

Monte Carlo 方法在碘盐监测抽样方法中的应用

郑建东 郑庆斯

【摘要】 目的 探索既能反映省级水平的碘缺乏病状况,又能发现地区级水平的非碘盐问题地区的新监测方案。方法 对甘肃省与福建省碘缺乏病综合干预试点研究的基线调查结果,运用 Monte Carlo 方法进行模拟抽样试验,分县级与地级两个层次进行。结果 在县级模拟中,一个县随机抽取 7 至 8 个村或 4 至 5 个乡,可较为准确地判定出全省的非碘盐问题县,真阳性率在 80% 左右,而对于非问题县的误判率较低,假阳性率在 20% 左右。在地级模拟中,每个地区选取 3 个县,可对全省的非碘盐问题地区较为准确地判定,真阳性率达到 100%。结论 建议全国碘缺乏病流行病学调查中碘盐的抽样方案覆盖各省所有地区,每个地区随机抽取 3 个县,每个县随机抽取 4 个乡,每乡随机抽取 2 个村,每村随机采取 10 户盐样,所获得资料经人口加权计算省级水平的碘盐覆盖率。

【关键词】 碘缺乏病; Monte Carlo 方法; 碘盐监测

Application of Monte Carlo sampling method in iodized salt monitoring ZHENG Jiandong, ZHENG Qingsi. Department of Iodine Deficiency Disorders, Institute for Infectious Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 100020, China

【Abstract】 Objective To explore a sampling method which could reflect iodine deficiency disorders (IDD) status at provincial level and discover risk areas with non-iodized salt problem. **Method** Baseline data of Iodized salt from Gansu and Fujian provinces were analyzed with Monte Carlo method both at county and prefecture levels respectively. True positive rate and false positive rate were also calculated. **Results** With data from 7-8 villages or 4-5 townships counties at risk could be discovered. The true positive rate was around 80% and false positive rate was around 20%. At prefecture level, when randomly selecting and checking 3 counties, the samples would satisfy the discovery of all the risk areas with non-iodized salt problem. **Conclusions** We suggested that the sampling method of iodized salt investigation in national IDD surveillance as follows: to randomly choose 3 counties at each prefecture, 4 townships at each county, 2 villages at each township and 10 salt samples by household survey. The coverage rate of iodized salt in a province could be calculated by post-weighted method with population number.

【Key words】 Iodine deficiency disorders; Monte Carlo method; Iodized salt monitoring

我国实施全民食盐加碘消除碘缺乏危害已经取得重大进展,监督与监测是持续消除碘缺乏危害的重要内容。按人口比例概率抽样法(PPS法)首先应用于免疫接种率的调查,后来逐步用于与免疫接种率调查相类似的其他公共卫生问题的调查^[1]。1995年以来已进行过3次碘缺乏病监测,采用PPS法抽样。PPS是一种很好的抽样方法,但有其局限性。在全国碘缺乏病监测中,它可对省级水平的碘缺乏病状况做出评价,但不能反映各项指标在省内各地县的分布情况;尤其是人口少的地区被选中的几率小,而这些地区往往非碘盐问题比较严重^[2]。

为了探索更为有效的监测方案用于全国监测,要求既能反映省级水平的碘缺乏病状况,又能发现省内的非碘盐问题地区。本研究采用 Monte Carlo 方法对甘肃省与福建省的调查数据资料进行模拟,并提出新的抽样方案。

资料与方法

1. 甘肃省与福建省碘缺乏病综合干预试点研究基线调查结果:调查覆盖全省所有地区、县、乡,每乡选取2个村,每村随机抽取10户食盐,采用半定量方法检测。

2. Monte Carlo 模拟:运用 Visual Foxpro 3.0 编制程序进行县级和地级两个层次的模拟,目的是:从一个县中随机抽取多少个村或乡可以较为准确地判定出该县为非碘盐问题县;从一个地区中随机抽取多

基金项目 联合国儿童基金会资助项目(O135YH101-12)

作者单位:100020 北京,中国疾病预防控制中心传染病预防控制所碘缺乏病室

少个乡或县可以较为准确地判定出该地区为非碘盐问题地区。包括 :县(县级行政单位)- 村,县(县级行政单位)- 乡,地(地级行政单位)- 乡和地(地级行政单位)- 县 4 个系列。

以县 - 村模拟为例,在省内所有县中各随机抽取 1 个村,以抽样村的基线调查结果来计算该县碘盐覆盖率,碘盐覆盖率低于 90% 则判定该县为非碘盐问题县,依据所有县的判定结果与其真实结果计算真阳性率(true positive rate, TP)与假阳性率(false positive rate, FP)。TP 和 FP 是筛查方法真实性的评价指标。TP 是指在实际的非碘盐问题县被正确判定的所占的比例,FP 是指碘盐覆盖率 $\geq 90\%$ 的县被判定为非碘盐问题县所占的比例。然后,在省内所有县中随机抽取 2 个村,进行第二轮判定,再计算相应的 TP 与 FP,随后是 3 个村、4 个村,直至 10 个村,分别计算 TP 与 FP。

