

人口普查

我国第四次人口普查漏报情况的重新估计^{*}
——基于第五次人口普查的分析

周 皓

【内容摘要】 文章对 2000 年第五次人口普查数据的调查时点进行调整后,重新估计了我国 1990 年第四次人口普查数据的漏报率,结论认为,1990 年四普的漏报率远高于四普公布的漏报率,其中低年龄段的漏报非常严重。因此四普数据的质量并没有我们以前所认为的那样好。

关键词: 人口普查;数据分析;漏报率

【作者简介】 周皓,男,1972 年生,北京大学社会学系博士后。北京:100871

数据质量是任何一次调查,乃至普查最为关心的问题之一。利用两次人口普查的数据,来评价前一次普查数据质量的方法,则是数据质量评价的一个重要方法。尽管第五次人口普查(以下简称五普)的数据质量被公认为是较差的,国家统计局公布的五普漏报率为 1.8%,比三普与四普的漏报率(2.55%和 0.6%)分别为高出几倍,但至少它为我们利用五普的人口性别年龄结构来评价四普数据的质量提供了人口性别年龄结构的基本框架。

然而,即便是在公认的五普漏报很严重的情况下,在利用五普的人口性别年龄结构,重新评价 1990 年第四次人口普查数据的质量时,发现四普数据的质量并没有我们以前所认为的那么好,其漏报率不仅高于国家公报数,而且其最保守的漏报估计可以与五普的漏报率几近相同。对这一结果,不论是在认识上,还是在情感上似乎都是无法接受的。因此本文将所得结果撰写成文,抛砖引玉,以期能够得到同行的指正批评。

1 数据、方法及几个基本假设

本文所用数据均来自于人口普查资料。其中 2000 年五普的人口性别年龄结构及死亡人口来自《中国 2000 年人口普查资料》的表 3-1 和表 6-4;按照对五普人口性别年龄结构的说明,出版的资料中所有数据均按实际调查到的 12.4 亿人口来汇总,而不包括漏报的 2800 万人(国家统计局,2002)。

本文所用 1990 年四普的人口性别年龄结构数据来自《中国 1990 年人口普查资料》。

为了能够使 2000 年五普的数据与 1990 年四普在进行年龄移算时具有可比性,有必要将五普人口的年龄从普查时点(2000 年 11 月 1 日)移至年中时点(2000 年 7 月 1 日)。在进行这一年龄移算时,需要作出这样一个假设,即出生人口与死亡人口在一年的各个月份上的分布是均匀的。在人口学中,这一假设事实是经常被用到的;而且根据 2000 年千分之一分的原始数据也证明出生人口在月份分布上是可以被视为均匀分布的。由此,年龄别人口数的移算公式为^①:

* 感谢郭志刚教授、陈卫博士的建议与意见。文责由作者自负。

① 这一计算公式事实上强烈要求,人口在各年龄组上的分布是均匀的,且死亡人口在各年龄与各时段上的分布也是均匀的。郭志刚教授曾对年龄别人口数的移算公式作过详细推导,并根据不同方法得到 2000 年年中时的人口年龄结构,尽管所得结果(即 2000 年年中的人口数与 1990 年四普的人口数之间的差异)在漏报的年龄组上有所不同,甚至于在部分年龄组上的数据有所减少,但总体的差异数却变得更大,从而使 1990 年的漏报率更加偏高。因此,本文未采用郭志刚教授的公式。对于年龄结构的移算公式将另文讨论。

$$P'_a = \frac{2}{3}P_{a+} + \frac{1}{3}P_{a+1} + \frac{5}{18}D_{a+} + \frac{1}{18}D_{a+1} \quad [1]$$

其中: P'_a 为年中时点上 a 岁的人口数; P_a 和 P_{a+1} 分别为普查时点上 a 岁和 $a+1$ 岁的人口数; D_a 和 D_{a+1} 分别为普查时点前一年中 a 岁与 $a+1$ 岁的死亡人口数。

