

· 动态 ·

捕获-标记-再捕获法的历史、理论及应用发展

胡伟江, 周安寿

(中国疾病预防控制中心职业卫生与中毒控制所, 北京 100050)

捕获-标记-再捕获方法 (capture-mark-recapture methods, 简称 CMR 法) 是根据 2 个或 2 个以上独立样本来估计生物群体大小的一种方法。它最早由野生动物学家用于估计限定区域内某种野生动物 (如鱼、鸟、昆虫等) 的数量, 现用于研究人类疾病和健康问题, 其理论随着生物统计学的发展而逐步完善。本文简单介绍 CMR 法的历史、基本原理及其在流行病学中的应用、存在问题与展望。

1 历史回顾

捕获-标记-再捕获方法根据资料来源途径的多少分为两样本 CMR 法和多重 CMR 法。其中两样本 CMR 法最为简单, 它最早应用于生态领域, 1894 年 Pertenson 首次将此法与鱼群标记相联系, 1917 年 Dalh 第一次真正将其用于捕鱼业, 1930 年 Lincoln 又用于估计鸭群的大小, 1949 年 Sekar 和 Deming 用来估计出生率、死亡率和登记的范围, 这标志 CMR 法已应用于人类健康领域, 但真正用于流行病学的研究则始于 1969 年, Wittes 和 Sidel 用于估计出生缺陷率, 同年 Lewis 用于医院内感染的研究。当时人们对此法没有很好地理解和应用, 直到 80 年代, CMR 法才在各个领域得到广泛使用。

在使用两样本 CMR 法的同时, 国外学者发现了此法的局限性, 尝试使用多种途径来源的样本来估计群体的大小, 最早于 1938 年 Schnable 估计鱼群大小时, 首次提出了多重 CMR 法, 随着生物统计学的发展, 多重 CMR 法逐步得到了完善。

2 基本原理

2.1 两样本 CMR 法

两样本 CMR 法是利用两个不同途径的资料 (样本) 来估计某一群体大小的方法。其原理是从总体为 N 的生物群体中随机捕获第一个样本, 然后对其进行标记并释放到原生物群体中去, 随后再从该生物群体中随机捕获第二个样本, 其中第二次捕获的样本中含有第一次捕获的个体, 最后用第一个、第二个样本的个数及第二个样本中含第一次捕获的个体数来估计该生物群体的大小。具体计算方法如下: 假设有 2 个不同途径来源的资料 A 和 B, n_{11} , n_{10} , n_{01} , n_{00} 分别代表在 2 个样本中都出现的个体数、仅在第一个样本中出现的个体数、仅在第二个样本中出现的个体数和未在 2 个样本中均未出现的个体数。假设 $r = n_{11} + n_{10} + n_{01}$, 则 $n_{00} = N - r$, n_A 为第一次捕获的个体数, 则 $n_A = n_{11} + n_{10}$, n_B 为第二次捕获的个体数, $n_B = n_{11} + n_{01}$, 如果所有个体有同等概率被上述 A 和 B 两个样本所捕获, 则 $n_A/n_B = n_{11}/n_{01}$, 所以 $N = n_A n_B / n_{11}$, 但只有当该生物群体很大时, 这

种估计比较准确, 而实际应用中由于总体的数量有限, 这种估计存在偏倚, 为此 Chapman 等提出了无偏估计公式, 即 $N = [(n_A + 1)(n_B + 1)/(n_{11} + 1)] - 1$, 其方差 $\text{Var}(N) = [(n_A + 1)(n_B + 1)(n_A - n_{11})(n_B - n_{11})]/[(n_{11} + 1)^2(n_{11} + 2)]$, 95% 的可信区间为 $N \pm 1.96 \sqrt{\text{Var}(N)}$ 。

CMR 法理论上可用于任何具有 2 个或以上个不完整资料的总体估计, 如医院记录、医生的病例记录、医生处方、各种伤害的保险或赔偿机构的登记资料及各种疾病监测系统的资料等等。为了使两样本 CMR 法能准确估计某生物群体的大小, 应用时应具备以下条件: (1) 在调查期间, 被调查人群保持不变 (即人群是封闭的); (2) 在捕获期间, 每一个被标记的个体标记没有丢失, 以确保第二次捕获的个体能与第一次捕获的个体相匹配; (3) 每一个个体都有同等机会被不同样本所捕获; (4) 2 个样本是相互独立的。在实际研究中选择限定的人群时, 第一个条件基本能得到满足, 而第二个条件能否得到满足取决于各个样本之间用于匹配的变量的惟一性, 即第二次被登记的代码必须与第一次相同, 否则估计可能出现偏倚。第三、四项条件难以得到真正满足, 第三项条件得不到满足会直接影响第四项条件, 即各个个体被不同样本捕获的机会不同, 两样本之间就可能存在依赖性, 若某一样本中的个体更可能被另一个样本所捕获, 则会低估总体的数量, 相反则会高估总体的数量。在这种情况下, 应用两样本 CMR 法应注意偏倚的出现。

