

国家人体生物监测项目抽样方法及误差估计

曹兆进 曲英莉 赵峰 刘玲 宋士勋 刘迎春 蔡嘉琦 施小明

100050 北京, 中国疾病预防控制中心环境与健康相关产品安全所环境流行病学室

通信作者: 施小明, Email: shixm@chinacdc.cn

DOI: 10.3760/cma.j.issn.0254-6450.2018.12.020

【摘要】 目的 探讨国家人体生物监测项目抽样方法, 并通过计算抽样误差来评估抽样设计。方法 该监测采用多阶段复杂抽样方法, 并以贵州省抽样结果为例, 应用多阶段不等概率抽样误差计算方法计算贵州省抽样误差及变异系数, 探讨多阶段抽样设计下误差估计方法。结果 该监测覆盖全国 31 个省(自治区、直辖市)的 152 个监测点, 抽取 21 888 名 3~79 岁城乡居民, 三个阶段置换率分别为 5.26%、6.35% 及 40.6%。经计算贵州省抽样误差为 3 207 594 人, 变异系数为 0.097。结论 根据多阶段不等概率抽样方法计算, 贵州省抽样变异系数较小, 精度较高, 但该方法未考虑人群缺失率、应答率等非抽样误差的权重调整, 大型公共卫生监测项目多阶段抽样误差计算方法有待继续研究。

【关键词】 人体生物监测; 多阶段复杂抽样; 抽样误差

基金项目: 中央财政经费项目-环境污染健康人体生物监测体系建设(131031108000150003)

Sampling methods and errors appearing in the China National Human Biomonitoring Program

Cao Zhaojin, Qu Yingli, Zhao Feng, Liu Ling, Song Shixun, Liu Yingchun, Cai Jiayi, Shi Xiaoming

Department of Environmental Epidemiology, National Institute of Environmental Health, Chinese Center for Disease Control and Prevention, Beijing 100050, China

Corresponding author: Shi Xiaoming, Email: shixm@chinacdc.cn

【Abstract】 Objective To explore the sampling method in China National Human Biomonitoring Program (HBP) and the related errors, so as to calculate and evaluate the study design in sampling. **Methods** The sampling method of HBP is of multistage nature. Taking the results of sampling method from Guizhou province as an example, results related to sampling error and variation coefficient were calculated, using the multistage unequal probability sampling error method. **Results** The HBP covered 152 monitoring sites in 31 provinces (autonomous regions and municipalities) and with 21 888 residents selected. The replacement rates at various stages were 5.26%, 6.35% and 40.6% respectively. The sampling error in Guizhou province was 3 207 594, and the coefficient of variation was 0.097. **Conclusions** According to the multi-stage unequal probability sampling method, the sampling coefficient variability appeared small with high precision, in Guizhou province. However, this method did not consider the weight adjustment of non-sampling errors such as population missing rate and response rate. Methods related to the calculation on multi-stage sampling error among large-scale public health monitoring projects need to be further studied.

【Key words】 Human biomonitoring; Multistage sampling; Sampling error

Fund program: Program of Central Financial Funds-Construction of Human Biomonitoring System for Environmental Pollution (131031108000150003)

多阶段抽样方法是目前国际上公共卫生领域大规模流行病学调查常用的抽样方法之一^[1]。多阶段抽样具有样本比较集中、便于调查及节省费用的优点, 但其抽样产生的样本具有复杂样本的特征, 各阶段抽样数据不独立, 易低估抽样误差或增加统计推断 I 类错误的风险^[2]。因此对多阶段抽样方法在项

目中的应用及其抽样误差的计算方法受到越来越多学者的重视。

人体生物监测是指通过对人体组织和体液内(血、尿、母乳等)环境化学物质及其代谢物的分析, 获取个体及群体暴露环境化学物质的类别、数量、负荷水平及变化趋势等广泛性的数据, 以评估人群环

