

# 对中国人口普查低龄人口数据的再认识

侯亚杰 段成荣

**【摘要】**文章利用 2017 年户籍登记数据对 1982~2010 年全国 4 次人口普查低龄组(0~16 岁)数据进行了评估。计算结果表明,“三普”至“六普”0~16 岁低龄组漏报率分别为 0.88%、2.87%、2.87%和 4.11%。“五普”至“六普”期间,青少年人口和中青年劳动年龄人口重报现象逐渐凸显,“五普”和“六普”14~49 岁的净重报人口数由 677 万上升至 1 851 万。最后,文章根据修正后历次普查低龄组数据重新计算了 1990~2010 年中国人口出生数量、出生性别比和生育水平,进一步证明由于低龄人口的漏报,以往对于这 3 个指标的估计存在一定程度的偏差,2005~2010 年中国的总和生育率可能在 1.6~1.7 之间。

**【关键词】**人口普查 户籍登记 人口漏报 人口重报 生育水平

**【作者】**侯亚杰 中国人民大学人口与发展研究中心,博士研究生;段成荣 中国人民大学人口与发展研究中心,教授。

## 一、问题提出与文献综述

1982~2010 年,中国进行了 4 次人口普查。在一个拥有近 14 亿人口的国家进行人口普查是一项艰巨的工作,普查过程中出现一些系统性问题,进而影响数据质量,是很难避免的。每次人口普查之后,都会有学者对历次普查同队列人口进行一致性比较。研究结果显示,相邻两次人口普查的同队列人口存在较为明显的不一致,主要表现在前次人口普查的 0~9 岁人口在后一次人口普查时(10~19 岁人口)存活率大于 1。众所周知,同一队列人口的存活率是不可能大于 1 的,这一统计现象的出现主要有两个原因,一是前一次人口普查低龄人口的漏报,二是后一次人口普查青少年人口的重报。目前相对一致的观点是,中国历次人口普查 0~9 岁儿童的漏报较为严重,其中 2000 年“五普”0~9 岁人口的漏报问题最为突出。有研究估算,1990 年“四普”0~9 岁儿童漏报 1 150 万~1 270 万,漏报率为 5%~6%(张为民、崔红艳,2003;周皓,2003;王金营,2003;李波、姜全保,2010)。而在 2000 年“五普”时,0~9 岁儿童漏报保守估计在 1 680 万以上,漏报率在 10%以上(崔红艳等,2013;张青,2005)。有些学者估计的漏报人数甚至高达 3 000 万以上(王金营,2003;李树苗等,2006)。与 2000 年“五普”相比,2010 年“六普”

0~9岁儿童漏报情况要好得多,估算的漏报率为5%~8%(崔红艳等,2013;王金营、戈艳霞,2013)。

在2000年以前的历次普查,除了极个别情况,人口重报的概率非常低(李若建,2013)。因而与低龄人口漏报的问题相比,关于人口重报问题的讨论在2000年“五普”之前非常少。但在2000年“五普”和2010年“六普”中,由于普查环境的变化和人口登记方式的改变,人口重报逐渐成为一个值得关注的问题。2000年“五普”以来,一些研究已指出16~21岁、30~39岁、50岁以上是人口重报比较集中的年龄段(王广州,2003),而2010年“六普”人口重报问题更是被反复提及(崔红艳等,2013;王金营、戈艳霞,2013;陶涛、张现苓,2013)。然而,2000年“五普”和2010年“六普”的相关年龄段究竟重报了多少?已有研究只进行了概数的估算,并没有进行详细的证明和精确的测算。

上文已提到,相邻两次普查同队列人口的存活率大于1,既可能是由于前次普查的漏报,也可能是后次普查的重报造成的。但是,大多数研究倾向于将这一问题归因于前次普查的漏报,而对于后次人口普查的重报问题又大多“语焉不详”。如果该问题是由后次普查的重报造成的,那么以往研究对于低龄人口漏报率的估计很有可能被高估,同时与相关年龄组数据质量密切相关的一些人口统计指标的估计(如生育水平估计)也可能存在偏误。鉴于此,本文对1982~2010年历次人口普查低龄组人口(0~16岁)的漏报问题及其他相关年龄组人口的重报问题进行系统的研究,并对历次普查低龄人口数据质量进行评价。另外,本文还试图对与低龄人口数据息息相关的生育问题进行再估计。

## 二、从人口发展连续性的角度审视人口普查低龄数据质量

在以往的认知中,1982年“三普”和1990年“四普”的数据质量比较理想。这主要得益于当时的普查环境。但30多年过去了,如果从“五普”和“六普”的视角来重新审视“三普”和“四普”,对历次普查同队列人口数据的认识,特别是低龄组人口数据的认识,是否会发生变化呢?为了回答这一问题,需要对历次普查数据进行以下调整和计算。

具体的推算方法主要包括两个步骤:(1)统一4次人口普查的普查时点。为了便于比较,本研究将“五普”、“六普”数据调整至当年7月1日,调整公式为: $P_{\alpha}' = \frac{2}{3}P_{\alpha} + \frac{1}{3}P_{\alpha+1} + \frac{5}{18}D_{\alpha} + \frac{1}{18}D_{\alpha+1}$ <sup>①</sup>。式中 $P_{\alpha}'$ 表示“五普”或“六普” $\alpha$ 岁年龄人口在7月1日时的数

<sup>①</sup> 该公式假设“五普”和“六普”相应年龄组人口的出生和死亡是均匀分布的。而实际上出生人口和死亡人口的月份分布并不可能满足这一假设,因此会导致不可忽视的调整误差(郭志刚,2004)。本文这里仅对“三普”“四普”低龄队列人口在之后普查的情况进行数量关系的讨论,因此没有考虑出生人口和死亡人口的月份分布。

