

流动人口的健康选择性

王伶俐¹, 周皓²

(1. 清华大学社会学系, 北京 100084; 2. 北京大学社会学系, 北京 100871)

摘要:流动人口的健康选择性一直是学界争论的焦点, 有的研究认为, 健康选择性会随着人口流动的普遍化而消失, 这是值得商榷的。因此, 文章利用中国家庭追踪调查(Chinese Family Panel Studies, CFPS)2010和2012年两轮数据, 先从截面数据和跟踪样本两个角度, 结合 Logistic 回归方法, 讨论流动人口的健康选择性; 再针对跟踪样本中的新增流动人口, 结合适合稀少事件的 firthlogit 回归和倾向得分匹配方法, 以自评健康为实验变量(treatment), 在剔除自评健康的内生性后, 从事实因果推论的平均实验效应角度, 讨论并检验了流动人口健康选择性的存在。结论认为, 流动人口健康选择性依然存在; 之所以在一般的统计模型中不显著的原因可能来自于自评健康的内生性和各种特征对健康选择性的替代作用。

关键词:流动人口; 健康; 自评健康; 选择性; 内生性

中图分类号: C92-05 **文献标志码:** A **文章编号:** 1007-0672(2018)06-0013-10 **收稿日期:** 2018-01-08

基金项目: 北京市流动人口的社会融合研究(编号: 13SHA001)。

作者简介: 王伶俐, 女, 安徽安庆人, 清华大学社会学系博士后, 研究方向: 流动人口、社会统计学; 周皓, 男, 浙江宁波人, 北京大学教授, 博士生导师, 北京大学中国社会与发展研究中心研究员, 研究方向: 人口迁移与流动。

一、引言

自雷文斯坦^[1]发表《迁移规则》以来, 流动人口的选择性一直是其研究的重要内容, 重点在于回答“谁在迁移”这一问题。尤其是正处在特殊社会转型时期的中国, 人口流动既是一个重要的社会现象, 更是社会转型的重要动力来源, 对这一问题的回答有着重要的社会意义。众多研究已说明我国人口迁移与流动在年龄、教育、婚姻状况等特征方面存在着显著的选择性^[2-13]。近年来, 有关选择性的讨论扩展到了健康问题^[14-18]。所谓健康选择性是指越健康的人更有可能流动。在中国的社会情境下, 移民的健康选择性是否存在及这种选择性是否会随着时间、特别是人口迁移与流动的普遍而消亡等问题仍在讨论中, 并未获得一致的结论。如同钰莹等^[14]认为随着社会发展, 流动人口的健康选择性逐步随着时间而消亡; 而有些研究则持相反意见, 认为流动人口仍存在健康选择性。本文关心的第一个问题是利用新的调查数据检验流动人口的健康选择性是否存在。个体健康会受到来自社会经济地位、行职业及个人特征等多方

面的影响。因此, 流动人口的健康选择性可能与其他选择性相重合, 或被其他各类选择性所替代, 从而使分析结果表现为不存在健康选择性, 但这并不能说明健康选择性无作用或者消失。因此, 如果想要真正地讨论健康与流动之间的关系, 线性回归或 Logistic 回归方法可能会由于内生性问题而得到有偏的结果, 需要利用其他方法来讨论或解决。因此本文想要讨论的第二个问题是内生性是否会导致健康选择性的消失。

本研究将利用“中国家庭追踪调查”两轮跟踪数据, 通过对比流动(特别是新增流动人口)与非流动人口, 真正检验了健康选择性在人口流动中的作用。首先, 从因果的时间顺序看, 截面数据中原因与结果通常都在一个时间点上, 从而无法真正判断两者在因果关系中的时间顺序; 而本文利用跟踪调查数据, 以前一期的健康作为原因, 探讨其对是否流动的影响, 才有可能真正讨论流动人口的健康选择性; 其次, 健康一般具有较强的时点性(即迁移或流动时刻的健康状况对其流动的影响), 本研究利用两轮调查之间新增的流动人口,

而不是所有流动人口,其优点在于:既避免了由于在流入地居住时间不同而产生的子群体差异(及由此带来的其他选择性),也避免居留时长带来的健康损耗问题^[19];再次,本文在讨论过程中利用各种方法控制了健康的内生性问题,进而相对有效地保证了估计的无偏性,有助于更好地理解流动人口的健康选择性。因此,本研究在因果关系、样本选择等方面都具新颖之处,分析结果可以检验以往某些理论与实证研究结论,因而具有重要的理论与现实意义。

二、文献综述与研究假设

(一)研究综述

国际移民研究中发现了移民的健康悖论现象(Hispanic Health Paradox 或 Epidemiologic Paradox)^[20],即移民虽然社会经济地位较低,但其健康状况要优于流入地本地居民。这种悖论现象不仅存在于墨西哥-美国移民群体以及其他国家中^[21-25],而且在人群死亡率、婴儿死亡率、自评健康、ADL等不同的健康指标^{[20][24][26-30]}上亦有所体现。

对这种悖论的理论解释主要有两种视角:一是移民的健康选择性,即健康移民假说(Healthy Migrant Hypothesis),即移民是经过正向健康选择的一批人,能够克服移民过程中的艰苦环境、适应流入地的工作生活环境以及高强度高体力的劳动;健康选择机制使得只有在流出地更健康的人才有可能进入迁移人口的行列。另一种是三文鱼偏误假设(Salmon Bias Hypothesis),即在人口迁移过程中,那些健康状况明显恶化的人往往无法长期滞留在迁入地,出于生活成本、社会保障需求等方面的考虑,这些人更可能返回迁出地^{[15][17]}。

国内有关流动人口健康问题的研究都试图检验健康移民假设和三文鱼偏误假设两种视角对中国流动人口健康选择的解释,总体上为中国流动人口的健康选择机制研究提供了基础^{[14-15][17][31]}。研究结果基本都表明:中国人口流动存在较明显的“健康移民”和“三文鱼偏误”选择效应^{[15][17][31]},且城乡流动对农村常住人口的总体健康状况具有重要