本文中最高年龄组定在 90 岁, 因此, 在进行年龄移算时, 对于 90 岁及以上组的人口并未予以调整, 即将年中时点上的 90 岁及以上组的人口数等同于普查时点上 90 岁及以上组的人口数。作出这种假设的原因是: 普查时点上所得到的 90 岁及以上组的总人口数为 971227 人, 其中男性为 285766 人, 女性为 685461 人。这部分人口对于总人口而言, 不会产生太大的影响。这是本文所作的第二个假设。因此, 在上式中, P_{91} 和 D_{91} 均为 0。

为了检验调整前后总人口的差异部分, 我们可以将 [1] 式的左边与右边同时各自相加, 可得:

$$\sum P'_a = \sum P_{a+} + \left[\frac{1}{3} \sum D_{a-} - \frac{1}{3} P_{0-} - \frac{1}{18} D_0 \right] \quad [2]$$

由式 [2] 可以看到, 2000 年五普的人口年龄结构从 2000 年 11 月 1 日移至 2000 年 7 月 1 日时, 总人口的主要差异在于死亡人口与 0 岁组的人口。而 0 岁组的人口为 13793798 人, 取 $1/3$ 的话即可达到 450 万人之多, 而调查所得的总死亡人口为 7313081 人, 其中 0 岁组死亡人口为 340085 人。对总死亡人口数取 $1/3$, 即为 244 万人左右, 0 岁组死亡人口数的 $1/18$ 将近 20 万人, 因此, 调整前后总人口的差异应该在 225 万左右。普查的总人口数约为 12.42 亿人, 因此, 2000 年 7 月 1 日的全国总人口应该是在 12.40 亿左右, 即比普查时少了 200 万人。这个结果与我们根据分年龄性别的人口数进行调整后的总人口数是完全一致^①。

本文的第三个假设与人口普查时的年龄误报(或称之为年龄堆积现象)有关^②。许多学者曾对四普的年龄误报、死亡率等作了评价。翟振武(1993)、李树茁(1994)等的许多教学对四普的死亡人口数据作了评价及调整。又如乔晓春、李建新(1995)曾利用 1982 年和 1990 年两次人口普查的数据, 讨论了人口普查中年龄误报和军队人口对两次普查间的存活率结果的影响作用, 并调整了 1990 年的人口年龄结构。但由于年龄堆积现象对于两次普查间人口的存活率会有一定影响, 因此, 本文假设: 年龄误报(或者是年龄堆积现象)在四普与五普这两次普查中是相同的, 即如果说四普中年龄堆积在 0 岁与 5 岁组上时, 那么这种情况同样会反映在五普的数据中。当然这仅仅是一种假设, 对于年龄五普的年龄堆积现象, 还应该再做深入的研究。事实上, 即使是三普与四普之间, 在年龄堆积现象上也存在着较大的差别(乔晓春、李建新, 1995)。

本文所用的方法是最为简单的年龄移算。事实上年龄移算就是一个人每过一年, 年龄都会增加一岁。那么相应于 1990 年四普中 a 岁的人口, 他们在 2000 年五普中应该是 $a+10$ 岁。当然, 这里的 $a+10$ 岁应该对应于 2000 年年中时点, 即 2000 年 7 月 1 日。据此, 我们会有一个简单的人口平衡方程:

$$D' = P_{a+10}^{2000} - P_a^{1990} \quad [3]$$

上式中, P_a^{1990} , P_{a+10}^{2000} 分别是 1990 年年中时 a 岁的人口数与 2000 年年中时 $a+10$ 岁的人口数, 他们应该属于同一出生队列; 而是指 1990 年年中为 a 岁的人口在 1990 年 7 月 1 日至 2000 年 7 月 1 日这十年中死亡的那部分人。