量,  $P_{\alpha}$  和  $P_{\alpha+1}$  表示“五普”或“六普” $\alpha$  岁和  $\alpha+1$  岁人口数量,  $D_{\alpha}$  和  $D_{\alpha+1}$  分别表示在“五普”或“六普”普查时点前 1 年之内  $\alpha$  岁和  $\alpha+1$  岁年龄的死亡人口数。(2) 将“三普”和“四普”0~16 岁分年龄低龄人口数按照 1982~2010 年的生命表存活率分别推算至“五普”和“六普”时点。推算所应用的生命表根据寇尔—德曼西区模型生命表和 1982~2010 年中国分性别人口平均预期寿命通过 Brass-logit 变换得到。在计算过程中, 历次普查中现役军人数量也分别回填到相应的年龄组中。

(一) 利用“四普”至“六普”数据评估“三普”低龄人口数据

图 1 给出了利用“四普”至“六普”三次人口普查数据估算的 1982 年“三普”分性别低龄组漏报率年龄分布情况。总的来看, 1982 年“三普”0~16 岁低龄组数据存在人口漏报, 三组漏报率指标差异并不大。若以“四普”至“六普”的同队列数据为参照, 1982 年“三普”0~16 岁人口大约漏报了 313 万~908 万人, 漏报率为 0.81%~2.37%; 男性漏报率为 1.00%~1.75%; 女性漏报率为 0.61%~3.02%。

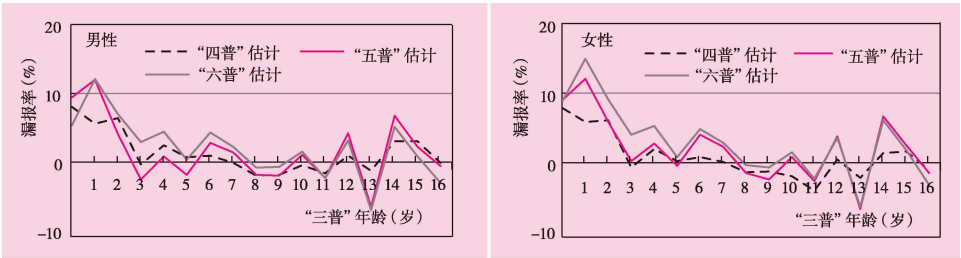


图 1 利用普查数据估计 1982 年“三普”低龄组人口的漏报率

从低龄组的分性别年龄漏报率来看, “三普”低龄人口数据的漏报主要集中在 0~5 岁年龄组, 其中“三普”男性 0~5 岁漏报人数为 231 万~320 万人, 漏报率为 3.83%~5.3%, 女性 0~5 岁漏报人数为 206 万~434 万人, 漏报率为 3.65%~7.49%。

(二) 利用“五普”和“六普”数据评估“四普”低龄人口数据

表 1 给出了利用“五普”和“六普”数据计算的“四普”低龄组漏报率。从中可以看出, “五普”和“六普”估计的“四普”低龄人口漏报率相差 1% 左右, 漏报人数相差 487 万人。利用“六普”数据估算的“四普”0~2 岁和 9~13 岁的漏报率高于“五普”估算结果, 而根据“五普”估算“四普”的 5~8 岁漏报率高于根据“六普”估算的结果。根据“五普”和“六普”的估算结果, “四普”0~2 岁人口分别漏报了 647 万和 984 万(漏报率为 8.51% 和 12.46%); 5~8 岁人口分别漏报了 336 万和 167 万(漏报率为 4.00% 和 2.05%); 9~13 岁人口分别漏报了 52 万和 342 万(漏报率为 0.54% 和 3.54%)。“四普”的 0~2 岁人口在 2000 年“五普”时为 10~12 岁, 依然属于低龄组, 因此这一年龄段人口在“五普”时仍然存在漏报现象, “六普”估计结果高于“五普”估计结果是可以理解的。但是 1990 年 9~13 岁人口到“五普”和“六普”时已分别为 19~23 岁和 29~33 岁, 已不再属于低龄人

表 1 1990 年“四普”低龄人口与 2000 年“五普”、2010 年“六普”相应数据的对比

“四普” 年龄(岁)	“五普”				“六普”			
	推算人数 (万人)	“五普”登记 人数(万人)	漏报人数 (万人)	漏报率 (%)	推算人数 (万人)	“六普”登记 人数(万人)	漏报人数 (万人)	漏报率 (%)
0	2268.41	2585.64	317.23	12.27	2256.73	2778.84	522.12	18.79
1	2293.55	2495.39	201.84	8.09	2281.00	2608.16	327.16	12.54
2	2387.24	2514.84	127.60	5.07	2373.19	2507.83	134.64	5.37
3	2404.82	2525.51	120.69	4.78	2389.52	2486.11	96.58	3.88
4	2122.78	2227.31	104.53	4.69	2108.24	2188.95	80.71	3.69
5	1983.12	2039.72	56.59	2.77	1968.74	1996.93	28.20	1.41
6	1892.35	2024.50	132.15	6.53	1877.93	1973.67	95.74	4.85
7	2006.15	2122.38	116.23	5.48	1990.27	2052.54	62.27	3.03
8	2186.79	2218.09	31.31	1.41	2169.00	2150.22	-18.79	-0.87
9	1796.35	1925.44	129.09	6.70	1781.27	1944.17	162.90	8.38
10	1908.95	1883.24	-25.71	-1.37	1892.41	1932.00	39.59	2.05
11	1916.65	1904.41	-12.24	-0.64	1899.74	1980.53	80.79	4.08
12	1871.37	1863.64	-7.73	-0.41	1854.61	1909.78	55.17	2.89
13	1919.97	1888.11	-31.86	-1.69	1902.52	1906.24	3.71	0.19
14	2026.99	2080.92	53.93	2.59	2008.21	2089.98	81.78	3.91
15	2143.02	2182.41	39.38	1.80	2122.54	2179.99	57.45	2.64
16	2325.56	2323.66	-1.90	-0.08	2302.50	2330.54	28.04	1.20
合计	35454.07	36805.21	1351.14	3.67	35178.44	37016.48	1838.05	4.97