境有毒化合物暴露的健康风险。因此,人群样本必须具有代表性。目前世界各国均已开展大型人体生物监测流行病学调查,并基于各国国情选择适合于本国的人群抽样方法。德国环境监测项目(GerEs)是全球最早开展的大型人体生物监测项目,其项目方案、抽样设计、生物样本采集及数据分析等技术性规范为其他国家生物监测工作的开展提供了非常重要的参考价值,其调查对象主要为成年人(GerEs I ~ III)和儿童(GerEs IV ~ V),采用二阶段随机抽样方法^[3],第一阶段将社区作为初级抽样单元(PSUs),在每个PSUs内按人口数量分为7~9个抽样点(sample points)后,按随机抽样的方法抽取150个抽样点;第二阶段按分层随机抽样方法抽取调查对象;美国国家健康与营养调查(NHANES)^[4]采取四阶段分层抽样的方法,第一阶段为从全美的郡组成的抽样框架中抽取初级抽样单元,第一阶段的初级抽样单元主要是郡,极少数情况下会合并邻近的郡,以保证初级抽样单元达到最小样本量要求;第二阶段为抽取街区或街区组合而成的片区;第三阶段为抽取居住区;第四阶段在抽取的居住区内选择个体与家庭。加拿大生物监测项目为满足健康调查的要求^[5],采用多级分层抽样方法,第一阶段采用加拿大劳动力调查抽样框,在全国设置了18个采样点;第二阶段结合人口普查的人口组成情况,将年龄分为6组,每个采样点的每个年龄组随机选择1户,通过询问的方式掌握家庭所有成员的名单,从该名单中选择调查对象。

人体生物监测在我国起步较晚。中国CDC环境与健康相关产品安全所于2015年在全国范围内启动人体生物监测项目现场调查,旨在评价我国居民人体生物组织中环境化学物质暴露水平及分布特征。该监测采用多阶段抽样方法,选择有代表性的人群样本,使样本具有全国代表性,同时兼顾经济有效的原则和实施的可行性。本研究旨在介绍国家人体生物监测的抽样方法与内容,并以贵州省为例计算抽样误差,探索该抽样方法的可行性与科学性。

对象与方法

1. 调查对象:国家人体生物监测抽样调查的主要指标为环境污染物人体暴露的总体均数。调查指标的置换度取95%($t=1.96$),要求对全国总体,环境污染物暴露总体均数相对误差为10%,各类环境污染物相对误差控制在2%~6%。确定样本量的主要参数为环境污染物暴露 \bar{x} 及 s (以铅为例, $s=10.75$,

$\bar{x}=24.98$)、相对误差(ε)、把握程度($z_{\alpha/2}$)和设计效应($deff$),全国设计效应估计为2。代入公式计算,得到:

$$n = \left(\frac{z_{\alpha/2} s}{\varepsilon \bar{x}} \right)^2 \times deff = 142 (\text{人})$$

考虑人群年龄、性别分配的均衡性、工作量及经费预算,在满足抽样要求的基础上,国家人体生物监测项目确定在152个监测点(区/县)内开展工作,每个监测点调查人数为144人,总样本量为21 888人。

2. 研究方法:采用三阶段抽样的方法,在全国抽取152个监测点;从每个被抽中的县级单位中抽取3个社区(村)作为调查单元;从每个被抽中的调查单元中抽取48名常住居民作为调查对象。

(1) 抽样框:第一阶段抽样框为全国2 869个县级单位(不包括新疆生产建设兵团);第二阶段抽样框为被抽中县级单位内所有的社区(村)。为保证抽样要求,若社区(村)人口数不足2 000人,则按位置相近、交通便利的原则将之与周围村进行合并为1个调查单元;第三阶段抽样框为被抽中调查单元内所有常住居民的人口学资料。常住居民指在调查前12个月内在监测地区居住6个月以上的中国籍居民,排除居住功能区中的居民,如军队、学校及养老院等。

(2) 抽样方法:

第一阶段抽样:首先根据第六次人口普查数据中31个省(直辖市、自治区)的总人口数,对各省(直辖市、自治区)内监测点个数进行分配。各省(直辖市、自治区)监测点个数按照以下公式进行计算,每个监测点平均代表人数约为900万人(280万~1 300万)。经计算后,各省(直辖市、自治区)分配监测点个数为2~8个,见表1。

$$\text{各省监测点个数} = \frac{\sqrt{\text{省人口数}}}{\sum \sqrt{\text{各省人口数}} / 152}$$

各省监测点个数确定后,以省(直辖市、自治区)为子总体,以城镇化率和第二产业就业率为指标在各省(直辖市、自治区)内对区/县进行聚类分层,层数与各省监测点数相等。然后在各省(直辖市、自治区)内按分层随机抽样的原则,在每层内随机抽取1个区/县作为监测点,全国共抽取152个监测点。

第二阶段抽样:在每个已抽取的监测点内,按PPS的方法抽取3个调查单元。首先根据各监测点的城镇化率确定其3个调查单元的城乡比例分配。城镇化率>75%的监测点,其3个调查单元全部为居委会;城镇化率为50%~75%,其调查单元为2个居委会和1个行政村;城镇化率为25%~50%,其调查单元为1个居委会和2个行政村;城镇化率为<

表 1 国家人体生物监测项目监测点地区分布

地区	监测点个数	人口数
广东省	8	104 320 459
山东省	8	95 792 719
河南省	8	94 029 939
四川省	7	80 417 528
江苏省	7	78 660 941
河北省	7	71 854 210
湖南省	6	65 700 762
安徽省	6	59 500 468
湖北省	6	57 237 727
浙江省	6	54 426 891
广西壮族自治区	5	46 023 761
云南省	5	45 966 766
江西省	5	44 567 797
辽宁省	5	43 746 323
黑龙江省	5	38 313 991
陕西省	5	37 327 379
福建省	5	36 894 217
山西省	5	35 712 101
贵州省	5	34 748 556
重庆市	4	28 846 170
吉林省	4	27 452 815
甘肃省	4	25 575 263
内蒙古自治区	4	24 706 291
上海市	4	23 019 196
新疆维吾尔自治区	4	21 815 815
北京市	3	19 612 368
天津市	3	12 938 693
海南省	2	8 671 485
宁夏回族自治区	2	6 301 350
青海省	2	5 626 723
西藏自治区	2	3 002 165

25%,其3个调查单元全部为行政村。确定城乡抽样比例后,每个监测点按城乡分2层,每层按照人口规模PPS抽样的方法进行抽取,共抽取3个调查单元。根据抽样方法,国家人体生物监测项目覆盖全国456个调查单元,其中248个社区及208个行政村,城乡比例为1.19:1。

第三阶段抽样:在每个已抽取的调查单元中,按照分层随机抽样的方法抽取调查对象。获得被抽取调查单元的常住人口学资料,按年龄分为3~、6~、12~、19~、40~、60~79岁组共6层,每层中再按性别分为2层,共计12层,按分层随机抽样的方法抽取样本4人,共计48人,进行问卷调查及生物样本采集。

3. 置换方法:

(1)第一阶段置换:监测点置换原则为所抽中监测点地理位置偏僻,交通极为不便,现场工作难以实施。根据置换原则确定是否需要进行监测点置换。符合置换条件的监测点返回其所在的分层中重新进行抽样,抽取新的监测点。监测点置换以全国为单

位,总置换率不得超过10%,若超过,则放弃本次抽样,重新抽取152个监测点。本项目全国152个监测点中,共有8个监测点符合置换原则,总置换率为5.26%(8/152)。

(2)第二阶段置换:调查单元置换原则:①监测点地理位置偏僻,交通不便,现场实施工作难度太大;②调查单元人口年龄、性别结构不符合生物监测抽样要求。本项目要求调查单元置换率不得超过10%。目前已完成第二阶段抽样的30个省(直辖市、自治区)441个调查单元中,共有28个调查单元提出并同意置换,总置换率为6.35%。