影响^[16]。

从流动人口的社会背景看,中国流动人口主要是从农村流向城市,以务工性流动为主,在城市中主要从事技术含量较低的体力劳动,面临着收入较低、工作风险较高、健康保障较差的现状^[32-33]。高强度的体力劳动及较差的生活环境^{[32][34]},对劳动者的身体素质有较高的要求,身体状况较差的流动人口将无法适应这类劳动,从而无法在城市中获得稳定的劳动收入。已有研究也表明:自评更健康的人,在流动后其自评健康指标仍优于城镇居民^[14-18]。这些背景与研究结果都表明健康选择性的存在。同时,亦有研究指出,随着流动时间的延长,流动人口的健康存在损耗^[35];健康耗损严重的农村外出务工人员回到农村的概率更高,造成了当前在城市务工的农村人员与城市本地工作人员相比看上去并未出现更为严重的健康耗损的假象^[19]。这从另一方面说明了“三文鱼偏误”效应。

但也有研究指出,中国流动人口健康选择机制的效应会随时期的变化而逐步消失;这种变化可能与中国经济改革背景下剧烈的社会经济、政策变动有关,因此流动人口健康选择机制的研究需要加入社会经济背景变迁等宏观维度因素的考察^[14]。而且,流动人口与流入地居民之间在长期健康状况上差异的显著性远远高于在近期健康状况上差异的显著性;不同健康指标的选取对实验结果的影响大于控制变量对结果的影响^{[18][3]}。

(二)研究评述

综上所述,可见尽管国内学术界已对健康与人口流动的关系有了较多的研究,当前在讨论流动人口的健康选择性时仍存在以下几点问题:

第一,数据问题。尽管已有很多流动人口(或农村劳动力转移)的调查数据,但各种来源的抽样调查数据,在抽样过程(如某些调查仅针对部分省份展开)与样本代表性(各种结构性偏差)等都需要予以检验与讨论,而不能“拿来主义”。且有关流动人口的调查不论是在流入地还是在流出地进行,由于涉及到研究组与参照组的选择而未必能

① 易龙飞等^[18]与同钰莹等^[14]的文章虽然使用的都是CHNS数据,但差异主要在以下三点:一、调查时点不同,前者使用了2006、2009和2011三轮数据;而后者则使用了1997~2009的五期数据;二、研究对象的年龄范围不同,前者是18~64岁,后者规定在16~35岁;三、健康的定义不同,前文使用各种客观指标,后者则只使用自评健康。因此从某种意义上,两篇文章说明的是各自的问题,而结果无法比较。但本文同意易文中提出的健康选择性依然存在的结论。

正确地探讨流动与健康间的关系。

第二,时间顺序问题。尽管已有研究注意到了健康与流动在因果关系上的时间顺序问题而采用滞后一期的健康测量结果^{[4][31]},但仍有部分研究是利用截面调查数据进行的分析。截面数据的问题就在于:原因与结果通常都在一个时间点上,从而无法在识别与判断因果关系时保证原因与结果在时间上的先后顺序。既无法判断原因的结果,也无法考察结果的真正原因。

第三,研究组与参照组的选择。这个问题既与调查数据有关,且也与因果的时间顺序相联系,从而影响到正确判断流动与健康间的关系。首先需要说明,从研究对象看,在讨论流动人口健康选择性时,应该以 $t-1$ 时刻的非流动人口为基础,将 $t-1$ 时刻至 t 时刻间新产生的流动人口作为研究组,相应的 t 时刻仍然是非流动的人口作为参照组,这才有可能检验健康对流动的选择性;而在讨论三文鱼效应时,研究对象应是 $t-1$ 时刻的流动人口,其中参照组是 $t-1$ 时刻和 t 时刻均为流动的人口,研究组则是在 $t-1$ 至 t 时刻之间返迁的原流动人口。即分析时应保证研究组与参照组都来自于同一群研究对象,这对样本就提出了很高的要求。

相应的,从调查数据来看,调查地点是获得合适的研究对象的重要基础。流入地调查的样本只包括了当前的流动人口和流入地的本地人口,但既没包括流出地的非流动人口,也没包括那些返迁的流动人口,因此,从某种意义上它无法说明流动人口的健康选择性。同样的,流出地调查尽管包括了返迁人口,却无法捕获已流出的人口(现有的流动人口),亦无法说明三文鱼现象。如果想要讨论流动与健康的关系,就必须结合流出地与流入地的调查,以保证研究组与参照组来自于同一群体。部分研究将所有的流动人口作为研究对象,将其与城市或农村常住人口相比较。这种比较显然不满足于上述对研究组与参照组的要求。

其四,健康选择性的时点性。任何人的健康状态都会随时间而发生变化,因此,要想真正讨论健康选择性,最理想的应该是测量流动人口在流动时刻(或返迁人口在返迁时刻)的健康状况。显然这并不现实。因此,在研究中会使用前一期的

测量结果作为流动/返迁时的健康状态。这只是一个近似结果,且仍然会要求前后两期测量的时间间隔不能太长。

同时,流动人口的健康随着在流入地居留时长而有所损耗^{[19][35]},因此,如果选择当前所有的流动人口作为研究对象时,一方面返迁人口存在健康的“三文鱼效应”,从而使流动人口的健康水平被高估(亦低估了非流动人口健康的平均水平);另一方面,计算所得的健康是以居留时长为权重的均值,而并非真正的流动决策时的健康状况。也就无法有效准确地估计健康选择性。

上述四个问题主要是从样本与健康作用的时点性来看的,但目前已有的研究在方法(如内生性与因果推论间的关系)和理论(研究结论的理论概括)等方面仍然需要继续深入。

(三)研究假设

综上所述,本文将主要讨论以下两个方面:首先,作为流动人口选择性的重要维度之一,健康选择性会长期存在并影响到人口的流动。尽管在控制了个体特征以后,健康对于流动的影响作用似乎并不存在,但有可能是因为健康选择性被其他因素的选择性所掩盖,而并不是说健康选择性不存在。其次,由于自评健康本身可能存在选择性,从而使分析过程中自评健康的系数可能产生选择性偏差(selection bias)。只有解决以上两个问题,才有可能对流动人口的健康选择性进行系统的检验。因此本文的研究假设是:流动人口存在健康选择性,即越健康的人越有可能流动。