从我国全国范围来看, 由于国际迁移的人数很少, 因此完全可以将全国人口视为封闭人口。在一个相对封闭的人口中, 除了 0 岁组的其他任意年龄段上, D' 都应该是恒小于 0。如果说, D' 是大于 0, 那么就有几种可能: 第一, 前一次普查在该年龄段上存在着漏报现象; 第二, 在该年龄段上有大批迁入人口, 且迁入人口的数量大于死亡人口; 第三, 由于前后两次普查在年龄误报上存在差异, 使前后两次普查在该年龄上无法对应。

① 之所以加入推导公式及这部分内容, 是因为担心读者会怀疑本文所用的数据在进行时点调整的过程中可能会出错。但事实上笔者对此数据作了严格仔细的检查, 并确认无错误的情况下再做后续工作的。

② 本文中, 笔者将年龄误报与年龄堆积现象视作同一问题, 因此在撰文过程中可能会相互替用。

2 利用五普数据,评价四普数据的质量

下面我们用 D' , 即相应年龄组(或同一出生队列)人口数在两次普查上的差异, 作为主要指标, 来比较两次普查之间的人口年龄结构之差异。在比较之前, 我们仍然要强调的是: 在下面的分析过程中, 我们都是以五普为标准, 来看四普的数据质量。

由图 1 可以看到, 在低年龄段(0~9岁)中四普的数据存在着较大的漏报现象。这种情况同样存在于在利用四普数据检验三普时。它与我国的计划生育政策有关, 对此是可以理解的。

首先我们来看五普与四普相应年龄组的人口数之差大于 0 的年龄分布。由于我们目前无法确定这种情况是否是由于四普中人口漏报引起的, 因此我们暂且称之为两次普查的差异, 而不称其为漏报。这种差异的年龄分布表现出的特点是:

(1) 在低年龄段 0~9 岁之间, 除了男性人口在 8 岁以外, 其他各年龄的差异值都大于 0。这一点与根据 1990 年四普与 1982 年三普、1987 年 1% 抽样调查的数据进行的比较结果完全相同, 但当时所指的年龄组是 0~6 岁(查瑞传等, 1996); (2) 总人口在 21 个年龄上有差异, 男性人口在 18 个年龄上有差异, 而女性人口则在 26 个年龄上有差异; (3) 所有在男性人口有差异的年龄上, 女性人口都存在差异(这是一种系统误差吗?); 但在某些女性人口有差异的年龄上, 男性人口并不存在这种差异。如 8、11、12、19、31、39 岁上; (4) 男女之间在差异的规模上也不相同, 但除了在 28 和 29 岁这两个年龄之外, 其他年龄上女性人口的差异都大于男性人口的差异。

那么, 在这种年龄分布上, 我们是否可以将这种差异值大于 0 的情况归因于年龄误报呢? 除了 0~9 岁以外, 在 11~40 岁组之间, 即 10 岁、20 岁、30 岁组之间, 除了 0 和 7 以外的其他 8 个数字都同时存在, 但在 3 个年龄组之间的分布不太相同。在 10 岁组中, 按照习惯, 年龄堆积现象可能会发生在 15 和 18 岁上(其中 15 岁是因为逢 5, 而 18 岁是成人年龄), 但在 11、12、14 和 19 岁上存在年龄堆积似乎很难解释。而在 20 岁组中, 男性与女性人口都集中在偶数的年龄上(除 29 岁以外), 但却未发生在 24 岁这一本命年上。而在 30 岁组上, 则是集中在 1、3、4、9 上。因此, 在这 3 个 10 岁一组的年龄组上这种差异性在年龄尾数上的分布, 即使是有年龄误报, 但这种年龄堆积现象也不可能真正地超过漏报的作用。更何况, 在本文的第一部分基本假设中, 已经假定这两次普查中年龄误报或年龄堆积现象在堆积年龄上完全相同。因此, 可以认为这种大于 0 的差异值是来源于漏报。

