

20 世纪90 年代中国人口死亡水平

黄荣清

【摘 要】 在人口普查中,人口死亡容易发生漏报。文章分析了中国人口死亡漏报的社会原因,利用两次普查的人口数据证明了人口死亡漏报确实存在。为了估计出实际的人口死亡水平,对两次普查的人口数据质量进行了推敲,根据登记的人口和死亡率数据的不同情况,提出了相应的修正方法。研究表明,20 世纪 90 年代男性的预期寿命为 67.97 岁,女性的预期寿命为 71.34 岁,均比人口普查登记资料算出的结果小 1.5 岁左右。

【关键词】 预期寿命 人口普查 死亡漏报

【作 者】 黄荣清 首都经济贸易大学人口经济研究所所长、教授。

目前,已有不少学者对 2000 年全国人口普查的结果进行了研究,但多集中于人口总量、生育水平方面。任强等(2004)根据“三普”至“五普”的结果,做了一定的修正,计算了中国人口死亡率、预期寿命等的变化。本文拟根据“四普”和“五普”的结果,估计 20 世纪 90 年代中国人口实际的死亡水平。

按照“五普”中的死亡人口和相关资料计算,1999~2000 年,中国人口预期寿命为 72.43 岁,其中男性为 70.65 岁,女性为 74.33 岁。由“四普”资料计算,1989~1990 年人口预期寿命为 70.06 岁,其中男性为 68.35 岁,女性为 71.91 岁^①。这 10 年间(严格地说是 10 年零 4 个月),人口预期寿命提高了 2.37 岁(男女合计)。对年龄存活概率进行几何平均可以算出,20 世纪 90 年代人口平均预期寿命男性为 69.46 岁,女性为 73.09 岁,男女合计为 71 岁左右(其生命表有关数据见表 1)。

如果人口普查中人口登记是准确的,死亡登记是完整的,上述结果可以表示中国 20 世纪 90 年代末、90 年代初和 90 年代平均的人口死亡水平。但如果登记的死亡人口有漏报,则上述结果就低估了中国实际的人口死亡水平,就需要重新估计人口实际死亡水平。

一、“五普”人口的死亡漏报

(一) 人口死亡漏报存在的原因

以往的经验 and 学者们的研究已经证实,人口普查中存在死亡漏报,“五普”也不会例外。人口普查中出现死亡漏报,不仅是普查员在调查时工作疏忽,在它后面还隐含着许多社会原

^① 按国家统计局公布,1989 年婴儿死亡率为 34.39‰,人口预期寿命为 68.40 岁,其中男性为 66.91 岁,女性为 69.99 岁(跨世纪的中国人□编委会,1994:187、197),这也是经过修正的,并不是按普查原始资料直接计算的结果。

表 1 20 世纪 90 年代中国人口生命表中有关指标(算法一)

年龄	男 性			女 性		
	年龄别 死亡率	死亡 概率	平均预期 寿命	年龄别 死亡率	死亡 概率	平均预期 寿命
0	24.30	23.81	69.46	31.13	30.31	73.09
1	1.90	7.55	70.15	1.98	7.88	74.37
5	0.75	3.75	66.67	0.54	2.70	70.94
10	0.57	2.86	61.92	0.41	2.05	66.13
15	0.92	4.59	57.09	0.68	3.39	61.26
20	1.33	6.62	52.34	0.97	4.85	56.46
25	1.40	7.00	47.67	1.01	5.02	51.72
30	1.71	8.52	42.99	1.14	5.71	46.97
35	2.25	11.17	38.33	1.42	7.08	42.23
40	3.19	15.84	33.74	2.01	10.01	37.51
45	4.73	23.41	29.24	3.08	15.31	32.86
50	7.46	36.66	24.88	4.91	24.28	28.33
55	11.99	58.36	20.72	7.73	37.97	23.97
60	20.30	96.91	16.84	13.10	63.57	19.81
65	33.30	154.50	13.36	21.67	103.24	15.97
70	56.15	246.84	10.31	37.91	173.79	12.50
75	86.93	357.72	7.86	60.89	265.53	9.58
80	140.98	513.88	5.83	103.32	409.11	7.10
85	202.31	650.33	4.49	157.44	557.35	5.32
90	274.00	1000.00	3.65	249.00	1000.00	4.02

注:本表是根据“四普”和“五普”资料计算得到的年龄别存活概率加以平均后的结果。平均值计算:设由 1990 年普查登记数据计算得到的死亡概率为 $Q_1(x)$, 2000 年普查登记数据计算得到的死亡概率为 $Q_2(x)$, 则平均的死亡概率 $Q(x) = 1 - \sqrt{[1 - Q_1(x)][1 - Q_2(x)]}$ 。

部分流动人口可能是常年在外出,居无定所。调查死亡人口是按人口死亡前常住地登记。如果一个常年在外出的人死亡,按人口普查的规定,其户口所在地不用报告(可能也不知道),而其所在的常住地(由于居无定所)可能也无人替他报告,这样便会出现漏报。而且,“五普”中常住人口是指居住半年及以上的人口,而死亡人口是指 1 年以内死亡的人口。作为死亡率的统计,分子和分母的定义不一致。例如,有一户来本地不满半年,但在过去一年内,该户有人死亡,因为不满半年,应按“暂住人口”登记,而暂住人口的登记中没有设死亡人口登记项。因为按规定,死亡人口应按死亡前的常住地登记。类似的情况,都是导致死亡登记混乱甚至漏报的原因。

(二)“五普”中人口死亡的漏报

为了证实人口死亡漏报确实存在,需要数据支持。下面用两次人口普查的平均死亡率做出的生命表中静止人口的留存率,与两次人口普查对应的同年出生人口之比的大小来说明。如果普查报告的数据是准确的,则生命表中静止人口的留存率应该等于两次普查对应的同年出生人口之比,即若设第一次普查 x 岁的人口为 $P(x, 0)$, 两次普查的间隔为 t 年,第二次普查 $x+t$ 岁的人口为 $P(x+t, t)$, 生命表中 x 岁的静止人口为 Lx , 则应有:

