

从“主要数据公报”看 第五次人口普查存在的问题

乔晓春

【提要】 本文通过对全国和各省人口普查第一、二号公报的数据和文字进行解剖和分析,提出人口普查存在的一些问题。本文认为第五次全国人口普查误差较大是有其客观原因的,但是公报并没有把普查客观上存在的问题反映出来,相反,从公布的指标和文字表述上却反映出了其他一些问题。单从公布的数据上看,数据的使用者对普查数据所反映的一些问题很难理解和解释清楚,需要人们进一步的讨论和有关部门的进一步澄清。然而,数据质量不高或涵义的不清会给普查资料的分析带来困难,处理不好会得出错误的结论,影响国家和地方政府的决策。

【作者】 乔晓春 中国人民大学人口研究所,教授。

中国第五次人口普查(以下简称“五普”)是一个十分宝贵的数据资源,开发和利用好这一资源将对制定国民经济和社会发展规划、促进和推动 21 世纪初中国社会经济的全面发展,产生积极的和巨大的作用。然而,在分析和使用数据之前,使用者有必要对数据的质量有一个一般性的认识。本文的目的是通过对人口普查主要数据的分析,使人们对普查数据存在的问题有初步的了解,从而能对人口普查数据所反映出的问题有一个客观的、科学的和准确的表述。

一、问题的提出

在人口普查正式开始登记以前,笔者曾根据国家统计局公布的历年年末中国大陆地区人口总数及年内出生率和死亡率推断,第五次人口普查得到的“比较合理”的中国大陆地区 2000 年 11 月 1 日人口总数应该为 12.70 亿人;由于国家公布的 90 年代历年人口出生率是在人口变动抽样调查直接得到的出生率的基础上经过最大可能的“上调”而得到的(贾同金,1995),从而导致公布的全国总人口比由实际调查推断出的总人口每年平均多了 200 多万人。根据历年人口变动抽样调查直接得到的数据估计,“五普”直接登记的人口应该在 12.46 亿左右;另外,根据各方面的信息来源发现,由于种种原因 90 年代中国人口登记的漏登率确实很高,为此笔者判断“五普”可能得到的人口总量上限为 13.00 亿人。笔者当时的结论^①是“五普”最能让人接受的总人口数值是在 12.70 亿左右;可以接受的人口总数应该在 12.46 亿~13.00 亿之间;由于现实情况只能导致人口的漏报率大于重报率,所以在可接受的区间内,普查得到的人口越多,说明普查质量越高;普查得到的人口越少,说

^① 笔者曾于 2000 年 9 月在美国北卡罗莱那人口中心所作的题为“中国人口数据质量与 2000 年人口普查”的学术报告中,对 2000 年中国人口普查可能得到的结果进行了估计。这里的判断和结论都是在当时的报告中所给出的。需要说明的是,这只是笔者个人的判断和评判标准。不同的人可能对人口普查数据质量的评判标准是不一样的。

明普查质量越差。超出 12.46~13.00 亿的范围,说明普查得到的人口总量数据是不可接受的。

2001 年 3 月 28 日国家统计局在“2000 年第五次全国人口普查主要数据公报(第一号)”中公布,全国总人口为 129 533 万人。其中祖国大陆 31 个省、自治区、直辖市和现役军人的人口共 126 583 万人^①。显然,这一数据与“最能让人接受的”结果很接近,或者说是一个比较“理想”的结果。

然而,在“2000 年第五次全国人口普查主要数据公报(第一号)”末尾的注中却看到了这样的说明:“普查登记结束后,全国统一抽取 602 个调查小区进行了登记质量的抽样调查。调查结果,人口漏登率为 1.81%。祖国大陆 31 个省、自治区、直辖市的总人口中已包括了据此计算的漏登人口。”这句话似乎告诉了我们三件事情:(1)人口漏登的程度;(2)漏登率是从 602 个调查小区的事后质量抽样调查直接得到的;(3)公布的祖国大陆总人口已经是修正后的结果。

我们也注意到,在 1982 年第三次人口普查和 1990 年第四次人口普查都曾在“主要数据公报”中公布过的几个重要数据,在“五普”公报中并没有出现,比如:按户口登记状况分的人口数和各部分所占的比例;普查前一年的出生人口和死亡人口,以及相应的出生率和死亡率;各类普查登记质量的抽样检验结果,其中包括人口数的重登率、漏登率、净差率、性别误差率、年龄误差率、出生人口漏报率和死亡人口漏报率。没有在“公报”中公布这些数据,自然有其原因。但肯定与普查数据的质量有关。