三、数据与方法

本文将使用有全国代表性的中国家庭追踪调查(China Family Panel Studies, CFPS)^①2010年和2012年两轮数据。

(一)变量的操作化定义

因变量:本文的因变量是个体的流动状态。我国的相关研究中,流动人口一般被定义为:在现居住地居住半年以上、且居住地不同于户口所在地的人。CFPS2010数据中只询问了户口所在地,但并没有询问在本地的居住时间长度。因此,2010年的“流动人口”定义为:现居住地与户籍登记地不同的人,而未考虑时间因素。在2012年的

① 有关该调查的详细情况请参见:<http://www.issn.edu.cn/cfps/>。

表1 关键变量的描述性统计量

	2010年	2012年	2010~2012新增流动人口
是否流动(0=未流动)	0.08(0.27)	0.05(0.22)	0.01(0.1)
自评健康(0=不健康)	0.47(0.5)	0.64(0.48)	0.47(0.5)
年龄	45.51(16.41)	43.83(16.9)	45.67(15.8)
样本规模	33,600	36,345	25,147

注:表中数字分别表示均值与标准差(括号内数字)。

跟踪调查中,增加了流出时间的问题。因此,2012年的“流动人口”定义在2010年的基础上加入了流出时间限制。在两轮调查数据处理中,分别对当年的“流动人口”标记为1;“非流动人口”标记为0;“新增流动”定义为10年调查时未流动而12年调查时为流动的人口(具体情况请见表1)。

自变量:

自评健康作为一个综合性的评价指标,能代表较为全面的个人健康状况,因此本文将使用自评健康这一指标作为研究自变量^①。该变量对应问卷中的问题“您觉得自己的健康状况如何”。在10年的数据中,将一般、比较不健康、不健康和非常不健康归为“不健康”(参照组),将“健康”归为一类;由此构成一个二分变量。在12年的回答中,将非常健康、很健康和比较健康归类合并为“健康”,将差和一般归为“不健康”(参照组)。

控制变量的选择包括个体和家庭两个层面,设置基本与同钰莹文相同。个体层次的变量包括:人口学特征(年龄和性别)、受教育水平、职业类型、居住地点等;家庭层次包括:父母年龄、配偶是否同住、是否有12岁以下孩子需要照顾、家庭规模、家庭年收入对数以及家庭人均收入等。同时为了控制地区间的差异对流动的影响,基于个人所在地省份设置了地区变量。

关键变量^②的描述性统计量请见表1。

(二)研究思路

本文将利用全国代表性追踪数据,讨论流动人口的健康选择性问题,并期与以往研究对话。因此,本文将首先从截面数据的角度,分别利用CFPS2010和2012两轮调查,考察调查时点的个体

健康对其当时的流动状态的影响。

其次,为了说明健康对流动在时间上的因果顺序,建立滞后效应(lag effect)模型,从2010年非流动人口的跟踪样本中,挑选出在2010年至2012年间新增加的流动人口(标记为1),以2010年和2012年均为非流动人口的跟踪样本为参照组;并以2010年时的变量为自变量,讨论2010年的健康状况对是否成为新增流动人口的影响作用。

上述两个步骤过程中,为了能够与以往结果对比,本文将首先设定与同文基本一致的模型;然后在此基础上加入年龄范围和年龄与健康的交互项建立模型,并与原模型进行比较,以说明两种模型间在结果上的差异。再次,将使用不同模型来检验流动人口健康选择性是否被内生性问题影响或其他选择性掩盖而产生偏差。最后,将自评健康视为实验变量,在控制自评健康的内生性问题的基础上,利用倾向得分匹配方法讨论自评健康对流动的影响作用。

(三)分析方法

本文所用的方法主要包括 Logistic 回归、对应于稀少事件(rare event)的firthlogit回归、以及倾向得分匹配方法。

所谓稀少事件,是指在一定的样本规模下事件发生的案例数极少、比例极低的现象。Logistic回归一般要求事件发生(因变量=1)的比例至少高于5%。一旦这种比例小于5%,Logistic回归的合理运用,不仅与事件发生比例有关,且也与样本总规模有关:如果样本总规模过少(如只有几十人),就不能用Logistic回归;如果样本总规模较大,则需要看发生事件数(绝对规模)。如果1000个样本中

① 本文在建立模型的过程中曾利用其他指标,包括身体健康(BMI、两周病伤和慢性病等)和心理健康(抑郁指数)等。但结果与自评健康基本相同。因此本文最终仅选择自评健康作为研究变量。

② 由于篇幅所限,这里仅展示关键变量的描述性统计结果,完整结果可联系作者。

表2 历次普查中流动人口的年龄结构变动

	1990年 ¹	2000年 ²	2010年 ³	CFPS2010 ⁵	CFPS2012	新增流动
平均年龄(岁)	27.99	27.05	31.40	40.78	36.67	33.63
中位数(岁)	24.80	25.65	29.38	38	33	28.5
16~35岁比例(%)	69.0(75.58) ⁴	68.71(78.07)	52.69(59.61)	44.03	56.64	34.92
36岁及以上比例(%)	22.3(24.42)	19.30(21.93)	35.70(40.39)	55.97	43.36	65.08

注:1.1990年数据来源:《中国第四次人口普查资料分析(下)》,P90;表6.5.4。该表第根据1990年人口普查1%抽样数据估计。该表以五岁分组,最高年龄为70岁及以上。平均年龄根据组中值计算;中位数基于累计百分比计算。“16~35岁组的比例”包括了15岁组(以五岁分组的15~19岁)。2.2000年数据根据五普0.95‰原始数据中的流动人口样本计算得到。3.2010年数据根据第六次人口普查“7-2全国按户口登记地、年龄、性别分的户口登记地在乡镇街道的人口”计算得到。其中,计算时剔除了“市辖区内人户分离”人口。4.括号中的数值系指:以16岁及以上流动人口作为总体,16~35岁和35岁以上流动人口各自所占比例。其余单元格内相同。5.CFPS数据根据原始数据计算得到。