因。(1)风俗习惯。一些地区和一些民族把家里死了人,特别是刚出生不久的婴儿死亡的现象看成是不吉利的事,不愿意向外人提及从而产生漏报。(2)顾忌。政府提倡火葬,但按中国传统习惯人死后“入土为安”,政府的意志和百姓的愿望不一致时,有些家庭为了逃避处罚,瞒报已经土葬的人口。(3)组织行为。一些企业在经济利益驱动下,往往忽视生产安全,出现伤亡事故时尽量不报或少报。而一些地方政府对发生的事故,也不认真追查。(4)技术原因。如判定婴儿死亡,需要有一定的医学卫生知识,需要进行详细的了解工作。全国人口普查调查工作动员了大约 2000 人,仅调查员就有 600 万人之多,不可能对这么庞大的人群进行非常详细的培训,作为普查员,要调查的项目很多,也不会会在某一项目上花费很多时间和精力。(5)其他原因。一些领养老金的人死亡后,其家庭成员为了继续领取其养老金而故意不报等。目前上述这些导致人口漏报的社会条件仍然存在,有些甚至更加严重,所以,发生人口死亡漏报的现象也是可以想象的。

与“三普”、“四普”相比,“五普”容易导致人口死亡漏报在于大量流动人口的存在和死亡人口、常住人口的统计规定。

$$P(x+t, t)/P(x, 0) = L(x+t)/L(x) \quad (1)$$

如果存在死亡人口漏报, 基于报告资料计算的静止人口留存率就会大于实际人口留存率, 则有:

$$L(x+t)/Lx > P(x+t, t)/P(x, 0) \quad (2)$$

由两次人口普查登记的人口留存率(简称人口留存率, 即 $P(x+t, t)/P(x, 0)$)和由两次普查人口死亡登记计算出的生命表静止人口留存率($L(x+t)/Lx$)之差反映了人口死亡漏报程度。静止人口留存率与人口留存率之差越大, 说明死亡漏报的程度越大。当人口死亡漏报非常小时, 由于人口年龄误报的存在和死亡率的随机变动, 人口留存率的点应紧靠着静止人口留存率曲线上下波动。由图 1 可以看出, 无论男女, 人口留存率的点和静止人口留存率曲线之间明显有一定距离。这说明, 至少在 65 岁以上(指 2000 年的年龄), “五普”的人口死亡漏报是存在的。并且, 女性人口留存率和生命表静止人口留存率的距离大于男性。说明女性人口死亡的漏报程度大于男性。

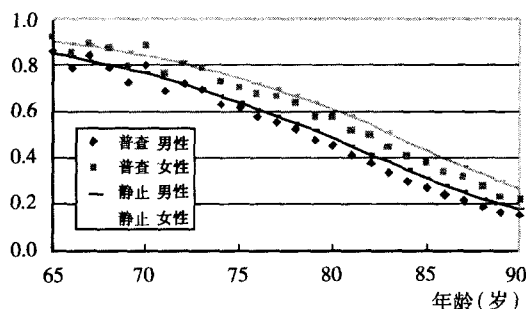


图 1 “四普”和“五普”的人口留存率与静止人口留存率的比较

注: 根据“四普”和“五普”资料计算。

二、两次普查人口和死亡人口数据

为了估计 20 世纪 90 年代中国人口的实际死亡水平, 我们首先要了解中国“四普”、“五普”的人口和死亡率数据情况, 因为估计方法本身是建立在可利用的数据资源之上的。

(一) 公布的数据情况

“四普”和“五普”公布的资料都提供了人口普查时前 1 年内死亡人口的分性别、年龄数据。其中, 1990 年的人口、2000 年的人口和死亡人口的年龄数据为 0~100 岁及以上(100 岁为半开区间的端点), 而 1990 年的死亡人口数据为 0~90 岁及以上(90 岁为半开区间的端点)。由人口和死亡人口数据我们可以计算出分性别、年龄的死亡率及编制出相应的生命表, 由于 1990 年只公布了 0~90 岁的分年龄数据, 我们也就只能算出以 90 岁为半开区间端点的分性别、分年龄死亡率。因此, 用两次普查公布的资料只能算出 1990~2000 年 0~90 岁及以上(90 岁为半开区间的端点)的平均死亡率及编制出对应的生命表。

(二) 人口数据的质量

由过去的经验和前面的讨论, 我们已经知道人口死亡有漏报, 下面主要讨论人口数据的质量问题。

从“四普”和“五普”对应年龄的人口比 $P(x+t, t)/P(x, 0)$ 的值来看, 在“四普”、“五普”相对应的各年龄中, 男性在 22 个年龄上、女性在 25 个年龄上该值大于 1。 $P(x+t, t)/P(x, 0)$ 的最后一个年龄为 51 岁, 其中在 0~9 岁的每个年龄(这里指“四普”时的年龄, 下同), $P(x+t, t)/P(x, 0)$ 的值都大于 1。对同期群的留存率来说, 在封闭人口的情况下(全国人口可以看做是封闭人口), 其值不可能大于 1。产生上述情况, 只能是“四普”人口 $P(x, 0)$ 少报, 或是“五普”人口 $P(x+t, t)$ 有重报。根据对“三普”和几次人口抽样调查的经验, 低年龄的人口漏报比较普遍, 所以可以认为在低年龄出现 $P(x+t, t)/P(x, 0)$ 大于 1 主要是因为“四普”0~9 岁年龄的人口漏报造成的。当然也不能排除“五普”16~19 岁年龄段人口有一定的重报。因为根据“三普”和“四普”两次普查的人口分析, “三普”在达到教育年龄以后的人口数是基本可信的。这样, 在 10 岁