口,但这一年龄组的人口在 2000~2010 年仍有近 300 万的增长,这一现象值得深入研究。

图 2 给出了“四普”分性别各年龄组的漏报率情况。对于“四普”0~16 岁女性人口来说,漏报率的差异主要体现在 0~5 岁和 9~13 岁人口“六普”估算结果显著高于“五普”结果,其中“六普”估算的“四普”0~5 岁漏报率尤其高。而对于“四普”0~16 岁男性人口来说,除了 0~2 岁、9~13 岁“六普”估计漏报率高于“五普”以外,根据“五普”数据估算的 2~9 岁男性漏报人数明显高于“六普”估算结果。根据“五普”数据估算,“四普”男性 2~9 岁人口大约漏报了 377 万人,而根据“六普”数据估算的漏报人数为 76 万人,前者比后者多出近 300 万人。

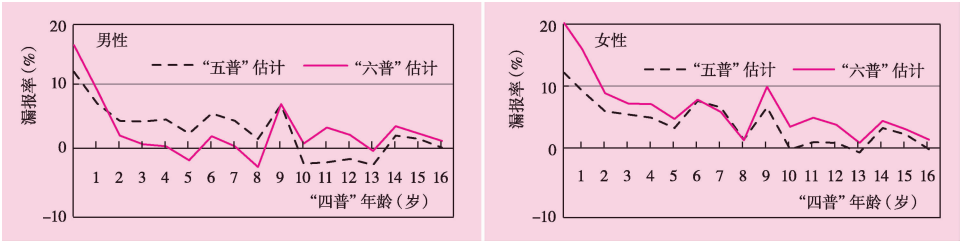


图 2 利用普查数据估算的 1990 年“四普”低龄组人口漏报率

### （三）问题与思考

1982~2010年,历次普查公布的净漏报率分别为0.071%、0.065%、1.81%和0.12%。虽然四次人口普查的漏报均处于较低水平,但近两次的人口普查的质量不如“三普”和“四普”已是基本共识。“三普”和“四普”是公认的数据质量较高的两次人口普查,虽然从总人口漏报率来看“四普”数据质量似乎优于“三普”,但就0~16岁低龄人口数据来说,“三普”好于“四普”。特别是0岁组人口的漏报率,“三普”漏报了10%左右,“四普”至少漏报了12%~19%。

除了人口漏报的问题外,通过对数据质量相对较好的“三普”和“四普”低龄人口数据的考察,本文还发现“五普”和“六普”中青年人口数据存在一些问题:(1)“三普”0~5岁女童数据的问题。根据“六普”28~33岁年龄组数据估算,该年龄段漏报人口为434万,而根据“五普”18~23岁数据估算的漏报人口为288万(对“四普”8~13岁人口漏报率的考察也存在同样的问题)。其原因是“三普”至“五普”相应年龄女性数据漏报,“六普”28~33岁年龄数据可能存在重报。(2)关于“四普”2~9岁男性人口漏报率的悖论。“四普”的2~9岁人口在“五普”和“六普”时已分别为12~19岁、22~29岁。根据低龄人口漏报的规律,“六普”估计的相应年龄的漏报率应当高于“五普”估计结果,且这种差异不应该过大。但估算出的“四普”2~9岁男性漏报人口“五普”估算结果比“六普”多出近300万,这可能表明中国“四普”至“六普”的低龄组和青年组数据的主要问题不仅是人口漏报,也存在人口重报问题,特别是“五普”男性青年组人口重报问题似乎更加突出。

根据1982~2010年历次人口普查同队列数据的校验结果,除了发现“三普”和“四普”低龄人口漏报问题外,还发现“五普”和“六普”青年组人口重报,这一现象也有学者指出过(胡耀岭、原新,2013;陶涛、张现苓,2013)。在目前的普查环境中,同一队列人口的漏报和重报极有可能作为一种系统性问题在相邻两次普查中连续出现。因此,要真正厘清这两个问题,仅依靠普查数据的证据是不够的。

## 三、利用户籍登记数据评估1982~2010年历次人口普查数据

目前,越来越多的学者将目光投向户籍登记数据(杨凡、赵梦晗,2013;陈卫,2014;翟振武等,2015)。本研究借助户籍登记数据来考察1982~2010年历次普查低龄人口数据质量,这可能会为澄清上述问题提供新的证据与支持。

### （一）户籍登记数据准确性评估

有学者指出,户籍登记数据对学龄儿童的统计比较完整,对16岁以上拥有身份证人口的统计基本能够全覆盖,这是基于户籍登记对于公民的作用而进行的判断(杨凡、赵梦晗,2013)。也有学者从户籍登记工作不断改进和数据获取渠道两方面证明户籍登记数据已具备“拿来就用”的条件(翟振武等,2015)。图3给出了2010年户籍登记数据



与“六普”尾数别修正惠普尔指数<sup>①</sup>的差异。2010年男性户籍人口各年龄尾数的堆积模式与“六普”男性基本吻合,而2010年女性户籍人口的在2和3年龄尾数上则出现了较明显的堆积和跨越,但仍在可接受的范围内。从图4可以看出2010~2017年的总和修正惠普尔指数<sup>②</sup>为0~0.5,女性的年龄堆积要比男性严重,2016和2017年的年龄数据质量要明显好于之前年份。有研究表明,“六普”的年龄数据质量好于“五普”,比“四普”略差(王金营、戈艳霞,2013),而2010年户籍登记年龄数据是近8年里较差的,因此有理由相信2016和2017年的户籍登记数据的年龄误报在可接受的范围内。

