

文章编号: 1006- 4346(2001)01- 0005- 07

# 利用人口普查原始数据对独生子女信息的估计

郭志刚

(北京大学 社会学系, 北京 100871)

**摘要:** 提出了一种利用人口普查原始数据估计各出生队列的独生子女- 孩次比例的方法, 并应用于1990年普查数据, 估计了部分队列的独生子女- 孩次比例, 量化地揭示了各出生队列中一孩从独生子女变为非独生子女的动态。对这一方法所得到的队列独生子女比例与其他来源的估计进行了比较。

**关键词:** 独生子女; 匹配; 孩次; 队列

**中图分类号:** C923      **文献标识码:** A

## Estimating Cohort Percentage of Only-Children Using Data of Population Census

GUO Zhi-gang

(Department of Sociology, Peking University, Beijing 100871, China)

**Abstract** This paper proposes a method to estimate the only-children/child order distribution by birth cohort using original data from population census. The results based upon the 1990 census data quantitatively depict the dynamics of the distribution by age. In addition, the results are compared to the estimates by other method and data source for their consistency.

**Key words** only-children; match; child order; cohort

20年前, 为了更有效地扼制人口迅猛增长的势头, 1980年9月中共中央关于控制我国人口增长问题致全体共产党员、共青团员的公开信“提倡一对夫妇只生育一个孩子”。后来, 一对夫妇只生育一个孩子也不同程度的具体体现在城乡各地的生育政策中。20年过去, 城市年轻人口中独生子女已经极为普遍, 农村中独生子女也有相当数量, 独生子女已经成为社会公众关心的热点问题, 这之中除了独生子女本身的成长与教育等问题外, 还有很多方面的应对政策也需要关于独生子女数量方面的信息, 其中特别是生育政策调整的时机与幅度问题和所谓“四二一”家庭结构问题也引发了很多的讨论。但是, 实际上关于独生子女数量的信息却极为贫乏, 也因此相关这方面的宏观定量分析也是凤毛麟角。这与众多的定性研究形成了强烈的反差。

收稿日期: 2000- 10- 15

作者简介: 郭志刚(1954- )男, 北京大学社会学系教授。

《市场与人口分析》2000年第4期发表杨书章、郭震威的文章<sup>[1]</sup>,比较系统地研究了独生子女现状及其对未来人口发展的影响。毫无疑问,这在关于独生子女定量化分析上迈出了重要一步。该文根据张二力、陈建利按1988年生育节育调查和1992年生育率调查所估计的数据<sup>[2]</sup>,进一步计算了中国1971~1995年期间出生的人口分城乡分性别的5岁一组的独生子女比例。该文还索引了不同来源的独生子女总量估计,并依据不同信息来源做出了自己的独生子女总量估计。

恰好我们也正在进行旨在研究生育政策调整及未来家庭结构的人口预测分析,对于这一特定目的的人口预测而言,实际上最重要的信息并不是现有独生子女的总量,而是预测中基期人口的分年龄的现存独生子女比例。目前最为全面而可靠的人口数据便是1990年人口普查,然而普查数据只提供了母亲一代按生育孩次的分布,并没有子女一代的独生子女属性和孩次属性,但这些信息对于上述研究却是必要的基础信息。因此,杨、郭二位所提供的出生队列独生子女比例对于我们的预测是一个重要的参考信息。由于我们事先并没有这一信息,所以直接采用了根据1990年普查的原始数据中的母子匹配来估计年龄别独生子女的比例的方法。应该说,这两种根据不同方法和数据来源的估计可以互为验证。并且我们相信,根据普查数据匹配估计队列独生子女比例的方法对于今后开发利用2000年第五次全国人口普查的数据也有参考意义。

## 1 数据与方法

我们进行队列独生子女比例估计的基础数据为1990年全国人口普查1%抽样数据(在进程序试验时,我们还对这一1%原始数据中的10%再抽样数据进行了匹配和更为细致的分析)。由于这一数据独立于计划生育部门的调查或报表,因此就数据来源角度而言,与杨、郭估计进行相互比较和验证便更加有效。

