

# 人口普查数据与行政统计数据偏离现象研究

——以中国四大城市为例

陈友华, 孙永健

(南京大学 社会学院, 江苏 南京 210023)

**摘要:** 人口统计数据是政策制定与规划编制的基础, 但中国来自不同渠道的人口统计数据之间却存在明显的差异, 而针对这类差异的纵向特征、形成原因与对策建议的研究相对不足, 且缺少区域性视角。借助地方政府历年发布的《国民经济和社会发展统计公报》、《统计年鉴》等来源资料, 梳理了2000—2020年间北上广深四大城市的人口统计数据, 通过构建“人口增长速度”和“行政统计偏差率”指标, 对四大城市历年人口普查数据与行政统计数据进行比较分析。研究结果显示: 四大城市行政统计人口值变动存在明显的非连续性, 每逢人口普查年, 人口会“突生性”地增长, 远远高于相邻非人口普查年份的人口增长速度。四大城市行政统计人口值常年小于人口普查修正值, 存在系统性低估。非人口普查年份期间, 行政统计值每年都会“遗漏”部分人口, 累积形成“数量缺口”, 而每逢人口普查年, 行政统计值便以普查值为准进行纠偏, “一次性修正”前9年中累积的“遗漏人口”。由此勾勒出人口行政统计“系统性遗漏”的偏离曲线, 揭示了人口普查与行政统计之间的偏离怪象, 并对其形成原因加以分析。基于此, 建议从三方面应对人口行政统计的系统性偏差: 一是转变人口控制的政策理念与实践, 重新审视由系统性遗漏所带来的政策偏差与风险; 二是辩证地看待行政统计的作用与限度, 充分理解行政统计的局限; 三是加强行政统计工作的科学管理。本研究成果为政策文本与科学研究中重视与规避人口统计偏差提供了某种警示作用与启发意义。

**关键词:** 人口普查; 行政统计; 系统性偏差

**中图分类号:** C921.2   **文献标识码:** A   **文章编号:** 1000-4149 (2022) 03-0086-11   **DOI:** 10.3969/j.issn.1000-4149.2022.00.005

收稿日期: 2021-07-15; 修订日期: 2022-01-24

基金项目: 国家社会科学基金重大项目“实现积极老龄化的公共政策及其机制研究”(17ZDA120)。

作者简介: 陈友华, 南京大学社会学院教授, 博士生导师; 孙永健, 南京大学社会学院博士研究生。

## 一、引言

人口是一切社会经济活动的基础, 人口统计数据是政府中长期规划与各项经济社会政策制定的基础与依据。我国人口统计数据的收集主要以国家统计局组织的人口普查为基础, 两次普查之间的1%人口变动抽样调查和每年一次的人口变动抽样调查作为收集人口数据的辅助手段<sup>[1]</sup>, 形成了定期人口普查、人口变动抽样调查与常规行政统计等多元人口统计体系。

然而, 我国不同渠道与口径下的人口统计数据之间却存在明显的差异。例如, 第七次全国人口普查公报一经公布, 便引起了社会各界的广泛热议, 特别是第七次全国人口普查数据与往年人口数据之间难以相洽而导致了所谓的“数据之惑”。其实, 以往研究早已捕捉到了人口普查数据与行政统计数据之间的某种偏离现象。钟水映在中国人口统计误差分析中提到, 常规的人口统计和人口普查之间存在人口规模上的差异<sup>[2]</sup>。杨贵军等也发现中国非普查年度人口总数常常与普查年度人口总数之间存在较大差异<sup>[3]</sup>。还有很多研究则从流动人口、出生人口、生育率等维度出发, 探讨了不同统计渠道的人口指标之间的区别与矛盾<sup>[4-6]</sup>。不过, 上述研究往往对这种差异本身的纵向特征、形成原因与对策建议讨论不足, 且缺少有针对性的区域视角的研究。

本文选取北京、上海、广州、深圳四大城市为研究对象, 旨在回顾与总结以往大城市人口普查数据与行政统计数据之间的差距, 描述两者之间的偏离程度和纵向变化特征, 进而分析偏差来源及其形成原因, 提出相应的治理对策, 以期在未来政策文本与规划编制中尽量规避统计偏差所引发的种种负面效应。

## 二、数据、指标与方法

### 1. 数据说明

人口普查是世界各国或地区人口统计的基本方法, 中国人口普查每十年组织一次, 并且在严谨调查设计基础上, 投入了大量的人力、物力和财力。尽管也有不少学者质疑人口普查本身的质量<sup>[1,7]</sup>, 但当其他类型人口调查与人口普查数据发生矛盾时, 一般还是以人口普查为基准来纠正、覆盖其他类型人口调查数据, 而不是相反<sup>[8]</sup>。人口普查相较于其他人口调查, 属于更全面、更权威的权威数据源, 通常被作为基期资料使用<sup>[9-10]</sup>。特别是第七次全国人口普查首次使用了电子登记与身份证验证等信息技术, 加之疫情防控与不经地方汇总的影响, “七普”数据质量得以极大提升。实际上, 许多西方国家和联合国人口统计司都会在人口普查数据公布之后, 对以往非人口普查年度的各类社会经济指标进行修订<sup>[3,11]</sup>, 中国自然也不例外。

