量评价结果表明,体格测量数据可靠度最高(0.992~0.996),实验室血脂检查数据次之(0.86~0.96),而血压数据较低(SBP 0.77~0.79; DBPⅣ 0.72~0.76; DBPⅤ 0.29),说明血压测量的质量控制比较困难,存在较大的测量误差,加强这方面的质量控制是必要的。根据WHO的规定,本文比较第Ⅳ和第Ⅴ相舒张压的可靠度,第Ⅳ相为0.72~0.76,第Ⅴ相仅为0.29,因而本研究分析时采用第Ⅳ相舒张压。为了客观评价本研究质量控制的效果,就此与美国Bogalusa研究结果[4]比较,各项测量误差和可靠度均接近。

等级数据可靠度评价结果表明,父母文化水平的可靠度最高(0.94~0.98),精神行为因素最低(0.29~0.45),说明精神行为评定的可靠度偏低,对此有待于进一步改进。

### 摘 要

本文应用“个体内”相关系数和Kendall相关系数作为估计定量数据和等级数据可靠度的指标,评价太湖、岳西中学学生血压流行病学研究数据的质量。结果表明,在定量数据中,体重、身高的可靠度最高(0.992~0.996),实验室数据次之(0.86~0.96),

血压较低(SBP 0.77~0.79; DBPⅣ 0.72~0.76; DBPⅤ 0.29)。在等级数据中,父母文化水平可靠度较高(0.94~0.98),家庭经济收入次之(0.61),精神行为因素较低(0.29~0.45)。本文建议医学研究论文中除应报告实验设计和观察方法之外,还有必要评价研究数据的质量。

### ABSTRACT

Measurement error and reliability of epidemiologic data in two middle schools in Taihu and Yuexi counties have been studied, in which intra-person standard deviation was used to estimate measurement error and intra-person correlation coefficient reliability respectively. In quantitative data, those of weight and height showed greatest reliability (0.992 for weight, 0.996 for height), followed by laboratory variables (0.86-0.96), and BP variables came last (0.77-0.79 for SBP, 0.72-0.76 for DBP<sub>IV</sub>, 0.29 DBP<sub>V</sub>). Among ranked data, those of the education level of the parents were highest (0.94-0.98) psychological and behavioral variables lowest (0.29-0.45), with family income in between (0.61). Accordingly, it is suggested that the quality of data should be estimated in research papers besides presentation of the design and methods of study.

### 参 考 文 献

1. Lilienfield AM: Foundations of Epidemiology, New York, Oxford, P 153, 1980
2. 连志浩等译: 流行病学导论, 人卫, 135页, 1982
3. 祖述宪: 安徽医学院学报, 17: 82, 1982
4. Fester TA et al: J Chron Dis, 33: 661, 1980
5. Gardner MJ et al: J Chron Dis, 26: 781, 1973
6. Shepard DS: J Chron Dis, 34: 191, 1981
7. Armitage P et al: Clin Sci, 30: 337, 1966
8. 徐端正: 相关分析的非参数方法, 科研中的统计方法, 19页, 1982
9. Siegel JM et al: Am J Epidemiol, 113: 117, 1981

## 请订阅《肾综合征出血热英、俄文资料》

为较系统了解肾综合征出血热国外研究动态,我们将1978~1984年《BIOLOGICAL ABSTRACTS》、《BIOLOGICAL ABSTRACTS/RRM》、《РЕФЕРАТИВНЫЙ ЖУРНАЛ 36 МЕДИЦИНСКАЯ ГЕОГРАФИЯ》、《РЕФЕРАТИВНЫЙ ЖУРНАЛ БИОЛОГИЯ》中有关肾综合征出血热论文摘要和题目索引编辑成册,制成复印本。英文版16开本,一套共14页,定价5元;俄文版16开本,一套共17页,定价6元。欢迎单位和个

人订阅。

订阅办法:

1、银行汇款:北京市昌平县人民银行沙河办事处,帐号880110,收款单位:流研所。附言栏请分别注明:订英文资料或俄文资料或英、俄文资料;

2、邮局汇款:北京昌平流字5号,出血热室,附言栏内容同上。

联系人:北京昌平流字5号出血热室 刘党女