2.2 多重 CMR 法

多重 CMR 法是利用 2 个以上来源的资料估计某生物群体大小的方法, 资料可利用对数线性模型进行分析。其基本方法: 将不同来源的资料列成不完整的 2^k 列联表 (k 为样本数), 其中有 $2^k - 1$ 个格子为已知数据, 即根据每一样本捕获的个体数进行简单计算得到, 有 1 个格子为空, 即此前 k 次捕获中均未被捕获的个体数, 根据已捕获样本的个体数, 通过对数线性模型来估计未知的个体数。不同来源的资料之间的依赖性通过对数线性模型中的交互作用项加以体现, 交互作用是否具有显著性通过模型的符合率及相似率加以检验。对数线性模型计算复杂, 可使用 GLIM 软件和 SAS 软件加以分析。多重 CMR 法与两样本 CMR 法相比, 前者无需考虑样本之间的依赖性, 优于后者, 能用于国家范围内的严重疾病的监测, 具有更广泛的应用前景。

3 CMR 法在流行病学研究中的应用

CMR 法技术由于其自身的特点, 已在各个领域得到了广泛应用, 为研究各种疾病的病因提供了坚实的技术支持, 目前利用此方法已在以下领域进行了大量的研究工作: (1) 出生缺陷, 主要用于估计唇裂、腭裂、脊柱裂, Down's 综合征、致命性酒精

收稿日期: 2004-10-29

作者简介: 胡伟江 (1976-), 男, 硕士, 研究实习员, 主要从事职业流行病学及职业病危害评价工作。

中毒的发病率或流行率,此外用于评估这类疾病监测系统系统的完整性;(2)癌症,采用两样本 CMR 法和对数线性模型用于检测某些地区癌症登记系统的完整性,以确保登记系统的完善,有部分研究用于估计某种乳腺癌筛查法的敏感度和灵敏度;(3)药物依赖,估计药物依赖和静脉注射吸毒的流行率,旨在准确估计某地区、国家中海洛因、鸦片等毒品的使用情况,为政府部门采取干预措施提供准确的数据;(4)传染性疾病,主要用于性传播疾病流行情况的研究,尤其是艾滋病,并评估此类检测系统的有效性和资料的完整性;(5)胰岛素依赖型糖尿病,该方法已被认为是全球范围内监测胰岛素依赖型糖尿病流行情况的标准方法,目前已被 WHO 用于儿童糖尿病多国计划;(6)伤害,主要用于交通事故、工伤、青少年伤害等意外事故流行情况的研究,目前已在国内外得到了广泛的应用,主要利用各种医疗、保险、赔偿登记资料来估计伤害事故的发生情况;(7)其他,用于估计精神疾病、Huntington's 病、心梗等疾病的发病率,同时用于流浪人群、孤儿等特殊群体的估计。

4 问题与展望

如果能系统、准确地使用 CMR 法,可大大提高各种疾病

监测系统的效率和准确性。但如前所述,在应用两样本 CMR 法时,须具备一些条件,如两样本之间存在依赖性会影响总体估计的准确性,同样用不同方法进行捕获时,标准的偏差也会影响其估计的准确程度。随着生物统计学的迅猛发展,CMR 法会得到进一步完善,并得到更广泛应用。在有多种资料来源的条件下建议采用多样本 CMR 法,以避免两样本 CMR 法的局限性,使估算结果更为准确。

参考文献:

- [1] International Working Group for Disease Monitoring and Forecasting. Capture-recapture and multiple-record systems estimation I: History and theoretical development [J]. Am J Epidemiol. 1995, 142: 1047-1058.
- [2] International Working Group for Disease Monitoring and Forecasting. Capture-recapture and multiple-record systems estimation II: Applications in Human Disease [J]. Am J Epidemiol. 1995, 142: 1059-1068.
- [3] Laporte RE, Dearwater SR, Chang YF, et al. Efficiency and Accuracy of disease monitoring systems: Application of Capture-recapture methods to injury monitoring [J]. Am J Epidemiol. 1995, 142: 1069-1077.
- [4] Morse T, Dillon C, Warren N, et al. Estimation of unreported work related musculoskeletal disorders in Connecticut [J]. Am J Ind Med. 2001, 39 (6): 636-642.