只有20个发生案例,那么Logistic回归就会有问题;如果是10000个样本中有200个发生案例(事件发生比例仍然只有2%),则Logistic回归是可以接受的;如果100000个案例中有2000个发生案例(比例仍然没变),那么Logistic回归的结果就没有任何问题^[36]。回应本文的样本情况,在2010年的25025个非流动人口中,只有246个新增流动人口,占比仅为0.98%。这个结果远小于上述经验案例情况。因此,利用Logistic回归和最大似然估计方法可能会产生一定的偏差,需要利用惩罚性似然估计方法(penalized likelihood)及其对应的firthlogit方法来估计(该方法可详细参阅King & Zeng^[36])。

倾向得分匹配方法已成为基于反事实因果推论、解决内生性问题后有效评估实验效应的重要方法之一(可详细参阅相关文献^[37-41])。

四、分析结果

(一)流动人口的年龄结构及其变动

同钰莹等^[14]在讨论健康选择性时,分析对象是16~35岁的人口。本文认为,尽管16~35岁组的流动人口占整个流动人口的比重确实较大,但事实上近年来流动人口的内部结构正逐步发生变化。流动人口不仅“流而不动”^[42],在城市中长期居留不再回到农村^[43];更重要的是内部年龄结构发生变化,流动人口的年龄中位数从1982年的23岁上升到2010年的29岁,且2000至2010年间,35岁及以上流动人口的增长率远高于较低年龄组^[43]。表2展现了第四、第五和第六次人口普查中流动人口

年龄结构及其变动状况。由表中的数据可看到,我国流动人口的平均年龄与年龄中位数都在不断提高;同时,35岁以上的流动人口比例也从1990年的20%左右,提高到2010年的35.7%(占16岁及以上人口的40.39%)。这一点也同样反映在CFPS的数据中。因此,从年龄结构来看,将年龄结构限定在16~35岁,绝不是“相对很少的流动人口”,而是可能遗漏了一大批的流动人口^①。而且,这种年龄结构的变化本身就体现了社会变迁的作用。因此,本文将分析扩展到全年龄段,但不同年龄段流动人口的迁移与流动的原因可能完全不同。

(二)截面数据与跟踪数据的分析结果

本部分使用2010年和2012年截面数据,以及跟踪样本数据,讨论流动人口的健康选择性问题。有关结果请见表3和表4。

表3中的模型(1)和模型(3)的设置(年龄与变量)与同钰莹等^[14]的文章完全相同;模型(2)和模型(4)将年龄扩展到全部年龄,并在变量设置中加入了年龄的平方项、年龄与自评健康的交互项^②。表4则是对跟踪样本的分析,其中的模型(5)的设置与同文相同,在后续模型中逐步扩展至全部年龄段,并加入年龄与自评健康的交互项。由于两期合并后的追踪样本中流动人口所占比例过低(共246人,占0.98%),无法使用常规的Logistic回归,因此代之以firthlogit回归方法。

2010年和2012年截面数据的分析结果都表明,在控制了各种其他因素以后,流动人口的自评

① 本文同意同钰莹等人关于提高年龄上限可能出现由于人口返迁及三文鱼现象的偏误问题的判断。

② 注意:2012年截面模型中,由于职业分类不同于2010年及中国健康与营养调查(CHNS),因此,在2012年的模型中未加入职业变量。

表3 截面数据分析结果

	2010年		2012年	
	(1)16~35岁	(2)所有年龄	(3)16~35岁	(4)所有年龄
自评健康(0=不健康)	0.0450(0.1086)	0.1501(0.1983)	-0.2600(0.2051)	-0.6716*(0.3485)
年龄	-0.0493*(0.0200)	-0.1000*** (0.0206)	-0.0173(0.0289)	-0.0684*(0.0370)
年龄*自评为健康		-0.0040(0.0047)		0.0177*(0.0069)
N	7917	25371	7234	16588

注:1.* p < 0.10, * p < 0.05, ** p < 0.01, *** p < 0.001;2.括号内数字为标准误;3.控制变量略,由于篇幅所限,这里仅提供研究变量的回归系数,下同。

表4 2010~2012年新增流动人口的firthlogit分析结果

	(5)	(6)	(7)
	16~35岁	所有年龄	所有年龄+交互项
自评健康(0=不健康)	0.2335(0.1973)	0.0946(0.1568)	0.9214*(0.4256)
年龄	-0.0541*(0.0318)	-0.0284*(0.0132)	-0.0183(0.0138)
年龄*自评为健康			-0.0261*(0.0124)
chi2	90.9397	255.1814	259.0153
ll	-571.5205	-921.9774	-915.1285
N	5620	19178	19178

注:1.* p < 0.10, * p < 0.05, ** p < 0.01, *** p < 0.001;2.括号内数字为标准误;3.控制变量略。

健康均不显著。但这一结果尚未真正说明流动人口的健康选择性的不存在,因为截面数据的原因变量与结果变量是在同一个时点上的,违背了因果关系的时间顺序。

为了改进这一问题,表4中的模型利用2010年的自变量,解释2010年至2012年间新增流动人口的决策机制。在表4的模型(5)和模型(6)中自评健康仍然是不显著的,说明作为原因的2010年健康状况不会影响2010至2012年间的流动决策,即流动人口的健康选择性不存在,这一结果与同钰莹等相同。但是模型(7)显示,在加入年龄与自评健康的交互项以后,不仅自评健康这一变量变成显著的,而且年龄与自评健康的交互项也是显著的。在模型(7)中,自评健康是正向的,即自评健康好的人,他们流动的可能性是不健康的人群的2.5128倍,相对高出了1.5倍;而年龄与自评健康的交互项则呈现出负向显著,即:在相同的健康状况下,年龄越大,越不会流动。这一结论可以说