(3)第三阶段置换:调查对象置换原则:①调查对象不在家;②调查对象拒绝调查;③调查对象无法接受调查;④调查对象未完成调查。第三阶段抽样完成后,调查员通过与调查对象预约沟通,确定是否需要置换并按照要求完成置换。调查对象抽样方法为在获得村级(或居委会)单位的人口学资料后,按年龄性别分为12层,每层按单纯随机抽样的原则抽取样本4人,并按年龄从低到高顺序进行编码(P1、P2、P3、P4),进行问卷调查及生物样本采集。抽取调查对象后,再随机抽取两组人群作为第一和第二预备置换对象。每组预备置换对象的每一层4人同样按照年龄从低到高顺序进行编码(C1、C2、C3、C4;或D1、D2、D3、D4),共抽取预备置换对象96人。若需置换调查对象,则用对应编码的第一预备置换对象进行置换(P2→C2),若置换对象也不符合项目要求,则按位置就近原则选择相邻预备置换对象(C1或C3,图1);若仍不符合,则在第二组预备置换对象中以同样的方法置换;通过以上置换若仍不能满足调查样本量要求,则在产生样本的同组人群中随机抽取。

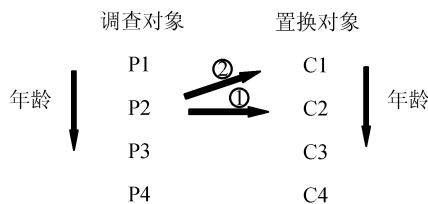


图1 调查对象置换方法

2016年已完成现场调查的66个监测点有9 365名调查对象,发生置换的共3 805名,应答率为59.4%,其中不同年龄段的应答率差异有统计学意义(P<0.05),男性应答率(60.5%)高于女性(58.2%);农村(65.1%)高于城市(55.0%)(表2)。

4. 抽样误差:抽样误差是由于随机抽样造成的

表2 不同特征人群置换情况比较

特征	未置换(%)	置换(%)	合计	χ^2 值
年龄组(岁)				<0.05
3~	998(63.9)	563(36.1)	1 561	
6~	947(61.0)	605(39.0)	1 552	
12~	861(55.6)	688(44.4)	1 549	
19~	832(53.1)	735(46.9)	1 567	
40~	936(59.7)	632(40.3)	1 568	
60~79	986(62.9)	582(37.1)	1 568	
性别				<0.05
男	2 840(60.5)	1 853(39.5)	4 693	
女	2 720(58.2)	1 952(41.8)	4 672	
城乡				<0.05
城市	2 932(55.0)	2 398(45.0)	5 330	
农村	2 628(65.1)	1 407(34.9)	4 035	
合计	5 560(59.4)	3 805(40.6)	9 365	

样本统计量与总体参数之间的差异,只要采用抽样调查,抽样误差就不可避免,估计量方差及标准差都是抽样误差的表现形式。抽样误差越小,说明样本统计量对总体参数进行估计值的精度越高。本监测项目采用多阶段分层抽样方法,其抽样设计相对复杂,产生的样本具有复杂样本特征,各阶段抽样数据不独立,因此不能采用简单随机抽样的误差估计公式计算抽样误差。考虑到该监测各阶段抽样采用的不等概率抽样,因此采用多阶段不等概率抽样的误差估计。

基本原理^[6]:设总体拥有 N 个初级单元,每个初级单元拥有 M_i 个二级单元,每个二级单元又拥有 K_{ij} 个三级单元,各阶段样本量分别为 n, m, k ,即抽样 n 个初级单元,在每个样本初级单元中,抽取 m 个二级单元,在每个样本二级单元中,抽取 k 个三级单元。

三阶段抽样每一阶段单元被抽中的概率为 Z_i, Z_{ij}, Z_{iju} ,且满足:

$$\sum_{i=1}^N Z_i = 1, \sum_{j=1}^{M_i} Z_{ij} = 1, \sum_{u=1}^{K_{ij}} Z_{iju} = 1;$$