1982 年和 1990 年全国人口普查各省、自治区、直辖市所发表的“主要数据公报”在格式上是统一的,所公布的指标也是一致的,甚至文字表述上都十分接近。然而,在“五普”各省、自治区、直辖市的“主要数据公报”中,有的省公布了出生率和死亡率的数值,而有的省却没有公布;有的省在公布直接调查到的人口数的同时,也公布了与国家“衔接”后的人口数;有的省公布了省内各市地的直接调查人口,有的省则只公布全省总人口而未公布省内各市地的人口;有的省公布了漏登率,有的省没有公布。这些都是“五普”本身所存在问题的反映。

本文将针对上述问题进行分析。事实上,这里提出的一些问题笔者并不能都给出答案,只能提出问题,但是从学术的角度我们还是有必要对其进行分析和讨论。

二、关于“五普”存在的问题

(一)“五普”的漏登率是低了还是高了

国家统计局根据“国际上曾有过共识”^②,认为普查登记误差在 2% 以下,数据是可信的^③。孙兢新(2001)指出,据联合国统计局资料,在 1985~1994 年间举行了人口普查事后质量抽查的 59 个国家中,误差率在 2% 以下的有 32 个,误差率在 2%~5% 的有 11 个,误差率在 5% 以上的有 16 个。从中国人口总量的误差率看,尽管比前几次人口普查增大了,但仍然在 2% 以下,这样的误差率属于普查质量较好的范围之内,普查数据是准确、可信的。在世界各国的人口普查中,中国的误差率是较低的。

仅仅从普查公布的误差率数值上看,与中国以往历次人口普查相比,“五普”的登记误差是最高

① 与以往普查公布的总人口数值到“个位”不同的是,“五普”公布的总人口只给到“万位”。

② 到目前为止,笔者并没有看到统计部门所提供的这种国际“共识”的证据,或者对这种“共识”的具体的和确切的表述。然而,即使是一个国际性机构或某“权威性”机构的认识,也未必可以成为国际上的共识。

③ 新华社 2000 年 3 月 28 日题为“朱之鑫说,第五次人口普查数据是可信的”一文中报道:朱之鑫在回答有关人口普查的漏登率问题时说,……国际公认漏登率低于 2% 的是非常可信的,2% 到 4% 是可以接受的,而超过 5% 的数据在使用上就要大打折扣。因此,第五次人口普查的数据是可信的(<http://www.cnmaya.com>)。

的。1953年人口普查登记的净漏报率为0.116%；1964年为0.001%；1982年为0.071%；尽管1990年人口普查也曾遇到过与“五普”同样的困难和问题，甚至事后质量抽查的登记对象同“五普”一样，也包括了“普查时不是本调查小区的登记对象，但人住本调查小区”的人口，但是普查事后质量抽查所得到的净漏报率也只有0.060%，而且组织这项工作的人也认为1990年中国人口普查的登记质量确实是高的，误差确实很小（杨日章，1993）。然而，“五普”人口的净漏登率竟是1990年人口普查的30倍。更令人难以置信的是，“人口普查质量控制和验收工作细则”在普查之前就已规定人口数净差错率要小于4%。同时还要求，“如差错率大于规定的要求，该普查区的普查表要全面返工重做，……如仍不合格，再返工重做，直到达到指标要求为止”。然而，按照这一要求，即使做的最不好，普查的净漏报率也只能达到0.4%，很难想象“五普”的最终漏报率会高达1.81%。果真如此的话，则说明全国绝大多数地区未能按照普查工作细则所规定的去做。这会使人怀疑普查的其他细则所做出的规定是否也都得到了遵守。

