

· 系列讲座 ·

# 捕获——再捕获方法及其应用

章扬熙

在公共卫生监测中,常遇到估计总体含量的问题。比如,一个地区有多少某种传染病人,一个区域有多少能感染钩端螺旋体的黑线姬鼠等。对于这种在一个封闭区域里估计某种个体的总体含量的问题,可以应用捕获——再捕获的方法(capture mark recapture method)求解。

[例 1] 为了调查一个地区有多少黑线姬鼠,可先随机捕捉 M 个该鼠(假设 M = 200),将每个鼠做上标志后放还到原地,再随机捕捉 n 个黑线姬鼠(假设 n = 100),观察这 n 个鼠中带有标志的鼠数为 m 个(假设 m = 30),则该地区的黑线姬鼠的总数 N 可用下式求得

$$\hat{N} = \frac{Mn}{m} \quad (1)$$

本例得

$$\hat{N} = \frac{200 \times 100}{30} = 667 \text{ (只)}$$

该地有 667 只黑线姬鼠,这是一个点估计值。对总体含量 95% 可信区间的估计,可用公式(2)计算。

$$\hat{N} \pm 1.96 \sqrt{\text{Var}\hat{N}} \quad (2)$$

式中  $\text{Var}(\hat{N})$  为  $\hat{N}$  的方差,用公式(3)计算。

$$\text{Var}(\hat{N}) = \frac{Mn(M-m)(n-m)}{m^2(m-1)} \quad (3)$$

本例黑线姬鼠总数的 95% 可信区间为

$$\text{Var}(\hat{N}) = \frac{200 \times 100 \times 170 \times 70}{30^2 \times (30-1)} = 9118.77$$

$$\hat{N} \pm 1.96 \sqrt{\text{Var}(\hat{N})} = 667 \pm 1.96 \sqrt{9118.77} = 480 \sim 854 \text{ (只)}$$

这说明该地黑线姬鼠总数在 480 ~ 854 范围内的可能性为 95%,也就是说做这样估计 100 次有 95 次所估计的值是对的。

从公式(3)可以看出,总体含量方差的大小受第 2 次查到的有标志的个体数 m 影响很大,当 m = 1 时,方差的分母为零,方差为 ∞,同时总体含量  $\hat{N}$  也

很大,这是不合理的。所以,当 m 较小时,尤其是当 m < 5 时,应用公式(4)来求总体含量  $\hat{N}$ ,应用公式(5)来求总体含量方差  $\text{Var}(\hat{N})$ 。

$$\hat{N} = \frac{(M+1)(n+1)}{m+1} - 1 \quad (4)$$

$$\text{Var}(\hat{N}) = \frac{(M+1)(n+1)(M-m)(n-m)}{(m+2)^2(m+1)} \quad (5)$$

应用此校正公式,本例总体含量 N 为

$$\hat{N} = \frac{201 \times 101}{31} - 1 = 654 \text{ (只)}$$

总体含量方差  $\text{Var}(\hat{N})$  为:

$$\text{Var}(\hat{N}) = \frac{201 \times 101 \times 170 \times 70}{32^2 \times 31}$$

黑线姬鼠总数的 95% 可信区间为

$$\hat{N} \pm 1.96 \sqrt{\text{Var}(\hat{N})} = 654 \pm 1.96 \sqrt{7610.32} = 483 \sim 825 \text{ (只)}$$

以上各总体含量计算公式的假定条件是:<sup>1</sup> 两次捕获在同一个有限的、静态的总体中进行,即总体量没有变化,或总体量的增加、减少的情况很少,可以不计;<sup>②</sup> 两次捕获都是随机样本,即每个个体都有同等的机会被捕获到;<sup>③</sup> 两次捕获都是独立的,即第 2 次每个个体被捕获的机会不受第一次捕获到否的影响。在进行捕获——再捕获调查时,要注意使调查符合这些假定条件。

捕获——再捕获方法还可以用来估计一个地区某种传染病人的总体含量,进而估计其现患率(Prevalence rate)或发病率(Incidence rate)。

[例 2] 某氏曾对马德里 15 岁以下的儿童糖尿病的患病率进行调查,第一个来源是三所大医院,发现儿童糖尿病患者 432 人,第二个来源是当地糖尿病学会,发现糖尿病患者 138 人,其中与第一个来源重复的有 119 人,试求该地儿童糖尿病患病率(调查期间平均人口数为 4433628 人)。

