

# 中国迁移流动人口的统计定义

## ——人口普查视角下的分析

周 皓

**【摘 要】**厘清概念、明晰定义有助于正确认识普查中有关人口迁移流动的数据。文章在强调明确登记对象和完备登记信息的基础上,着重讨论户口标准中流动时间及其可能带来的问题、迁移人口的重要性及无法识别问题、5年前常住地样本的混合属性问题。文章认为:(1)七普登记的离开户口登记地时间使登记对象变成一次“新”的不同于以往3种统计口径的迁移流动,可能导致信息陈旧、低估城—城流动比例、无法估计小区域的新增流动人口等问题,普查题项最好能增加离开户口登记地后的流入地及最近一次迁移流动时间。(2)忽略迁移人口将低估中国人口迁移流动的规模和活跃程度。中国情境下的研究应重视区分人口迁移与人口流动两个概念,并从普查题项设计、统计定义到数据分析等全方位加强对中国迁移人口的调查与研究。(3)5年前常住地样本是现有流动人口、返迁人口和迁移人口等多种属性人口组成的混合体;研究时应注意混合样本属性并进行分别讨论。文章建议加强流出地视角的研究,充分开发出生地信息,并进一步开放普查数据的使用。

**【关键词】**人口普查;迁移;流动;统计定义

**【作者简介】**周皓,北京大学社会学系教授。

**【原文出处】**《中国人口科学》(京),2022.3.17~30

**【基金项目】**本文为国家社会科学基金项目“中国流动人口调查的比较研究”(编号:19BRK039)的阶段性成果。

### 一、引言

随着2020年第七次全国人口普查数据的发布与数据开发的启动,包括人口迁移流动研究在内的中国人口科学将迎来一个新阶段。然而,在利用七普数据进行中国人口迁移流动研究之前,应注意以下两个问题:(1)普查题项设置与统计定义问题,以及由此引起的一系列相关后果。例如,离开户口登记地的时间具有重要意义,但其不一定对应“最近一次迁移”,相关信息(如迁移原因、流出地城乡属性等)可能不再是最新的信息,登记的流动时间无法作

为重要的辅助信息来识别迁移人口等。这些问题与普查登记的题项有关。(2)以往研究中存在概念混淆和测量不准等问题,如有的研究不进行区分地使用迁移人口与流动人口的混合样本,得到的分析结果是两类人群的混合效应。上述问题不仅导致数据不清、对象不明等动摇研究基础,而且影响研究结果的准确性,更可能误判未来中国人口流动的发展趋势,削弱政策制定的科学性和有效性,乃至影响民生、就业、公共服务供给等一系列公共政策。

上述问题的根源在于忽略了基本概念与研究对

象的定义,未考虑概念与研究对象之间、研究对象与研究问题之间的对应关系,只是简单直接地使用数据,从而导致人口迁移流动研究存在数据误用的“乱象”。段成荣、孙玉晶(2006)曾讨论2005年及以前历次人口普查与1%人口抽样调查中流动人口的识别标准,但此后对普查视角下中国迁移流动人口的统计口径问题鲜有讨论。在即将进入七普数据深入开发阶段之前,厘清概念、准确理解是正确使用普查数据的基础。同时,反思人口普查中关于人口迁移流动的题项设置,将有益于2025年1%人口抽样调查和第八次人口普查表的设计与数据收集。为此,本文将围绕七普数据开发,厘清人口迁移流动的定义与统计口径,提出迁移流动人口的统计口径应以明确登记对象(“某次”迁移行为)和记录完备信息(“何时从何地到何地”)为标准,从普查题项设计角度讨论户口标准的统计定义,强调迁移人口的重要性,讨论5年前常住地标准的混合样本属性问题。由于七普和六普在题项设计上相近,因此,除必要说明外,本文均以七普为例,结论同样适用于六普。

本文中迁移人口指普查时点“人在户在”、但其户口登记地在规定时间内发生过跨越乡镇街道改变的人口;流动人口是指户口登记地未发生改变,但现居住地跨越乡镇街道(不同于户口登记地)且在本地居住半年以上(第二款人),或者在本地居住不满半年但离开户口登记地半年以上(第三款人)的人口。显然,这个流动人口定义对应人口普查中的第二、第三款人。在上述标准下还需剔除市内人户分离人口。

## 二、人口迁移流动的统计口径与登记信息的完备性

人口普查中有多种识别人口迁移流动的统计口径,一般包括:(1)出生地信息(可以反映终身迁移的情况);(2)一定时间(如1年或5年)前的常住地;(3)最近一次迁移流动(Siegel等,2004)。不论哪种

统计口径均包含时间维度与空间维度。如图1所示。图1中A点在空间维度表示登记的出生地点,在时间维度表示出生时间。由于事件的空间和时间必然是重合的,因此,图1中将空间维度与时间维度合并于点A上。同理,D点在空间维度表示现居住地信息,在时间维度表示普查时点。B点和C点也同理。图1的上半部分表示统计定义,如以A为起点、D为终点,则两点之间在空间维度(登记信息)的异同表示个体从出生到普查时点是否发生过迁移。如果登记的出生地与现居住地相同,则该个体未发生过迁移流动;否则可认为该个体有过迁移流动行为。5年前常住地和最近一次迁移标准也同理,只是对应的时间不同。3种统计口径对应3种迁移行为,且对应的普查题项不相同。