从数值上看，这样高的误差似乎很难让人相信和接受，但是，如果人们了解了该误差是怎样得来的，也就可以理解这样高的漏报率有其“一定道理”。

武洁（2002）在对人口普查事后质量抽查和误差计算办法时说，普查对总人口登记质量的估计采用常住人口和暂住人口相结合的方法，就全国范围来讲，无论是常住人口，还是暂住人口，都是普查要登记的对象，因此，事后质量抽查实际调查的是现有人口。崔红艳（2002）也提到抽查时，对前一天在该调查小区过夜的所有人都进行登记。然后分常住人口和暂住人口。将常住人口与原普查表登记的人口一一匹配；对暂住人口，按照本人申报的11月1日登记地或常住地归类后，由国家统一组织对其常住地的普查表进行逐人查询，检查普查表上是否已经登记了这个人。最后，将常住人口和暂住人口的漏报情况合并，计算出漏登（率）。从这里可以看出，“五普”事后质量抽查的核查对象和漏报率的统计比1982年普查有了“实质性”的改进，但却与1990年普查所用的办法相近，即将以往只检验常住人口的登记质量扩大为将常住人口和暂住人口的登记质量一起进行检验。当然，“五普”对“暂住人口”的核查会比1990年普查更容易、也更有效，原因是“五普”要求各普查区不仅要登记离开户口所在地半年以上的“常住人口”，也登记了离开户口所在地不到半年的“暂住人口”。从“五普”事后质量抽查结果看，在漏报的总人口中，有近80%的人口是由于暂住人口漏报造成的（武洁，2002）。暂住人口漏报率比1990年普查有大幅度增加，一方面说明流动人口规模的大量增加和流动人口漏报比较严重，另一方面也说明对暂住人口的核查效果更高了。如果“五普”事后质量抽查，按照1982年普查的办法，只核查常住人口登记误差，不检验暂住人口登记误差，那就将有占近80%的暂住人口漏报被忽略掉，其人口漏报率很可能在0.4%左右。为此可以说，正是由于“五普”将暂住人口作为调查对象并引入检验对象，才使漏报率得以大幅度上升。当然，客观原因带来的登记困难，也是造成人口漏登率升高的重要原因之一。我们现在的的问题是，1990年普查也同样将暂住人口漏报作为事后质量抽查的对象，为什么就没有得到如此高的漏报率？如果把“五普”漏报率高，仅仅解释为人们不愿意或回避登记，似乎证据不足，因为“五普”所遇到的困难和问题1990年普查时同样遇到过。

（二）将暂住人口的漏登率引进总漏登率是否合适

从理论上来说，人口普查事后质量抽查对象仅为常住人口，并以此作为检验全国人口登记误差的办法并没有错。因为普查区内的常住人口既包括了人户一致的人口和人在本地、户口在外地并已离开户口所在地半年以上的外来人口，同时也包括了户口在本地而且离开本地不到半年的外出人口。后者既是本地的常住人口（在本地的常住人口表上登记），又是外地的“暂住人口”（在外地的暂住人口表上登记）。因此，在对登记质量进行核查时，检验本地常住人口实际上已经包括了“户口在本地而且离开本地不到半年的外出人口”登记误差的检验，而这部分人口同时又是外地的“暂住人

口”。然而，“五普”的事后质量抽查则是将本地的常住人口和暂住人口的漏报情况合并，那么从理论上讲，就存在着重复计算暂住人口漏报率的情况。暂住人口的漏报率越高，重复计算的比例也会越大，总漏报率同时也会越高。

当然，实际情况不会像理论上说的那样严重。原因是从常住地角度去核查“户口在本地而且离开本地不到半年的外出人口”是很困难的，因为这部分人根本就不在普查登记地，很难准确地去核对他们是否有重登或漏登，特别是对那些全户外出的家庭。相反，从现住地更容易找到当事人，并且可以直接了解他们的普查登记情况，与他们的户口所在地进行核对。另外，由于事后质量检验是一种抽样调查，那么“户口在本地而且离开本地不到半年的外出人口”与“暂住人口”的交叉就不会很大。尽管如此，利用这种方法去核对普查登记质量，在理论上是有问题的，而且肯定会存在一定程度的重复计算漏报率的情况。

我们相信，在中国现有的情况下，出生人口漏报和流动人口漏报都是比较普遍的。但是利用普查事后质量抽查在全国范围内得到这样高的漏登率也是相当不容易的。因为普查时人们避免或回避登记应该登记人口的理由，在下一轮登记时同样存在。

（三）“五普”到底登记了多少人

严格地说，无论普查还是抽样调查，公布总量数据时，都应该是将直接调查或登记到的数据和相应的误差率分别公布，而不是只公布修正后的总量数据。但“五普”却破了例，所公布的祖国大陆总人口是经过修正的数据。然而我们很想知道到底“五普”漏登了多少人，以及直接登记到的人口有多少。国家统计局将事后质量抽查得到的 1.81% 的漏登率，“作为这次普查总人口的漏登率，并对普查登记的人口进行调整，即国家统计局发表的普查总人口 126 583 万人中，已经包括了漏登的人口 2 246 万（登记人口 124 337 万，其中解放军 250 万）”^①。这说明，“五普”实际登记到的祖国大陆 31 个省、自治区、直辖市和现役军人的人口应该为 124 337 万人。这意味着，按照国际惯例或按照 1982 和 1990 年人口普查公布总人口的办法，“五普”直接调查到的或应该公布的全国 31 个省、自治区、直辖市和现役军人总人口为 124 337 万人。