本文选取四大城市行政统计人口数和普查修正人口数作为关键研究变量。行政统计人口数来自于各地方政府历年发布的《国民经济和社会发展统计公报》, 该系列数据由地方统计局或国家统计局地方调查队负责收集, 数据公布较为及时, 但仅涉及总人口及其年龄与城乡构成、出生、死亡、流动等少数几个主要人口数据。

经人口普查数据修正后的人口数值主要来源于各级统计部门发布的《统计年鉴》。该类

《统计年鉴》为按年连续出版的大型统计资料书，大部分城市<sup>①</sup>的《统计年鉴》会根据最新人口普查数据对以往年份常住人口行政统计数据进行了修正，以此来纠正先前的数据记录。例如，《北京统计年鉴 2021》在“常住人口（1978—2019年）”数据集中就注明“2006—2009年常住人口、出生率、死亡率等数据根据2010年人口普查数据进行了修订；2011—2019年数据根据2010年、2020年两次人口普查结果进行了修订”，而此前根据《北京市国民经济和社会发展统计公报》汇总的人口数据则被更正。

## 2. 指标构建与方法设计

第一步，本文借助“人口增长速度”作为辅助考察指标。人口增长速度是指该年的行政统计人口增加数与上一年度人口数之比，计算方法如下：

$$\text{第 } N \text{ 年人口增长速度} = \left( \frac{\text{第 } N \text{ 年行政统计人口数}}{\text{第 } N - 1 \text{ 年行政统计人口数}} - 1 \right) \times 100\% \quad (1)$$

可以用上述公式考察行政统计公布的相邻两年的人口增长之间的非连续性或跳跃性，该指标更加直观地反映行政统计值的变化规律，由此判断人口统计数值的合理性。

第二步，本文以人口普查修正后的人口数为准线，构建了行政统计人口数的偏差率指标，进而对各地区不同年度的人口普查与行政统计展开差距分析，计算方法如下：

$$\text{行政统计偏差率} = \left( \frac{\text{行政统计人口数}}{\text{普查修正人口数}} - 1 \right) \times 100\% \quad (2)$$

如果该指标大于0，意味着行政统计人口数大于人口普查修正数，非人口普查年度行政统计人口数高估了人口规模，反之则意味着行政统计人口数小于人口普查修正数，非人口普查年度行政统计人口数低估了人口规模。相对于其他社会经济指标而言，人口规模是一个惰性与累积变量，即使考虑到城际间的人口流动，人口规模在短时间内的变动幅度一般也很小。此外，北上广深等超大城市的人口体量庞大，哪怕是1%的偏差也可能导致十万级的人口误差。鉴于此，本文在定义偏差率的判别标准时给出的阈值都较低，即判定数值都相对很小：以0为临界值，当该指标的绝对值小于1%时，表示行政统计人口数较为准确；当该指标的绝对值介于1%到2%时，表示行政统计人口数存在轻微偏差；而当该指标的绝对值大于2%时，则表示行政统计人口数出现了严重偏差。

## 三、人口普查与行政统计之间的系统性偏离

### 1. 2000—2020年行政统计口径下的人口增长速度

表1汇总了2000年以来北上广深四大城市在行政统计口径下的人口增长速度，结果表明：第一，四大城市作为人口大规模迁入地，城市人口规模长期保持增长。第二，受到种种人口与政策因素，特别是超大城市人口调控政策的影响，自2015年后，北京和上海部分年份的人口规模开始下降。第三，人口变动存在明显的跳跃性或非连续性，具体表现为每逢人口普查年，人口会“突生性”地增长，远远高于相邻非人口普查年份的人口增长速度。由此可以推断，大城市非人口普查年份的人口增长速度存在被严重低估的倾向，并且这种低估

<sup>①</sup> 也有部分省市事后不对人口行政统计数据进行了修正，但我们也不能因此推断其行政统计数据必然准确。

具有系统性与长期性, 期间“遗漏”的部分人口会以人口普查年份“突生性增长”的方式得到“补偿”。

## 2. 2000—2010 年普查与行政统计间差距分析

如果行政统计数据质量较高, 则行政统计人口数与人口普查修正值之间应该比较接近, 行政统计偏差率理应围绕 0 上下浮动, 系统性高估或低估均表明行政统计出现了较严重的偏差。表 2 呈现出四大城市在“五普”与“六普”期间 (2001—2009 年) 的行政统计人口数及偏差率, 揭示出某些人口是如何在非普查年份被系统性遗漏, 又如何在“六普”结果公布后被一次性修正。具体而言, 北上广深四大城市的行政统计值在人口普查年会定时回归人口普查值, 但在非人口普查年份均系统性地低于人口普查修正值, 从而长期低估了城市的人口规模, 其中上海最为严重, 广州、深圳和北京次之。