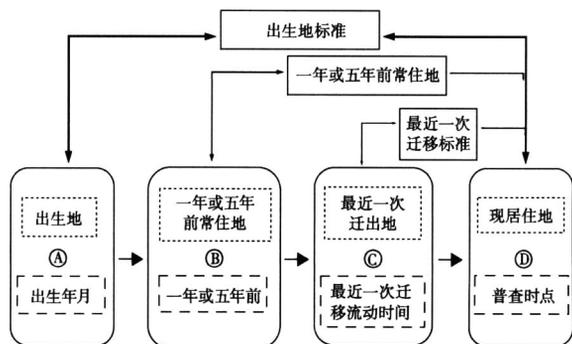


图1 国际通用的人口迁移的3种统计口径

普查登记信息的完备性包括登记对象和信息完备。其中,登记对象指记录哪次迁移流动行为(如上述3次不同的迁移行为或特定的某次迁移行为);信息完备是指普查应完备地记录该次迁移流动行为“何时从何地到何地”的信息,包括流动时间、流出地与流入地3个方面。完备性是人口迁移流动调查的基本原则。如果将现居住地视为迁入地,则出生地或5年前常住地可以视为迁出地,从而判断其是否有终身迁移或5年间迁移。虽然无法了解其迁移流动的时间,但考虑到普查项目的简约,也基本能够满足上述原则。最近一次迁移可以反映一段时期内(如5年内)的人口迁移流动情况,而且该时间有助于粗略估计5年间历年的人口迁移流动情况,

也有助于区分迁移与流动,因此,最近一次迁移流动行为及其时间是人口迁移流动研究中的重要因素之一,有必要详细记录其时间与迁出地(现居住地为迁入地)。

### 三、户口标准中的流动时间问题

本文中户口标准是指根据普查表中户口登记地信息、结合流动时间来识别流动人口的统计口径。这是三普以来最经典的识别流动人口的统计口径。这一口径涉及户口登记地与流动时间两个方面,其中关键在于“时间”。流动时间是普查过程中流动人口登记的重要维度之一,是数据分析过程中对流动人口的识别标准和决定迁移流动登记对象的重要标准,在分析与研究中具有重要的辅助作用。六普和七普的时间题项设置完全不同于五普,但在过去十几年间几乎没有讨论过这一问题。本文将讨论普查登记的“离开户口登记地时间”的重要性和缺陷,以及这种缺陷对后续研究的影响。

(一) 户口标准相关题项设置在七普与五普间的差异

户口标准相关题项设置的改变主要为:(1)登记方式的改变。七普分别记录户口登记地与离开户口登记地的时间,地址信息不再包括时间信息,明确区分了空间标准与时间标准,相当于第二、第三款人标准被分解成空间和时间两个题项;五普的户口登记状况同时包含空间信息与时间信息。(2)登记地点的变化。七普登记了两个地址,分别是户口登记地和普查时点居住地;五普只有户口登记状况(且无具体空间地址信息)。(3)空间尺度的改变。七普的空间范围下沉到村(居)委会,包括本村居委会和本乡镇街道的其他村居委会。(4)时间标准的改变。七普登记的是“离开户口登记地时间”,五普登记的是“何时来本乡镇街道居住”,即来本地居住的时间。这是完全不同的两个时间点。

与人口迁移流动相关的其他题项设置的改变为:(1)迁移原因的改变:这与登记时间的改变有关,

七普是“离开户口登记地原因”,五普则是“迁移原因”(对应来本地居住时间);(2)来源地城乡属性的改变:七普登记了户口登记地类型(乡镇街道),而五普登记迁出地类型(对应来本地最近一次迁移)。(3)户口性质的改变:五普和六普中的“户口性质”(包括农业和非农两类),七普改为“是否有农村土地承包经营权”,这是人口普查历史上第一次取消登记户口性质,具有重要的标志性意义。2005年和2015年两次1%人口抽样调查都包括“1年前常住地”题项;后者还包括“在本市居住时间”题项。

(二) “离开户口登记地时间”的意义及问题

“离开户口登记地时间”是一个非常重要的时间点。从普查登记看,该时间点可以弥补七普户口登记地只有空间信息、不包括时间维度的不足,从而完成流动人口的识别并满足普查登记原则。从个体流动经历看,离开户口登记地时间所对应的流动行为可能是除出生地标准以外的个体第一次流动,它是未来可能再流动的基础。在研究上,根据该时间不仅可以测量流动人口外出的时间长度及相应的留存率,还可以研究初次流动时的各种特征,如初次流动的真实原因、流动人口流出地的城乡属性等,并可以作为重要辅助信息反映流动人口各种状况及其历时变化。虽然离开户口登记地的这一次迁移流动及其时间对相关研究具有重要意义,但从登记对象和信息完备性看,离开户口登记时间及其相应题项不完备。