我们还发现漏登的 2 246 万人，是用登记的 124 337 万人乘以漏登率计算出来的。那么，这里隐含的关系应该是：人口漏登率 = 漏登人口 / 实际登记人口。

在这里，作为分子的“漏登人口”并不包含在分母的“实际登记人口”中，那么这一指标是“比”，而不是“比例”（查瑞传，1991）。事实上，人口漏登率应该是“比例”的概念，即人口登记率 + 人口漏登率 = 1^②（假定没有其他误差，或漏登率为净误差率）。如果是这样的话，人口漏登率应该等于漏登人口除以修正后的总人口（等于实际登记人口加上漏登人口）。利用这一方法，修正的总人口应该这样计算：

修正总人口 = 实际登记人口 / (1 - 漏登率) = 124 337 / (1 - 0.0181) = 126 629 万人。

（四）为什么“五普”只有人口“漏登”而没有人口“重登”

任何人口调查都应该既有人口的漏登，又有人口的重登；重漏相抵以后，才可以得到人口的净误差率（或是净漏报，或是净重报）。但是，“五普”公报却只给出了人口的漏登率。问题是：人口普查登记不可能没有重报，那么重报率是多少？如果普查公报公布的“人口漏登率”不是人口的净漏报

① 张为民、徐钢：《2000 年人口普查实施效果和提出的挑战》，提交“第八次全国人口科学大会”论文（未刊稿），2002 年 6 月。

② Judith T. Lessler and William D. Kalsbeek 合作的 Nonsampling Error in Surveys（《调查中的非抽样误差》）一书是这样定义调查回答率的：回答率 = 对报告单位完成的访问数 / 样本中合格的报告单位数；该书指出，无回答率是“……在样本中（无回答的）个体所占的比例……与回答率互补。”

率^①,那么用此数据来修正普查总人口是错误的;如果这一数据是净漏报率,那么漏报率又是多少?而且漏报率肯定大于1.81%的净漏报率^②。

1990年普查公报公布对普查登记质量抽查结果是人口数重登率为0.1%,漏登率为0.7%,重漏相抵,人口数净差率为0.6%;性别误差率为0.14%;年龄误差率为3.07%;出生人口漏报率为1.03%;死亡人口漏报率为4.9%。“五普”在“各阶段质量控制和验收工作细则”中对普查登记阶段的各类误差率提出了要求,并给出了各指标的合格标准,比如,人口数净差错率要小于4%;人口数(毛)差错率小于5%;出生人口差错率小于5%;死亡人口差错率小于5%;性别项差错率要小于4%;年龄项差错率要小于10%(国务院人口普查办公室,2000)。遗憾的是,普查公报没有公布这些数据,我们也无从了解到“五普”相应指标的误差率到底是多少。这样,在分析人口的性别年龄结构,以及出生和死亡人口时,我们对这些数据的质量是不了解的。

(五) 分配给各省不同规模“衔接”人口的依据是什么

由于全国公布的祖国大陆总人口已经是加入了漏报人口后的结果,各省直接登记到的人口之和并不等于全国公布人口,因此,必须将估计出的漏登人口分解到各省区去,以使各省公布人口之和尽可能地与全国人口一致。

从表1给出了公布的全国总人口,以及公布的各省、自治区、直辖市和直接登记的总人口之差,可以看出:“五普”全国共“衔接”了2 072万人,其中分配给省里1 967万人,未能分配下去的105万人。我们发现,根据这一结果计算的人口漏登率为1.64%,而不是1.81%。另外,从表1中还可以看出,分配给各省的“衔接”人口,以及各省的“漏登率”存在着相当大的差异。分配给江苏和河南省的人口都为134万,尽管二者都属于人口大省,但是从人口流动的角度看,江苏属于人口的净迁入省,河南则属于人口净迁出省。江西和广西都不属于人口大省,但它们被“衔接”的人口也在100万。从各省的“漏登率”上也很难发现规律。青海的“漏登率”最高,为6.95%。如果人口漏登主要是由于出生漏报和流动人口漏报的话,那么青海在这两方面的问题都不突出;其次是新疆,为4.10%;再往后依次是海南(3.94%)、江西(2.42%)、广西(2.32%)和宁夏(2.31%)。另一个奇怪的现象是,西藏和贵州的人口漏登率均为0。既然在解释全国存在人口漏登率是“不可避免”的,那么为什么这两个地区却“避免”了人口的漏报?