以上 $P(x+t, t)/P(x, 0)$ 出现大于 1, 可以认为是由于“五普”人口重报造成的。虽然从 40 岁开始在大多数年龄 $P(x+t, t)/P(x, 0)$ 的值已经小于 1, 但直到 55 岁以前, 大多数年龄 $P(x+t, t)/P(x, 0)$ 的值还大于 $L(x+t)/L(x)$, 说明在 55 岁以前, 存在人口重报的现象。

下面, 笔者以“四普”的人口和“四普”、“五普”的平均死亡率做出的生命表为基础, 推算出 2000 年人口与“五普”报告的人口进行比较来估计人口报告数据正确的年龄区间。设 x 岁的估计人口为 $P(x)$, 则:

$$P(x) = \begin{cases} P(x, 00) & (x > 90) \\ P(x-10, 90) L(x+t)/L(x) & (x < 90) \end{cases} \quad (3)$$

即在 90 岁以上(由于“四普”公布的死亡资料没有 90 岁及以上分年龄死亡资料, 所以以 90 岁作为截止年龄) $P(x)$ 等于 2000 年人口, 而 90 岁以下的人口以 1990 年人口乘以生命表静止人口留存率来推算, 为了减少人口的年龄波动, 计算从高到低的各年龄人口之和, 令:

$$S_0(x) = \sum_{\infty}^x P(x, 00), S(x) = \sum_{\infty}^x P(x) \quad (4)$$

$$\Delta(x) = S(x)/S_0(x) - 1$$

由于存在死亡人口的漏报, 所以基于普查资料算出的静止人口留存率应大于实际留存率, 即:

$$S(x) > S_0(x), \Delta(x) > 0 \quad (5)$$

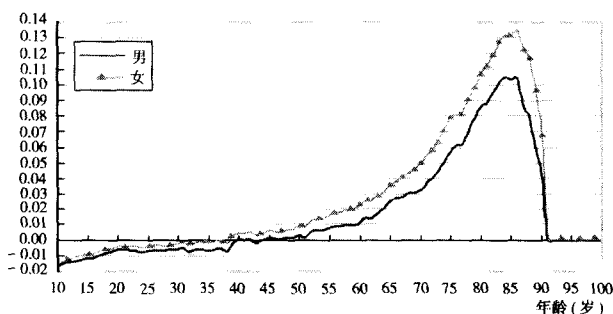


图2 2000年调查人口与估计的人口比较

注: 根据“四普”和“五普”资料计算。

从图 2 可以看出, 40 岁以下, $\Delta(x) < 0$, 所以 40 岁以下的人口数据肯定有误, 不能作为推算实际死亡水平的基础数据。40 岁以上, 虽然有 $S(x) > S_0(x)$, 但通过比较年龄组数据: $S(x) = \sum_x^{x+4} P(x)$ 和 $S_0(x) = \sum_x^{x+4} P(x, 00)$, 在 40~44, 45~49, …… 60~64 年龄组, $S(x) \leq S_0(x)$, 说明在 40~64 岁, “五普”人口也存在重报现象, 直到 65 岁(2000

年的年龄)以后, 才保持了 $S(x) > S_0(x)$ 的倾向。所以, 如果“四普”10 岁以上人口是基本准确的话, 那么只能认为, “五普”人口数据直到 65 岁以后才基本准确。

65 岁以后人口数据基本准确, 并不能认为 65 岁以后各年龄报告的人口都准确无误。根据经验我们知道, 高龄老人容易发生年龄误报。下面我们来看一下 1990 年 80 岁以上的年龄和 2000 年对应年龄的人口比。从图 3 看, 高龄男性对应年龄的人口比很不正常。一般来说, 留存率随着年龄的提高应该减小。但高龄男性在 93 岁(2000 年的年龄)以后, 留存率不但没有减小, 在一些年龄反而增大。出现这种情况, 往往是由于年龄报告存在问题, 可以推测是由于“五普”年龄较低的高龄人口登记到更高年龄的缘故。

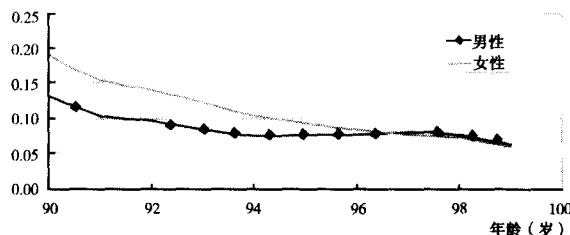


图3 1990~2000年高龄人口的留存率

注: 根据“四普”和“五普”资料计算。

总之, 65 岁以下的人口数据, “四普”低年龄人口存在漏报, “五普”存在人口重报。65 岁以上的, “五普”男性高龄老人存在年龄误报。登记的平均死亡率只能算 90 岁以下, 由于人口死亡漏报的存在, 平均死亡率要小于实际死亡率。

登记的平均死亡率只能算 90 岁以下, 由于人口死亡漏报的存在, 平均死亡率要小于实际死亡率。

三、20 世纪 90 年代人口死亡水平的估计

应该说,如果死亡报告和人口报告同时有误的话,任何估计都将失去根据。因为,为了证明使用估计方法可靠和估计结果正确,总需要有一个判断的标准,如果存活人口数据和死亡人口数据都有误,则哪一个也不能作为判断标准。没有标准,也就无从判断修正结果正确与否。好在中国人口普查的死亡报告和人口报告虽然同时有误,但并不是所有年龄上人口报告都有误,在部分年龄区间,可以认为人口报告是正确的,我们可以以这部分人口为基础,推测死亡报告的误差。