#### 1. 登记对象

七普记录的离开户口登记地的这一次迁移流动,可能不同于“最近一次迁移”。假设如图2所示,某个体在2014年9月10日(对应图2中的O地)离开户口登记地前往C地,并在C地居住了4年,然后于2018年9月20日离开C地到现居住地D地。仅按照户口标准识别,其户口登记地在O地,居住地在D地,因此其属于流动人口,再结合时间标准,其“离开户口登记地时间”应该按照2014年9月10日登记,因此被识别为流动人口。显然这个时间点不是该个体真正到本

地(现居住地D地)的时间,也不同于“最近一次迁移或流动”时间。虽然不知道这种情况的发生比例有多高,但“存量”流动人口中有多次流动经历的情况比较普遍。2015年1%人口抽样调查数据同时提供了离开户口登记地时间和来本市居住时间,对这两个时间交互分析后发现,两个时间不相同的流动人口占13.54%。随着居住时间的延长,该比例呈上升趋势。不满半年时该比例为8.90%,3~4年和4~5年的比例分别高达18.06%和17.83%。由于七普表现出流动人口规模剧增,以及流动时间延长,这一比例可能还会增加。因此,七普事实上记录一次除上述3种统计口径之外的“新”的流动经历,这是流动人口识别与分析中的关键点。

## 2. 信息完备性

即使作为一次“新”的迁移流动行为,它登记的信息仍不完备。按照“何时从何地流到何地”的原则,上述题项只记录户口登记地(流出地)及其时间,没有记录相应的流入地。仍从上例看,该流动人口在2014年9月10日离开户口登记地,其实际流入地假设为图2中的C点,而普查表中没有相应的题项,只能将离开户口登记地后的流入地默认为现居住地。现实情况可能是,对有多次流动经历的人而言,离开户口登记地后的流入地并不等同于现居住地。可见,七普的题项设置未能完备地记录离开户口登记地这一次流动的所有信息。

### (三) 登记的流动时间可能导致的问题

除了信息不完备外,“离开户口登记地时间”由于不同于最近一次迁移时间而带来以下几个主要问题。

#### 1. 相关信息无法与最近一次迁移相对应

七普登记了“离开户口登记地”这一次新的流动,

而不一定是“最近一次流动”;因此,相关的信息(如迁移原因、来源地城乡属性等)对应的都是这次“新的”流动。从原因看,七普登记的迁移原因并不一定是最近一次流动的真正原因。如上例中,该个体离开户口登记地的原因可能是婚姻嫁娶(图2中的O点到C点的流动),但最近一次流动(图2中的C点到D点)的原因可能是务工经商或工作调动。七普登记的原因应该是婚姻嫁娶,而非务工经商,即由于登记对象并非最近一次迁移流动而导致七普登记得到的可能是“过期”的原因。而且由于登记对象的不同,七普登记的迁移原因与五普结果并不具有可比性,毕竟登记的是两次不同的迁移流动行为。这在后续的分析中需要注意。

中国流动人口的城乡属性结构对中国人口流动阶段性及未来发展趋势的研判具有重要意义。由于是从户口登记地角度的考察,因此七普的来源地城乡属性似乎可以真实地揭示流动人口的城乡属性结构。如果深究,这种统计定义可能会高估乡—城流动的比例、低估城—城流动的比例。首先,随着人口迁移流动强度的提高,多次流动、循环流动等多种流动形式及比例都会随之提高;而多次流动更可能是城—城流动。由于登记的是离开户口登记地的迁移流动行为(且可能更多的是乡城流动),而非多次流动等情况,虽然登记的来源地城乡属性确实反映了流动人口真正的最初流出地,但未必是最新的来源地属性,因此当前的结果相当于有一部分最新的城城流动被旧的乡城流动所代替,从而可能会低估城—城流动所占的比例。其次,如果考察2000年以来历次普查或1%人口抽样调查的结果,2000、2005、2010、2015和2020年乡城流动的比例分别为52.20%、61.40%、63.20%、48.90%

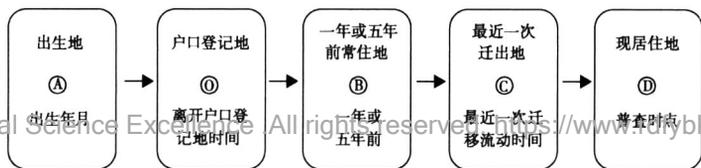


图2 七普中离开户口登记地时间的示意图

和 66.26% (段成荣等, 2019; 周皓, 2021)。2000 年以来, 乡—城流动人口的比例不降反升, 甚至七普比六普高 3 个百分点, 比五普高 14 个百分点。按照朱宇等 (2016) 对中国人口流动的阶段性判断, 中国乡城流动人口比例在逐步下降。理论判断与现实之间的差距有多大? 是现实如此还是由于题项设置问题? 显然, 这个问题的回答关乎对中国未来人口迁移流动发展趋势的判断。在题项上几次普查之间存在差异, 五普登记的是最近一次流动时间及其对应的流出地城乡属性, 而六普和七普登记的是离开户口登记地时间及户口登记地的城乡属性。这种理论与现实间的差距只能等未来的数据予以检验。

类似的还有一些衍生指标, 如流动人口社会融合研究可能更关心他们在本地的居留时长及其分布, 这时只能利用“来本地居住的时间”, 而非“离开户口登记地”的时间。当然离开户口登记地的平均时间同样有重要的社会涵义, 但流入地相对可能更关心已在本地居住多长时间。这种相关信息或陈旧, 或悖论等都与登记对象并非最近一次迁移有关, 而其背后则是登记的时间问题。