其中, $i=1, 2, \dots, N; j=1, 2, \dots, M_i; u=1, 2, \dots, K_{ij}$; 记 Y_{iju} 为总体单元指标值, y_{iju} 为样本单元观测值,则总体总量表示为:

$$Y = \sum_{i=1}^N \sum_{j=1}^{M_i} \sum_{u=1}^{K_{ij}} Y_{iju}$$

对总体总量的无偏估计为:

$$\hat{Y} = \frac{1}{nmk} \sum_{i=1}^n \frac{1}{Z_i} \sum_{j=1}^m \frac{1}{Z_{ij}} \sum_{u=1}^k \frac{y_{iju}}{Z_{iju}}$$

其方差为:

$$V(\hat{Y}) = \frac{1}{n} \left(\sum_{i=1}^N \frac{Y_i^2}{Z_i} - Y^2 \right) + \frac{1}{nm} \sum_{i=1}^N \frac{1}{Z_i} \left(\sum_{j=1}^{M_i} \frac{Y_{ij}^2}{Z_{ij}} - Y_i^2 \right) + \frac{1}{nmk} \sum_{i=1}^N \frac{1}{Z_i} \left[\sum_{j=1}^{M_i} \frac{1}{Z_{ij}} \left(\sum_{u=1}^{K_{ij}} \frac{Y_{iju}^2}{Z_{iju}} - Y_{ij}^2 \right) \right]$$

式中, $Y_{ij} = \sum_{u=1}^{K_{ij}} Y_{iju}, Y_i = \sum_{j=1}^{M_i} \sum_{u=1}^{K_{ij}} Y_{iju}$

$V(\hat{Y})$ 的一个无偏估计为:

$$v(\hat{Y}) = \frac{1}{n(n-1)} \sum_{i=1}^n (\hat{Y}_i - \hat{Y})^2 = \frac{1}{n(n-1)} \left(\sum_{i=1}^n \hat{Y}_i^2 - n \hat{Y}^2 \right)$$

式中, $\hat{Y}_i = \frac{1}{mz_i} \sum_{j=1}^m \frac{1}{z_{ij}} \left(\frac{1}{k} \sum_{u=1}^k \frac{y_{iju}}{z_{iju}} \right) (i=1, 2, \dots, n)$

从公式可以看出,对于三阶段抽样,方差的主要项为第一阶段,因此为了提高抽样精度,降低方差,应尽可能增加第一阶段样本量。但实际应用中还要考虑实施调查的费用问题,一般来说,第一阶段抽取的单元较多,会导致样本分布较散,实施调查的成本增加,因此实施多阶段抽样时要权衡调查成本和估计效率这两个要素。

实例分析

以本监测项目中贵州省抽样数据为例进行实证分析。2014年贵州省共有34 748 556人,根据第一阶段抽样方法,确定贵州省内监测点个数为:

$$\text{贵州省监测点个数} = \frac{\sqrt{\text{贵州省人口数}}}{\sum \sqrt{\text{各省人口数}/152}} = 5(\text{个})$$

根据国家统计局数据,截止2014年,贵州省共有88个区(县、地级市),将88个区(县、地级市)以城镇化率和第二产业就业率为变量进行样本聚类,并抽取5个区(县)作为监测点,分别为水城县、花溪区、桐梓县、玉屏县及汇川区。根据抽样方法,在已抽取的各监测点内抽取调查单元及调查对象,并根据各阶段单元数及样本量,计算出各阶段抽样比(表3)。

$$\hat{Y} = \frac{1}{nmk} \sum_{i=1}^n \frac{1}{Z_i} \sum_{j=1}^m \frac{1}{Z_{ij}} \sum_{u=1}^k \frac{y_{iju}}{Z_{iju}} = \frac{1}{5 \times 3 \times 48} \left(88 \times \frac{909\,000}{7\,585} \times 4\,696 \times 48 + 88 \times \frac{909\,000}{5\,820} \times 3\,796 \times 48 + \dots + 88 \times \frac{506\,485}{3\,673} \times 4\,121 \times 48 \right) = 33\,124\,398(\text{人})$$