表 1 2001—2020 年四大城市行政统计人口增长速度 %

年份	北京	上海	广州	深圳
2001	1.44	0.34	—	—
2002	2.88	0.68	—	—
2003	2.33	5.29	—	—
2004	2.49	1.81	—	—
2005	3.03	2.07	—	—
2006	2.80	2.08	2.71	2.26
2007	3.29	2.37	2.99	1.79
2008	3.80	1.64	1.36	1.77
2009	3.54	1.74	1.50	1.64
2010	11.79	19.85	22.98	16.38
2011	2.89	1.95	0.33	0.92
2012	2.51	1.40	0.69	0.76
2013	2.20	1.46	0.68	0.77
2014	1.74	0.44	1.19	1.41
2015	0.88	-0.43	3.22	5.56
2016	0.11	0.18	4.02	4.66
2017	-0.10	-0.06	3.24	5.21
2018	-0.76	0.23	2.80	3.98
2019	-0.03	0.18	2.69	3.16
2020	1.66	2.43	22.02	30.67

数据来源: 行政统计人口数来源于各级政府发布的《国民经济和社会发展统计公报》中的常住人口数。

表 2 2001—2009 年四大城市人口行政统计及系统性低估

万人, %

年份	北京			上海			广州			深圳		
	行政统计值	普查修正值	偏差率	行政统计值	普查修正值	偏差率	行政统计值	普查修正值	偏差率	行政统计值	普查修正值	偏差率
2001	1383.3	1385.1	-0.13	1614.0	1668.3	-3.26	—	—	—	—	—	—
2002	1423.2	1423.2	0.00	1625.0	1713.0	-5.14	—	—	—	—	—	—
2003	1456.4	1456.4	0.00	1711.0	1765.8	-3.11	—	—	—	—	—	—
2004	1492.7	1492.7	0.00	1742.0	1835.0	-5.07	—	—	—	—	—	—
2005	1538.0	1538.0	0.00	1778.0	1890.3	-5.94	949.7	949.7	0.00	827.8	827.8	0.00
2006	1581.0	1601.0	-1.25	1815.0	1964.1	-7.59	975.5	996.7	-2.13	846.4	871.1	-2.83
2007	1633.0	1676.0	-2.57	1858.0	2063.6	-9.96	1004.6	1053.0	-4.60	861.6	912.4	-5.57
2008	1695.0	1771.0	-4.29	1888.5	2140.7	-11.78	1018.2	1115.3	-8.71	876.8	954.3	-8.12
2009	1755.0	1860.0	-5.65	1921.3	2210.3	-13.07	1033.5	1187.0	-12.93	891.2	995.0	-10.43

数据来源: 行政统计人口数来源于各级政府发布的《国民经济和社会发展统计公报》中的常住人口数; 普查修正后的人口数来源于各级政府发布的《统计年鉴》中的常住人口数。

注: 广州、深圳等市在 2005 年以前的《国民经济和社会发展统计公报》中缺失常住人口项或统计质量存在严重问题, 因此本文对其 2001—2004 年的行政统计不予考察。

从偏差率的变动趋势来看: 北京 2001—2005 年常住人口的低估程度相对较小, 而越接近 2010 年低估的程度越严重, 偏差率的绝对值呈逐年放大之势, 反映出系统性偏差的累积效应。上海 2001 年常住人口的低估程度已较为明显, 而越接近 2010 年低估的程度越严重, 偏差率的绝对值同样呈逐年放大之势。自 2005 年起, 广州和深圳的偏差率也呈现类似的变

动规律。

### 3. 2010—2020年普查与行政统计间差距分析

表3同样以北上广深四大城市为例,考察其“六普”与“七普”之间(2011—2019年)行政统计人口数及偏差率,再次揭示了常住人口“从系统性遗漏到一次性修正”的偏离怪象。相较之下,此阶段各市人口的行政统计均存在系统性低估问题,且愈发明显。其中,广州和深圳最为严重,广州最大偏差率高达-17.11%,深圳更是达到了-21.81%。上海、北京次之,上海最大偏差率为-2.14%,而北京则不超过-2%,这或与近年来北京市、上海市人口调控政策实施有关<sup>①</sup>,也可能受到生育政策逐步放开的少许影响<sup>②</sup>。