2. 登记的流动时间无法作为省或小区域内流量与存量流动人口的识别标准

人口迁移流动研究有必要加强“流量”与“存量”人口间的比较 (Qi 等, 2021; 周皓、刘文博, 2022)。根据研究问题区分并选用合适的“流量”或“存量”样本, 以解决因果变量间的时间顺序和结构性偏差等方法论问题, 完成不同研究目的与任务, 揭示不同研究问题所对应的社会现象。虽然从全国而言, 七普登记的流动时间可以满足对全国流动人口“流量”与“存量”的区别与估计, 但对分省而言, 各省无法真正了解在普查前 5 年内 (及 5 年期间历年的) 新增流动人口情况, 甚至可能会产生误判。从上例看, 该个体从 O 点 (户籍登记地假设为安徽) 离开后到 C 点 (假设为上海), 然后又从 C 点流动到 D 点

(假设为江苏)。那么对于江苏来说, 按照七普的标准, 由于其 2014 年离开户口登记地, 因此江苏直接将其视为存量流动人口, 而事实上他是 2018 年流入本地 (江苏), 按照最近一次迁移流动口径, 应为 5 年内新增的“流量”流动人口。可见七普中流动时间题项可能无法满足分省或更小区域单位流量流动人口的识别与监测, 不利于各省或更小区域单位的相关研究和政策制定。从现实与政策需求的角度看, 在了解与掌握全国流动人口状况的同时, 各省或更小区域同样非常需要且十分有必要掌握区域内流动人口的基本情况, 特别是新增的“流量”流动人口的情况。

3. 无法作为重要辅助信息用于迁移人口的识别

迁移人口的识别与估计在当下中国仍具有重要意义。从普查统计口径看, 迁移人口是“人在户在”的人口, 仅根据户口登记地无法判断其户口是否迁入, 还需要结合户口迁来本地的时间。如果能知道某个体来本地居住的时间, 也可以间接判断其是否为迁移人口。这种判断只是大概的, 还可能混杂着返迁人口。但七普登记的时间因不涉及“人在户在”而无法作为重要辅助信息用于迁移人口的识别。

上述问题是普查题项设置的问题, 虽然不会影响中国流动人口规模等大数的估计, 但会在数据分析中影响其他相关信息 (特别是与五普的比较) 的结果和迁移人口识别。在已完成普查的背景下, 没有任何方法可以改进估计结果。本文提出上述问题并不是否定该题项, 而是提请注意以离开户口登记地为标准和以最近一次迁移流动为标准是存在的差异。鉴于此, 本文建议针对普查表中户口标准的相关题项设置进行以下修订: (1) 在修订迁移流动时间题项之前, 首先需要明确登记对象, 即哪一次迁移流动行为更需要被关注。离开户口登记地这一次流动和最近一次迁移流动这二者的取舍决定于各自的意义。离开户口登记地不仅可以了解流动人口总体的

流出地情况,还可以弥补户口登记地题项中不包含时间信息的问题,而“来本地居住”则对应最近一次迁移流动,可以真正考察近5年内的流量人口。如果可能,同时登记这两次流动,既可以有效识别最近一次迁移流动(以便与通行的3种统计口径相对应),还可以进一步丰富人口迁移流动的普查信息,更详细地了解多次迁移流动及返迁的情况(这种多次迁移与返迁的情况未来会随着迁移流动的逐步增强变得更加普遍)。当然同时登记两次迁移会使题项数有所增加。如果非要二选一,则应调查最近一次迁移流动行为。因为该次迁移流动行为对应的时间可以用于迁移人口的识别,并与其他许多信息(如流动原因等)关联,同时能更真实地反映普查前5年内历年的人口流动情况,与真正的流量人口相对应,弥补5年前常住地信息的不足。(2)信息完备性要求登记每一次迁移行为的发生时间、流出地与流入地。因此,针对离开户口登记地这一次流动行为应加入离开户口登记地后的流入地信息;对最近一次迁移流动,则需加入“从何地来本地居住”和“何时来本地居住”这两个信息,以保证登记信息的完备性。

#### 四、识别迁移人口的重要性

迁移人口又指户籍迁移人口,是指普查时点“人在户在”、但其户口是在普查前一段时间内迁入的人口。魏京生(1984)提出以户口登记地是否改变为标准区分流动人口与迁移人口,这对概念一直被学界所公认和使用。然而,在人口普查中,不论是题项设置,还是普查公报(只提供城市化水平与人户分离情况)均未包含迁移人口的相关内容(相应的统计口径和统计数据),显然对迁移人口缺乏一定的关注与重视。甚至有些研究使用“迁移人口”概念对应包括流动人口在内的非纯粹迁移人口的样本,有些“流动人口”样本包含迁移人口。对此,本文认为,中国的人口迁移流动研究不仅要重视区分人口迁移与人口流动两个概念,还要从普查题项设计、统计定义到数据