估计的基本思路:因为1990年普查数据提供15~64岁妇女的分性别的曾生子女数和存活子女数,考虑到在绝大多数情况下未成年子女都与母亲生活于同一家庭户中,所以有可能根据母亲与子女分别与户主的相对关系将其匹配起来,并根据母亲的子女数信息判断其子女是否独生子女。并且,当母亲的子女数量与性别均与同户生活的子女完全对应时,还能够进一步确定这些子女的孩次,并因此得以重建该母亲的生育史。

由于根据普查资料只能进行同一家庭中母亲与其现有孩子(不含死亡或已离家的子女)的匹配,所以匹配采用存活子女口径。因此,所匹配出的孩次与常用的出生孩次概念有所不同,而与计划生育部门<sup>[3]</sup>所用孩次概念有相近之处。尽管这对于重建妇女的生育史损失了一定的有效性,然而对于预测未来人口结构,以及规划生育政策调整和分析家庭结构变化反而更为合适,因为这些研究实际上都应以现有存活子女为基础。

具体操作方法简述如下:

(1)独生子女指户中关系与母亲匹配的存活子女中那些根据母亲申报没有其他存活子女的孩子。对于二代户中(或多代户户主在倒数第二代时)户主为父亲且母亲不在本户的,则寻找户主的子女。当子女数等于1时,这一子女作为独生子女。否则,所有子女均作为非独生子女。

(2)孩次的鉴别是在匹配子女与母亲申报的存活子女在数目及性别上完全吻合时按照各个孩子的年龄排序确定的,否则孩次为未知。

(3)对于集体户中的年轻人口和家庭户中与户主为非直系关系(即“其他亲属”或“非亲属”)的年轻人口则根本无法进行母子匹配,因此规定这些年轻人口的独生子女属性为未知,自然孩次也未知。

(4)母子匹配中的另一个问题是抱养或过继的子女。由于“第四次全国人口普查表填写说明”关于妇女曾生、存活子女数的填写要求虽然一方面定义这两种子女数“都是指亲生子女。不包括丈夫前妻留下的子女和过继的子女,原则上也不包括领养的子女。”但另一方面又允许“有的妇女对领养关系一直保密的可以作为本人生育的子女和现在的存活子女。”<sup>[41]</sup>因此,妇女在申报这些项目时,既可以按前面严格的定义申报,也可以按后面宽容的定义申报。与本研究目的一致,在匹配的操作中,如果母亲申报子女数大于1或户内子女数大于1,所有子女均作为非独生子女,也就是说,这里汇总的非独生子女属性将领养子女考虑在内。如果母亲申报子女数等于0且户内子女数等于1时,这个子女当然是领养的,但也作为独生子女对待。

另外,由于匹配时鉴别子女孩次所使用的年龄是周岁,没有用到普查数据所提供的出生月份信息,所以如果一个妇女一年有两次生育,就会被当作双胞胎处理。这种情况并不多,而我们的主要目的是鉴别孩次,而不是恢复母亲的生育史,所以这样做并不会造成明显误差,却可以简化匹配程序和运行时间。

于是,通过匹配给原始数据中每一0~14岁个人建立独生子女属性与孩次属性两个新附加的变量。

## 2 匹配的初步结果及分析评价

根据对少儿人口匹配后形成的独生属性和孩次属性,分别汇总出城乡性别年龄别的未知属性比例来评价匹配结果的可用性(见表1)。匹配后仍属性未知表示匹配不成功,于是属性未知的比例越高,匹配结果的可用性越差。