(六) 为什么很多省不愿意接受国家分配的“衔接”人口

“五普”事后质量抽查是自下而上进行的,因此在国家获得了全国的人口漏报误差的同时,各省也应该有相应的人口漏报误差。然而,事实是各省、自治区、直辖市均按国家要求不公布本地区事后质量抽查得到的误差率,而是接受了国家根据1.81%漏报率分配给各地的“衔接”数字,并依据分配来的数据对各省直接调查的总人口进行了上调。可能分配下来的“衔接”数据与各地掌握的漏报率有一定的出入^③,一些省的普查公报则流露出了他们不情愿的情绪。云南省的普查公报写道,全省普查时点实际登记的总人口数为4 235.9万人,按照1.81%的漏登率增加各省总人口,这样省普查时点总人口为4 287.9万人。宁夏回族自治区普查公报指出,全区总人口包括,直接普查登记的人口548.55万人;根据全国数字衔接的需要及漏登率估算的13万人。由于各省对衔接的态度不

① 至少在普查公报中没有进行说明。所以很容易让人理解为就是漏登率,而不是净漏登率。

② 因为重登率不可能等于零。

③ 比如宁夏回族自治区在普查公报中公布了自己得到的漏报率。公报写道,普查登记结束后,全区统一抽取10个调查小区进行了登记质量的抽样调查。抽查结果,人口漏登率为1.24%。而按国家分配的数据,其误差率为2.31%。据了解,北京市共抽取了18个调查小区(国家点4个)进行了登记质量的抽样调查,得到的漏报率为0.6%,而国家分配的漏报率为1.81%。

同,所以发表公报的内容和形式也不尽相同。为了回避省内各市(地)直接登记人口之和与国家分配给省里的总量不一致的问题,到目前为止,一些省仍不公布市(地)人口数,如吉林、黑龙江、江苏、浙江、安徽、江西、湖南和四川省等。为了不出现数据不一致的情况,一些省干脆把误差再分解到各市(地),并最终公布了与省级数据相一致的分市(地)总人口,如天津、内蒙古、陕西、青海等省(区)。而多数省在公布国家分配的省总人口的同时,未将国家分配的误差数分解到各市(地),而是公布了省内各市(地)直接登记到的总人口。并对数据不一致的情况做了“说明”。

当各省、自治区、直辖市作为任务接受了国家“衔接”的人口后,一些省并没有将普查公布的人口作为估计该省 2000 年底和 2001 年底的“基数”。相反,他们则把直接登记的人口作为“基数”来推断未来的人口,从而导致这些省 2000 年和 2001 年年末人口仍然少于普查公布的 2000 年 11 月 1 日的人口。比如,辽宁省 2000 年 11 月 1 日人口普查公布的总人口为 4 238 万;公布的 2000 年底人口为 4 184 万(直接登记的 4 182 万+后两个月的自然增长 2 万人);公布的 2001 年年底总人口为 4 194 万人(根据 2001 年人口变动抽样调查推

断),仍少于普查公布的 2000 年 11 月 1 日的 4 238 万人。由于这些省的人口增长率目前仍处于正增长阶段,所以 2001 年人口数少于 2000 年人口数,是不正常的。表 2 列出了这样一些省。很明显,除了广东省以外,所有这些省 2001 年年末人口都大于人口普查直接调查的人口,而小于普查公布的人口。广东省所公布的 2001 年年末人口数甚至低于普查登记到的人口,这不能不使人怀疑广东省

表 1 全国分地区的直接登记人口和“漏登率”