分析等全面加强对中国迁移人口的调查与研究,从而全面理解中国的人口迁移流动。

迁移人口是中国人口迁移流动中的重要组成部分。迁移人口与流动人口一样,首先是发生地理空间位置改变,即常住地跨越一定地域行政界线。因此,迁移人口本身就是人口迁移流动大潮中的一部分,是人口活跃程度的重要来源之一。忽略了迁移人口,调查得到的中国人口迁移流动的规模与强度在一定程度上会被低估,进而影响对中国人口迁移流动的整体状况及其发展趋势的研判。

一段时期以来,国家与政府一直在努力倡导公共服务均等化、不断深化户籍制度改革,许多省及某些中小城市逐步放开户籍,正在逐步还原户籍登记制度本身的属性。七普第一次不再登记户口性质,代之以“是否有农村土地承包经营权”。这是社会发展进步与户籍制度改革不断深化的重要信号之一。但在现阶段乃至今后一段时间内,由于城市化道路的选择、城市发展的差异性,以及城市体系的客观存在,户籍制度仍然将是中国的基本社会制度之一;即使将来真正地实现了全面的户籍制度改革目标,将户籍制度与所附带的社会福利制度相剥离,户籍制度仍将作为其本原的人口经常性登记制度而存在。因此,户籍迁移人口将是这种社会制度背景的重要表现形式之一。人口普查作为反映国情国力的重要基础数据来源,应对此有所体现。人口普查中的迁移人口,既可以与普查中的流动人口相结合,共同从总体上反映中国人口在空间上的迁移流动状况,也可以与户籍登记制度系统中的户籍迁移人口相互比照,更可以反映户口变动(迁移)人口的分布及变动等整体状况、区域差异及发展趋势。可见,迁移人口的普查登记具有重要的现实意义。

在强调迁移人口重要性的同时,需要重视迁移人口与流动人口之间的区别。虽然二者都是地理空间上跨越一定行政界线的移动,但除了户口登记地

是否改变这一特征性差异外,中国人口迁移与人口流动是社会属性完全不同的两种行为。一是迁移人口与流动人口的各种人口学与社会特征不同,且迁移人口具有高度的选择性(如受教育水平的选择性等)。这些差异既是个体选择的结果,也是制度选择的结果。例如,虽然迁移流动的原因都是工作调动,但迁移人口的工作调动与流动人口的工作调动背后的动因、机制等可能完全不同。即二者在进入上存在较强的选择性。二是从结果看,户籍制度背景下,迁移人口与流动人口在迁入地所面对的问题与困难截然不同。以社会融合为视角,二者不仅在社会融合的经济、社会与文化、结构和认同等各维度上可能存在差异,而且在方向、途径等融合过程及融合的最终结果也不尽相同。比如,二代移民的教育机会问题。对于迁移人口而言,其适龄的子代可以直接入读公立学校,但流动人口的子代就读公立学校的机会在各城市、各地区之间仍然存在较大差异。再如,两类人口的居留意愿。迁移人口更倾向于长期,甚至一辈子都居住在户口迁入地,而流动人口不仅长期居留意愿相对较低,且其所谓的长期性也相对更短(通常以5年为测量标准)。从动因到过程再到结果,迁移人口与流动人口之间可能存在本质差异,需要特别关注,并在研究中得到体现,才有可能真正揭示两类人口不同的机制与问题,也才有可能真正反映中国的人口迁移流动状况,而不是以纯粹的“流动人口”为代表。

如上文所述,七普题项中只有离开户口登记地的时间及户口登记地性质,未包括识别迁移人口所必要的信息,因此,本质上七普数据无法真正识别迁移人口。如果要利用普查数据从“人在户在”的人口中间识别迁移人口,并满足登记信息的完备性,普查表需要增加户口迁移时间和来源地信息两个题项。这是比较理想的题项设置,但不是最简约的。若加入“来本地居住时间”,既可以从时间维度保证识别的可能性,还与最近一次迁移时间一致,并保证简约

性。这种简约方法下得到的只是近似的估计,样本中可能还包括返迁人口。这也是5年前常住地信息无法用于识别迁移人口的原因。

### 五、5年前常住地:混合样本属性

5年前常住地是人口普查中与人口迁移流动相关的另一个重要题项,是国际通行的人口迁移统计口径之一。它反映的是两个时点(普查前5年与普查时点)居住地的改变,代表5年间的迁移流动情况,类似于5年间“流量”的概念。然而,这一口径除自身固有的不足外,在中国还有混合样本属性问题,需要在数据分析与研究过程中特别注意。

#### (一)5年前常住地选项的变化

七普中5年前常住地的选项为:(1)本县(市、区、旗);(2)其他地区(请填写具体省市县地址码)。以5年前常住地为标准,选项1对应本地人口,选项2对应5年间的迁移流动人口。由于提供了省市县三级地址编码,可以区分省际和省内县际两种跨越不同行政界线的“迁移”形式。如果数据允许,甚至可以得到分别以地区和县为基本单位的迁移矩阵。以往普查的该信息只能得到省级的迁移矩阵,无法得到详细的地区层级或县级的迁移矩阵。可见,七普中5年前常住地的选项更加详细丰富。