明,健康与年龄可能共同决定了个体的流动决策。因此健康选择性可能仍然是存在的。

此外,模型(6)和模型(7)中的其他变量也呈现出一定的差异。例如加入交互项后,年龄变量的负向作用仍然存在但是显著性消失,这也进一步说明年龄与自评健康的共同作用机制。性别变量一直呈现负向显著,即女性比男性更易于流动;婚姻状况在16~35岁人群中具有的显著性,在全年龄段中则呈现出不显著;教育的选择性呈现出稳定的作用,即受教育水平越高,越容易流动;个人职业、家庭结构及经济收入等均不显著。这些结果,有些与以往的研究结论相符(如受教育水平等),有些却是相悖的(如性别、婚姻状况等)^①。

这一结果引发的思考是^②:流动人口的健康选择性是否可能被其他因素所替代,还是由于自评健康可能存在内生性问题影响结果呢?因此接下来将讨论健康选择性的替代问题。

(三)健康选择性的替代问题

① 为节约篇幅,本文主要关注自评健康与年龄这两变量。对模型中的其他变量不再予以特别的解释。

② 同时可能影响结果的另一个问题来自于样本。由于两次调查之间的样本规模较小,使其对两年间的新增流动人口的代表性较低。由于CFPS是全国代表性数据,因此,本文在此不再讨论有关样本问题。

表5 2010~2012年新增流动人口影响因素的分步firthlogit回归结果(全部流动人口)

	(8)	(9)	(10)	(11)	(12)	(13)
	健康	健康+年龄	健康*年龄	健康+教育职业	(11)+家庭+地区	(12)+健康*地区
自评健康(0=不健康)	0.5339*** (0.1304)	-0.0101 (0.1362)	0.8484* (0.3729)	0.7174* (0.3784)	0.9214* (0.4256)	1.4994** (0.5076)
年龄		-0.0656*** (0.0052)	-0.0527*** (0.0071)	-0.0372*** (0.0080)	-0.0183 (0.0138)	-0.0187 (0.0137)
年龄*自评为健康			-0.0264* (0.0106)	-0.0226* (0.0108)	-0.0261* (0.0124)	-0.0254* (0.0124)
自评健康 *地带						-0.9306* (0.4011)
自评健康 *地带						-0.6729* (0.4080)
chi2	16.7754	173.0543	185.9742	282.2422	259.0143	263.0189
ll	-1368.4012	-1266.1754	-1258.4571	-1200.9180	-915.1285	-910.0378
N	25016	25016	25016	24967	19178	19178

注:1.† p < 0.10, * p < 0.05, ** p < 0.01, *** p < 0.001;2.括号内数字为标准误;3.控制变量略。

为了讨论流动人口健康选择性的替代问题,我们通过逐步回归的方法重构表4中的模型(7)。具体结果请见表5。仅使用自评健康进行单变量回归时,结果呈现出高度的正向显著,即表示自评越健康的人,其成为流动人口的概率是不健康者的1.7倍。如果不考虑其他因素,则健康选择性是成立的。在此基础上,加入年龄变量后则自评健康不再显著,而年龄却呈现出负向的作用,即年龄每增加一岁,成为流动人口的概率会下降7%左右($1-\exp(-0.0656)$)^①。这两个模型之间的差异,说明健康对流动的作用可能会被年龄所解释。如果再考察健康与年龄之间的关系,则可以发现,年龄对于自评健康有着显著的负向关系,年龄越大,自评健康状况越差。这一方面说明自评健康可能会有内生性问题,另一方面则说明年龄可能存在着替代健康的选择性问题。

加入年龄与自评健康的交互项以后,自评健康的作用不仅显著性发生变化,而且方向也发生了变化,由模型(9)中的负向不显著,变成了模型(10)中的正向显著,进一步说明了健康选择性的存在。该交互项在模型(10)呈现出负向显著的作

用,即在自评健康较好的人群中,年龄越大越不容易流动。但如果要考察在相同年龄之下自评健康的作用时,就需要将该交互项的系数与自评健康这一变量的系数相加(即 $0.8484-0.0264=0.8220$),则相对于而言,自评健康越好的人成为流动人口的可能性是自评不健康的人的2.2750倍,即高出将近1.3倍。

此后的几个模型中,自评健康以及自评健康与年龄的交互项这两个变量的作用几乎没有发生任何实质性的变化,都呈现出显著的作用,且方向一致。但在模型(12)和(13)中加入家庭层次的变量以后,年龄的作用却变得不显著了。这种情况有可能是因为年龄存在着与教育类似的阶段性的影响作用^[44]。

之所以在模型(11)的基础上,继续加入了自评健康与地带的交互项,希望考察健康选择性的地区差异。结果表明存在地区异质性作用,在健康的人群中,东部地带的人更不容易流动(=-0.9306);而中西部地带则未呈现出显著的选择性。

(四)自评健康对流动的平均实验效应的估计
以上分析结果对本文提出的研究假设提供了

① 要注意年龄的阶段性作用,即不同年龄段,其作用可能是不同的。这里既有代际(如老一代流动人口与新生代流动人口)的关系问题,也有不同年龄段对应不同的流动原因与结果的可能。因此,年龄的作用尚需要进一步的讨论。

表6 利用倾向得分匹配方法的估计结果(全部流动人口)

Sample	设置方法	Treated	Controls	Difference	S.E.	t	参数设置下的回归分析结果	
未匹配		0.012845	0.006865	0.005979	0.001304	4.58	treated	0.0060***
ATT	1. 缺省设置	0.012845	0.008626	0.004219*	0.001968	2.14		(0.0013)
ATT	2.1:100	0.012845	0.009452	0.003393*	0.001501	2.26	常数项	0.0069***
ATT	3.1:1000	0.012845	0.009217	0.003628*	0.001473	2.46		(0.0009)
ATT	3.radius	0.012846	0.006865	0.005980***	0.001327	4.51	N	22610