利用两次人口普查数据用死亡年龄分布资料估计成人死亡率(Preston 等,1982)或从死亡年龄分布中估计成人死亡率(Bennett 等,1981)等都要有一些条件。例如,要求有较准确的年龄人口,或者有较准确的死亡人口资料,或者死亡漏报在各个年龄程度相同,或者要求人口是稳定人口。对于中国的人口资料,这些条件常常难以满足。如出生数有波动,死亡率在下降,稳定人口的条件不具备。因为存在死亡人口漏报,所以要重新估计死亡水平,并且在各个年龄,死亡漏报的程度并不相同,如婴儿死亡率是最容易漏报的。所以还要探讨用其他方法来估计实际的死亡水平。

本文在估计 20 世纪 90 年代中国人口的实际死亡水平时按这样的原则:(1)在有可靠的人口数据的年龄区间内,以普查资料计算出的死亡率为基础并对它进行修正,使修正后的死亡率与两次普查对应年龄的人口相一致。与其他可信的调查结果(主要指婴儿死亡率)基本一致。(2)对没有数据验证(即 2000 年 65 岁以下)的年龄区间,参考其他资料,在(1)的基础上进行估计。由于中国人口和死亡率数据的复杂性,用一种方法或一个式子可能无法解决全部问题,必须分情况进行不同的处理,以下分步来阐述处理过程。

(一) 对已有的死亡率和人口数据的初步修正

在修正死亡率时,一般地说,对原始数据,尽可能不要去改动。但当我们确实有根据证实原始数据有误,不修正这些数据又将对最终结果有很大影响时,不得不对原始数据做一些修正。笔者在估计 20 世纪 90 年代人口死亡率时,对原始数据做了一些修正。

1. 对年龄别死亡率进行光滑处理。由于估计是以原有的死亡率数据为基础的,按照死亡率变化的规律,过一定年龄后,死亡率应该是随年龄的增加单调上升。但直接由普查资料计算出的死亡率,在一些年龄上可能出现波动的“异常”现象。从“五普”死亡率数据看,全国男性出现“异常”的次数有 11 次,女性出现“异常”的次数共 8 次(黄荣清,2003)。这种异常,可能是由于死亡发生的随机性产生的,更大的可能是人口和死亡人口报告的年龄有误,如果原封不动地用这样的数据,得到的估计死亡率曲线肯定也是“凹凸不平”,所以需要这样修正:保持 5 岁组的存活概率(死亡概率)不变并以此为节点值,利用三次样条函数进行插值,求出单岁的存活概率,这样就可得到一条光滑的死亡率曲线。处理后的年龄别死亡率,虽然对原始数据做了一些小的改动,但总体死亡水平(预期寿命)保持不变。

2. 高龄男性人口数据的修正。从图 3 上看,女性高龄人口数据质量尚可,我们可以利用女性人口数据来修正男性人口。修正方法是:以 90 岁、91 岁和 100 岁以上女性 1990~2000 年的留存率为节点,男性和女性在 90 岁、91 岁和 100 岁以上留存率之比作为节点的函数值,利用有理函数插值方法,求得男性在 90 岁以上的留存率,从而估计 2000 年男性在这些年龄上的人口,经过调整后的人口与原数据比较见表 2。从表 2 可以看出,在 90 岁以上的人口总数并没有改变,但增加了 95 岁以下的人口,减少了 95 岁以上的人口。其实际意义是假定部分 95 岁以下的人高报了年龄。

表 2 2000 年 90 岁以上人口数的修正

年龄	原数据	调整后人口数	差
90	88128	88128	0
91	53460	53460	0
92	40460	40965	505
93	27802	29767	1965
94	19908	22689	2781
95	16617	16583	-34
96	12543	11646	-897
97	10068	8167	-1901
98	7599	5884	-1715
99	4546	3842	-704
100	4635	4635	0

(二) 55~90 岁的死亡力估计

55~90 岁是人口数据资料质量较好,且有两次普查登记死亡率做参考的年龄段。

笔者曾经提出,设报告死亡力为 $u_0(x)$,实际死亡力为 $u(x)$,假定 $u(x)$ 和 $u_0(x)$ 有线性关系:

$$u(x) = a \cdot u_0(x) + b \quad (6)$$

对公式(6)两边积分:

$$\int_0^x u(y) dy = a \int_0^x u_0(y) dy + bx$$

得到: $U(x) = aU_0(x) + bx$ 或

$$\ln l(x) = a \ln l_0(x) - bx \quad (7)$$

若两次普查的时间间隔为 t ,有:

$$\ln[l(x+t)/l(x)] = a \ln[l_0(x+t)/l_0(x)] - bt \quad (8)$$

由于 $tRx = P(x+t, t)/P(x, 0) = L(x+t)/Lx \approx l(x+t+0.5)/l(x+0.5)$, 令死亡报告得到的生命表的存活概率为 $l_0(x)$, 两次普查得到的人口留存率为 tRx , 由公式(3), 得到:

$$\ln(tRx) = a \ln[l_0(x+t+0.5)/l_0(x+0.5)] - bt \quad (9)$$

如果上面的假定成立, 两次普查得到的实际人口留存率的对数与报告静止人口存活率的对数呈线性关系。如果两次普查的人口数据基本正确, 则由此求出的留存率就反映了实际的死亡力。利用公式(9)估计出参数 a 和 b , 再用公式(7), 由报告的 $l_0(x)$ 可算出代表实际死亡水平生命表中的 $l(x)$, 进一步可以算出死亡概率等其他指标(黄荣清, 1994)。

表 3 不同数据的回归结果

时间(年)	性别	b	a	决定系数	年龄范围
1982~1990	男	-0.00012	1.14967	>0.99	(16, 90)
	女	-0.00003	1.17373	>0.99	(16, 90)
1990~2000	男	-0.00121	1.11502	>0.99	(65, 90)
	女	-0.00056	1.15601	>0.99	(65, 90)