表1 匹配结果的评价

年龄	独生属性未知比例(%)				孩次属性未知比例(%)			
	城镇		农村		城镇		农村	
	男	女	男	女	男	女	男	女
0	6.79	7.63	5.85	6.37	8.86	10.23	5.93	10.35
1	6.79	7.36	4.69	5.54	9.02	10.07	7.90	9.49
2	7.44	8.05	4.20	5.14	9.82	10.74	7.86	9.75
3	7.15	7.68	3.85	4.63	9.78	10.80	7.97	9.60
4	7.51	7.78	3.48	4.28	10.65	10.86	8.42	10.02
5	7.96	8.10	3.31	4.06	10.96	11.70	9.18	10.59
6	7.43	7.97	3.17	3.89	10.21	11.11	9.96	11.10
7	7.37	7.77	3.12	3.72	10.66	10.87	11.07	11.81
8	6.87	7.19	3.09	3.59	10.35	11.06	12.58	13.02
9	6.26	6.79	3.18	3.71	11.21	11.87	14.08	14.46
10	5.69	6.20	3.75	4.15	13.21	13.45	18.09	17.98
11	5.98	5.96	4.12	4.90	16.85	17.42	22.36	22.07
12	6.31	6.38	5.01	6.03	20.36	20.12	26.99	27.05
13	7.15	7.05	6.22	7.22	24.54	24.42	32.99	32.26
14	8.35	8.81	7.72	8.92	29.99	30.36	39.61	39.33
合计	7.01	7.40	4.33	5.10	13.57	14.19	15.14	16.26

从表1中合计一行可以看出,无论是按城乡或是按性别,独生子女属性的未知比例显著小于孩次属性的未知比例。独生子女属性未知比例在所有年龄别单元中都在9%以下,各年龄合计的相应比例则更低,其中农村明显低于城镇。总的来说,通过普查数据的匹配已经把握了绝大部分少儿人口的独生属性。

对于匹配后仍未知独生属性者,经过数据检查发现:这些人主要是三代及以上户中申报为户主的“孙子女”和“子女”的人,共占63.11%;其次为各种家庭户类型中的“其他亲属”,占18.24%;再次为假三代户中的“孙子女”,占8.33%;另外比较显著的部分还有生活于集体户的人,比

例为 5.2%。可见, 匹配还有可能适当改进。特别是 2000 年普查对“与户主的关系”的分类定义做了调整, 可以提供更多的有用信息, 更便利了应用这一方法匹配母子关系。

从表 1 还可看出, 各类孩次属性的未知比例都随年龄增长而显著上升, 而独生属性的未知比例却相对比较稳定。这是因为孩次的判断只有在所有存活兄弟姐妹都在本户时才能判断, 而独生属性只需要母亲申报的存活子女数便能判断。因此独生属性未知比例要比孩次属性未知比例低。

我们曾用普查 1% 数据中再抽取 1/10 的较小样本进行过上限为 25 岁的年轻人口属性统计, 但结果显示从 15 岁组开始独生属性未知的比例已经超过 10%, 并且在随后的年龄组增加极快。显然, 这时匹配效果已经很不好了。为了将匹配误差尽量缩小, 我们对普查 1% 数据进行母子匹配时只取 0~14 岁年龄段的匹配结果来估计总人口相应年龄段的独生子女比例和孩次比例。在 1990 年普查 1% 数据样本中, 0~14 岁的人口共有 3145656 人。

匹配结果表明少儿人口的独生属性结果是可用的, 但孩次属性结果则显得很不好。经过分析, 我们发现有可能利用独生属性的已知信息来弥补孩次属性信息的缺损。

将匹配所得的每个孩子的独生属性结果与孩次结果进行交互分析, 得到以下结果(表 2):

表 2 少儿人口中独生属性与孩次属性的交互分布(%)

总%	独生属性	孩次				合计
		未知	1	2	3+	
城 镇	未知	7.20				7.20
	独生		48.84			48.84
	非独	6.67	12.89	17.62	6.78	43.96
	合计	13.87	61.73	17.62	6.78	100.00
农 村	未知	4.70				4.70
	独生		15.37			15.37
	非独	10.98	23.40	27.62	17.94	79.93
	合计	15.68	38.76	27.62	17.94	100.00

虽然表中农村部分未知孩次情况的合计比例约占 16%, 但其中的 70% (即表中的 10.98%) 实际上已知为非独生子女。匹配出的农村少儿人口中独生子女比例为 15.37%, 即使将未知独生子女属性的比例(4.7%)全部作为独生子女来对待, 农村少儿人口中的独生子女比例也仅为 20%。