表3 2011—2019年四大城市人口行政统计及系统性低估

年份	北京			上海			广州			深圳		
	行政统计值	普查修正值	偏差率	行政统计值	普查修正值	偏差率	行政统计值	普查修正值	偏差率	行政统计值	普查修正值	偏差率
2011	2018.6	2023.8	-0.26	2347.5	2355.5	-0.34	1275.1	1346.3	-5.29	1046.7	1123.3	-6.81
2012	2069.3	2077.5	-0.39	2380.4	2398.5	-0.75	1283.9	1415.5	-9.30	1054.7	1196.3	-11.83
2013	2114.8	2125.4	-0.50	2415.2	2448.4	-1.36	1292.7	1472.2	-12.20	1062.9	1257.3	-15.46
2014	2151.6	2171.1	-0.90	2425.7	2467.1	-1.68	1308.1	1529.0	-14.45	1077.9	1324.1	-18.59
2015	2170.5	2188.3	-0.81	2415.3	2457.6	-1.72	1350.1	1595.0	-15.35	1137.9	1414.2	-19.54
2016	2172.9	2195.4	-1.02	2419.7	2467.4	-1.93	1404.4	1678.4	-16.33	1190.8	1501.5	-20.69
2017	2170.7	2194.4	-1.08	2418.3	2466.3	-1.94	1449.8	1746.3	-16.98	1252.8	1587.3	-21.07
2018	2154.2	2191.7	-1.71	2423.8	2475.4	-2.08	1490.4	1798.1	-17.11	1302.7	1666.1	-21.81
2019	2153.6	2190.1	-1.67	2428.1	2481.3	-2.14	1530.6	1831.2	-16.42	1343.9	1710.4	-21.43

数据来源:行政统计人口数来源于各级政府发布的《国民经济和社会发展统计公报》中的常住人口数;普查修正后的人口数来源于各级政府发布的《统计年鉴》中的常住人口数。

从偏差率的变动趋势来看,广州、深圳、上海和北京的偏差率的绝对值呈逐年放大态势,同样出现了系统性低估的累积效应。

### 4. 偏离曲线:从系统性遗漏到一次性修正

总体而言,我国诸多大城市出于对人口总量控制、流入人口限制以及计划生育的响应,这些城市的行政统计人口数常年低于人口普查修正数,存在系统性低估。大城市流动人口占流入地人口比重大,对流入地人口指标的影响显著<sup>[12]</sup>,加之北上广深等特大城市作为重要的经济或政治中心,其人口数量问题常年饱受社会各界热议。因此,本文聚焦于此类人口集聚与经济发达城市的人口统计问题,由此勾勒出人口行政统计“从系统性遗漏到一次性修正”的偏离曲线,呈现出人口普查与行政统计之间的偏离怪象,如图1所示。

分析偏离曲线可知:①非人口普查年份,大城市人口的行政统计值系统地低于普查修正值,期间每年都会“遗漏”部分人口;②随着时间的推移,行政统计“系统性遗漏”的人口会累积,造成人口普查年度的前一年(如 $n+9$ 年、 $n+19$ 年……)产生明显的“数量缺口”;③每逢人口普查年,人口的行政统计值便以普查值为准进行纠偏,“一次性修正”前9

① 自2014年以来,北京、上海一方面采取严厉的人口调控政策,防止外来人口过度聚集,另一方面与其他一二线城市一同加入“人才争夺战”之中,利用其超大城市优势试图吸引大量人才。这些效果卓著但方向相反的人口政策最终使得北京、上海人口快速增长的态势逐渐放缓,甚至偶有下降,也打破了人口自然与机械增长的某种连续性。

② 党的十八大以来,国家先后对生育政策进行了三次调整,分别于2014、2016与2021年实施“单独两孩”、“全面两孩”与“三孩”生育政策,进而减少行政统计人口系统性低估程度。

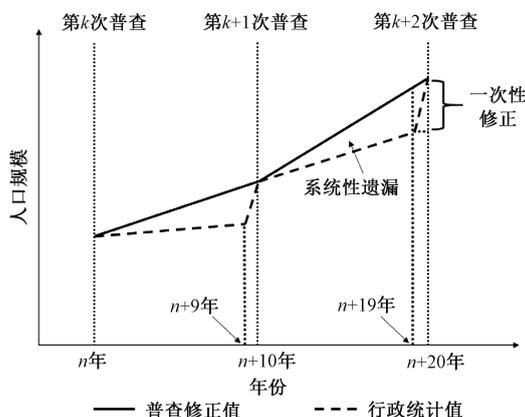


图1 大城市人口行政统计数据“从系统性遗漏到一次性修正”的偏离曲线

年中累积的“遗漏人口”，两条曲线发生短暂重合；④行政统计口径的人口增长速度会明显小于普查口径的人口增长速度，并且前者的纵向变化具有明显的非连续性。

需要说明的是，偏离曲线的勾勒是基于对典型大城市的实证考察与理论抽象而形成的，实际的行政统计人口数与人口普查修改人口数之间的关系要比偏离曲线复杂得多。这里的偏离曲线十分类似于韦伯（Weber）所说的“理想类型”<sup>[13]</sup>，在现实生活中可能并不完整存在。换言之，现实人口的行政统计值与人口普查修正值均不是完美的线性曲线，常出现波动式变化。此外，并非所有大城市的人口行政统计均符合偏离曲线的变动规律，部分城市在某些时间节点上可能具有理想类型的主要特征，但其余时间段则表现不明显。