注: 这里的年龄指两次普查中后一次普查年龄。

根据笔者的验证, 在多数情况下, 利用公式(9)估计出的实际死亡力有很高的准确度。例如, 把 1982~1990 年和 1990~2000 年生命表各年龄的留存率对数作为自变量, 两次普查相对应的同期群人口比作为因变量, 对已知数据回归, 就会得到回归系数 a 、 b 和决定系数(见表 3)。

如果忽略 b , 公式(6)的意义表示实际死亡力是估计死亡力的 a 倍, 换句话说, 每个年龄死亡率的相对误差为 $100(a-1)\%$ 。例如, 由表 3 可知, 1982~1990 年的年龄死亡力的误差, 男性为 14.97%, 女性为 17.37%。1990~2000 年的年龄死亡力的误差, 男性为 11.50%, 女性为 15.60%。1990~2000 年实际人口的留存率对数与相应的直线回归值如图 4 和图 5 所示。

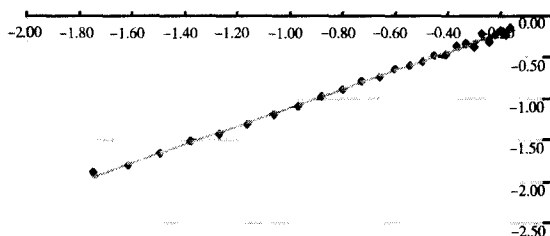


图 4 1990~2000 年男性留存率的回归

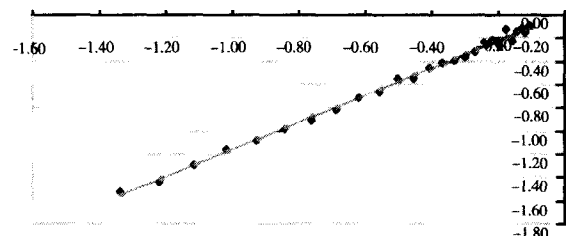


图 5 1990~2000 年女性留存率的回归

在公式(6)基础上,还可进一步改进。把公式(6)中的常数 b 改成函数 $b(x)$,即:

$$u(x) = a \cdot u_0(x) + b(x) \quad (6')$$

借用构造死亡模型时常用的词,在(6')中, a 表示水平, $b(x)$ 表示模式。

与公式(7)类似,有:

$$U(x) = a U_0(x) + B(x) \quad (7')$$

这里 $B(x) = \int_0^x b(y) dy$ 。同样,与公式(9)类似有:

$$\ln(tRx) = a \ln[L_0(x+t+0.5)/L_0(x+0.5)] - B(x+0.5) + B(x-0.5) \quad (9')$$

一般不必过多考虑 $b(x)$ 的形式,可假设 $b(x)$ 具有多项式形式: $b(x) = b_0 + b_1x + b_2x^2 + \dots$ 。

由两次普查的存活率数据 tRx ,利用线性多元回归方法,估计出参数 a, b_0, b_1, b_2, \dots ,再利用公式(7'),估计出实际死亡水平。

在 55~80 岁(1990 年的年龄),利用“四普”、“五普”的人口数,由公式(9')有:

$$\ln[P(x+10, 00)/P(x, 90)] = a \ln[L_0(x+10)/L_0(x)] + B(x-0.5) - B(x+0.5) \quad (x=55, 56, 57, \dots, 80)。$$

这里 $P(x, 90)$ 指“四普” x 岁的人口, $P(x+10, 00)$ 指在 2000 年年中和 $P(x, 90)$ 相对应的人口。 $L_0(x)$ 为由普查资料算出的 x 岁静止人口, $B(x) = \int b(y) dy$, $b(x) = b_0 + b_1x + b_2x^2 + \dots$ 。

由 $\ln[P(x+10, 00)/P(x, 90)]$ ($55 \leq x \leq 80$) 的值,利用最小二乘法,可决定参数 a, b_0, b_1, b_2, \dots ,并由此估计“四普”和“五普”的 55~90 岁各年龄的留存率并由此计算该年龄区间内各年龄的存活概率及其他一些死亡指标。

(三) 55 岁以下的年龄死亡力估计

对于 55 岁以下的年龄死亡力,由于没有可靠的数据,估计有一定的随意性。由实际死亡力与报告死亡力模型的构造, a 值的意义表示实际死亡力 $\mu(x)$ 是报告死亡力 $\mu_0(x)$ 倍数,是个常数,不依年龄变动而变动,可以用它来估计 55 岁以下的年龄死亡力(除了 0 岁外),即令: $U(x) = aU_0(x)$ ($x < 55, x \neq 0$)。

在本文中,根据公式(9')的计算结果($b(x)$ 取 4 次多项式),男性的 a 值为 1.086,女性的 a 值为 1.127。

由公式(6)或(6'), a 的意义表示实际死亡力对登记死亡力的倍数,没有考虑模式差异,所以上式的估计只能说是比较谨慎的、相对粗略的估计。

(四) 高龄人口死亡率的估计

由于“四普”公布的死亡资料没有 90 岁及以上分年龄的死亡资料,所以无法用上述方法估计 1990~2000 年 90 岁及以上人口平均死亡率。虽然“三普”和“五普”提供了 90 岁及以上分年龄的死亡资料,其数据质量不高,需要修正。由前面算出的 90 岁以下的静止人口, $L(x) = L(x-10) \cdot P(x, 00)/P(x-10, 90)$ ($x=90, 91, \dots, 99$)。

为了求得半开区间 $[100, \infty)$ 的人口数,设 1990 年 90 岁及以上人口和 2000 年 100 岁及以上人口之比为 R ,假定它对应于静止人口 $T(90)$ 和 $T(100)$ 之比,则:

$$T(90)/T(100) = \sum_{90}^{\infty} P(x, 90) / \sum_{100}^{\infty} P(x, 00) = R \quad (10)$$

$$T(90) = T(100) + \sum_{90}^{99} L(x) \quad (11)$$

由于公式(11)右边第二项已由前面求出,则由公式(10)和(11)的关系,可估计 $T(100)$,即: $T(100) = \sum_{x=90}^{99} L(x) / (R-1)$, $l(x) = 0.5[L(x-1) + L(x)]$ 。