#### (二)5年前常住地样本的混合属性

许多从地理空间角度关注人口迁移流动的研究常使用5年前常住地的OD矩阵(迁出—迁入矩阵)讨论迁移流动人口的空间分布与机制问题。但由于5年前常住地指标本身固有的局限,据此得到的迁移流动人口样本与理论定义(户口标准下)的迁移流动人口存在一定差异,且不具有可比性,而且该样本内部至少包含迁移人口与流动人口两种不同类型的人口,可以视为两种类型人口的混合体。图3在一定程度上说明了5年前常住地样本可能存在的问题。5年前常住地标准下的样本与户口标准下的流动人口样本,既有重合又有差异。这两种口径各有目标,定义及样本也有所不同。

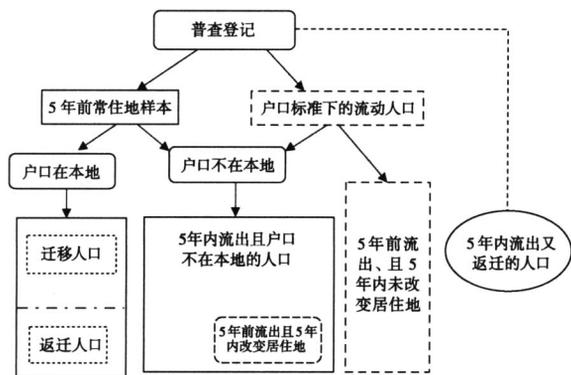


图3 5年前常住地标准下样本混合属性示意图

首先,5年前常住地标准可能忽略部分事实上的流动人口。以图2为例,某个体在5年中进行了一次流动(从B地到C地),然后又在普查之前从C地返回B地(假设D地与B地重合)。那么,该个体在5年内的这次迁移流动可能无法被真正记录(即图3中5年内流出又返迁的人口),无法记录这5年中的多次流动经历。另一种情况是,某个体在5年前作为流动人口生活于普查时居住地,且5年内未发生任何流动,那么,其5年前常住地与现居住地完全相同。在户口标准下该个体为真实的流动人口,但5年前常住地标准下为非流动人口(对应图3中5年前流出且5年内未改变居住地的人口)。这是5年前常住地标准和户口标准之间的主要差异。后一种情况属于存量流动人口,与5年前常住地标准的测量目标(流量流动人口)有一定区别,但这两种情况在一定程度上说明5年前常住地标准可能会低估人口迁移流动的规模与强度。这是5年前常住地标准固有的问题。

其次,5年前常住地样本的混合属性。假设两个个体5年前常住地均为图2中的B地,但其中个体1在普查前5年内户口迁入D地(对应图3左侧的迁移人口),而个体2则流动到D地(对应图3中5年内流出且户口不在本地的人口)。按照5年前常住地标准,这两个个体均为5年内的迁移流动人口,但前者为迁移人口,后者为流动人口。因此,5年前常住地样本事实上是由户口在本地的迁移人口与户口

不在本地的流动人口共同组成的,这就造成5年前常住地样本具有混合样本属性。当然可以将5年前常住地与现驻地、户口登记地三者相结合以进一步区分迁移与流动,即将5年前常住地不同于现居住地、且户口登记地在本乡镇街道的视为5年内的迁移人口;将那些户口登记状况为第二、第三款的人视为流动人口。这也是目前流动人口研究中通常的做法。

然而,这种交互式识别标准仍可能存在问题。试考虑以下两种情况:(1)返迁人口。按5年前常住地在外地、普查时户口在本地的迁移人口(对应图3左侧户口在本地的人口)可能包含返迁人口。户口登记地在本地并无法表明其户口是否有过迁移。如某个体户口在本地,5年前作为流动人口居住在外地,在5年内的某个时点返回户口登记地(现居住地)生活,那么事实上他的户口登记地在整个5年期间并没有发生变化,而5年前常住地与现居住地却不同。显然这类人事实上是返迁的流动人口,而非迁移人口。因此,结合户口登记地得到的所谓迁移人口样本可能是迁移人口与返迁流动人口的混合体。(2)5年前常住地为其他地方,且(户口标准下)户口登记地为第二、第三款的人,未必是真正的“流量”流动人口。5年前常住地标准的测量对象是5年内新增的流量流动人口。如果5年前常住地为户口登记地,且不同于普查时点居住地,则5年前常住地样本能够较好地反映新增流量流动人口的情况。如果该个体5年前已是流动人口(如图2中的B地),在普查前5年内的某时刻又从B地(如省内其他地方)流动到普查时的登记地D地(对应图3中5年前流出并且5年内改变居住地部分),则在5年前常住地标准下,该个体应为新增的省际流动人口。事实上,其在5年前已经离开了户口登记地。虽然从各省或更小区域看,他们属于新增的流量流动人口,但从全国总体的角度看,他们则属于存量流动人口。该标准下的样本未必是真正的“流量”流动人口。

由此可见,5年前常住地样本存在以下问题:(1)可能遗漏严格“户口标准”下真正的流动人口;(2)可能同时包含迁移人口、返迁人口与流动人口等,具有混合样本属性;(3)可能错误地将存量流动人口识别为新增流动人口。因此,在使用和分析过程中应慎重地考察5年前常住地标准下的流动人口样本。

### (三)5年前常住地样本的实证描述

为展示上述讨论提及的可能性,本文使用五普和六普数据,对5年前常住地和户口标准下的流动人口类别进行简单的交互分析,以揭示5年前常住地样本具有的混合样本属性。其中,户口标准下的流动人口被分为本地人口、省内流动与省际流动三类(见表1)。