地 区	普查公布数 ^① (万人)	直接调查数 ^② (万人)	“衔接”人口 ^③ (万人)	“漏登率” ^④ (%)
北 京	1 382	1 357	25	1.81
天 津	1 001	985	16	1.60
河 北	6 744	6 668	76	1.13
山 西	3 297	3 247	50	1.52
内 蒙 古	2 376	2 332	44	1.85
辽 宁	4 238	4 182	56	1.32
吉 林	2 728	2 680	48	1.76
黑 龙 江	3 689	3 624	65	1.76
上 海	1 674	1 641	33	1.97
江 苏	7 438	7 304	134	1.80
浙 江	4 677	4 593	84	1.80
安 徽	5 986	5 900	86	1.44
福 建	3 471	3 410	61	1.76
江 西	4 140	4 040	100	2.42
山 东	9 079	8 997	82	0.90
河 南	9 256	9 124	134	1.43
湖 北	6 028	5 951	77	1.28
湖 南	6 440	6 327	113	1.75
广 东	8 642	8 523	119	1.38
广 西	4 489	4 385	104	2.32
海 南	787	756	31	3.94
重 庆	3 090	3 051	39	1.26
四 川	8 329	8 235	94	1.13
贵 州	3 525	3 525	0	0.00
云 南	4 288	4 236	52	1.21
西 藏	262	262	0	0.00
陕 西	3 605	3 537	68	1.89
甘 肃	2 562	2 512	50	1.95
青 海	518	482	36	6.95
宁 夏	562	549	13	2.31
新 疆	1 925	1 846	79	4.10
地区合计	126 228	124 261	1 967	1.56
现役军人	250	250	0	0.00
地区与军人合计	126 478	124 511 ^⑤	1 967	1.56
国家级“衔接”			105 ^⑥	0.08 ^⑦
总误差率	126 583		2 072	1.64

注:① 数据取自“2000 年第五次全国人口普查主要数据公报(第二号)”。

② 数据取自《2000 年第五次全国人口普查机器汇总数据集》。③ 等于各省的普查公布数减去直接调查数。④ 等于衔接人口除以普查公布人口,可以看成是国家级人口漏报率。在“主要数据公报(第二号)”中被解释为“常住地待定的人口”。⑤ 这一数字与前面给出的普查直接登记的 124 337 万人不一致。原因是此数据是机器汇总的结果,而另一数据是手工汇总数据,二者有所不同。

⑥ 为公布的祖国大陆总人口与 31 个省、自治区、直辖市和现役军人的人口之差。⑦ 此数据由 105 除以 126 583 而得到。

表2 部分2001年末人口数少于
普查公布人口数的省(区) 万人

地区	普查公布数 (2000.11.1)	直接调查数 (2000.11.1)	2001年 年末人口数*
河北	6 744	6 668	6 699.1
山西	3 297	3 247	3 271.63
辽宁	4 238	4 182	4 194
吉林	2 728	2 680	2 690.8
江苏	7 438	7 304	7 354.9
浙江	4 677	4 593	4 613.4
福建	3 471	3 410	3 440
山东	9 079	8 997	9 041
湖北	6 028	5 951	5 974.56
广东	8 642	8 523	7 783.41
云南	4 288	4 236	4 287.4
新疆	1 925	1 846	1 876.19

* 数据引自于2001年国民经济和社会发展的统计公报(<http://www.stats.gov.cn/tjgb/index.htm>)。

(七) 修正后的总人口能否分解

目前已经公布的全国以及各省2000年第五次全国人口普查主要数据中,除总人口外,其他各种构成的数据,相对数使用的是直接登记人口的结构,绝对数则是按直接登记的相对数以全国公布的全国或省级总人口为总体进行推算。这样计算的数据实际上隐含了一个假定,即假定误差人口的所有人口、社会、经济等结构与实际登记人口的相应结构完全一致。很明显做这种假定是不合适的。因为如果说“五普”有这样程度的漏报,那么漏报人口将主要集中在16~30岁的流动人口和0~5岁的婴幼儿。漏报人口的结构与实际调查到的总人口在结构上差异相当大。因此,用实际调查到的人口结构来代替调整后人口的年龄、性别和其他人口的社会经济结构,其结果是有偏的,或者说不反映事实的。

那么,在人们对“五普”资料进行分析时,就面临两难的困境。如果认为经过修正的人口总量是合理的话,那么我们可能无法得到科学的与这一人口总量相一致或相协调的人口结构;相反,如果认为直接调查到的人口结构是合理的,而且准备用这一数据进行人口分析的话,我们只有放弃得到与修正后人口总量有关的各种人口结构,即使我们认为修正后的数据更为准确。

(八) 何为“外来人口”和“常住地特定人口”

“五普”的“主要数据公报”中还出现了一些以往普查公报中从来没有出现过,并且是一些让人难以理解的术语。比如在公布普查快速汇总的各省、自治区、直辖市人口数的“主要数据公报(第二号)”的注1中这样写道:“祖国大陆31个省、自治区、直辖市的人口,是普查登记的2000年11月1日0时的数据(包括外来人口,不包括外出人口)。”可能是国家的统一要求,各省“公报”对全省经过修正的总人口的说明,都强调了“包括外来人口,不包括外出人口”,这种解释很让人费解。