至此,我们已估计出 1990~2000 年生命表中全部年龄的 $l(x)$,以此可估计生命表的其他函数。

(五) 婴儿死亡率的估计

婴儿死亡率在数据采集和生命表的编制中比较特殊,要结合实际调查数据来估计。

对于“三普”时的婴儿死亡率,周有尚等(1989)根据卫生部在1986年进行的婴儿死亡调查结果,发现“三普”报告的婴儿死亡率过低。若假定从1981~1986年婴儿死亡率无明显变化,则在被调查地区,“三普”报告的婴儿死亡,农村少报了约44%,城市少报了4%,婴儿死亡率估计为50.4‰。美国学者鲍思顿(1992)使用公开出版的中国人口地图集上的2378个县、市中的2306个县、市(除西藏自治区中的72个县、市)的婴儿死亡率数据分析后估计,中国的婴儿死亡率在1981年为56.2‰。

1988年全国2‰妇女生育节育调查结果显示,20世纪80年代中国的婴儿死亡率为39.33‰(顾江等,1993)。1991~2000年全国疾病监测网的数据报告显示,20世纪90年代中国婴儿死亡率逐步下降,1991年婴儿死亡率为50.2‰,1995年为36.4‰,2000年为32.2‰。根据联合国人口年鉴公布的数据,20世纪80年代初,中国婴儿死亡率为50‰左右,90年代上半期,中国婴儿死亡率为46.7‰。2000年,在32‰左右。

另外,据世界银行估计,20世纪80年代初,中国婴儿死亡率在40‰以上,到2000年,降至30‰。笔者根据2001年167个国家的人均国民收入和婴儿死亡率数据的回归方程估计,2001年中国婴儿死亡率为30‰。

上述数据由于来源不同,使用的估计方法不同,在不同年代,婴儿死亡率的估计有较大差别。但从不同的数据中,我们还是可以得到这样的结论:婴儿死亡率随年代推移在不断下降,男婴死亡率下降要快于女婴,且女婴死亡率高于男婴。根据不同数据判断,估计20世纪80年代初中国婴儿死亡率在50‰以上,90年代初在40‰,2000年在30‰。80年代婴儿死亡率平均在45‰左右,90年代平均在35‰左右,不会有太大的误差。

根据上述婴儿死亡率水平和几次全国人口普查中婴儿死亡率的性别差别,参考最常用的Coale和Demeny的分区模型生命表(1966)和联合国发展中国家的模型生命表,笔者估计出男女婴儿死亡率。

(六) 2000年生命表和年中人口数的估计

上面作为推算基础的2000年人口 $P(x,00)$,是年中人口,但“五普”提供的是11月1日的人口,与年中7月1日差4个月,这样,“五普” x 岁人口与“四普” $x-10$ 岁人口,从相同的出生时间上看,并无确切的对应关系,为此,需要将“五普”人口数调整到2000年年中时的人口。调整方法为:设2000年生命表的 x 岁静止人口为 $L(x)$,普查时点 x 岁人口为 $P_0(x)$,则在2000年普查时点对应1990年有相同出生时间的 $P(x-10,90)$ 人口为: $P_1(x)=2P_0(x)/3+P_0(x+1)/3$,要把这部分人口从11月1日调整到7月1日,考虑到死亡因素影响,其人口数还将发生变化,设: $L_1(x)=2L(x)/3+L(x+1)/3$,则在2000年年中,与 $P(x-10,90)$ 对应的人口为: $P(x,00)=P_1(x) \cdot L(x)/L_1(x)$ 。

虽然“五普”提供了死亡人口资料,由此可计算出普查时点前一年(1999年11月1日至2000年10月31日)的生命表,并以此作为2000年生命表的估计,但我们在前面的讨论中已经说明人口普查有死亡漏报,由人口普查数据计算的死亡力要低于实际死亡力。这样,用普查的生命表估计出的 $P(x,00)$ 可能会少于实际对应的人口,从而过低地估计两次普查间的人口死亡力并由此带来其他误差。为了解决这一问题,先设定一个很小的数作为误差界限,以普查资料做出的生命表为初值,通过它估计两次普查间的死亡力和2000年的死亡力,把它与初始的生命表进行比较,如果两个生命表的差别小于规定的误差,则停止计算,否则,则对初始的生命表加以修正,重新开始计算直到满足规定的误差为止。其计算框架如图6所示。

计算的基本想法是：从普查登记的死亡率出发，根据普查的人口数据对其修正，分别估计出 1982~1990 年和 1990~2000 年两次普查间的死亡力。假设 1982~1990 年，1990~2000 年到 x 岁的存活概率分别为 $l_1(x)$ 和 $l_2(x)$ ，设 $l_1(x)$ 和 $l_2(x)$ 有关系： $\ln \ln[1/l_2(x)] = A + B \ln \ln[l_1(x)]$ 。