其次, 看四普的最低漏报率。在 1990 年四普时为 0~9 岁的低年龄段人口是在 1981~1990 年间出生的人群。而当时正值我国计划生育工作抓得较紧的时候, 因此人口漏报成为出现这种差异情况的主要原因应该是可以接受的。将 0~9 岁漏报的人口相加得到 10,321,746 人, 即为四普中低年龄段漏报的最低人口数(因为这里未考虑死亡), 而四普总人口数为 1,133,709,738 人, 因此, 1990 年四普的总人口漏报率至少在 0.91% 以上。即使不再加上四普本身公布的 0.2% 的漏报率, 所得的结果显然已经远高于我们的想象, 也远高于四普公布的漏报率了。

如果再加上其他年龄的漏报人数, 那么根据总人口的数据, 所有年龄的漏报人口数为 15,474,826 人, 四普的总人口数为 1,133,709,738 人, 因此漏报率达 1.365%。而如果根据分性别的人口数来看,

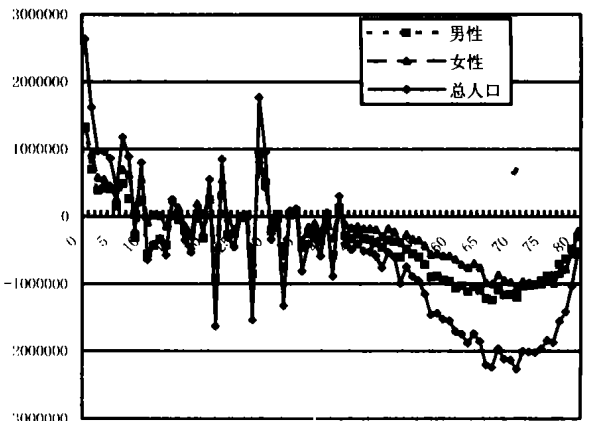


图 1 四普与五普人口数在 7 月 1 日时点上的比较
(2000~1990)

对图 1 的几点说明: ①首先将 2000 年普查时点上的数调整到年中(2000 年 7 月 1 日的); ②将 1990 年普查时 0 岁的人移至 10 岁, 即每个年龄组都后移的数, 但数值不变; ③然后在图 1 中将相应年龄(指 1990 年 a 岁对应于 2000 年 $a+10$ 岁)的两个数相减(2000 年 $a+10$ 岁人口数 - 1990 年 a 岁人口数)。

男性人口漏报 6,591,692 人,女性人口漏报 9,176,487 人,那么两性合计的漏报人数达 15,768,179 人,占四普总人口的 1.391%。这也就是说,在不考虑各年龄在这十年中的死亡人口的话,1990 年四普的漏报率至少是在 1.365% 以上。这个数据已经非常接近于五普公报中的漏报率(1.8%)。

我们应该注意到,假如说总人口的漏报率为 1.365% 时,0~9 岁的低年龄段漏报人口占总人口的 0.91%,相当于 3/4 的漏报来自于低年龄段。如果对比三普与四普的数据,我们也可以看到 1982 年三普中同样存在低年龄段人口的漏报情况(乔晓春、李建新,1995)。因此,可以认为在低年龄段存在漏报是我国人口普查中的一个普遍现象。

但至少在这里,我们尚未考虑几个非常重要的因素:

第一,四普与五普之间十年中的死亡情况。如果再考虑死亡的因素,那么相当于在 2000 年普查的人口数上,再继续增加一部分人口,相应的,也就是将 1990 年四普的人口再减去死亡的人口数,这样一来,四普的漏报率还要提高。

第二,五普的人口漏报情况。如果再结合五普的人口漏报情况,也就相当于是目前在计算的这个漏报人数上再加上五普中的漏报人数,从而会使四普的实际漏报率更高。按照国家统计局公布的数据,五普的漏报率为 1.8%,大约为 2800 万人。那么,按照 1990 年 0~9 岁低年龄段漏报的情况,这批人的漏报率约为 0.9% 的话,10 岁及以上年龄中的漏报率至少还应该 0.9%。由于 2000 年中 0~9 岁的出生人口是在 1990~2000 年之间出生的,他们不会影响到 1990 年的人口数,因此,扣除这部分 0.9%,还有另外一半 0.9% 就应该加到 2000 年普查中,这也就相当于是 1990 年的漏报率在 1.365% 的基础上,再加上 0.9%,就达到了 2.265%。