县 - 乡与地 - 乡模拟的抽样与上相同,分别进行 10 次。地 - 县层次的抽样截至 5 个县。因为县一般有 10 个以上的乡,而地区少有 10 个以上的县。

3. 统计分析 采用 SPSS 统计图 Excel。

结 果

1. 甘肃和福建两省县级层次的模拟结果 :该层次模拟的目的是确定每县抽取多少个村或乡可以较为准确地判定出非碘盐问题县。在甘肃和福建两省内,每县随机抽取 7 至 8 个村,TP 值均可以达到 80% 左右,FP 值在 20% 左右,在此基础上增加村的数目并未显著提高判定效率(图 1)。

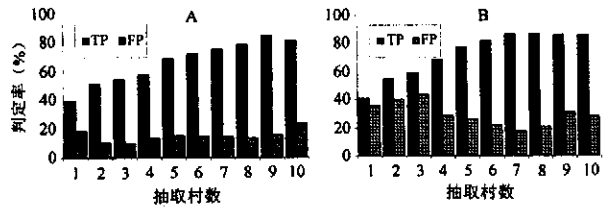


图1 应用 Monte Carlo 方法对甘肃省(A)和福建省(B)县 - 村级碘盐监测抽样模拟结果

在两省内每县随机抽取 4 至 5 个乡,TP 值均可以达到 80% 左右,FP 值在 20% 左右,在此基础上增加乡的数目也未能显著提高判定效率(图 2)。

2. 甘肃和福建两省地级层次的模拟结果 :该层次模拟的目的是确定每地区抽取多少个乡或县可以较为准确地判定出非碘盐问题地区。甘肃与福建两省的资料在地 - 乡模拟中出现较大差异,可能与碘

盐分布均匀程度有关。甘肃省碘盐分布相对不均匀,每地区抽至 8 个乡 TP 值达到 80%,而福建每地区抽至 4 个乡 TP 值已达到 100%(图 3)。

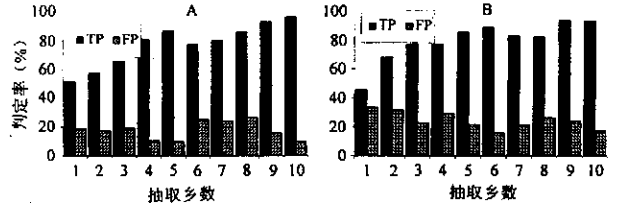


图2 应用 Monte Carlo 方法对甘肃省(A)和福建省(B)县 - 乡级碘盐监测抽样模拟结果

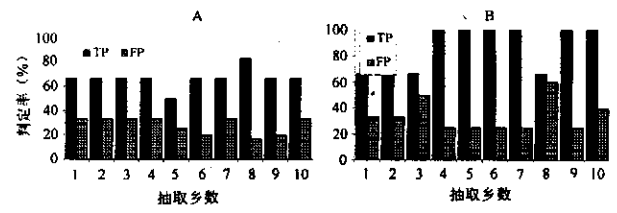


图3 应用 Monte Carlo 方法对甘肃省(A)和福建省(B)地 - 乡级碘盐监测抽样模拟结果

两省内一个地区随机抽取到 2 个县,对问题地区的判定准确率达到 100%。抽取到 3 个县时,福建省的结果显示,FP 值为 0(图 4)。

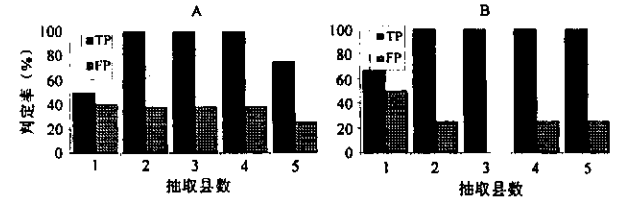


图4 应用 Monte Carlo 方法对甘肃省(A)和福建省(B)地 - 县级碘盐监测抽样模拟结果

讨 论

1. 我国于 1995、1997 和 1999 年进行过 3 次全国范围的碘缺乏病流行病学调查,抽样方法采用 PPS 法,以省为抽样单位,每省抽取 30 个群(县)结果可对省级水平的多项指标做出评价,包括甲状腺肿大率,人群尿碘水平以及碘盐覆盖率等。PPS 法有其局限性,以省为单位进行 PPS 抽样,只能反映省级水平,不能反映省内地区间的差异或分布的不均匀程度,尤其是人口少的地区被选中的几率小,而这些地区往往非碘盐问题较为严重,因此需要探索一种解决上述问题的简便高效的抽样方法。本研究中,我