表1中的结果表明,五普和六普中5年前常住在外省的人中,按照户口标准为省际流动人口的比例分别为83.86%和76.05%;另有13.79%和22.70%被识别为本地人口,这些人既可能是返迁人口,也可能是迁移人口。因此,根据5年前常住地标准得到的省际流动人口样本是返迁人口、迁移人口和流动人口这三类人口的混合体,至少他们与户口标准下的省际流动人口之间存在一定的差异。

两种标准下同时为省际流动人口的比例从五普的83.86%下降到六普的76.05%,而户口标准下为本地人口的比例则正好相反,从五普的13.79%提高

到六普的22.70%(见表1)。这种此消彼长的原因,既可能来自流动人口的返迁,也可能由于近年来户籍制度的改革,户籍迁移人口有所增加。尽管五普数据可以粗略估计迁移人口与流动人口,但七普数据无法真正识别迁移人口。因此,对于这些“人在户在”的人口无法进一步区分是迁移人口还是返迁人口。另外,五普中5年前常住地为省内的人口仍有54.65%的人被户口标准识别为省内流动人口,但有43.66%的人被视为本地人口(省内返迁与省内迁移人口的混合)。六普中仅有10.18%的人依据户口标准被识别为省内流动人口,这一比例的大幅度下降尚需深入考察。

综合上述分析发现,5年前常住地样本尽管从某种意义上可以反映近5年人口迁移流动的情况,但该样本事实上是现有的流动人口、返迁人口和迁移人口等多种属性人口的混合体<sup>①</sup>,与户口标准下的迁移人口或流动人口均无法对应。如果利用这种样本进行分析,那么混合样本属性必然影响分析结果,且无法真正揭示流动人口与迁移人口的差异。为此,在实际使用中要注意这种混合样本属性,尽量结合户口标准进一步区分迁移与流动两个子样本,并分别予以分析与讨论(尽管迁移样本仍然是迁移人口与返迁人口的混合)。这将有助于理解流入地选择机制或类似研究中所存在的矛盾现象(周皓、刘文博,2022)。

表1 五普和六普中5年前常住地与户口标准分类

5年前常住地	2010年六普户口标准				2000年五普户口标准			
	本地人口	省内流动	省际流动	合计	本地人口	省内流动	省际流动	合计
省内(人)	3157550	367747	86604	3611901	38942	48749	1509	89200
列比例(%)	87.42	10.18	2.4	100.00	43.66	54.65	1.69	100.00
省外(人)	44979	2486	150724	198189	4396	748	26735	31879
列比例(%)	22.70	1.25	76.05	100.00	13.79	2.35	83.86	100.00
合计(人)	3202529	370233	237328	3810090	43338	49497	28244	121079
列比例(%)	84.05	9.72	6.23	100.00	35.79	40.88	23.33	100.00

注:根据五普、六普原始微观数据计算得到。

已公布的历次普查数据一般只包括现住地与5年前常住地的交互表,均未提供按照户口登记状况再细分的5年前常住地识别结果。上述这种对5年前常住地样本再结合户口标准区分出迁移与流动人口样本的方法,需要使用普查的微观数据。

## 六、结论与思考

### (一) 主要结论

本文依照最近一次、5年前、出生地3种统计口径的逻辑,在讨论人口迁移流动的统计口径与登记信息完备性的基础上,着重讨论户口标准的流动时间及其可能带来的问题,迁移人口的重要性及无法识别问题、5年前常住地样本的混合属性问题。主要结论为:(1)迁移流动人口的统计定义包括两个要点:明确的登记对象和完备的登记信息,即完备记录“某一次”人口迁移流动的“何时从何地到何地”的信息。(2)户口标准下,普查登记的流动时间是离开户口登记地时间;登记对象是一次“新”的不同于以往3种统计口径的迁移流动;未登记离开户口登记地后的流入地,导致登记信息不完备;由此带来诸如信息陈旧、低估城—城流动比例、无法估计小区域的新增流动人口等问题。建议增加离开户口登记地后的流入地,以及最近一次迁移流动时间与流出地。(3)目前普查数据无法识别迁移人口。户籍制度是中国最重要的社会制度之一,迁移人口是中国人口迁移流动的重要组成部分,忽略迁移人口将低估中国人口迁移流动的规模和活跃程度。中国情境下的研究应重视区分人口迁移与人口流动两个概念,并从普查题项设计、统计定义、数据分析等方面加强对中国迁移人口的调查和研究,以便全面理解中国的人口迁移流动。(4)5年前常住地标准下的样本是现有流动人口、返迁人口和迁移人口等多种属性人口组成的混合体。分析研究应注意混合样本属性并分别讨论。同时,5年前常住地样本与户口标准下的流动人口样本的重合部分所占比例逐步下降,说明迁移与返迁人口的比例可能在逐步上升,这从某种