其中:

$$\hat{Y}_i = \frac{1}{mz_i} \sum_{j=1}^m \frac{1}{z_{ij}} \left(\frac{1}{k} \sum_{u=1}^k \frac{y_{iju}}{z_{iju}} \right) = \frac{88}{3} \times \frac{909\,000}{7\,586} \times \left[\frac{1}{48} \times 4\,696 \times 48 + \dots + \frac{909\,000}{2\,847} \times \left(\frac{1}{48} \times 1\,403 \times 48 \right) \right] = 47\,037\,114(\text{人})$$

表 3 贵州省点抽样各阶段抽样比

监测点		调查单元		调查对象			
名称	Z _i	名称	Z _{ij}	抽样比	Z _{ijk}	抽样比	
水城县	Z ₁	1/88	俄嘎村	Z ₁₁	7 586/909 000	Z _{111,112...1 148}	1/4 696
			水库村	Z ₁₂	5 820/909 000	Z _{121,122...1 248}	1/3 796
			花地村	Z ₁₃	2 847/909 000	Z _{131,132...1 348}	1/1 403
花溪区	Z ₂	1/88	黄河社区	Z ₂₁	11 890/272 137	Z _{211,212...2 148}	1/11 236
			濛江社区	Z ₂₂	4 637/272 137	Z _{221,222...2 248}	1/4 418
			金农社区	Z ₂₃	2 590/272 137	Z _{231,232...2 348}	1/2 590
桐梓县	Z ₃	1/88	太白社区	Z ₃₁	14 723/566 759	Z _{311,312...3 148}	1/14 391
			中山村	Z ₃₂	2 718/566 759	Z _{321,322...3 248}	1/1 230
			八一村	Z ₃₃	2 543/566 759	Z _{331,332...3 348}	1/1 947
玉屏县	Z ₄	1/88	街上社区	Z ₄₁	2 480/119 810	Z _{411,412...4 148}	1/2 701
			新店村	Z ₄₂	2 183/119 810	Z _{421,422...4 248}	1/2 149
			大屯村	Z ₄₃	2 211/119 810	Z _{431,432...4 348}	1/2 211
汇川区	Z ₅	1/88	乌江社区	Z ₅₁	3 878/506 485	Z _{511,512...5 148}	1/1 034
			新舟路社区	Z ₅₂	5 812/506 485	Z _{521,522...5 248}	1/6 503
			三联村	Z ₅₃	3 673/506 485	Z _{531,532...5 348}	1/4 121

依次类推,计算: $\hat{Y}_2 = 23\ 131\ 962$; $\hat{Y}_3 = 36\ 502\ 031$;

对样本总量的无偏估计为:

$$\hat{Y}_4 = 20\ 801\ 724; \hat{Y}_5 = 48\ 149\ 157;$$

由此计算方差为:

$$v(\hat{Y}) = \frac{1}{n(n-1)} \sum_{i=1}^n (\hat{Y}_i - \hat{Y})^2 = \frac{1}{5(5-1)} \\ (47\ 037\ 114 - 33\ 124\ 398)^2 + \dots + (48\ 149\ 157 - \\ 33\ 124\ 398)^2 = 51\ 443\ 300\ 578\ 832(\text{人})$$

则抽样的标准误差为: $\hat{\sigma}_{\bar{Y}} = \frac{\sqrt{v(\hat{Y})}}{\sqrt{n}} = 3\ 207\ 594(\text{人})$

$$\text{变异系数}(CV) = \hat{\sigma}_{\bar{Y}} / \hat{Y} = \frac{3\ 207\ 594}{33\ 124\ 398} = 0.097$$

根据方法学部分计算得出,国家人体生物监测项目中贵州省抽样标准误差为 3 207 594 人,意味着从估计的总人数与实际人数平均相差 3 207 594 人,抽样误差的 CV 为 0.097。抽样阶段置换率分别为 5.26%, 6.35% 及 40.6%。