第三,在与上表中的年龄相邻近的其他年龄上面,有些相应年龄的两次普查间的差异值仅为几百人。如 35 岁组的女性两者仅相关 -476 人,这也就是说,在 1990 年 35 岁的 9,328,991 人中,在这十年间仅死亡了 476 人。这怎么可能呢?因此,如果再考虑这些年龄组上的漏报情况(事实上我们无法估算这些年龄,或者是任何年龄上准确的漏报情况,而只能给出一个最低的漏报率),那么事实上,1990 年四普的人口漏报率就会更高。

第四,目前的估算仅仅是这些年龄上的漏报人数之和,如果再加上四普本身公布的 0.2% 的漏报,那么四普的人口漏报率的最低线就会在 1.565% (= 1.365% + 0.2%) 以上。

在上述四个方面未被考虑在内的重要因素中,对于后三者的原因是无法在现在的情况下来完成。对于五普数据的漏报情况,我们只能根据以后的第六次人口普查来进行检验;而对第三个原因,或许可以通过将年龄组合并(即以 5 岁或 10 岁为一组)而进行描述。而对于第四个原因,即 0.2% 的漏报率是否应该加入到目前的最低漏报率上就已经显得不再重要了。目前我们能够考虑作出修正的是将死亡因素考虑在内。

3 考虑死亡因素后四普的漏报情况

张为民、崔红艳(2002),于学军(2002)都考虑死亡因素的情况下利用 2000 年五普的数据,重新推算了 1990~2000 年间各年的总和生育率。他们都运用了模型生命表。同时,由于在历次普查中,特别是在第四、第五次普查中,死亡人口的漏报情况是比较严重的(这一点都基本上是公认的),因此,在第五次人口普查中死亡人口肯定会比实际的死亡人口少。而且,在 1990~2000 年的十年间平均预期寿命有不同程度的提高,而年龄别死亡率在历年间则有一定的差异。但即便如此,它也为我们提供了一个大概的框架。而且由于本文的主要目的不是对死亡数据的漏报作估计。因此这里我们先做这样一个假设:即在 1990~2000 年的十年间,死亡模式与死亡水平都与五普所得的情况完全相同。为此,根据五普中死亡人口的分性别年龄结构,可以反推得到 1990 年时的人口年龄结构。注意,在这种假设下,反推所得到的人口年龄结构必然会低估 1990 年人口年龄结构的数据,因为十年间的死亡人口数必然大于五普中调查所得的死亡人口。因此,这里所估计的四普人口的漏报率应该说是最低限。

根据五普人口年龄结构及死亡人口的数据反推得到 1990 年时的人口年龄结构(以下简称反推数据),与四普的人口年龄结构(简称原数)相比较得到图 2。

由于考虑了死亡人口,因此,所得到的结果从理论上讲,两者应该是完全相等的。但如果反推数据减去原数大于0,说明原数有漏报的可能;相反,如果反推数减去原数小于0,那么情况就比较复杂,数据既有可能是正确的,也有可能是包含了死亡漏报的因素,也有反推的基数(即2000年人口普查)漏报的原因等等。我们先来看反推数据与原数的比较结果。

比较图1和图2,可以发现,三条曲线分别都有提高的趋势,即每个年龄组由于加入了死亡人口,而使反推数与原数之差,要大于五普与四普相应年龄组的差异。其道理十分简单。

如果不考虑年龄误报等其他因素,而仅将差异数看成是1990年四普中的漏报人数,那么,四普的漏报情况就更加严重了。总人口的差异数之和达21,153,473人,占四普0~80岁人口组(1,127,528,151人)的1.876%;占四普总人口的1.866%;而分性别的差异数之和达21,587,007人,占四普0~80岁人口组的1.915%、总人口的1.904%。这两个漏报的数据比国家统计局公布的2000年人口普查的漏报率还要高。