注:1.† p < 0.10, * p < 0.05, ** p < 0.01, *** p < 0.001;2.括号内数字为标准误。

支持,即在社会的变迁和流动人口群体内部结构变化的背景下,流动人口的健康选择性依然存在。但自评健康变量内生性导致的估计偏误还没有被讨论。因此,本文将自评健康为实验变量,利用倾向得分方法,解决内生性偏误后再考察流动与自评健康间的关系问题。分析结果请见表6。该表包括了两个部分,左侧部分是各种参数设置之下的平均实验效应的分析结果;右侧则是在各种参数设置之下 Stata 按回归分析方法输出的结果(由于在各种设置之下的结果基本相同,因此,此处仅给出一个结果作为代表)。其中,用于构建自评健康倾向得分的协变量包括:年龄、性别、是否已婚、教育、行业、居住地、个体收入、家庭收入、地带、BMI 指数、肥胖状况以及精神健康状况。

首先看右侧结果。该结果表明,不论何种匹配方法,在控制了自评健康的内生性问题以后,平均实验效应仍然是显著的正向作用,即越健康的人越容易流动。而且这一结果在各种参数设置条件下都是相同的。

其次,再来看左侧不同参数设置条件下的分析结果。这些结果都表明,自评健康对于流动的实验效应都是显著的正向:在缺省设置的条件下,估计的平均实验效应为0.004219;而在其后的两种设置下,其实效效应则分别为0.003393和0.003628,其中的差异主要来自于匹配案例数的不同。在半径匹配中,该实验效应则为0.005980,且显著性增加。

如果将这些系数与前两部分的分析结果中的系数对比(尽管 Logit 回归分析中各模型间的系数可能无法对比^[45-46]),可以看到,平均实验效应明显下降,从原来的0.9214下降为现在的0.006。这种较大幅度的下降,其根源还是在于自评健康的选择性。即如果不考虑 logistic 回归中的特殊性问

题,那么在上述回归分析中关于自评健康对流动的影响作用可能会由于自评健康的内生性问题而导致偏差。同时,这一结果也说明流动人口健康选择性的存在。

五、结论与讨论

本文利用中国家庭动态调查数据,利用各种方法讨论并回答流动人口健康选择性是否存在这一问题。分析结果表明:(1)流动人口内部年龄结构正在发生变化,这既是社会变迁的结果,也是流动人口健康选择性时需要考虑的问题;(2)在排除了变量的时间因果顺序后,跟踪样本中的新增流动人口仍存在健康选择性;(3)健康选择性可能与年龄选择性同时存在,从而可能会使年龄选择性(或其他可能的选择性)掩盖了健康选择性的真实存在;(4)即使剔除了自评健康的内生性(或选择性)问题以后,自评健康对流动决策的实验效应仍然是正向的,这说明健康选择性仍存在于流动人口的流动决策之中。研究结果说明流动人口的健康选择性并未随着社会环境与背景的变化而消失,相反,在考虑了社会变迁(如调查时点的变化和流动人口年龄结构的变化)以及自评健康的内生性问题后,健康选择性依然存在;越健康的个体越有可能进行流动。这在某种程度上说明,健康,类似于某些人口学(如年龄)或社会特征(如婚姻、教育等),总是存在选择性的。只是这种选择性可能会与其他选择性特征或是重合或是替代,从而在实证分析结果中呈现出不显著的作用。但这种统计上的不显著,并不能够真正地说明健康选择性的消失。有可能的是,不论在何种社会背景之下,其作用呈现出的强弱之分,或是显性或隐性之分。因此,重新回到迁移的选择性理论中,则健康同样是迁移与流动的选择性的重要维度之一。

当然上述结果仍是阶段性的,其中样本、分析

过程与分析方法等都还有进一步商榷的地方。

首先,本文中所考察的样本是针对所有年龄段的人口,结果表明流动人口的健康选择性是存在的。但如果仅对16~35岁的人群,健康选择性并不存在。这一点与同钰莹等的结果是相同的。这两个完全不同的结果告诉我们应该用变迁的眼光看待这种健康的选择性问题。一方面,社会变迁提高了所有年龄段人群的流动强度(即流动人口的年龄结构发生的变化),特别是大于35岁年龄组的人群的流动率,从而使原有的年龄分组有所局限;并使健康的选择性可以扩展并适用于所有年龄段。另一方面则是社会变迁使得16~35岁的人群普遍健康与普遍流动,年龄的选择性从某种意义上掩盖了健康的选择性,从而使该年龄段的健康选择性似乎不存在了;即健康选择性被年龄选择性所掩盖,但并不能说健康选择性消失了。

其次,从样本的角度看,本研究尽管利用了跟踪样本,且从跟踪样本的非流动人口中寻找出两次调查期间的新增流动人口。但分析过程并未判断这些新增流动人口的流出地,而是以2010年的全部非流动人口作为参照组。这时的参照组可能是流出地人群和流入地人群这两个部分的混合,而并非真正的“潜在流动人口”。当然这是从严格意义上说的。在实际操作中,流出地与流动地总是相对的(有部分人流入本地,而有部分人从本地流出),从总体上很难判断哪个是流出地,哪个是流入地。因此,从某种意义上,这种健康的选择性又可以被看成是健康悖论的一种,即流动人口在更恶劣的环境(包括居住环境、从业环境以及家庭环境等)之下,其健康状况仍然优于本地人口。同时,限于样本规模的问题,我们无法真正做到对流出地与流入地的判断,进而更严格地区分流出地的“潜在流动人口”和流入地的“本地人口”。这将是今后研究中需要十分强调的问题之一。

再次,本文未讨论健康对流动距离的影响。事实上,流动距离的长短必然会与其健康状况有关,但本文仅考察了全部跨县域及以上的流动人口,而未包括县内流动人口;同样也因篇幅与样本量问题而未给出省际与省内流动之间的差异。这同样也是将来有关健康与流动关系研究中的重要问题之一。本文显然没有考虑这些新增流动人口的以往流动历史。

最后是研究方法问题。本文运用了firthlogit模型,主要是考虑到样本(特别是新增流动人口的)规模,且这种方法可以作为今后类似问题的处理方法。但从研究自变量的角度来看,本文选择自评健康作为健康的代表性测量指标,一方面其测量本身可能存在差异性与内生性问题(如测量锚点与评价标准等),进而使目前的研究结果仍然存在偏误(测量的内生性问题可能无法剔除);另一方面,以自评健康作为实验(treatment),来考察自评健康对流动的实验效应,是否可行仍然可能值得商榷。除了测量问题以外,诸如流动本身的比率较低、难以捕捉真正的实验效应、自评健康的多分类等都是需要进一步考虑的。