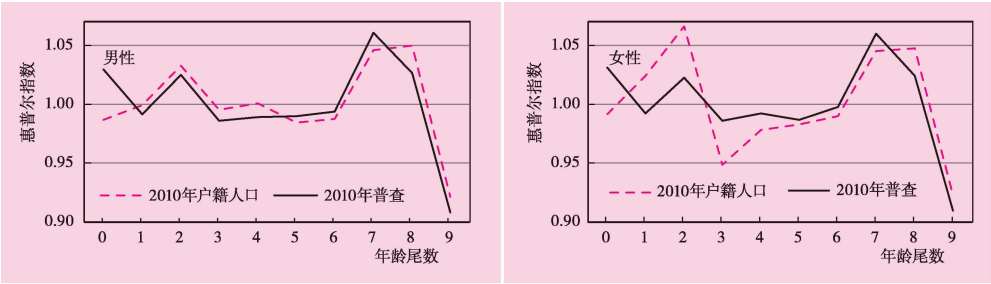


图3 2010年户籍人口与“六普”人口分性别尾数别修正惠普尔指数比较

表2给出了2010~2016年6~45岁户籍登记人口数据。总的来看,2010年6~25岁户籍登记人口数据有随着时间的推移数量逐渐增长的趋势,但持续增长的时间逐渐减少。其中2010年6~10岁户籍登记人口保持增长的时间最长,但2015年前后开始稳定,数量约为7280万人。

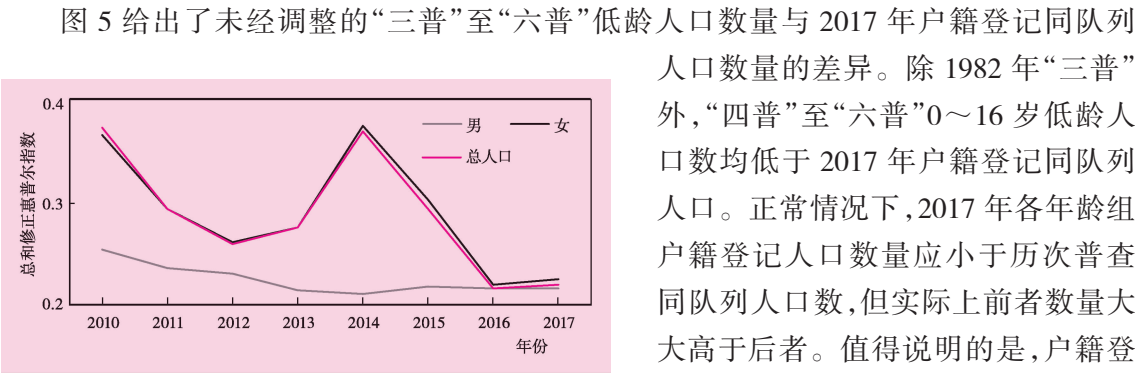


图4 2010~2017年户籍人口总和修正惠普尔指数

图5给出了未经调整的“三普”至“六普”低龄人口数量与2017年户籍登记同队列人口数量的差异。除1982年“三普”外,“四普”至“六普”0~16岁低龄人口数均低于2017年户籍登记同队列人口。正常情况下,2017年各年龄组户籍登记人口数量应小于历次普查同队列人口数,但实际上前者数量大大高于后者。值得说明的是,户籍登记数据不会受人口迁移流动与人口

① Noubissi(1992)提出尾数别修正惠普尔指数,在没有年龄偏好和回避的情况下,各尾数别修正惠普尔指数都为1。  
② Spoorenberg等(2007)对尾数别修正惠普尔指数进行了再次修正与整合,通过将尾数别修正惠普尔指数与1的离差的绝对值求和,由此得到一个总和修正惠普尔指数,若某一人口的年龄分布在任意尾数均不存在堆积和跨越的情况下,则总和修正惠普尔指数为0。

表 2 2010~2016 年同队列户籍登记人口数据比较 万人

2010 年 登记年龄(岁)	年 份						
	2010	2011	2012	2013	2014	2015	2016
6~10	7029.27	7140.14	7198.19	7236.84	7278.06	7260.68	7275.10
11~15	7886.37	7924.51	7925.99	7914.57	7875.69	7771.55	7768.39
16~20	10121.20	10122.60	10082.28	10057.23	9997.60	9893.73	9914.39
21~25	12183.64	12199.25	12119.62	12099.77	12034.89	11919.15	11942.91
26~30	10091.70	10078.86	9963.21	9921.16	9840.18	9735.38	9746.96
31~35	10057.48	10029.16	9876.53	9815.68	9712.49	9593.15	9594.48
36~40	12550.13	12516.96	12343.04	12281.33	12182.87	12049.37	12044.83
41~45	12498.81	12463.57	12310.96	12260.08	12180.68	12066.72	12054.72

分离的影响。同时除“六普”若干队  
列外,历次人口普查 0~16 岁人口在  
2017 年时已不再属于低龄范畴,因  
而在户籍登记数据中出现大规模人  
口重报、漏报的可能性不大。因此,  
户籍登记人数与普查同队列人数的  
差异主要是由于历次普查低龄组人  
口漏报。

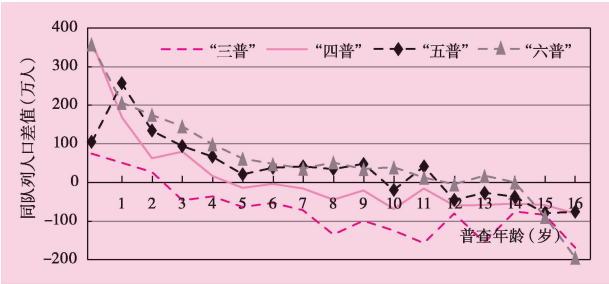


图 5 1982~2010 年历次人口普查 0~16 岁人口与  
2017 年户籍登记同队列人口差异

图 6 给出了 2010 年 0~50 岁户  
籍登记人口在 2017 年时的队列存活率与模型生命表存活率,从中可以发现以下特点:  
(1)由于新生婴儿及低龄儿童延迟申报户口的现象较多,户籍登记 5 岁及以下年龄组  
人口存活率异常偏高;(2)户籍登记 16~50 岁劳动年龄人口存活率与模型生命表存活  
率耦合程度较高。

通过对多个统计时点户籍登记数据的比较,可以看出 6~50 岁户籍登记人口数据  
在数量上具有较高的一致性和稳定性,同时 6~50 岁户籍登记人口数据的变化过程基