如果我们承认低年龄段(0~9岁)的漏报现象比较严重,而其他年龄段的这种差异可能是由于四普或五普时年龄误报等其他因素引起的,那么,1990年四普中0~9岁组的漏报率亦高达1.031%(0~9岁的漏报人口数为11,628,306人),比上面未考虑死亡因素的漏报率(0.91%)要高出0.12%,占总漏报率(1.876%)的54%。

根据上面粗略的估算,我们完全可以认为,1990年四普的漏报率在1%以上。这里既只考虑了0~9岁组人口的漏报情况,也只考虑了按照2000年普查所得的年龄别死亡人口因素。再一次强调,这里所考虑的死亡因素本身仍然存在着漏报现象。

上面只是对0~9岁组中反推数与原数之差的比较。但事实上,我们没有理由怀疑在两者之差大于0的多数年龄上会存在年龄误报的现象。年龄在25~29岁、31~35岁和41~46岁这三个主要的年龄组上,几乎每个年龄的差异数都大于0;这也就是说,即便是有年龄误报存在,也不可能会引起如此巨大的差异。因此,我们宁可将这些年龄上的差异数看成是漏报数,而不是将它们作为年龄误报等其他因素而引起的偏差。

4 结论及问题

在写这个结论时,确实比较为难,因为四普数据的质量在以前的十年中得到了广泛的肯定,现在这种“漏报严重”的结论,似乎是我们很不愿接受、也是很难接受的。然而,在对这两次普查的数据及中间的计算过程进行了反复的检查以后,笔者不得不再重提本文最初的一句话:四普数据的质量并没有我们所想象的好。

本文利用2000年第五次人口普查的数据,重新评价了1990年四普的数据。主要结论如下:

(1) 1990年四普的数据存在着较大的漏报现象,漏报程度远比四普公报中的漏报率要高。按照最为保守的估计,四普的漏报率应该在1.866%以上,与五普的漏报率非常相近。

(2) 在历次普查中,低年龄段人口(0~9岁)的漏报情况比较严重,漏报率至少占总人口漏报率的50%以上。这一现象在历次普查,特别是1990年四普中也同样存在,它与我国计划生育的政策环境有着密切的关系。由于造成这种漏报的社会条件还将继续存在,因此可以相信2000年五普中同样存在着较为严重的低年龄段的漏报。

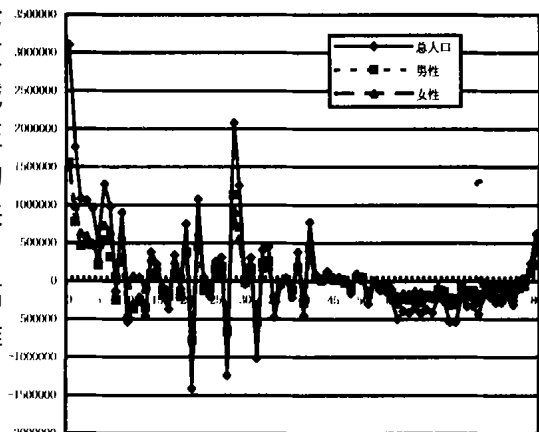


图2 考虑死亡因素以后五普与四普数据的比较
(反推数-原数)

对图2的说明:①反推数是指2000年普查的人口数,结合2000年普查的人口数,推算得到的1990年年中的人口数。其中,2000年普查的人口数所在时点已移至2000年年中;②原数是指1990年四普的人口年龄结构。

(3) 年龄误报等其他各种因素可能会引起本文推算过程中的一些误差,但由于两次普查间相应年龄的人口数之差在部分年龄组上分布均匀,因此,可以认为年龄误报所引起的偏差对于漏报率的总体评价不会产生太大的影响。