#### 四、人口普查与行政统计之间偏离的原因与对策建议

##### 1. 人口普查与行政统计之间偏离的成因

大城市行政统计人口数与人口普查修正人口数之间发生偏离有其复杂的成因，既有客观层面的因素，这是任何社会科学调查研究都难以避免的，属于内在固有因素；也有主观层面的成分，即会受到制度、经济和社会等诸多因素的影响，它们属于外在干扰因素。

第一，人口调查对象具有复杂性和不可重复性。任何形式的人口调查，无论调查方案设计得如何科学、调查过程控制得如何严密、资料处理得如何规范，都仍将受制于调查者和被调查者自身的主观倾向、客观条件、调查目的等因素的影响与制约。因此，在种种现实条件束缚下，人口统计中出现的某些偏差本就不可能彻底消除。

第二，相对于人口普查而言，行政统计所需的人力、财力、时间等基础条件相对缺乏。政府统计工作者和基层单位统计人员承担的调查任务非常繁重，除了定期开展的各项社会经济普查与抽查，还包括其他规模不等的调查和定期统计报表等<sup>[14]</sup>。加之，行政统计填报时间短、调查频率较高、经费相对不足、实际调查难度大，种种因素致使行政统计对人口总量的精准统计更加棘手。

第三，人口迁移与流动为人口统计调查带来了极大挑战。随着对人口迁移流动限制的逐

步松动,大规模的人口迁移流动成为常态,2020年第七次全国人口普查资料显示,中国的迁移流动人口数量仍在持续增加。人口迁移流动更可能导致区域人口统计指标失真<sup>[12]</sup>。更为重要的是,人口普查与行政统计有关流动人口的调查口径在流动时间、空间单元和登记方式等方面并不完全统一,进而导致统计结果之间的偏差。例如,针对流动人口数量庞大的现实,“六普”采用了双口径方法,流动人口在户籍地和常住地同时登记,再由人口普查部门进行“去重”,大大降低了漏登率。而大城市在行政统计过程中,一方面仅能从本地(即人口流入地)展开调查,难以实施双口径法,另一方面对于常住人口的判定也更为严格,结果促成了“系统性的遗漏”。

第四,与人口普查相比,人口行政统计的规范性和客观性有待提升。郭志刚对历年统计公报中人口指标的研究发现,在没有严格、详细的人口推算与统计测算的前提下,个别指标经历了有针对性的明显调整,结果是历年公报的人口统计与实际调查结果之间出现了一定的差异<sup>[15]</sup>。而一旦统计年报中的人口统计值被大大低估,由此进行各项人均指标换算与检验时,就会产生人均占有数高估的现象,最典型的是北上广深等流动人口流入较多的大城市的人均GDP、人均收入等统计指标可能被大大高估。

## 2. 人口行政统计偏差的治理对策

(1) 辩证看待人口行政统计的作用与限度。本文所揭示出的系统性遗漏这一客观事实,无疑促使我们重新审视与反思人口行政统计的作用。

一方面,政府与学界应当充分正视行政统计的偏差所在,在相关政策制定与学术研究中要特别重视大城市常住人口被低估以及其后果。行政统计得到的人口年报是中央和地方政府决策时的重要依据,也是事后考核国民经济和社会发展成果、居民生活水平的关键指标。各区域人口是最基本、最重要的国情国力统计,严重偏误的人口数据不仅对国民经济与社会发展规划极为不利,也将误导后续政策制定与变革,无益于诸多社会经济问题的化解。因此,在对区域人口(特别是大城市人口)进行抓取与分析时,应当主要以人口普查数据为准,非普查年则以事后人口普查修正值为准。若暂时无法获取最新《统计年鉴》或还未等到新一轮人口普查数据,则应对行政调查数据有甄选、有条件地使用,必要时可进行适当的调整与修正。

另一方面,也不可全盘否定人口行政统计的价值。行政统计存在的意义本就是为了弥补人口普查耗时耗力的缺陷,从而在有限的时间与资源约束下,为人们连续性地了解人口数量及其变化提供某种可能。人口普查与行政统计的定位在根本上存在差异,因此不能以前者的高标准来要求后者,要辩证地看待行政统计的作用和限度。

(2) 转变人口控制为主的政策理念与实践。首先,深刻反思过去以人口控制为主基调的人口政策理念。我国人口政策长期以来是以人口预测为基础,以人口与计划生育政策为干预手段,以实现预期的人口控制目标<sup>[16]</sup>。无论是全国层面持续数十年的严厉计划生育政策,还是区域层面“控制大城市规模,合理发展中等城市,积极发展小城市”的政策指导,都使人们对待人口数量问题特别敏感与谨慎,对于人口集中与产业集聚的大城市而言更是如此。及时改变人口控制思想,有利于人口统计工作回归科学范畴,也可促使人口普查与行政