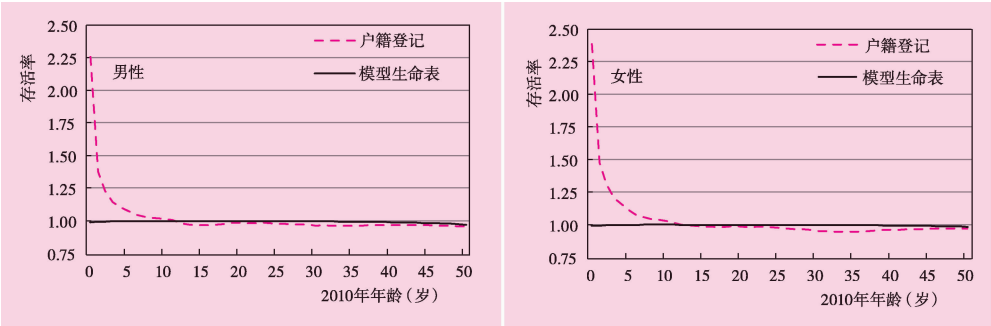


图 6 2010 年 0~50 岁户籍登记人口在 2017 年的存活率与模型生命表存活率比较

本符合中国人口数量变化的规律。6~50岁是经历入学、升学、就业等人生重要阶段,户籍登记对这一年龄段人口来说不可或缺,因此出现大规模漏登的可能性较小。另外,由于户籍人口统计口径与常住人口统计口径不同,人口重报的可能性在户籍登记数据中也相对较小。基于以上事实,本文认为,中国6~50岁户籍登记数据是可靠的,而2017年相应年龄户籍登记数据的可靠性要高于之前年份。为了保证本研究估算的准确性,同时也为了使估算基础数据最大限度地涵盖历次普查的低龄组,本文使用2017年7月1日和11月1日6~50岁户籍登记数据分别计算几次普查数据中低龄人口的漏(重)报情况。

(二) 利用户籍登记数据估计“三普”至“六普”低龄人口数据的漏报率

表3给出了以2017年户籍登记数据为标准计算的历次人口普查低龄组的数据漏报情况。估算结果显示,1982~2010年历次人口普查中国0~16岁低龄人口共漏报327万、1 034万、950万和1 101万,漏报率分别为0.88%、2.87%、2.87%和4.11%。由此可见,“三普”至“六普”0~16岁低龄组人口漏报问题逐渐严重。从分年龄漏报率看,“三普”和“四普”低龄人口的漏报主要分别集中在0~2岁和0~4岁年龄组;而“五普”和“六普”低龄人口数据在10岁以前均存在漏报,15~16岁组存在人口重报的现象。

图7反映了1982~2010年历次人口普查分性别分年龄的人口漏报率情况。除“三

表 3 1982~2010 历次人口普查低龄数据漏报情况

普查 年龄 (岁)	“三普”		“四普”		“五普”		“六普”	
	漏报 (万人)	漏报率 (%)	漏报 (万人)	漏报率 (%)	漏报 (万人)	漏报率 (%)	漏报 (万人)	漏报率 (%)
0	169.20	7.85	442.74	16.47	131.77	8.88	373.02	21.49
1	118.13	6.60	231.93	9.27	274.13	19.49	218.99	12.36
2	90.05	4.86	119.45	4.81	151.16	9.85	184.57	10.63
3	16.71	0.87	130.22	5.19	108.77	7.07	152.65	9.14
4	23.51	1.29	59.17	2.74	81.28	5.11	105.06	6.48
5	-1.83	-0.10	24.15	1.22	36.05	2.10	66.58	4.34
6	13.37	0.67	32.74	1.72	52.85	3.14	51.71	3.38
7	0.16	0.01	22.55	1.13	57.30	3.13	39.34	2.85
8	-51.95	-2.29	-1.70	-0.08	52.00	2.72	55.49	3.91
9	-9.83	-0.41	16.71	0.93	66.44	3.23	38.15	2.61
10	-29.06	-1.21	-27.50	-1.48	6.23	0.24	42.61	2.87
11	-49.43	-1.92	27.09	1.41	66.79	2.61	16.42	1.17
12	29.91	1.17	-16.37	-0.90	-19.56	-0.81	-0.46	-0.03
13	-29.50	-1.11	-10.78	-0.57	1.65	0.06	21.13	1.37
14	39.81	1.67	-2.59	-0.13	-10.85	-0.48	4.94	0.31
15	30.38	1.39	0.32	0.02	-54.78	-2.79	-82.13	-4.79
16	-32.32	-1.35	-14.02	-0.62	-51.03	-2.61	-186.92	-11.09
合计	327.31	0.88	1034.12	2.87	950.20	2.87	1101.16	4.11



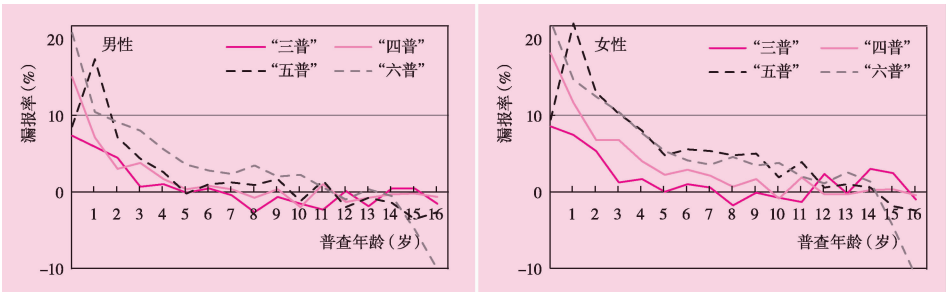


图 7 根据 2017 年户籍登记人口数据推算的 1982~2010 年历次人口普查低龄组漏报率

普”外,在其他各次人口普查中,中国低龄女性的漏报率水平明显高于低龄男性。其中“五普”的低龄女性人口漏报问题尤为突出。与男性相比,“六普”15~16 岁女性重报问题更加严重。