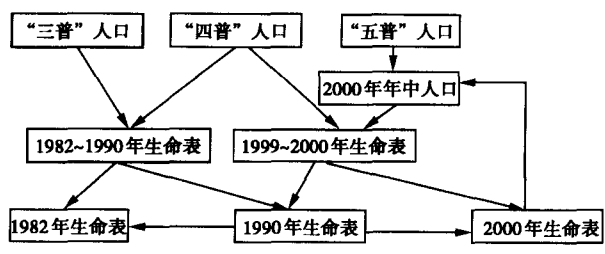


图 6 20 世纪 90 年代中国人口死亡力的估计计算框架

这里 A、B 可由各年龄的 $l_1(x)$ 、 $l_2(x)$ 直接估计出。把 1982~1990 年的死亡水平看做相当于 1986 年的死亡水平，1990~2000 年的死亡水平看做相当于 1995 年的死亡水平，则利用内插的方法，可估计 1990 年的死亡水平： $\ln \ln[1/l_{90}(x)] = 4 \ln \ln[t(x)]/9 + 5 \ln \ln[1/t(x)]/9$ 。根据估计的 $l_{90}(x)$ 和 $l_1(x)$ 估计 1982 年的死亡力、2000 年的死亡力为： $l_{82}(x) = l_1(x)l_1(x)/l_{90}(x)$ ； $l_{00}(x) = l_2(x)l_2(x)/l_{90}(x)$ 。

由 $l_{00}(x)$ 计算 $L_{00}(x)$ ，把它和开始的 2000 年 $L(x)$ 估计值相比较，如果两者的差在规定的范围内，则停止，否则以两者的平均值再重新计算，通过反复计算，最后可同时得到 1982~1990 年、1990~2000 年、1982 年、1990 年和 2000 年的生命表。

(七) 各年龄段死亡率的整合

在估计全年龄的死亡时，由于不同年龄段的数据情况不同，笔者相应地使用了不同方法。这样，在使用两种不同方法的交接年龄，死亡率曲线会出现不“光滑”。为此，还要对死亡率曲线做光滑处理，处理方法类似于死亡率数据的预处理，这里不再赘述。

通过上述步骤的计算，可求得各年龄的 $l(x)$ 并由此编制相应的生命表（见表 4），得到的结果是，在 20 世纪 90 年代，中国人口死亡的实际水平男性预期寿命为 67.97 岁，女性预期寿命为 71.34 岁。均比人口普查登记资料算出的结果低 1.5 岁左右。其他有关的死亡指标和生命表中的指标见表 4。

表 4 20 世纪 90 年代中国人口生命表中有关指标(算法二)

年龄	男性			女性		
	年龄别死亡率	死亡概率	平均预期寿命	年龄别死亡率	死亡概率	平均预期寿命
0	34.07	33.17	67.97	40.02	38.78	71.34
1	2.09	8.31	69.30	1.85	7.36	73.21
5	0.84	4.19	65.86	0.72	3.59	69.74
10	0.59	2.96	61.13	0.44	2.20	64.98
15	1.00	4.98	56.30	0.76	3.81	60.12
20	1.44	7.18	51.57	1.10	5.46	55.34
25	1.52	7.59	46.93	1.13	5.66	50.63
30	1.86	9.24	42.27	1.29	6.43	45.90
35	2.44	12.13	37.64	1.60	7.97	41.18
40	3.47	17.19	33.07	2.27	11.27	36.49
45	5.14	25.40	28.60	3.47	17.23	31.88
50	8.10	39.76	24.28	5.53	27.30	27.39
55	13.02	63.21	20.17	8.75	42.87	23.08
60	21.88	104.04	16.35	14.04	67.99	19.00
65	34.85	161.02	12.94	23.55	111.69	15.19
70	59.12	258.68	9.92	42.23	191.97	11.76
75	93.93	380.55	7.47	68.82	294.88	8.93
80	152.88	545.11	5.53	117.47	451.95	6.58
85	216.56	670.30	4.31	172.24	588.00	4.99
90	258.13	729.24	3.69	251.30	723.92	3.82
95	288.51	770.45	3.18	271.93	747.69	3.41
100	450.03	1000.00	2.22	379.67	1000.00	2.63

注：本表数据按本文中介绍的方法计算。

四、小结与讨论

按估计的 1990~2000 年生命表，在这样的死亡力作用下，由 1990 年人口推算 2000 年 11 月 1 日的人口与普查人口之差如图 7 所示。

由图 7 可知，推算人口与“五普”人口在 65 岁以后非常接近。实际上，推算的 65 岁以后人口与“五普”人口，男性差 10531 人，女性差 2791 人，合计为 13322 人；而 1990 年 55 岁以上人口，男性为 6800 万人，女性为 7070 万人，合计为 13870 万人。换句话说，1.38 亿人经过 10 年后的人口数和推测的人口只差 1.33 万人。可见，推算的误差非常小。

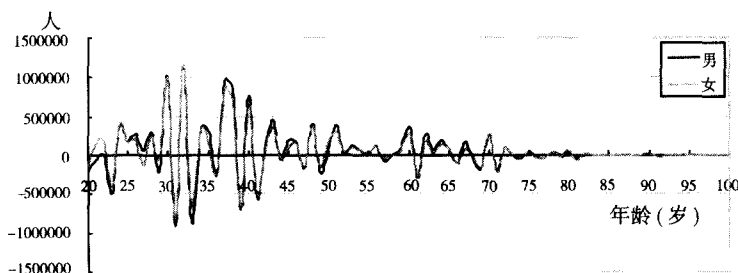


图 7 2000 年登记人口与估计人口之差

本文在推测实际死亡率时,是用已估计出的 55(65)岁以上的实际死亡力和报告死亡力的关系来估计 55 岁以下的死亡力,在估计时不考虑年龄的影响,是一个谨慎的估计。如果加入年龄影响,人口死亡力可能会更高一些,或变小一些。

从经验上也可以知道,经济发

展水平对人口健康提高有促进作用。利用美国人口咨询局《世界人口数据表》(2003 年)的数据,把一些明显不好的数据和特殊的数据删除,各国(共 160 多个国家)按购买力平价计算的人均国民收入和预期寿命进行相关分析,证明两者存在明显的正相关关系。人均国民收入越高,平均预期寿命就越高,而且这种相关具有统计学意义上的显著性。同时,当人均收入和平均寿命达到一定水平后,人均收入对平均寿命提高的作用变得很小。人均收入和平均寿命的这种关系,类似于 Logistic 函数。若设人均收入为 I ,平均寿命为 E , E 为 I 的函数,对人均收入和预期