首先,应用公式(1)、(2)、(3)计算该地儿童糖尿病病人总数的点估计值及区间估计值。

$$\hat{N} = \frac{432 \times 138}{119} = 501$$

该地糖尿病病人总数的点估计值为 501 人。其

95% 可信区间为

$$\text{Var}(\hat{N}) =$$

$$\frac{432 \times 138 \times (432 - 119) \times (138 - 119)}{119^2 \times (119 - 1)} = 212.17$$

$$\hat{N} \pm 1.96 \sqrt{\text{Var}(\hat{N})} = 501 \pm 1.96 \sqrt{212.17} =$$

472 ~ 530 人。该病人总数在此范围内的可能性为 95%。

该地于该期间的患病率点估计值为

$$\text{PR} = \frac{\hat{N}}{P} \times 100000/10 \text{ 万} = \frac{501}{4433628} \times 100000/$$

10 万 = 11.3/10 万

总体患病率的 95% 可信区间为

$$\frac{472}{4433628} \times 100000/10 \text{ 万} \sim \frac{530}{4433628} \times 100000/$$

10 万 = 10.6/10 万 ~ 12.0/10 万

由于这两个来源的病人数都是旨在掌握该地的总情况,故可计算各次的发现率和两次的总发现率。

第一来源的发现率 FR,用公式(6)计算

$$\text{FR} = \frac{M}{N} \times 100\% \quad (6)$$

第二来源的发现率 SR,用公式(7)计算

$$\text{SR} = \frac{n}{N} \times 100\% \quad (7)$$

两个来源的总发现率 TR,用公式(8)计算

$$\text{TR} = \frac{M + n - m}{N} \times 100\% \quad (8)$$

本例,求得

$$\text{FR} = \frac{432}{501} \times 100\% = 86.22\%$$

$$\text{SR} = \frac{138}{501} \times 100\% = 27.54\%$$

$$\text{TR} = \frac{432 + 138 - 119}{501} \times 100\% = 90.02\%$$

在疾病监测点的传染病居民回顾性抽样调查中,于年末或次年初回忆一年来居民新患传染病的情况,所得到的是一年内的新发病例数,与已报告的病例卡片核对,可知重复发现病例数,据此可估计该年该监测点实际新发病例数,进而可求该年该监测点的估计发病率。应当指出,若应用监测点的漏报率来估计全地区的发病率,严格地说是欠妥的,因为抽样调查的总体是监测点,不是全地区,除非客观上监测点的居民传染病发病的频率与分布和全地区相同。这一点要引起注意。

[例 3] 某监测点于 1997 年初做传染病居民回顾性抽样调查,发现 1996 年新发法定传染病病人计 65 人,经与该点 1996 年传染病报告卡片 1 456 张核对,重复的有 40 人,该点居民 1996 年平均人口为 318 462 人,试求该点 1996 年估计发病率。

本例 M = 1 456, n = 65, m = 40, 应用公式(1)求得该点 1996 年新发传染病病人数的点估计值为

$$\hat{N} = \frac{1456 \times 65}{40} = 2366(\text{人})$$

估计年发病率 IR 为

$$\text{IR} = \frac{2366}{318462} \times 100000/10 \text{ 万} = 742.95/10 \text{ 万}$$

再应用公式(2)、(3)求该点 1996 年新发传染病病人总数的 95% 可信区间,得

$$\text{Var}(\hat{N}) = \frac{1456 \times 65 \times (1456 - 40) \times (65 - 40)}{40^2 \times (40 - 1)} = 53690(\text{人})$$

传染病病人总数的 95% 可信区间为

$$2366 \pm 1.96 \sqrt{53690} = 1912 \sim 2820(\text{人})$$

这说明,该点 1996 年实际传染病病人总数在 1 912 ~ 2 820 人范围内的可能性为 95%,也就是说作这样 100 次推断,其中 95 次是对的。

进而可求估计年发病率 IR 的 95% 可信区间为

$$\frac{1912}{318462} \times 100000/10 \text{ 万} \sim \frac{2820}{318462} \times 100000/10 \text{ 万} = 600.39/10 \text{ 万} \sim 885.51/10 \text{ 万}$$

如果在疾病监测两次抽样的、独立的主动监测中,发现现患病例(不是一年的新发病例),对该年的年发病率估计,情况略复杂一些。

[例 4] 某市 1997 年 3 月对全市汽车司机同时进行了两次独立性的随机抽查,甲组发现淋病现患 120 人,乙组发现 160 人,两组重复者 80 人,淋病平均病程为 15 天,该市共有司机 5 736 人,假设淋病患者在一年均内均匀分布,无季节性,试估计该年汽车司机的淋病发病率。