(三) “五普”和“六普”人口重报问题

低龄人口的漏报问题是全世界所有人口普查均无法避免的问题,即便是公认的数据质量较好的“三普”和“四普”也存在低龄人口漏报。此外,“五普”以来 14~16 岁人口重报问题日益突出。图 8 展示了未考虑死亡因素的情况下“四普”至“六普”与户籍登记数据同队列人口的数量对比。从中可以发现,“五普”和“六普”14~16 岁人口是同队列数据中数量最多的,不仅高于前一次普查同队列人口,也高于后一次普查及 2017 年的户籍登记同队列人口。在考虑了低龄人口漏报因素之后,户籍登记 31~33 岁和 21~23 岁人口分别与“四普”和“五普”的 4~6 岁人口数量大致相当在逻辑上是说得通的,而 14~33 岁年龄段人口的死亡水平显然不支持图 8 中该队列规模的下降。因此可以断定,“五普”和“六普”14~16 岁人口存在一定程度的重报。

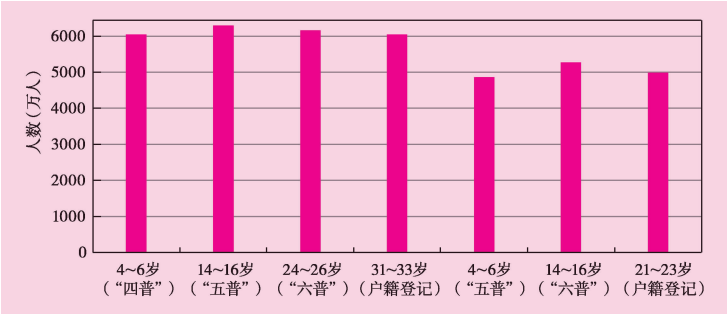


图 8 同队列“五普”、“六普”与 2017 年户籍登记人口数据

表 4 “五普”和“六普”14~49 岁人口的漏(重)报情况

普查 年龄 (岁)	“五普”		“六普”	
	漏(重)报人数 (万人)	漏(重)报率 (%)	漏(重)报人数 (万人)	漏(重)报率 (%)
14~19	-329.54	-2.70	-771.19	-7.15
20~24	-41.70	-0.45	-331.47	-2.66
25~29	-57.02	-0.50	-165.56	-1.67
30~34	-134.70	-1.10	-334.69	-3.58
35~39	-159.31	-1.54	-189.94	-1.65
40~44	-44.41	-0.59	-44.90	-0.37
45~49	89.83	1.16	-13.16	-0.13
合 计	-676.85	-0.95	-1850.91	-2.42

“五普”和“六普”14~16岁人口数量存在重报的现象,那么在16岁以上劳动年龄人口中这一问题是否更加突出?从表4给出的利用户籍登记人口数据估算的“五普”和“六普”14岁以上人口的重(漏)报人数和重(漏)报率可以看出,“五普”和“六普”14~44岁人口重报问题比较突出,并且呈加重的趋势。值得注意的是,14~34岁人口重报的问题在“五普”至“六普”期间愈发凸显,其中14~24岁和30~34岁年龄组变化尤其显著,而这一年龄段恰恰是升学、参加工作及婚育等重大人口事件发生的时期,也是人户分离、迁移流动频发的关键年龄段。

图9给出了“五普”和“六普”14~49岁男性和女性人口重(漏)报情况的差异。总体来看,2000年“五普”男性和女性14~49岁人口重报率差异并不大(男性为1.42%,女性为0.44%),人口重报主要集中在15~20岁,其中男性和女性分别重报了227万和174万,重报率分别为3.77%和3.06%。“六普”时,14~49岁男性和女性人口分别重报了749万和1102万,重报率为1.91%和2.95%。分年龄来看,“六普”男性人口的重报主要集中在15~21岁年龄组,重报了492万,重报率为6.57%;而在23~27岁年龄组主要表现为男性人口漏报,漏报人口为89万,漏报率为1.6%;在30~35岁男性人口重报重新显现,但人口重报的严重程度要小于15~21岁年龄组,重报人数为180万,重报率为3%。在“六普”14~49岁女性中,15~22岁女性人口重报问题最突出,重报人数为649万,重报率为8%,25~35岁女性人口重报的问题虽有所减弱,但仍较明显,重报人数为390万人,重报率为3.8%。

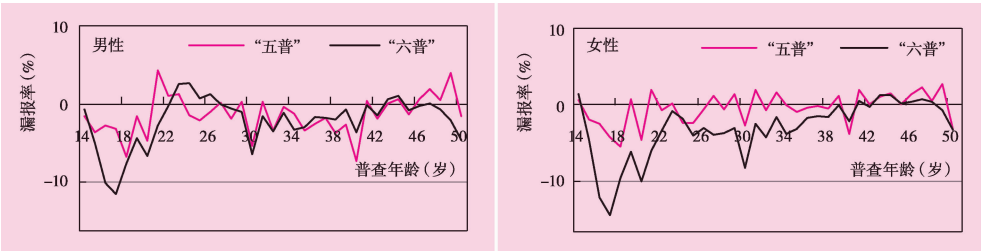


图9 根据2016年户籍登记数据计算的“五普”、“六普”14~49岁以上人口的漏(重)报率

#### 四、数据修正与生育相关问题初探

人口普查低龄数据的准确性直接影响相关人口问题的澄清,其中最为重要的是1982~2010年中国出生人口相关问题。本文将对1982~2010年生育相关问题进行简要探讨。

##### (一) 1982~2010年中国出生人口数量估计

图10给出了利用普查修正数据计算的1982~2010年出生人口数量及与其他来源数据计算结果的差异。根据国家统计局公布的出生率计算,中国出生人口数量从20世

纪 80 年代中后期开始持续下降,21 世纪出生人口数进入了相对平稳状态,2010 年中国的出生人口为 1 600 万左右。值得注意的是,国家统计局公布的 1982~2010 年的出生人口峰值和利用普查数据直接估算的出生人口峰值出现在 1987 年,出生人口数量为 2 500 万~2 600 万人。而利用上文历次普查低龄组漏报率修正后的普查数据估算的这期间的出生峰值出现在 1990 年,约为 2 780 万人;1987~1990 年历年的出生人口数量均在 2 500 万以上。20 世纪 90 年代后,国家统计局公布的出生人口数量开始高于普查估算数量。估算结果显示,1992~2005 年国家统计局公布的出生人口高估的数量均在 100 万人以上,其中 1997 年出生人口数量高估了近 460 万。2006 年以后,国家统计局公布的出生人口又开始有所低估,其中 2007~2010 年的出生人口被低估的数量均在 100 万以上。最后,利用修正后普查数据估算结果显示,中国出生人口数量在经历了 20 世纪 90 年代开始的 10 多年下降后,于 2004 年开始回升,2007 年以后的出生人口数在 1 700 万以上。