寿命按 Logistic 函数拟合。则男性的 Logistic 函数为: $E(I) = \frac{81.30}{1 + 59.32 \exp(-0.06613 \cdot I)}$, 女性的 Logistic 函数为: $E(I) = \frac{86.56}{1 + 79.56 \exp(-0.7003 \cdot I)}$ 。

按照世界银行的计算,2001 年中国人均国民收入,按购买力平价为 3950“元”。若按照上面拟合的函数计算,则 2001 年男性平均寿命估计为 65.13 岁,女性为 70.12 岁(黄荣清、庄亚儿,2004)。根据图 6 的框架计算,中国 2000 年人口的预期寿命男性在 69 岁左右,女性在 72 岁左右,都超过了上述估计的预期寿命。与女性相比,中国男性预期寿命比拟合函数的值高得多,而女性的实际值则和拟合函数的值比较接近。这也从一个侧面证明,本文对中国人口实际死亡水平的估计并没有高估。

参考文献:

1. 任强等(2004):《20 世纪 80 年代以来中国人口死亡的水平、模式及区域差异》,《中国人口科学》,第 3 期。
2. 跨世纪的中国人口编委会(1994):《跨世纪的中国人口(综合卷)》,中国统计出版社。
3. 周有尚等(1989):《中国婴儿死亡率分析》,《中国人口科学》,第 2 期。
4. 顾江等(1993):《中国婴儿死亡率及其影响因素分析》,《中国生育节育抽样调查北京国际研讨会论文集》,中国人口出版社。
5. 鲍思顿(1992):《中国婴儿死亡率模式》,载于翟振武主编:《人口数据分析方法及其应用》,外文出版社。
6. 黄荣清(1994):《中国 80 年代死亡水平的研究》,《中国人口科学》,第 3 期。
7. 黄荣清(2003):《年龄别死亡率数据异常的检验与讨论》,《人口与经济》,第 6 期。
8. 黄荣清、庄亚儿(2004):《人口死亡水平的国际比较》,《人口学刊》,第 6 期。
9. Bennett, N., and S. Horiuchi(1981), Estimating the Completeness of Death Registration in Closed Population, *Population Index*, Vol. 47, No. 2.
10. Coale, A. and P. Demeny(1966), *Regional Model Life Tables and Stable Population*, Princeton University Press.
11. Preston, S. H. and A. J. Coale(1982), Age Structure Growth Attrition and Accession: A New Synthesis, *Population Index*, Vol. 48, No. 2.

(责任编辑:朱 萍)

ABSTRACT

Problems Pending Solutions in Macro-control Policies

Liu Jianxin • 2 •

With China's reform going in depth, a number of economic phenomena unseen or rarely seen in other countries have emerged as the result of conflicts in the deep structure of the national economy. To explain such phenomena is beyond the power of traditional economic theories. It is therefore impossible to fundamentally solve the problems by continuing the existing macro-control policies without any innovative measures. In this essay, weaknesses of the macro-control policies themselves and the deviations of the policy effects from what expected by the efforts are analyzed. Hypotheses of Keynes' demand management theory are revised on a theoretical basis. Based on the theory of short-term supply shortage and long-term production element shortage, demand management is not a viable approach to full employment. In addition, a number of measures are proposed for the prevention of stagflation.

Mortality Level of the 1990s in China

Huang Rongqing • 11 •

The death of population is often underreported in census. This paper analyses social determinants of death underreporting in China, and proves the existence of death underreporting by analyzing two census data. It also examines the quality of census data, and suggests a correction method according to registration population and mortality, which is important for estimating the factual mortality. The study shows that life expectancy is 67.97 for male and 71.34 for female in 1990s, 1.5 years less than reported by census data.

The Underreporting and Sex Ratio of Lower Age Group in the Fifth Census: Investigating School Enrolment Data

Zhang Qing • 21 •

This paper finds that 20.42 million under age 17 were underreported in the 5th census, by comparing school enrolment data. The total population in 2000 should be between 1253.15-1260.87 million then, which is 4.96-12.68 million less than published data, the net underreporting is between 8.04-15.76 million, and the rate of underreporting is 0.65-1.27%. Meanwhile, it is estimated that sex ratio of age group 0-4 is about 114 in the fifth census.

Urbanization Effects and Gender Preference

Gu Shengzu Chen Lai • 30 •

With an analysis on gender preference, this paper discusses the causes of imbalanced sex ratio at birth and suggests that promoting rural-urban migration (or urbanization) and introducing urban style of marriage/living arrangement and social security system to rural areas would improve the social-cultural environment, under which the imbalanced sex ratio at birth can be eliminated.

Direct Determination of High Sex Ratio at Birth in Rural China: A Cohort Study

Wu Zhuochun and Others • 38 •

A cohort analysis was performed to examine the direct determinants of high sex ratio at birth in rural China, utilizing regular record data from rural family planning system. The evidence-based study found that the sex selective-induced abortion is the most important determinant of high sex ratio at birth, and the second determinant is underreporting of live birth of girls, and the third is poorer care of female newborn. The three factors contribute 70%, 20% and 10% respectively to the reported high sex ratio within one week of birth.

The Design of Dual Basement Collection about Basic Pensions

Li Min Ma Lijun • 44 •

The purpose of dual basement collection lies in distributing pension burden among the enterprises and ensuring the collecting rate and amount. The basic pensions currently comprise of true-basic pensions and cost of reform. When the "dual basement" collection is put into force, we can collect the true-basic pensions at the rate of 7-8% on the base of total wages, and the cost of reform at 2-3% on the profits before taxed. The pensions coming from the profits are regarded as the mandatory loan which the government borrows from enterprise, the government should repay this debt gradually in the future.