(4) 尽管本文考虑了死亡因素,但所采用的历年死亡人口的年龄结构是完全按照 2000 年五普的死亡人口年龄结构,因此,这种假设会导致所估计的十年间死亡人口总数与各年龄的死亡人口数,与实际的死亡人口数相比,都会有不同程度缩小,从而使估计所得的四普的漏报率仅仅是最低线。

(5) 由于本文所用的五普数据是国家统计局公布的五普人口年龄结构的数据。在推算过程中并未考虑五普公报的 1.8% 的漏报率(即 2800 万)。如果再结合五普的漏报率,那么四普的漏报率可能还要高。

据上所述,1990 年第四次人口普查漏报率的最低水平应该在 1.866% 以上。但针对本文的研究方法、过程及结论而言,都可能会存在许多值得商榷或者是今后还有待于进一步深入讨论的地方。我想非常有必要在这里一并提出^①。

(1) 有关年龄移算方法的探讨。由于我国人口年龄结构的凹凸起伏较大,因此在进行时点移算的过程中,会使结果受到一定的影响。同样的,死亡人口的分布对人口年龄结构移算的结果也存在一定的影响^②。

(2) 在本文中,是用 2000 年的五普作为标准来检验四普数据的质量。但有学者提出,用本身质量相对较差的五普数据,来评价四普的数据质量,是否存在问题;因此,是否可以将 1990 年四普的数据从 1990 年 7 月 1 日移算至 1990 年 11 月 1 日,然后再来比较两次普查间的差异。我个人觉得这也是另外一种检验方法。

(3) 在本文中,曾作出了一个强烈的假设,即四普与五普的年龄误报模式是相同的。这里存在两个问题:一是五普数据的问题,即这次普查中有些数据可能是错误的(即有些是十几岁的人在申报时仅为几岁);二是这种年龄误报模式发生在相同的出生队列、不同普查的不同年龄组中。如四普中 10~20 岁组的人,在五普中已经是属于 20~30 岁的人,那么,尽管他们是属于同一出生队列,但在发生年龄误报时却属于不同的年龄组,从而可能会有不同的年龄偏好,进而导致相同队列的人口有一定的偏差。

(4) 漏报在性别上的差异。事实上不仅仅是漏报的问题,在男性与女性人口的年龄误报模式上同样可能存在差异。这方面的研究可能会更有利于揭示我国总人口性别比偏高的另一方面原因。

参考文献:

- 1 中国 2000 年人口普查资料. 中国统计出版社, 2002
- 2 中国 1990 年人口普查资料. 中国统计出版社, 1993
- 3 李树茁. 80 年代中国人口死亡水平和模式的变动分析—兼论对 1990 年人口普查死亡水平的调整. 人口研究, 1994; 2: 40
- 4 查瑞传等编. 中国第四次全国人口普查资料分析. 高等教育出版社, 1996
- 5 翟振武. 1990 年婴儿死亡率的调整及生命表估计. 人口研究, 1993; 2: 9
- 6 乔晓春, 李建新. 中国人口年龄结构再分析及其调整. 人口与经济, 1995; 3: 29~33
- 7 张为民, 崔红艳. 对 2000 年人口总数的初步评价. 人口研究, 2002; 4
- 8 于学军. 对第五次人口普查数据中总量和结构的估计. 人口研究, 2002; 3

(责任编辑: 段成荣 宋 严 收稿日期: 2002-11)

① 这些问题原先写得比较简单仅提到了第一点和第三点。林富德教授、杜鹏教授、刘金塘副教授、孟向京博士、宋健博士、王金营博士等对本文的不足之处提出了许多有益的建议与意见,因此在这里作出稍微详细一点的描述。在此一并致谢! 但文责由作者自负。

② 出生人口的月份分布同样也是个问题。根据五普资料,郭志刚教授曾检验了出生月份的分布,结果与一直以来所做的假设基本相符,即在每个月份上的分布基本上是均匀的。