人们都知道,人口普查的登记对象是常住人口,其时间标准为“半年”,因此各地的总人口肯定就是普查规定的“常住人口”。确切地说是“人在本地、户口在外地、并已离开户口所在地半年以上”的人记入本地人口,而“户口在本地、人在外地、并已离开本地超过半年”的人未记入本地人口。这在

“直接调查人口”的可信性。

为什么一些省不愿意接受普查公布的数据?令人费解。既然事后质量抽查采用计算机随机抽样和普查员异地调查的办法进行,全国共抽取了602个调查小区,约16万人;抽查时,各地以县级人口普查办公室为单位,进行了异地调查,那么,各省级地区就应该有各自的人口漏登率,而且各地的漏登率与全国的结果也应该是统一、协调的;各省也就没有理由不接受他们自己本身存在的漏报率。即使省级事后质量抽查的抽样样本规模较小,样本的代表性比较差一些,但一些人口规模比较大的省,其代表性是不成问题,何况很多省为了增加抽样的代表性,还自行增加了一些抽样样本。退一步说,即使有样本代表性问题,有依据总比没有依据要好,各省也不会从根本上拒绝普查公布的数据。1982年和1990年人口普查事后质量抽查,所抽取的样本规模与“五普”规模相近,当时它不仅被各省接受了,而且在各省的普查主要数据公报中也发表了。

常住人口登记对象的“五款人”中早已明确,而且没有丝毫异议。但是当看到在公布的某省总人口后面注明“包括外来人口,不包括外出人口”,人们很容易认为,这个省的总人口包括了所有的“外来人口”(无论他们来本地多长时间),未包括所有的“外出人口”(无论他们离开本地多长时间),即普查时人在哪里就在哪里进行登记。这就使人误以为普查是以“现有人口”为登记对象。这种注释不仅不能让人更准确的理解普查所涵盖的人口范畴,反而使人误解普查登记对象。因此说,这种注释不仅是不必要的,而且是错误的。

另外,“主要数据公报(第二号)”注2中还写道:“各省、自治区、直辖市和中国人民解放军现役军人的合计与祖国大陆总人口之差,为常住地待定人口。”由于公布的全省总人口与省内各市(地)合计人口不等,很多省在公报的注中也做了同样的解释。这里的问题是:何为“常住地待定人口”。

普查对象的五款人中只有“户口待定”,从来没有“常住地待定”这一项。很明显这是指“五款人”以外的又一种人。如果按字面理解,“常住地待定人口”应该指被调查者本人不知道自己的常住地,或调查者无法为被调查者“安排”一个常住地;而且“常住地待定人口”除了常住地这一个项目不明确外,其他的登记项目,如出生年月,性别,民族等都被登记了。进一步看,既然全国和省级都有常住地待定人口,那么,全国的常住地待定人口应该指的是“不知道是哪个省的人”;而某省的常住地待定人口指的就是“知道是该省的人但不知道是该省哪个市的人”。对于不了解普查的人来说,很难理解这些是什么样的人。

事实上,了解情况的人都知道,这部分人就是普查“漏登”的人口。如果在普查公报的注释中不用“常住地待定人口”,而用“常住地漏登人口”,涵义就更清楚了。

(九) 在公报中未公布的几个指标的结果如何

普查“主要数据公报”没有公布户口登记状况、出生率、死亡率及详细的关于普查登记误差的检验结果。根据人口普查机器汇总结果,我们可以初步地估计前两项,而登记误差在汇总结果中是得不到的。

根据第五次人口普查直接登记得到的结果,2000年11月1日全国总人口为1 242 612 226人^①。其中:(1)居住本乡镇街道、户口在本乡镇街道的人口为1 089 412 261人,占全部登记人口的87.67%;(2)居住本乡镇街道半年以上,户口在外乡镇街道的有134 286 452人,占10.81%;(3)在本乡镇街道居住不满半年,离开户口登记地半年以上的有10 104 305人,占0.81%;(4)居住本乡镇街道,户口待定的人有8 052 482人,占0.65%;(5)原住本乡镇街道,现在国外工作学习,暂无户口的有756 726人,占了0.06%。

根据前面对外来人口的定义,“五普”直接登记到的全国外来人口(即二三款人口之和)共有14 439万人,占全国登记总人口的11.62%。这说明,在2000年11月1日零时,全国有11.6%的人口在户口所在地以外的地方居住。这个数字反映了中国常住的外来人口总规模。

尽管全国公报没有公布出生率和死亡率数据,但某些省还是公布了,如浙江、湖南、广西、上海、青海等省(区);大部分省没有公布出生率和死亡率数据,原因可能是出生和死亡人数偏少,或难于决定是用直接登记总人口还是用修正后的人口数作为计算出出生率和死亡率的分母。