首先,估计淋病现患人数, M = 120, n = 160, m = 80, 应用公式(1)得

$$\hat{N} = \frac{120 \times 160}{80} = 240$$

再应用公式(9),求年发病人数

$$\text{年发病人数} = \text{患病人数} \div \text{病程(年)} \quad (9)$$

本例,估计淋病年发病人数为

$$240 \div (15/365) = 5840(\text{人})$$

淋病年发病率为

$$\frac{5840}{5736} \times 100\% = 101.81\%$$

也就是说,平均司机一年要患 1.0181 次淋病。

还可以进一步作区间估计,应用公式(2)、(3)计

$$\text{Var}(\hat{N}) = \frac{120 \times 160 \times (120 - 80) \times (160 - 80)}{80^2 \times (80 - 1)} = 121.52$$

95% 可信区间为

$$240 \pm 1.96 \sqrt{121.52} = 218 \sim 262(\text{人})$$

年发病人次数的 95% 可信区间为  
 $218 \div (15/365) \sim 262 \div (15/365) = 5305 \sim 6375$   
 (人次)

汽车司机淋病年发病率的 95% 可信区间为  
 $\frac{5305}{5736} \times 100\% \sim \frac{6375}{5736} \times 100\% = 92.49\% \sim$   
 111.14%

## 浙江省温州爆发流行嗜酸性细胞性脑膜脑炎

王小同 黄汉津 董其昌 林燕 陈欣欣 上官云强 林李淼 应斌宇

1997年11月,我们收治了19例嗜酸性细胞性脑膜脑炎。在国内大陆地区,爆发流行嗜酸性细胞性脑膜脑炎尚属首次发生。现报告如下。

临床资料:本组系住院病人,共19例,男性11例,女性8例,年龄27~53岁。潜伏期:从1~24天。1例生食淡水螺肉后出现皮肤瘙痒、丘疹。1例呕吐一次、无腹泻。发作期:发作十分突然。2例有咽痛、全身乏力症状。14例有头痛,程度剧烈,常伴夜间痛醒,其中3例伴恶心呕吐。4例有肋间神经痛;6例表现上肢痛,7例下肢痛,8例有腹痛、背痛、腰痛。疼痛呈持续性,阵发性加剧。其受累区域皮肤呈烧灼感,不能触摸,否则疼痛加剧,但无压痛。严重者微风吹过也感到不适。12例有低热,1例高热,39.5℃,持续2天。体检:1例短暂性嗜睡,1例近事记忆遗忘。1例左侧周围性面瘫。3例克氏征可疑阳性。其他神经系统体检阴性。肌肉无压痛、眼睑无水肿,皮下无结节、肿块。无皮疹。辅助检查:周围血像:WBC4.2~1.2×10<sup>9</sup>/L, E0.04~0.65。腰穿:CSF压力升高4例,1例大于0.92kPa。F检查:色清无黄变,WBC24~1400×10<sup>6</sup>/L, E0.06~0.72。无红细胞。糖大于

50mg%。1例氯化物为9.3mmol/L,其他均正常。潘氏反应±~+,蛋白0.22~1.20g/L,CSF培养阴性,未找到霉菌、结核菌。4例血中旋毛虫、囊虫、包虫、肺吸虫、华支睾吸虫、曼氏裂头蚴抗体均阴性。血、脑脊液中弓形体、巨细胞病毒、风疹病毒、单纯疱疹病毒抗体阴性。头颅CT平扫无异常。治疗和预后:经肠虫清、糖皮质激素治疗,疗效佳。无死亡。流行病学调查:本组病人在发病前均有进食盐渍淡水螺肉史,集体进餐,集体发病。在同一来源的螺肉中找到广州管圆线虫第一期幼虫、在当地褐家鼠的肺动脉内找到广州管圆线虫的成虫。

讨论:本组病人周围血像、脑脊液中嗜酸性细胞增高,故嗜酸性细胞性脑膜脑炎可以确诊。嗜酸性细胞性脑膜脑炎常见于寄生虫感染。本组病人血清中囊虫、肺吸虫、包虫、旋毛虫抗体均阴性,故可排除上述寄生虫病。由于没有发现游走性皮下肿块、肌肉无压痛、脑脊液无红细胞、无黄变,故也不考虑软腭口线虫病。广州管圆线虫病是引起人类嗜酸性细胞性脑膜脑炎的主要原因之一,本组病人在发病前有生食淡水螺肉。根据临床资料和流行病学调查,可以推断本组病因应为广州管圆线虫病。

作者单位:浙江省温州医学院附属第二医院 325000

(收稿:1998-02-06)