一起 CS₂ 泄漏爆燃致学生群体不适反应的报告

蔡丽红¹, 赵清波¹, 郭洪涛¹, 张国东¹, 亚云珠²

(1. 沈阳市卫生监督所, 辽宁 沈阳 110014; 2. 沈阳市疾病预防控制中心, 辽宁 沈阳 110031)

2002 年 4 月 27 日 8 时许, 某农药厂 CS₂ 泄漏后遇热发生爆燃, 造成附近一学校 257 名学生出现不同程度头晕、恶心、咽干等症状, 现报告如下。

1 事故调查

该农药厂建于 60 年代, CS₂ 备料间位于厂区东南侧, 高 4.5 m, 占地约 20 m², 门窗无玻璃, 直接与外部相通。CS₂ 储罐在室内一角半地下低温放置, 罐口高出地面 0.5 m。上料管与供热蒸气管在一侧墙壁上平行走向, 供热蒸气管高出 CS₂ 上料管 0.6 m, 高出地面 3.6 m。室内无任何机械通风设施。CS₂ 备料岗采用一班制, 定员 1 人。事发当日操作员打开 CS₂ 上料阀上料后离去, 没有及时发现 CS₂ 上料管补漏的焊缝处再次出现漏点, 使 CS₂ 顺漏点喷出 (估计当时泄漏量约 5 kg), 恰好喷射到上方的供热蒸气管上, 导致 CS₂ 遇热立即发生爆燃。大量 CS₂ 蒸气顺门窗外逸扩散。在 CS₂ 备料间东侧约 200 m 处有一小学校, 为东西朝向三层建筑。事发当时为西南风 2~3 级, 恰好将爆燃产生的大量 CS₂ 蒸气顺风吹到下风向的小学校。学生正在上课, 大部分教室开着窗。

事发后, 采用活性炭吸附采样, 气相色谱法检测。在 CS₂ 备料间内储罐旁测得 CS₂ 短时间容许浓度 66.4 mg/m³; 在 CS₂

备料间与教学楼之间距备料间 6 m 处测得 CS₂ 短时间容许浓度 10.3 mg/m³。

2 临床资料

事发当日下午, 相继有 20 多名学生开始出现头晕、恶心、咽干症状。此后一周内, 共有 257 名学生 (约占全校学生的 1/3) 陆续出现上述症状, 绝大部分学生经当地职业病救治机构临床检查, 心率、血压、体温、跟腱反射、胸透、脑电图等无异常改变, 休息 2~3 d 后症状消失; 仅有 6 名学生脑电图的 Q 波频段功率增强, 经临床医学观察, 营养脑细胞治疗一周, 脑电图上述改变消失。事发时生产现场及周围无生产作业人员, 故厂内没有人员出现中毒或不良反应。

3 分析与建议

3.1 该厂是一家生产农药近 40 年的老企业, 因资金有限, 生产作业场所没有进行过技术改造, 厂房简陋, 设备陈旧, 管路年久腐蚀, CS₂ 上料管多次出现漏点, 而且上料管与供热蒸气管距离较近, 造成 CS₂ 泄漏后遇热立即发生爆燃。

3.2 建立健全岗位责任制。一是实行设备定期检修和每班生产前检查制; 二是严格遵守操作规程, 不能擅自脱岗。

3.3 消除隐患。将经过 CS₂ 备料间内的供热蒸气管移出, 改变走行路线, 使其远离热源; 同时更换上料管, 不再采用焊接补漏方法, 避免再次泄漏。

3.4 目前, 该企业的东面是小学, 南、北、西均为住宅, 企业被围在中间。为防止类似事故再次发生, 企业应尽快迁走, 另择符合《工业企业设计卫生标准》(GBZ1-2002) 要求的厂址, 重新建设生产。

本事故提示, 学校、住宅选址应严格执行《中小学校建筑设计》(GBJ99-86) 等相关法律法规和标准, 禁止在污染源周围建学校和居民住宅区及公共建筑。避免因选址不当, 造成学校和居民区等周围环境被有害物质污染。