根据机器汇总数据,可以看出从1999年10月31日至2000年11月1日,全国共出生14 114 535人,死亡7 313 081人,自然增长6 801 454人。在出生、死亡分布均匀的假设下,按公布总

① “第五次全国人口普查主要数据公报(第一号)”公布的祖国大陆31个省、自治区、直辖市和现役军人的人口共126 583万人。这一数据包括了根据1.81%的漏登率计算的漏登人口。而经过机器汇总的直接调查的全国各省总计人口为1 242 612 226人。这个数据与快速或手工汇总的结果是不同的。这里所用数据均为直接调查到的人口数,而不是用修正过的数据。

人口推断的期中人口为126 243万;按直接调查总人口推断的期中人口为124 171万。按公布总人口推断,在1999年10月31日至2000年11月1日期间,中国人口的出生率为11.18‰,死亡率为5.79‰,自然增长率为5.39‰;按直接调查总人口推断,出生率为11.37‰,死亡率为5.89‰,自然增长率为5.48‰。从估计的结果看,出生率和死亡率都偏低,而且出生率偏低的程度更大一些。根据统计局已经公布的结果,1998年出生率、死亡率和自然增长率分别为16.03‰、6.50‰和9.53‰;1999年出生率、死亡率和自然增长率分别为15.23‰、6.46‰和8.77‰(国家统计局,2000)。如果1998和1999年的人口出生率是合理的话,1999年10月31日至2000年11月1日期间的出生率不可能下降到11‰。实际上,根据直接调查数据推算的1999年10月31日至2000年11月1日期间的总和生育率也只有1.22。很明显,如果按国家已经公布的1998年和1999年出生率作为标准,人口普查的出生人口漏报是比较严重的。

三、结束语

尽管“五普”出现过这样或那样的问题,但其成功远远大于它所存在的问题。“五普”所面临的困难是在中国人口普查历史上从未遇到过的,普查的组织者和工作人员所付出的辛苦也比任何一次普查都多。在这样的不利条件下,“五普”能取得目前的成绩,已经相当不容易了。然而,我们这里讨论“五普”存在问题的目的是为了今后的人口普查能搞得更好,也为了能更客观、更科学地开发人口普查数据资料创造条件。由于笔者没有参与和接触过人口普查的工作,对普查的设计和登记过程不甚了解,只是根据对以往中国人口普查的了解和研究,提出一些个人的看法。提出的问题肯定有不当地方,希望对此有兴趣并有一定研究的学者和普查实际工作者共同参与讨论。

参考文献:

1. 贾同金等:《人口变动抽样调查现状与问题分析》,《人口研究》,1995年第5期。
2. 杨日章:《事后质量抽查工作的改进与质量控制》,《中国1990年人口普查——国际讨论会论文集》,中国统计出版社,1993年。
3. 孙兢新:《来之不易的低误差率》,《人民日报》,2001年3月29日。
4. 武洁:《人口普查中的事后质量抽样调查》,“第八次全国人口科学大会”论文(未刊稿),2002年。
5. 崔红艳:《对2000年人口普查人口总量的初步评价》,“第八次全国人口科学大会”论文(未刊稿),2002年。
6. 查瑞传主编:《人口普查资料分析技术》,中国人口出版社,1991年。
7. 乔晓春:《中国人口普查研究》,中国人口出版社,1995年。
8. 国务院人口普查办公室、国家统计局编:《中国第四次人口普查的主要数据》,中国统计出版社,1991年。
9. 国家统计局编:《中国统计摘要(2000)》,中国统计出版社,2000年。
10. 国务院第五次全国人口普查办公室、国家统计局编:《第五次全国人口普查工作手册》,中国统计出版社,2000年。
11. 国务院人口普查办公室、国家统计局编:《2000年第五次全国人口普查主要数据》,中国统计出版社,2001年。
12. 国务院人口普查办公室编:《第五次全国人口普查普查员手册》,中国统计出版社,2000年。
13. 北京市第五次人口普查领导小组办公室:《北京市第五次人口普查工作手册》,2000年。
14. 各省、自治区、直辖市人口普查办公室:《人口普查公报》,中国统计信息网,2002。
15. Judith T. Lessler and William D. Kalsbeek:《调查中的非抽样误差》,金勇进译,中国统计出版社,1997年。
16. Bernard Benjamin (1970), The Population Census, A Social Science Research Council Review, Heinemann, London.

(责任编辑:朱 犁)