城镇中匹配出的独生子女比例占 48.84%。如果将未知是否独生子女的 7.2% 权且全部作为独生子女, 城镇中的独生子女比例将达到 56%。比较之下, 可以

看出 80 年代生育政策及其执行情况的城乡差异。

在上表的城乡两个部分中, 实际上各有两个交互单元的案例没有确定的孩次, 其中各有一个交互单元同时还未知独生属性, 但比例都相对较小。我们可以按照一定的假设对他们进行估计。

第一步, 假设孩次未知的非独生子女在孩次分布上与其他已知孩次的非独生子女的孩次分布相同, 便可以把他们分配到各个孩次中去。

第二步, 对于孩次未知且独生属性也未知的案例, 先要假定他们的独生属性分布与其他已知独生属性的案例的分布相同, 将其分配到独生或非独生的类型中去。然后, 分配到独生子女类型中的那一部分自然都作为一孩; 其次, 我们采用第一步估计完成后所得到非独生子女孩次分布作为分配标准, 将未知独生子女属性的案例中分配到非独生子女类型中去的那一部分再进一步分配到各个孩次中去。这样, 我们便可以得到分为四档的综合分布类型, 即: 1、独生子女/一孩; 2、非独生子女/一孩; 3、非独生子女/二孩; 4、非独生子女/三孩及以上。

表 3 1990 年少儿人口独生-孩次类型比例的估计结果

类别	年龄	独生一孩	非独一孩	二孩	三孩+	合计
城镇男性	0	76.90	0.55	18.06	4.49	100
	1	74.14	1.93	18.35	5.57	100
	2	69.85	5.02	19.54	5.59	100
	3	66.29	7.71	20.54	5.46	100
	4	65.79	10.41	18.38	5.42	100
	5	64.92	12.73	16.33	6.01	100
	6	67.08	13.70	13.39	5.83	100
	7	66.63	13.63	13.85	5.90	100
	8	62.57	14.36	17.18	5.89	100
	9	52.86	17.70	21.53	7.91	100
	10	40.29	17.55	30.25	11.92	100
	11	32.04	19.19	32.58	16.18	100
	12	25.67	23.10	34.37	16.86	100
	13	18.36	29.23	34.73	17.68	100
	14	12.73	35.00	34.52	17.75	100
合计		53.89	14.26	22.86	8.99	100
城镇女性	0	79.64	0.51	16.06	3.80	100
	1	75.92	2.07	17.63	4.38	100
	2	70.48	6.71	18.06	4.75	100
	3	66.11	9.77	19.23	4.90	100
	4	64.83	13.47	17.03	4.67	100
	5	62.77	16.85	15.02	5.36	100
	6	63.20	19.03	12.51	5.26	100
	7	62.41	19.97	13.26	4.37	100
	8	57.10	21.12	16.26	5.52	100
	9	47.49	25.22	20.61	6.67	100
	10	34.39	24.31	30.96	10.34	100
	11	26.96	25.23	32.95	14.86	100
	12	20.95	28.31	34.03	16.71	100
	13	14.63	33.90	34.03	17.43	100
	14	10.12	39.29	32.52	18.07	100
合计		51.26	18.66	21.87	8.21	100
农村男性	0	45.27	0.35	34.97	19.41	100
	1	39.53	4.08	36.21	20.19	100
	2	28.27	15.46	36.28	19.99	100
	3	21.82	22.43	36.32	19.43	100
	4	17.92	28.11	35.06	18.91	100
	5	14.21	32.27	32.94	20.58	100
	6	12.99	34.55	30.62	21.83	100
	7	13.33	34.22	30.34	22.11	100
	8	12.60	33.90	30.60	22.90	100
	9	9.57	35.90	31.00	23.53	100
	10	7.73	33.62	32.65	25.99	100
	11	6.41	32.77	32.91	27.90	100
	12	5.08	34.43	33.30	27.19	100
	13	3.82	35.62	34.14	26.42	100
	14	2.81	37.70	35.01	24.48	100
合计		17.11	26.27	34.10	22.51	100
农村女性	0	48.11	0.28	34.36	17.24	100
	1	42.46	4.79	34.63	18.12	100
	2	28.97	18.56	34.25	18.21	100
	3	21.12	26.48	34.76	17.64	100
	4	16.24	32.41	33.69	17.65	100
	5	11.77	37.58	31.47	19.18	100
	6	9.70	40.78	29.52	20.00	100
	7	8.44	41.51	29.78	20.27	100
	8	6.97	41.68	29.86	21.50	100
	9	4.93	42.11	30.76	22.20	100
	10	3.61	38.40	33.10	24.89	100
	11	2.87	38.00	32.17	26.96	100
	12	2.27	37.99	33.03	26.70	100
	13	1.87	38.60	34.00	25.52	100
	14	1.49	39.78	34.70	24.04	100
合计		15.05	30.85	33.03	21.06	100