统计之间的偏差分布呈现方向随机性而非单一性。

其次,尽快推动户籍制度的改革与完善。人口变动有其自身的发展变化规律,并不完全取决于人口政策规定。例如,不少学者考察了北京市人口规划工作的发展历史,发现北京近30年来的人口发展是一个不断突破总规化目标的过程,人口规模预期指标长期低于实际人口<sup>[17-20]</sup>。其实,人口本就存在一种基于理性选择下的自合理分布机制<sup>[21-22]</sup>,但户籍制度及在此基础上的包括教育、就业、医疗与社会保障在内的一整套制度安排却事实上制约着劳动力的自由流动,使得许多大城市常住居民迟迟未能享有同等的劳动力市场机会、社会保障和公共服务。因此,流动人口很难在流入地定居下来并获得当地的户籍,取而代之的是频繁的短期流动,而这在春节等时期表现得尤为明显,不仅给人口的年度统计带来了巨大困难,而且还可能因人口流动导致疫情扩散等社会问题。

最后,重新审视由系统性遗漏所带来的政策偏差与风险,并采取相应的补救措施。对大城市而言,系统性遗漏可能会导致政府规划对城市规模的预计不足,引发公共服务供给不足,所提供的土地、住宅、基础设施、公共服务将不能满足实际人口需要。如若政府不能意识到行政统计对人口的低估并及时纠偏,反而将各类“大城市病”归咎于外来人口,就会引发新一轮对人口规模的担忧与低估。因此,借助人口普查及修正值,应重新评估与反思过往的部分政策实践,对于由系统性低估造成的公共资源配置不足,要及时补充与完善,更好地平衡人口与经济社会之间的关系。

(3) 加强人口行政统计的科学管理。首先,明确与统一人口指标(特别是迁移流动人口)的概念与口径。西方国家由于实行自由迁徙,并且建立了覆盖全国的社会保障体系,因此对人口迁移、流动的概念界定相对单纯<sup>[23]</sup>。基于中国的户籍制度,统计、公安、卫生健康等部门出于不同的行政目的,常常对同一群体展开调查,过程中采用的概念分类与统计口径不尽相同,其结果会造成不同来源渠道数据之间的偏差。故而,为了提高人口行政统计的准确性与可比性,应尽快明确和统一不同区域、各级政府的人口调查中的指标使用,尽可能以人口普查的概念界定与统计口径为指导标准。

其次,进一步明晰人口统计工作的责任主体,建立信息共享机制。尽管统计局是中国各级政府的官方统计机构与官方权威发布机构,但人口行政统计主要有国家统计局、公安部、国家卫生健康委员会(主要是出生数据)三大数据来源。此外,也与民政、教育、人社、街道社区、企业的信息系统密切关联。人口行政统计质量的提高有赖于责任制的健全,应将责任落实到少数且具体的机构,为后续偏差追责与质量监管提供明确主体。

再次,保障行政统计的经费投入,提高统计队伍的政策水平和工作能力。要确保人口统计数据质量,需要有充足的资金辅助运转,进而才能够在保证统计工作顺利完成的基础上提升人口统计数据的质量。但与此同时,人口统计也要注重提高资金利用率,将经费使用情况纳入年终考核目标,防止财政资源的错配与浪费。在“财力”有所保障的前提下,不断强化“人力”支持,加强统计队伍建设,加大统计工作培训力度,不断夯实年度人口统计基础工作。

最后,充分利用现代信息技术与手段。随着科技进步,我国已经步入数字信息时代,人

口统计调查也需要紧跟时代步伐。现代信息技术可以帮助人口调查简化工作流程、提高工作效率、改善数据质量。例如,大数据与云计算技术便能够搜集全国统一且人人独有的鉴别码来提高人口数据匹配的速度和准确性,从而更好地剔除统计中的重报<sup>[24]</sup>，“七普”人口数据质量的显著提升便得益于此。作为经济发达地区,大城市具备更优越的科技支持与数字条件,因此其人口行政统计与信息技术展开融合会更为便捷,也更容易向“七普”技术看齐。

## 五、结论与讨论

本文汇总了2000—2020年间中国四大城市的人口普查与行政统计数据,通过构建“行政统计偏差率”这一指标,对各市历年人口普查修正值与行政统计值进行差距分析。结果显示,北京、上海、广州、深圳等经济发达地区的行政统计人口数常年低于人口普查数,存在系统性低估。基于对偏差率的纵向分析,本文勾勒出人口行政统计“从系统性遗漏到一次性修正”的偏离曲线,揭示了人口普查与行政统计之间的偏离怪象。这种系统性偏离,既源于社会调查固有局限等客观因素,也受到制度、经济和社会等诸多干扰因素的影响。