图 11 给出了 1982~2010 年中国出生性别比的情况。由于国家统计局公布的出生数据中并没有分性别的出生信息,因此,本研究将根据普查数据直接计算结果和利用历次人口普查低龄组漏报率修正过的结果进行了比较。修正后普查数据估算结果显示,1982 年以来,中国的出生性别比一直低于普查数据直接计算结果。2000 年以后,中国人口出生性别比上升的速度得到一定程度的遏制,但仍处于较高水平。2000~2010 年人口出生性别比为 115~118。根据上文历次人口普查低龄组漏报率情况,本文认为,1982~2000 年中国人口出生性别比可能有所高估(其中,1982~1990 年中国人口出生性别比可能在正常范围的上限附近,为 105~108。但由于女童漏报导致这一时

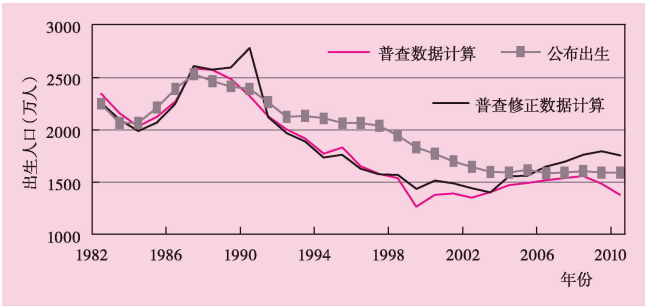


图 10 1982~2010 年不同来源出生人口数量比较

注:1982~1989 年普查修正数据和普查数据估计出生人数的统计时点为 7 月 1 日;1990~2010 年为 11 月 1 日;国家统计局公布出生人口的统计时点为历年的调查时点。

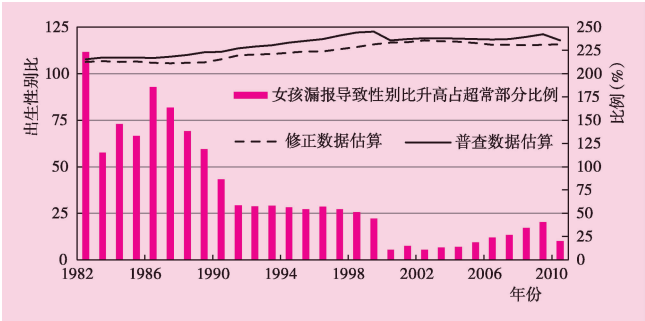


图 11 1982~2010 年中国出生性别比估算结果

注:女孩漏报导致性别比升高占超常部分比例的计算公式为:(普查数据估算性别比-修正数据估算性别比)/(普查数据估算性别比-107)。

期的人口出生性别比的估计值在107~112之间)。

(二) 1990~2010 年中国生育水平估计

总和生育率是衡量一个国家和地区生育水平高低的重要指标。1990~2010 年非普查年份的育龄妇女生育完成情况可根据“四普”、“五普”和“六普”的情况线性内插得到。需要说明的是,上文已证明中国近两次普查数据低龄组存在较为严重的人口漏报,而中青年组存在较为明显的人口重报,因此本文计算的 1990~2010 年中国总和生育率高于利用历年公布数据计算的总和生育率。

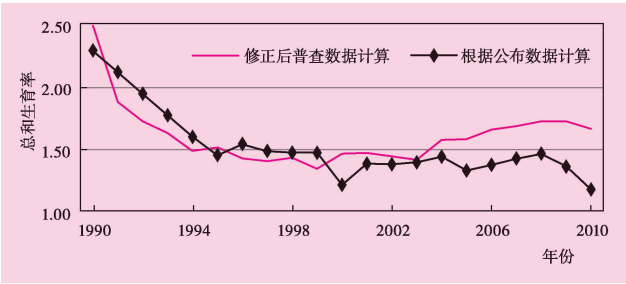


图 12 1990~2010 年利用不同数据和方法估算的总和生育率

注:(1)修正后普查数据计算结果的统计时点为历年 11 月 1 日,根据公布数据计算结果的统计时点为历年调查时点。(2)公布数据来自《中国人口统计年鉴》和《中国人口与就业统计年鉴》,其中人口普查年份和 1%人口抽样调查的年份为人口普查和 1%人口抽样调查结果。(3)由于 1991~1993 年国家统计局未公布生育率情况,故以线性内插结果替代。

左右,这与直接利用国家统计局公布数据计算的生育水平(1.42)相差不大。2005~2010 年为第三阶段,利用修正后普查数据估算的中国生育水平有所回升,平均总和生育率为 1.68,其中 2008 年和 2009 年的总和生育率均在 1.7 以上。利用历年公布数据计算的 2005~2010 年的平均总和生育率仅为 1.36,并没有出现明显的上升趋势。

五、结论与讨论

本研究利用 2017 年户籍登记数据重新估计了 1982~2010 年 4 次人口普查 0~16 岁低龄人口的数据情况,得出以下结论。

第一,1982~2010 年中国历次人口普查 0~16 岁低龄组人口数据均存在漏报的问题并且有逐渐加剧的趋势,但是,2000 年“五普”低龄组的漏报水平似乎并没有之前估计的那么高。“三普”至“六普”0~9 岁低龄人口的漏报率分别为 1.85%、5.11%、6.43%和 8.82%,漏报人数为 368 万、1 077 万、1 011 万和 1 296 万。与其他学者的估算结果相比,