讨 论

多阶段抽样方法具有抽样框获取方便、提高抽样误差等优点,因此目前世界各国生物监测项目均采用多阶段抽样方法抽取调查对象^[4-5]。本项目采用三阶段抽样方法,并在每阶段内采用聚类、分层,PPS 等抽样方法,增加层间差异,减小层内差异,减少抽样误差,使抽取的调查对象具有全国代表性。

应答率是指调查中应答单元占全部样本单元的比重,它是评价调查执行优劣的重要指标,调查人群的应答率低则影响调查的随机性与代表性。随着经

济的发展和信息交流增长,人口流动性增加及人群社会安全意识的增强,导致公共卫生调查的应答率越来越低,20 世纪 80 年代,健康检查调查(Health Examination Surveys, HESs)的应答率可达 70% ~ 80%,而目前各国健康调查人群应答率均普遍降低为 40% ~ 60%^[7-8],如芬兰健康调查项目 25 ~ 64 岁男性人群应答率由 1978 年的 84% 降至 2002 年的 59%; 2007—2012 年欧洲健康调查项目中荷兰、德国、意大利及芬兰人群应答率分别为 45%、54%、55% 及 63%^[9]。2011 年韩国生物监测问卷调查及体格检查应答率为 76.1%, 2015—2016 年美国 NHANES 体格检查应答率为 58.7%^[10]。本项目 2016 年第三阶段人群总应答率为 59.4%(5 560/9 365),城市应答率低于农村,原因可能为城市人口流动性高、城市居民因社会安全意识较强导致配合度较差。调查人群应答率低影响项目调查的随机性及代表性,且易造成抽样误差的增大,其已成为公共卫生调查的一个不容忽视的问题。

随着复杂抽样技术尤其是多阶段抽样在大型流行病学现况调查中不断推广应用,针对其复杂样本的抽样误差也日益受到关注。胡桂华等^[11]通过多阶段不等概率抽样误差计算方法,对 2015 年全国人口普查采用的二阶段随机抽样方法进行验证,结果表明抽样标准误差为 118 976 人,变异系数为 0.098,二阶段抽样方法适合于人口数估计。李镒冲等^[1]利用 2010 年中国慢性病监测数据,通过多阶段模拟抽样方法,对不同抽样方法的误差估计进行比较,结果显示初级单元(PSU)入样比对误差影响较大。联合国人口调查咨询专家认为,抽样调查估计值的相对误

差控制在10%以内,数据质量好;控制在15%以内,数据质量较好,控制在20%以内,数据可用。一些国家在进行国家级抽样调查时规定估计值的相对误差超过30%不能公布^[12]。本研究以贵州省为例计算多阶段不等概率抽样的抽样误差,结果表明抽样误差为3 207 594人, CV=0.097,贵州省抽样误差较小,精度较高,但该计算方法未考虑人群缺失、置换等非抽样误差的权重调整,大型公共卫生监测项目多阶段抽样误差计算方法有待继续研究。

志谢 感谢中国人民大学统计学院金勇进教授、山西医科大学公共卫生学院王彤教授及中国CDC信息统计中心于石成教授对抽样设计的指导和帮助

利益冲突 无

参 考 文 献

[1] 李镡冲,赵寅君,王丽敏,等. 考虑多阶段抽样设计的误差估计[J]. 中华流行病学杂志, 2016, 37(3): 425-429. DOI: 10.3760/cma.j.issn.0254-6450.2016.03.028.
Li YC, Zhao YJ, Wang LM, et al. Variance estimation considering multistage sampling design in multistage complex sample analysis [J]. Chin J Epidemiol, 2016, 37(3): 425-429. DOI: 10.3760/cma.j.issn.0254-6450.2016.03.028.

[2] Heeringa SG, West BT, Berglund PA. Applied survey data analysis[M]. Boca Raton, FL: CRC Press, 2010.

[3] Schulz C, Conrad A, Becker K, et al. Twenty years of the German Environmental Survey (GerES): Human biomonitoring-Temporal and spatial (West Germany/East Germany) differences in population exposure [J]. Int J Hyg Environ Health, 2007, 210(3): 271-297. DOI: 10.1016/j.ijheh.2007.01.034.