意义上说明户籍制度的深入改革和流动人口返乡现象并存。

### (二) 两点思考

1. 充分利用普查数据信息,加强流出地视角的研究

七普的登记方式和登记地点发生了根本性变化。事实上,2005年的1%人口抽样调查以来,1%人口抽样调查和普查同时登记了户口登记地和普查时点居住地。尽管设置两个题项的目的是为了“逢人必登”,却为人口迁移流动研究提供了新视角。如果以户口登记地为标准(流出地),则普查时点居住地在某种意义上可被视为流入地,因此七普可同时从流出地和流入地两种视角考察人口迁移流动,而且从流出地视角的研究可以真正考察“流出地”的“推力”,特别是家庭户角度的研究,这是因为这种流出地的家庭户特征才是真正促使人口流动的家庭原因,而普查数据包含且可汇总相应的家庭户特征。研究视角的改变可能可以揭示以往研究所未发现的内容。

从全国人口总体看,上述两种视角下的流动人口(不论是规模还是结构)应基本一致(需要结合普查登记时间与空间标准)。从微观数据看,由于普查登记原则、微观数据的抽样等原因,上述两种视角下的结果会存在一定的差别。实际分析需要注意可能的结构性差异及其影响,并慎重讨论流出地视角下的流动人口。

### 2. 充分深入地开发出出生地信息

2000年五普以来,普查表均包含出生地信息,对应的选项为:(1)本县、市、区,(2)本省外县、市、区,(3)省外(仅在“省外”时要求填写外省省份)。其中选项1理论上对应非迁移或流动人口(仍有可能是返迁人口),选项2或选项3则表明其自出生到普查时点为止至少发生过一次省内或省外迁移流动。出生地信息从终身迁移角度反映个体的迁移活跃程度。段成荣(2000)利用1988年全国生育节育抽样

调查数据估计全部被调查者中有 85.10% 的人申报其居住在出生地所在的县。五普中第一次加入出生地信息,普查时仍然居住在出生地所在县(市、区)的人口比例分别为五普的 85.20% (张善余,2004;刘金塘等,2004)、六普的 81.22% (桂宾,2014) 和七普的 77.35%。这一比例的下降说明中国人口迁移的活跃程度正在逐步增强,但总体上中国人口仍然相对静止。

“出生地”题项所能提供的信息相对较少,使用或受关注程度一直较低。特别是该信息既忽略了中间迁移,也无法提供迁移的具体时间点,无法满足“何时从何处到何处”的信息完备原则;同时也无法估计迁移时间与年龄,无法考虑死亡等因素对迁移规模与比例的影响。这些不足限制了出生地信息的使用。在众多人口迁移流动研究文献中,仅有少数文献使用或分析出生地信息。尽管出生地信息存在上述问题,但从队列视角考察中国人口的终身迁移情况,包括迁出地与迁入地的分布、规模与强度等仍十分有益,也可以进行国际比较。因此,出生地信息亟待充分使用与深入开发。

总之,对任何一个研究问题的讨论与回答都应建立在正确恰当的研究对象的基础上。利用普查数据进行人口迁移流动研究时,应详尽理解普查题项设置、明确各种统计口径、认识各统计口径自身及结果样本可能存在的问题。人口普查数据是推动人口科学及相关学科发展的重要基础,只有充分利用、深入开发才能突其宝贵之处。尽管六普数据已在部分数据实验室公开使用,但约束条件较多,限制了普查数据的充分使用。本文认为,在强调数据保密、保证个体隐私安全的前提下,尽可能地开放数据,尽可能便利地使用数据,以使普查数据得到更大程度、更深层次的开发与利用,并促进包括人口迁移流动研究在内的人口学及相关学科的进一步繁荣与发展。

#### 注释:

①5年前常住地在外省的人口如果再进一步按照户口标准区分并予以比较,可以发现,户口在本地的常住人口、省内流动人口与省际流动人口这三类人口在年龄结构、性别结构、婚姻状况和受教育水平等方面存在着显著差异。

#### 参考文献:

- [1]段成荣(2000):《利用出生地资料进行人口迁移分析》,《人口学刊》,第3期。
- [2]段成荣、孙玉晶(2006):《我国流动人口统计口径的历史变动》,《人口研究》,第4期。
- [3]段成荣等(2019):《中国人口的迁移转变》,《人口研究》,第2期。
- [4]桂宾(2014):《中国人口终身迁移状况探索》,《改革与开放》,第6期。
- [5]刘金塘等(2004):《中国人口终身迁移状况分析》,《人口研究》,第3期。
- [6]魏津生(1984):《国内人口迁移和流动研究的几个基本问题》,《人口与经济》,第4期。
- [7]张善余(2004):《基于出生地的中国人口迁移态势分析》,《市场与人口分析》,第3期。
- [8]周皓(2021):《中国人口流动模式的稳定性及启示——基于第七次全国人口普查公报数据的思考》,《中国人口科学》,第3期。
- [9]周皓、刘文博(2022):《流动人口的流入地选择机制》,《人口研究》,第1期。
- [10]朱宇等(2016):《国内人口迁移流动的演变趋势:国际经验及其对中国的启示》,《人口研究》,第5期。
- [11]Qi W., Abel G. J., Liu S. (2021), Geographic Transformation of China's Internal Population Migration from 1995 to 2015: Insights from the Migration Centerline. *Applied Geography*. 135:102564.
- [12]Siegel J. S., Swanson D. (2004), *The Methods and Materials of Demography* (Second Edition), Academic Press.