本研究估算的“四普”和“五普”0~9岁人口的漏报率和漏报人数相对较少,其中“四普”0~9岁漏报人口的估计数量比其他学者(张为民、崔红艳,2003;周皓,2003;王金营,2003)的估算少了70万~190万。对于“五普”数据,现有研究普遍认为“五普”低龄人口的漏报率是最高的,0~9岁人口的漏报率超过10%(崔红艳等,2013)。但本文认为,由于“六普”青少年人口(特别是16~19岁年龄组)数据中存在较为严重的人口重报现象,因此利用“六普”数据估计“五普”低龄人口漏报数量很可能存在高估,而“五普”低龄人口数据的实际漏报率可能与“四普”相差不大。由于可利用的资料有限,现有研究对“六普”0~9岁数据漏报情况的估计比较保守。

第二,劳动年龄人口的重报问题在“五普”“六普”中愈发凸显。“五普”和“六普”14~44岁的净重报人口数由766万上升至1838万。其中14~29岁的重报人口分别由“五普”的428万上升至“六普”的1268万。此外,“六普”14岁以上人口的漏(重)报模式存在明显的性别差异。“六普”的人口重报问题不仅体现在青年组,而是会持续至40岁,并在女性人口中更加突出。

第三,20世纪90年代中期以后中国开始进入中低生育率阶段,但由于出生人口的漏报和育龄妇女人数的重报,21世纪前10年中国的生育水平可能被低估。本研究认为,1990~1995年中国的总和生育率为2.5~1.5;1996~2004年为1.45左右;2005~2010年逐渐上升至1.7附近。

第四,由于人口普查和人口抽样调查数据低龄人口的漏报,中国1982~2010年的出生人口性别比在一定程度上被高估,其中20世纪80年代初至90年代初尤为突出。21世纪以来,中国人口出生性别比上升速度得到遏制,但仍无法改变中国出生性别比偏高的客观事实,2000~2010年中国的出生性别比估计值为115~118。

值得一提的是,尽管本研究基于户籍登记数据对人口普查中低龄人口数据进行评估是一个十分有益的尝试,但本文的概念界定、估算方法及研究结果中也存在一些值得商榷之处:(1)低龄组的人口的界定。以往研究大多将低龄组人口定义为0~9岁,而本研究将低龄组定义为0~16岁。这主要有两个考虑。一是16岁以下非劳动年龄人口出现漏(重)报的可能性更大;二是青少年人口漏(重)报模式可能会与儿童有很大的不同。为了便于与其他研究比较,本研究最后也计算了0~9岁人口的漏(重)报率。(2)户籍登记数据中的年龄误报。虽然本研究认为户籍登记年龄数据的质量在可以接受的范围内,但是户籍登记人口的年龄误报是客观存在的,目前尚没有研究对这一问题进行系统分析。本研究利用户籍登记单岁组数据进行漏(重)报率的测算,一定程度上会受到年龄误报的影响,如果对几个年龄组合并计算可能可以减少户籍登记数据年龄漏报的影响。(3)年龄移算的假定。一般进行年龄移算时通常假定一年内出生人口和死亡人



口是均匀分布的,对于特定的年龄组而言,该假定可能对出生人口的估计产生一定影响,且影响方向不确定,这一问题还有待深入探讨。

虽然户籍登记数据中依然存在一些问题,但是户籍登记数据质量在稳步提升。考虑到户籍登记不同于人口普查的数据收集方式,在很大程度上克服了人口漏报的问题。将户籍登记数据于人口普查数据相结合,可以从不同的角度对人口普查数据,特别是低龄组人口数据进行考察,同时,户籍登记数据的连续性也要优于普查,这对于及时准确的评估低龄组人口的漏报也有一定的帮助。

参考文献:

1. 陈卫(2014):《2000年以来中国生育水平评估》,《学海》,第1期。  
 2. 崔红艳等(2013):《对2010年人口普查数据准确性的估计》,《人口研究》,第1期。  
 3. 郭志刚(2004):《对“五普”和“四普”队列人口一致性的再检验》,《中国人口科学》,第6期。  
 4. 胡耀岭、原新(2013):《1982~2010年期间四次全国人口普查数据一致性研究——基于出生人口队列的分析》,《人口研究》,第3期。  
 5. 李波、姜全保(2010):《基于五普数据对四普数据质量的估计》,《西北人口》,第3期。  
 6. 李若建(2013):《历次人口普查中低年龄组人口漏报研究》,《中山大学学报(社会科学版)》,第2期。  
 7. 李树苗等(2006):《“五普”人口总量和结构的分析与调整》,《人口学刊》,第5期。  
 8. 陶涛、张现苓(2013):《六普人口数据的漏报与重报》,《人口研究》,第1期。  
 9. 王广州(2003):《对第五次人口普查数据重报问题的分析》,《中国人口科学》,第1期。  
 10. 王金营、戈艳霞(2013):《2010年人口普查数据质量评估以及对以往人口变动分析校正》,《人口研究》,第1期。  
 11. 王金营(2003):《2000年中国第五次人口普查漏报评估及年中人口估计》,《人口研究》,第5期。  
 12. 杨凡、赵梦涵(2013):《2000年以来中国人口生育水平的估计》,《人口研究》,第2期。  
 13. 翟振武等(2015):《现阶段中国的总和生育率究竟是多少?——来自户籍登记数据的新证据》,《人口研究》,第6期。  
 14. 张青(2005):《用小学入学人数检验“五普”低龄组的人口漏报和性别比》,《中国人口科学》,第3期。  
 15. 张为民、崔红艳(2003):《对中国2000年人口普查准确性的估计》,《人口研究》,第4期。  
 16. 周皓(2003):《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》,《人口研究》,第2期。  
 17. Ewbank D.C.(1981), *Age Misreporting and Age-Selective Underenumeration: Sources, Patterns, and Consequences for Demographic Analysis*. Washington: National Academy Press.  
 18. Spoorenberg T. and Dutreuilh C.(2007), *Quality of Age Reporting: Extension and Application of the Modified Whipple's Index*. *Population*. 62(4): 729-741.

(责任编辑:李玉柱)