但是,在实际过程中,我们不能忽视性别年龄组之间的差别。因此,上述估计过程是按各性别年龄组分别完成的。

### 3 最终估计结果与分析比较

根据 1990 年全国人口普查 1% 抽样原始数据对少儿人口的独生-孩次比例的估计结果列在表 3 中。

经常见到的一种误解是混淆独生子女与一孩之间的差别。实际上这两者既有关联又有差别。在 0 岁组时,这两者之间差异很小。差异主要在于双胞胎(或多胞胎)的出生,一孩出生后不久即丧失独生子女属性。另外一种情况是,母亲在一年之内的两次生育。所以,一般而言,一孩出生时虽然等同于独生子女,然后其中会有相当的部分随着时间(即孩子的年龄)由于其母亲后来的生育而变化成非独生子女。而在不考虑其他子女死亡条件下,孩次不会变。尽管表 3 只是时点数据,而不是真正意义上的出生队列动态,但从中仍能反映出这种变化过程。

需要特别说明的是,这一少儿人口中的独生子女比例只是普查时点上的统计,随着时间推移,还会有一部分转变为非独生子女,对较年幼的出生队列,尤其是这样。严格地说,队列独生子女比例的统计不仅要有出生队列的明确限定,还要有统计时间的明确限定。

杨书章、郭震威发表的独生比例估计似乎已经作为各队列的终结情况,因为文中已经使用这些比例来估计未来年代这些队列的婚配概率。因此,本文匹配出来的估计与他们的估计口径上有所不同。此外,他们估计的参照时间是 1995 年,而本文匹

配结果的参照时间却是1990年。尽管如此,两套估计仍然具有可比性,因为假定独生子女与非独生子女死亡水平相同,那么各队列中的独生子女比例在达到婚配年龄之前大体是一个随时间单调下降的过程。并且,我们匹配所涉及的队列正好与他们估计队列中的三个相吻合。换句话说,我们的估计值如果高于他们的估计值,两者便可认为具有一定程度的一致性,否则便是相违的。

表4同时列出两套估计进行比较。可以看到,对城镇1976~1985年期间出生的两个5年队列,无论男女,均是杨、郭的估计值较高,其中较早出生的1976~1980队列相差了5~6个百分点。而这部分人正是当前进入婚配年龄段的人口。按照杨、郭二位的估计,城镇这一队列双方为独生子女的婚配概率为83.8%,而至少一方是独生子女的婚配概率为49.72%。如果按照本文匹配的估计,则相应的婚配概率分别只有5.49%和41.59%。应该说,两者虽然有一定出入,但并不是非常大。至于匹配估计的1986~1990队列的独生子女比例高于杨、郭二位的估计比例,倒是反映了逻辑上的一致性。

比较农村独生子女比例的两套估计,均是匹配估计高于杨、郭二位的估计。

此外,两套估计中都大体反映出在同队列中男性独生子女比例高于女性。但是,值得注意的是,其中有两个例外是1986~1990队列的匹配估计在城乡同时出现男低于女的现象,而背后的原因尚不清楚。