针对大城市人口行政统计的系统性遗漏问题,本文提出相应的治理对策:一是辩证地看待人口行政统计的作用与限度,既要充分正视行政统计的偏差所在,在相关政策制定与学术研究中要重视大城市常住人口的低估及其后果,也要理解行政统计自身的局限,不可吹毛求疵;二是转变人口控制为主的政策理念与实践,尽快推动户籍制度及其在此基础上的一整套制度的改革与完善,重新审视由系统性遗漏所带来的政策偏差与风险;三是加强人口行政统计的科学管理,明确与统一人口指标的概念与口径,明晰人口统计工作的责任主体,保障行政统计的经费投入量与资金利用率,提高统计人才队伍的政策水平和工作能力,充分结合现代信息技术与手段。

“中国究竟有多少人”、“大城市究竟能容下多少人”等均是严肃且重要的议题,我们要充分认识到几个百分点的偏差也可能造成相去甚远的结论,任何政策制定与科学研究对于人口数据的使用都应该审慎为之。本文在有限数据资料的基础上对人口普查与行政统计之间的偏离现象进行了初步研究,揭示出行政统计对大城市人口的系统性误判及其成因。尽管人口统计调查中存在不可回避的本源性缺陷,但很多人干扰因素是能够通过制度规范得到解决的。本文无意于否定人口行政统计的意义,反而希望未来各项人口调查质量能有所提升。希冀更多的人加入相关研究中来,以推动中国人口统计更好地回归科学价值与政策指导意义。

### 参考文献:

- [1] 乔晓春, 江益. 对近年来中国人口统计学学科发展状况的评述 [J]. 人口与经济, 2008 (6): 49-55.
- [2] 钟水映. 对中国人口总量统计误差的研究应该慎之又慎——兼与梁中堂教授商榷 [J]. 中国人口科学, 2004 (2): 76-80.
- [3] 杨贵军, 孟杰, 李楠. 中国省域人口总数修订 [J]. 统计与信息论坛, 2016 (7): 35-41.
- [4] 于学军. 对第五次全国人口普查数据中总量和结构的估计 [J]. 人口研究, 2002 (3): 9-15.
- [5] 郭志刚. 近年生育率显著“回升”的由来——对2006年人口和计划生育调查的评价研究 [J]. 中国人口科学, 2009 (2): 2-15.

- [ 6 ] 吕利丹, 段成荣. 对我国流动人口统计调查的总结与思考 [J]. 南方人口, 2012 (3): 75-82.
- [ 7 ] 乔晓春. 从“主要数据公报”看第五次人口普查存在的问题 [J]. 中国人口科学, 2002 (4): 48-56.
- [ 8 ] 梁中堂. 2000 年中国人口总量和妇女生育率水平研究 [J]. 中国人口科学, 2003 (6): 9-16.
- [ 9 ] 胡桂华, 刘维娜. 人口普查数据: 修正还是不修正 [J]. 中国统计, 2013 (3): 52-53.
- [ 10 ] 韦艳, 张力. “数字乱象”或“行政分工”: 对中国流动人口多元统计口径的认识 [J]. 人口研究, 2013 (4): 56-65.
- [ 11 ] PACE M, LANZIERI G, GLICKMAN M, et al. Revision of the European standard population [M]. Luxembourg: Publications Office of the European Union, 2013: 11-23.
- [ 12 ] 梁秋生. 人口大流迁时期区域人口统计指标的效度失真 [J]. 中国人口科学, 2010 (4): 23-30.
- [ 13 ] 韦伯. 韦伯方法论文集 [M]. 张旺山, 译. 台北: 联经出版事业公司, 2013: 59, 244.
- [ 14 ] 胡桂华, 韦建英. 行政记录的统计使用 [J]. 中国统计, 2009 (11): 48-50.
- [ 15 ] 郭志刚. 中国的低生育率与被忽略的人口风险 [J]. 国际经济评论, 2010 (6): 112-126.
- [ 16 ] 陈友华, 孙永健. 人口预测和人口规划系统性偏差之特征、原因与应对 [J]. 探索与争鸣, 2021 (7): 80-89.
- [ 17 ] 段成荣. 要发展地看待城市人口容量问题 [J]. 人口研究, 2005 (5): 55-58.
- [ 18 ] 崔承印. 对北京人口规模的反思与认识 [J]. 北京规划建设, 2006 (5): 67-69.
- [ 19 ] 胡兆量. 北京人口规模的回顾与展望 [J]. 城市发展研究, 2011 (4): 8-10.
- [ 20 ] 陈义勇, 刘涛. 北京城市总体规划中人口规模预测的反思与启示 [J]. 规划师, 2015 (10): 16-21.
- [ 21 ] 陈友华. 迁徙自由、城市化与贫民窟 [J]. 江苏社会科学, 2010 (3): 93-98.
- [ 22 ] 陈友华, 孙永健. 非均衡发展: 人口发展理论的批判与建构 [J]. 学海, 2021 (4): 47-55.
- [ 23 ] 杨雪. 再论人口“迁移”的概念与统计口径界定 [J]. 西北人口, 2004 (1): 19-21.
- [ 24 ] 张广宇, 顾宝昌. 人口重报: 人口普查面临的新挑战 [J]. 人口与经济, 2018 (3): 1-12.