们使用 Monte Carlo 方法,利用甘肃和福建两省综合干预项目基线调查的资料数据进行模拟试验,能获得较准确判定非碘盐问题地区的抽样方法。

Monte Carlo 方法,也称作计算机随机模拟方法,是计算机模拟中最为重要的一种,是一种基于随机数的计算方法,通过随机试验去求解众多类型的问题。随着计算机技术的飞速发展,计算机模拟作为一种新的研究方法得到更多的应用。Monte Carlo 方法通过计算机程序,对真实的过程进行微观模拟,得出有关规律。Monte Carlo 模拟是一种通过随机变量函数的概率模拟、统计试验或随机抽样来求解现实问题的近似解的计算方法。Monte Carlo 试验是从有限总体中抽样,并通过观察抽样试验的结果来间接得到实际问题的解答。它是通过反复随机抽样来进行,结果也是通过随机抽样试验得到。本研究即是一次 Monte Carlo 试验,一个省内采集的所有盐样作为一个有限总体,在有限总体中反复随机抽取不同数量的盐样,直至接近真实值。

2. 本研究应用西部省份甘肃和沿海省份福建的资料对反映非碘盐问题有较好的代表性。两省均存在非碘盐冲击问题,甘肃省的非碘盐问题比较严重而福建省非碘盐问题县多集中在沿海地区。为了准确反映两省的非碘盐分布状况,1999 年启动的综合干预基线调查覆盖省内所有地区、县、乡,每个乡随机抽取 2 个村。甘肃和福建两省的基线调查的碘盐覆盖率分别为 81.7% 和 91.3%,所有县中非碘盐率在 10% 以上的分别占 37.9% 和 24.1%。此次调查覆盖面广,抽样点密,样本量大,接近普查,调查结果能较真实反映省内的非碘盐分布情况,因此在本研究中被用来作为计算 TP 和 FP 的参照值。

本次 Monte Carlo 试验的结果表明:在一个县中随机抽取 7 至 8 个村或者 4 至 5 个乡,就可以较为准确地判断出全省的碘盐问题县(TP 值在 80% 左右),并且对于非问题地区的误判率也不高(FP 值在 20% 左右),再增加抽取的乡、村数只是增加工作量,而效率并未明显提高。地-县的模拟结果显示,在每个地区内随机选取 2 个县对全省的碘盐问题地区的判定结果很好,TP 值达到 100%。为确保数据的代表性和更好的判定准确率,可以考虑每个地区抽取 3 个县,此时福建省的 FP 值为 0。因此建议新的全国监测抽样方法覆盖每省所有地区,以地区为抽样单位,每个地区随机抽取 3 个县,每个县随机抽取

4 个乡,每乡随机抽取 2 个村,在每个村内采用入户方式随机采取 10 户盐,将能得到一个对地区级比较理想的判定结果。

3. 在模拟试验中,甘肃省地-县第 5 次模拟出现 TP 值下降的结果,这可能与 Monte Carlo 方法固有的特点有关。Monte Carlo 试验具有 4 个特点:第一是随机性,因为是随机试验,所以结果不可避免地含有随机成分,同样的试验次数得到的结果一般不会完全相同;第二是简单性,大量重复的抽样试验并通过观察试验结果经过简单计算即可得到结果;第三是灵活性,可设计不同的抽样试验来实现求解实际问题的目的;最后是近似性,结果是通过随机试验统计得到的,真值不可得知,基于统计理论,随着试验次数的增加所得到的解会以某种方式逐步接近于真实解。由于 Monte Carlo 方法固有的特点,可能会出现模拟结果图中些许不规律的现象,如前述出现抽样单位增多而 TP 值下降的现象。

4. 本模拟的结果对于判定一地区是否非碘盐问题地区具有相当高的效率,但存在是否适于直接计算率的问题。我们对一个地区随机抽取 3 个县作地区模拟试验,发现对于碘盐分布不均匀的地区直接计算的结果稳定性较差,变异较大,不能真实反映该地区的碘盐实际水平,不宜直接计算率,应经过人口加权统计处理。经过后加权,即抽样县的碘盐覆盖率用本县人口数进行加权,根据每个地区 3 个县的抽样密度计算所得率对全省应具有代表性。

根据研究结果所提出的碘盐监测抽样方案,抽样点相对于原方法有所增加,但所获得的信息大大超过原方法。原方法只能获得省的碘盐覆盖率,而本方案可较准确地判定地级及县级是否存在非碘盐问题,在进行人口加权后可获得省、地、县甚至乡的碘盐覆盖率。操作和计算简便,易为基层专业人员掌握,有利于推动我国碘缺乏病防治的深入开展。

(本研究运用甘肃省与福建省的基线调查资料,美国流行病学教授 Jonathan Gorstein 博士对本研究进行技术指导,中国疾病预防控制中心薛禾生研究员为 Monte Carlo 方法提供咨询,一并致谢)

参 考 文 献

- 1 曾光,主编.现代流行病学方法与应用.第 1 版.北京:北京医科大学中国协和医科大学联合出版社,1994.61-74,237-249.
- 2 孙树秋,王守德,苏明,等.PPS 法的流行病学原理及其在碘缺乏病监测中的应用.中国地方病学杂志,1996,15:115-117.

(收稿日期 2002-02-25)

(本文编辑:张林东)