因此,本文的结论仍是初步的,还需要今后更深入地讨论。✱

参考文献:

- [1] Ravenstein E. G. The Laws of Migration[J]. Journal of the Statistical Society of London, 1885, 48(2): 167-235.
- [2] 张善余. 关于我国人口迁移的几个问题[J]. 人口研究, 1992(2): 16-22.
- [3] 杨云彦. 中国人口迁移年龄模式及其应用[J]. 人口学刊, 1992(4): 7-11.
- [4] 杨云彦. 人口迁移年龄模式的分解与合成[J]. 人口研究, 1992(4): 15-22.
- [5] 李树茁. 80年代中国人口迁移的性别差异研究[J]. 人口学刊, 1993(5): 16-21.
- [6] 张善余, 张茂林. 差别人口迁移与性别构成地区差异的扩大化[J]. 人口学刊, 1996(1): 3-10.
- [7] Zhao Y. Leaving the Countryside: Rural-To-Urban Migration Decisions in China [J]. The American Economic Review, 1999, 89(2): 281-286.
- [8] 段成荣, 韩荣炜, 刘岚. 人口迁移选择性及其度量[J]. 南京人口管理干部学院学报, 2001(2): 22-24.
- [9] 李玲. 改革开放以来中国国内人口迁移及其研究[J]. 地理研究, 2001(4): 453-462.
- [10] Liang Z., Ma Z. D. China's Floating Population: New Evidence From the 2000 Census[J]. Population and Development Review, 2004, 30(3): 467-488.
- [11] 严善平. 地区间人口流动的年龄模型及选择性[J]. 中国人口科学, 2004(3): 32-41+81-82.
- [12] 唐家龙, 马忠东. 中国人口迁移的选择性: 基于五普数据的分析[J]. 人口研究, 2007(5): 42-51.
- [13] 牛建林. 改革开放以来乡城劳动力流动对农村居民教育的选择性及其变迁[J]. 劳动经济研究, 2014(4): 121-140.

- [14] Tong Y., Piotrowski M. Migration And Health Selectivity in the Context of Internal Migration in China, 1997~2009 [J]. Population Research and Policy Review, 2012(4):497-543.
- [15] 齐亚强, 牛建林, 威廉·梅森等. 我国人口流动中的健康选择机制研究[J]. 人口研究, 2012(1):102-112.
- [16] 牛建林. 人口流动对中国城乡居民健康差异的影响[J]. 中国社会科学, 2013(2):46-63+205.
- [17] Lu Y., Qin L. Healthy Migrant and Salmon Bias Hypotheses: A Study of Health and Internal Migration in China[J]. Social Science & Medicine, 2014, 102:41-48.
- [18] 易龙飞, 亓迪. 流动人口健康移民现象再检验: 基于2006~2011年CHNS数据的分析[J]. 西北人口, 2014(6):36-42.
- [19] 周小刚, 陆铭. 移民的健康: 中国的成就还是遗憾? [J]. 经济学报, 2016(3):79-98.
- [20] Markides K. S., Coreil J. The Health of Hispanics in the Southwestern United States: An Epidemiologic Paradox[J]. Public Health Reports, 1986, 101(3):253-265.
- [21] Lu Y. Test of the 'Healthy Migrant Hypothesis': A Longitudinal Analysis of Health Selectivity of Internal Migration in Indonesia[J]. Social Science & Medicine, 2008(8):1331-1339.
- [22] Lu Y. Rural-Urban Migration and Health: Evidence From Longitudinal Data in Indonesia[J]. Social Science & Medicine, 2010, 70(3):412-419.
- [23] Rechel B., Mladovsky P., Ingleby D., et al. Migration and Health in An Increasingly Diverse Europe[J]. Lancet, 2013, 381(9873):1235-1245.
- [24] Padovese V., Egidi A. M., Melillo Fenech T., et al. Migration and Determinants of Health: Clinical Epidemiological Characteristics of Migrants in Malta (2010-11)[J]. Journal of Public Health, 2014, 36(3):368-374.
- [25] Nauman E., VanLandingham M., Anglewicz P., et al. Rural-to-Urban Migration and Changes in Health Among Young Adults in Thailand[J]. Demography, 2015, 52(1):233-257.
- [26] Jasso G., Massey D. S., Rosenzweig M. R., et al. Critical Perspectives on Racial and Rthnic Differences in Health in Late Life[M]. Edited by Anderson N B, Bulatao R A, Cohen B. Washington (DC): National Academies Press (US), 2004.
- [27] Palloni A., Arias E. Paradox lost: Explaining the Hispanic Adult Mortality Advantage[J]. Demography, 2004(3):385-415.
- [28] Hummer R. A., Powers D. A., Pullum S. G., et al. Paradox Found (Again): Infant Mortality Among the Mexican-Origin Population in the United States[J]. Demography, 2007, 44(3):441-57.
- [29] Turra C. M., Goldman N. Socioeconomic Differences in Mortality Among U.S. Adults: Insights Into the Hispanic Paradox[J]. The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 2007, 62(3):S184-92.
- [30] Goldman N., Pebley A. R., Creighton M. J., et al. The Consequences of Migration to the United States for Short-Term Changes in the Health of Mexican Immigrants[J]. Demography, 2014, 51(4):1159-1173.
- [31] 秦立建, 陈波, 余康. 农村劳动力转移的健康选择机制研究[J]. 南方人口, 2014(2):62-71.
- [32] 牛建林, 郑真真, 张玲华等. 城市外来务工人员的工作和居住环境及其健康效应——以深圳为例[J]. 人口研究, 2011(3):64-75.
- [33] 赵建忠, 管冬竹, 原凌云等. 北京市丰台区13家用人单位流动人口职业健康监护的现状[J]. 职业与健康, 2009(13):1362-1363.
- [34] 王桂新, 苏晓馨等. 城市外来人口居住条件对其健康影响之考察——以上海为例[J]. 人口研究, 2011(2):60-72.
- [35] 和红, 曹桂, 沈慧等. 健康移民效应的实证研究——青年流动人口健康状况的变化趋势及影响因素[J]. 中国卫生政策研究, 2018(2):1-9.
- [36] King G., Zeng L. Logistic Regression in Rare Events Data [J]. Political Analysis, 2001, 9(2):137-163.
- [37] Dehejia R. H., Wahba S. Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs [J]. Journal of The American Statistical Association, 1999, 94(448):1053-1062.
- [38] Morgan S. L., Harding D. J. Matching Estimators Of Causal Effects: Prospects and Pitfalls in Theory and Practice[J]. Sociological Methods & Research, 2006, 35(1):3-60.
- [39] Morgan S. L., Winship C. Counterfactuals and Causal Inference: Methods and Principles For Social Research[M]. New York: Cambridge University Press, 2007.
- [40] 郭中阳, 弗雷泽·马克·W. 倾向值分析: 统计方法与应用[M]. 重庆: 重庆大学出版社, 2012.
- [41] 郭中阳. 因果分析高级讲座: 倾向值分析及相关模型[Z]. 北京: 北京大学社会学系暑假班, 2014.
- [42] 巫锡炜, 郭静, 段成荣. 地区发展、经济机会、收入回报与省际人口流动[J]. 南方人口, 2013(6):54-61+78.
- [43] 郑真真, 杨舸. 中国人口流动现状及未来趋势[J]. 人民论坛, 2013(11):6-9.
- [44] 周皓. 流动人口的政治参与——制度与教育[J]. 人口与社会, 2016(4):21-35.
- [45] 洪岩璧. Logistic模型的系数比较问题及解决策略: 一个综述[J]. 社会, 2015(4):220-241.
- [46] 许琪. 利用小普查数据对“随机婚配”假定的再检验——兼论logistic回归系数的可比性问题[J]. 人口与经济, 2016(3):43-51.