表4 本匹配估计与杨郭估计的比较

出生年代	匹配估计的独生子女比例(%)		杨郭估计的独生子女比例(%)	
	城镇男性	城镇女性	城镇男性	城镇女性
1976~1980	25.78	21.30	31.55	26.55
1981~1985	63.08	58.82	66.59	59.65
1986~1990	70.59	71.37	67.82	61.51

  

出生年代	匹配估计的独生子女比例(%)		杨郭估计的独生子女比例(%)	
	农村男性	农村女性	农村男性	农村女性
1976~1980	5.13	2.41	4.74	2.00
1981~1985	12.59	8.39	10.54	5.19
1986~1990	30.70	31.44	13.18	9.58

注:杨郭估计见《市场与人口分析》2000年第4期第11页表1。

估计。

与杨书章、郭震威二位用另外的方法和数据所得到的估计相比,发现在可比性最强(即作为最终分布有效性最强)的1976~1980出生队列,匹配方法得到的城镇男女独生子女比例均低于他们的估计比例约5~6个百分点;比较1981~1985出生队列时也是匹配估计的比例略低。因此,两种估计之间存在逻辑上的一致。而匹配方法得到的农村男女独生子女比例却都略高于他们的估计比例。但是必须说明,以上比较只能从一致性角度来相互验证两套估计,目前没有更可靠的信息来确定哪一套估计更为准确。

鉴于目前得不到直接调查的队列独生子女比例信息,因此只能通过间接方法加以估计。所以,我们将自己进行匹配估计的方法及有关结果公布出来,希望得到批评和建议,以便加以改进与完善。

#### 4 小结

本文提出一种新的方法根据人口普查原始数据来估计队列独生子女比例及孩次信息,讨论了有关概念定义及操作步骤,并应用这一匹配方法于1990年普查数据,得到1976~1990年队列的独生-孩次比例

注释:

国家计划生育委员会(1997)在1997年全国人口与生殖健康调查中所用的定义为:“孩次——指活产子女出生时在当时家庭所有存活子女(包括抱养、过继等)中的排行。”本文所用的孩次定义与其不同之处在于,不是以该子女出生的时点作为参照时间,而是以普查时点作为参照时间,即孩次的操作定义指该子女在普查时所有存活兄弟姐妹中的排行。

当实际上父亲身边的子女并不是独生子女时,这种假设会高估独生子女比例。应该说,他们对其估计的方法没有足够明确的描述。

参考文献:

- [1] 杨书章、郭震威 中国独生子女现状及其对未来人口发展的影响[J] 市场与人口分析, 2000, 6(4): 10-17.
- [2] 张二力、陈建利 遵循现行生育政策一对夫妇终身生育的孩子数[Z] 1998
- [3] 国家计划生育委员会 技术文件——1997年全国人口与生殖健康调查[Z] 1997.
- [4] 国务院人口普查办公室/国家统计局人口统计司 中国1990年人口普查资料(第四册)[Z] 北京: 中国统计出版社, 1993 517- 518 [责任编辑: 陈功]

## 稿 约

在新的一年里,本刊欢迎广大作者围绕以下论题惠赐稿件:

1. 稳定低生育水平的宏观态势研究;
2. 稳定低生育水平的地域差异研究;
3. 稳定低生育水平的政策选择研究;
4. 中国人口发展目标研究;
5. 西部开发中的人口问题研究;
6. 五普资料的分析研究;
7. 人口、资源、环境与可持续发展研究;
8. 人口老龄化趋势、影响与对策研究;
9. 人口老龄问题、老龄产业、养老保险的理论与应用研究;
10. 高龄老人与健康长寿的理论与应用研究;
11. 婚姻、家庭的人口学研究;
12. 社会主义生育文化理论与实践研究;
13. 流动人口管理研究;
14. 人力资源开发、就业与劳动力市场研究;
15. 生殖健康与计划生育的理论研究;
16. 民族人口研究;
17. 人口思想与人口学科建设与发展研究;
18. 人口学及其分支学科的方法学研究;
19. 计划生育工作实践与交流;
20. 重要人口与计划生育学术会议综述与动态。