[4] Johnson CL, Dohrmann SM, Burt VL, et al. National health and nutrition examination survey: sample design [R]. National Center for Health Statistics, Vital Health Stat, 2014, 2(162).

[5] Government of Canada. Third report on human biomonitoring of

environmental chemicals in Canada: results of the Canadian health measures survey cycle 3 (2012-2013) [R]. Ottawa: Minister of Health, 2015.

[6] 金勇进. 抽样:理论与应用[M]. 2版.北京:高等教育出版社, 2016.
Jin YJ. Sampling: Theory and Application [M]. 2nd ed. Beijing: Higher Education Press, 2016.

[7] Atriotic BK, Bates N, Burt G, et al. Nonresponse in U.S. government household surveys: consistent measures, recent trends, and new insights [J]. J Off Stat, 2001, 17(2): 209-226.

[8] Tolonen H, Helakorpi S, Talala K, et al. 25-year trends and socio-demographic differences in response rates: Finnish adult health behaviour survey [J]. Eur J Epidemiol, 2006, 21(6): 409-415. DOI: 10.1007/s10654-006-9019-8.

[9] Mindell JS, Giampaoli S, Goesswald A, et al. Sample selection, recruitment and Participation rates in health examination surveys in Europe-experience from seven national surveys [J]. BMC Med Res Methodol, 2015, 15(1): 78. DOI: 10.1186/s12874-015-0072-4.

[10] National Health and Nutrition Examination Survey. Unweighted Response Rates for NHANES 2015-2016 by Age and Gender [R]. National Center for Health Statistics, Vital Health Stat, 2017.

[11] 胡桂华,孙晓宇,康颖,等. 人口抽样调查方案设计研究[J]. 徐州工程学院学报:自然科学版, 2017, 32(4): 24-29. DOI: 10.3969/j.issn.1674-3571.2017.04.005.
Hu GH, Sun XY, Kang Y, et al. Study on the design of population sampling survey scheme [J]. J Xuzhou Inst Technol: Nat Sci Ed, 2017, 32(4): 24-29. DOI: 10.3969/j.issn.1674-3571.2017.04.005.

[12] 胡英. 抽样调查方法应用中的问题[J]. 中国统计, 2003, (7): 5-6.
Hu Y. Problems in the application of sampling survey methods [J]. China Stat, 2003, (7): 5-6.

(收稿日期:2018-07-26)

(本文编辑:万玉立)

中华预防医学会流行病学分会第七届委员会名单

(按姓氏笔画排序)

主任委员	李立明										
副主任委员	刘天锡	杨维中	吴凡	何耀	汪华	胡永华	姜庆五	詹思延			
常务委员	王岚	叶冬青	余宏杰	汪宁	沈洪兵	陆林	陈坤	周晓农	赵根明	段广才	
	贺雄	唐金陵	曹务春	崔萱林							
委员	于雅琴	么鸿雁	王岚	王蓓	王开利	王文瑞	王定明	王素萍	王效俊	仇小强	
	叶冬青	冯子健	毕振强	吕筠	庄贵华	刘天锡	刘殿武	闫永平	许汴利	严延生	
	杜建伟	李丽	李琦	李凡卡	李申龙	李立明	李亚斐	李俊华	李增德	杨维中	
	吴凡	吴先萍	邱洪斌	何耀	何剑峰	余宏杰	汪宁	汪华	沈洪兵	张晋	
	张颖	陆林	陈坤	陈可欣	陈维清	岳建宁	周宝森	周晓农	单广良	孟蕾	
	项永兵	赵亚双	赵根明	胡东生	胡代玉	胡永华	胡志斌	胡国良	段广才	俞敏	
	施榕	施国庆	姜晶	姜庆五	贺雄	贾崇奇	夏洪波	栾荣生	唐金陵	曹广文	
	曹务春	崔萱林	董柏青	程锦泉	詹思延	蔡琳	戴江红	魏文强			
秘书长	王岚										
副秘书长	吕筠										