(下转31页)

科学,2013(1):1-4.

[17] 都阳、蔡昉、屈小博、程杰. 延续中国奇迹: 从户籍制度改革中收获红利[J]. 经济研究, 2014(8):4-13.

[18] 尹秀芳. 劳动力转移的结构红利效应研究[J]. 经济问题探索, 2016(1):33-41.

[19] 魏杰, 汪浩. 结构红利和改革红利: 当前中国经济增长潜力探究[J]. 社会科学, 2016(1):28-33.

[20] 伍山林. 农业劳动力流动对中国经济增长的贡献[J]. 经济研究, 2016(2):97-108.

[21] 钟水映, 李魁. 人口红利、空间外溢与省域经济增长[J]. 管理世界, 2010(4):14-23.

[22] 姚树洁, 冯根福, 开蕾. 外商直接投资和经济增长的关

系研究[J]. 经济研究, 2006(12):35-46.

[23] 葛顺奇, 罗伟. 外商直接投资与东道国经济增长——基于模仿与创新的研究[J]. 世界经济研究, 2011(1):56-60+89.

[24] 周少甫, 王伟董, 登新. 人力资本与产业结构转化对经济增长的效应分析——来自中国省级面板数据的经验证据[J]. 数量经济技术经济研究, 2013(8):65-77.

[25] 何景熙, 何懿. 产业——就业结构变动与中国城市化发展趋势[J]. 中国人口·资源与环境, 2013(6):103-110.

[26] 张同斌. 从数量型“人口红利”到质量型“人力资本红利”——兼论中国经济增长的动力转换机制[J]. 经济科学, 2016(10):5-17.

Demographic Dividend, Structural Bonus and Regional Economic Growth

ZHAO Chun-yan

(Development institute, Yunnan University, Kunming Yunnan 650091, China)

Abstract: Based on the panel data of 30 provinces in China in the past 2000 ~ 2016 years, based on the per capita output and the neo classical endogenous growth theory, we constructed a model based on the neoclassical endogenous growth theory, and applied the GMM method to analyze the impact of demographic dividend and structural bonus on China's Provincial Economic growth. The empirical results show that the direct effect from the point of view, the demographic dividend, dividend Chinese structure directly promote the provincial economic growth and demographic dividend effect more; within the region, the rise of the elderly dependency burden will erode the demographic dividend to bring down the burden of children showed significantly to the eastern and central regions, the western region is not significant; and structure the bonus to promote economic growth in the East, in the west, decreasing trend. From the indirect effect, child dependency ratio decreased, the elderly dependency ratio rising of human capital investment through increased formation of significant positive effect on economic growth. The child dependency ratio decline can promote human capital investment to increase the impact on regional economic growth is the East > middle > West; elderly dependency ratio can rise through the promotion of human capital investment to increase the impact on the economic growth is the East > Western > middle; agricultural employment structure transfer through human capital investment to increase the impact on economic growth is the middle > west > east.

Key words: Demographic dividend; Structural bonus; Human capital; Economic growth

(上接 22 页)

Migration and Health Selectivity

WANG Ling-xin¹, ZHOU Hao²

(1. Department of Sociology, Tsinghua University, Beijing 100084, China;

2. Department of Sociology, Peking University, Beijing 100871, China)

Abstract: Using two waves national-representation data - Chinese Family Panel Studies (CFPS), this paper examine the 'healthy migrant hypothesis' in a the social context of contemporary China. The Logistic regression models are used to the cross sectional and followed samples at first. Then, an appropriate comparison is used between the new migrant during 2010 and 2012 and the non-migrants at two waves among the followed sample, while propensity score matching methods is used to eliminate the selection bias of the self-rate health, which is regarded as the treatment, based on the counterfactual causal inference framework. Consistent with research from both international migration contexts and other internal migration settings, this paper finds that migrants are positively selected on the basis of health, and the relationship between health and migration will not diminishes across time. The non-significant relationship in regression models maybe caused by the endogeneity of self-rate health and the exchangeable replacement of other migration selection characteristics to self-rate health.

Key words: Migration; Health; Self-rate health; Selection; Endogeneity