## Research on the Bias between Population Census Data and Administrative Statistics Data in Metropolises: Taking Four Mega Cities in China as an Example

CHEN Youhua, SUN Yongjian

(School of Social and Behavioral Sciences, Nanjing University,  
Nanjing 210023, China)

**Abstract:** Demographic data is the basis of policy-making and planning, but there are obvious differences among demographic data from different channels in China. However, the researches on the vertical characteristics, causes and countermeasures of such differences are relatively insufficient, and which is a lack of regional perspective. With the aid of data sources such as *the statistical bulletin of national economic and social development* and *the statistical yearbook* issued by local governments over the years, this paper sorts out the demographic data of the four major cities of Beijing, Shanghai, Guangzhou and Shenzhen from 2000 to 2020, and constructed “population growth rate” and “administrative statistics bias rate” indicators. This paper makes a comparative analysis of census data and administrative statistics data in four major cities. The results show that: 1) There is obvious

discontinuity in the change of administrative statistical population value in the administrative statistics of the four major cities. The population growth in each census year is “abrupt”, which is much higher than the population growth rate in adjacent non-census years. 2) The administrative statistical population value of four mega cities is always lower than the revised value of the census, which is systematically underestimated. During non-census years, the administrative value will omit part of the population, forming a “quantity gap”. Every census year, the administrative value will “miss” part of the population every year, resulting in a cumulative “quantity gap”. In every census year, the administrative statistics will correct the “missing population” accumulated in the previous nine years. This paper outlines the deviation curve of “systematic omission” in population administrative statistics, reveals the strange phenomenon of deviation between population census and administrative statistics, and analyzes its causes. Based on this, the paper proposes to deal with the systematic bias from three aspects: First, change the policy concept and practice of population control, reexamine the policy deviation and risk caused by systematic omission. Second, dialectically treat the function and limit of administrative statistics, fully understand the limitations of administrative statistics. Third, strengthen the scientific management of administrative statistics. The results provide some warning and enlightening significance for policy text and scientific research to attach importance to and avoid demographic bias.

**Keywords:** population census; administrative statistics; systematic bias

[责任编辑 刘爱华]

LUO Zhihua<sup>1,2</sup>, WU Ruijun<sup>1,2</sup>, JIA Zhike<sup>3</sup>

(1. The Center for Modern Chinese City Studies, East China Normal University, Shanghai 200062, China; 2. School of Social Development, East China Normal University, Shanghai 200241, China; 3. College of Philosophy and Sociology, Hebei University, Baoding 071002, China)

**Abstract:** Given the low fertility rate, the high cost of child-rearing have already emerged as the dominant constraint that hinders the conversion of fertility intention into actual fertility behavior. Based on the data of “the Survey of Child-rearing Costs and Fertility Intentions” in Xi’an in 2019, employing the method of variance analysis and Poisson Regression, this research discusses the impact of child-rearing cost on the child-bearing intention of couples with one child, which is mainly derived from theoretical perspective of the new family economics and family life cycle theory. The findings are as followed: the average number of intended children of couples with one child is 1.35. Meanwhile, the average numbers of children planned by families with children aged 0-3, 4-6 and 7-12 years old are 1.43, 1.32 and 1.29 respectively. The direct economic cost of a couple with child is mainly related to education and food costs, among which food costs are mostly for families with child aged 0-3, while education costs account for majority as child aged 4-6 and 7-12 years. Indirect costs are mainly related to opportunity cost, time cost and the labor cost, among which the families with child aged 0-3 years old pay more time costs, while families with children aged 4-6 and 7-12 years old pay more in opportunity and labor costs. There is significant heterogeneity in the effect of rearing costs on the number of children that the couples with one child intend to have. Direct education costs, recreation costs and indirect opportunity costs and labor costs will reduce the number of children that the couple with one child plan to. The planed numbers of families with child aged 0-3 years old are mainly influenced by opportunity costs and manpower costs, meanwhile, planed numbers of families with child aged 4-6 years are mainly influenced by education costs and opportunity costs, moreover, planed numbers of families with child aged 7-12 are mainly constrained by recreation costs and manpower costs. These findings will help improve us to deepen the understanding of the relationship between the costs of child-rearing and fertility intentions, as well as provide us with a new idea to reduce family costs of child-rearing separately, improving fertility intention of families with different life cycles precisely.

**Keywords:** one-child couples; family costs of child-rearing; fertility intention; new family economics; family life cycle