

被访者驱动抽样在农民工调查中的应用:实践与评估*

刘林平 范长煜 王 娅

提要:针对长期以来农民工调查缺乏抽样框的困境,本研究引入被访者驱动抽样(Respondent-Driven Sampling)对全国7个省13个城市(区)农民工进行了较大规模的问卷调查。本次调查针对中国农民工的实际情况,采用非固定点调查法,并对原有方法有所改进。本文是对RDS在农民工调查中的效果评估。数据分析结果表明,此次对农民工的RDS抽样所获样本达到了均衡,且在较短招募批次内快速收敛,所获样本代表性较好。本文认为,RDS抽样虽然不是对目标群体的完全随机抽样,但可以在努力改进和严格执行抽样程序的基础上获得一个渐进无偏的样本,并对样本偏差做出估计。RDS可以作为农民工及其他缺乏抽样框群体抽样的有效的替代性方法。

关键词:被访者驱动抽样 农民工 均衡收敛

一、抽样困境和被访者驱动抽样的引入

在农民工抽样调查中,抽样框始终是困扰研究者的一大问题。众所周知,针对频繁流动的农民工,就目前的抽样条件而言,在调查过程中建立一个完整的抽样框,进而获得具有代表性的样本是非常困难甚或不可能的事情。以往的农民工问卷调查在实际操作中往往采用街头拦访、“滚雪球”、判断抽样等非概率抽样方法。概率抽样的目的是为了准确地估计抽样误差,只有这样统计推论才能适用于总体(巴比,2005)。如何从实际情况出发,使抽样结果更接近概率抽样,这是农民工抽样调查要解决的关键问题之一。

* 本文是教育部2013年度哲学社会科学重大课题攻关项目“户籍限制放开背景下促进农民工中小城市社会融合的社会管理和服 务研究”(项目号:13JZD018)的阶段性成果之一。本文实际上是课题组集体努力的成果,在写作过程中,毕先进、王开庆、李潇晓、雍昕、朱艳婷、刘昱君、沈宫阁、王佳、马天琳、王震和周玲玲等人参与讨论、翻译资料,做出了重要贡献,特此致谢! 本文的修改完善也得益于匿名审稿人的意见,特此感谢!

在农民工调查中,无论是流出地抽样还是流入地抽样,都面临着农民工流动和调查者经费、时间有限等因素制约。就流出地调查来说,可以用农村的户籍登记数据建立抽样框,但由于农民工已经外出打工,要在流出地对他们面对面进行调查基本不可能。即使是在一些特殊时段,比如春节期间,也不能保证所有农民工都会返乡。而电话访谈的拒访率过高,也不适合较长的问卷。当然,理论上可以使用农村户籍登记数据建立抽样框,然后到城市去追踪农民工进行面访,但由于农民外出打工不太可能完全聚集在一个城市,一个分散的追踪过程要耗费大量人力、财力,不是普通研究者所能承受的。就流入地来说,虽然现行制度要求流动人口到流入地派出所登记并领取一个临时居住证明,但仍然有大量外来工没有登记,难以据此建立一个完整的抽样框。当然,其他抽样方法,比如地图抽样,理论上可以进行,但操作中难度很大,对调查者的经费要求也很高。

因此,有学者提出在外来农民工抽样中采用“被访者驱动抽样”(Respondent-Driven Sampling,简称RDS),也叫“受访者驱动抽样”、“应答者推动抽样”或“同伴驱动抽样”(Qiu et al., 2012)。在此之前,一般的“滚雪球”抽样得到了广泛的应用,但“这种方法的明显缺陷是,所得样本并不是从目标群体中通过等概率抽样获得的……这种抽样中选取的前几个对象往往会影响所得样本的属性。因此,几乎不可能通过这样的样本来对总体的情况做有效概括”(特雷曼等,2012:410-411)。1997年以来,美国康奈尔大学的社会学家道格拉斯·赫克索恩(Douglas Heckathorn)对滚雪球抽样做了重要改进,并在数理上证明,在满足特定(强)假设时,这种改进了的滚雪球抽样能产生对总体情况的渐进无偏估计(Heckathorn, 1997)。所谓被访者驱动抽样,就是“滚雪球抽样”与一个给样本赋予权重的数学模型相结合、以弥补非随机抽样缺陷的抽样方法。^① RDS的两个主要创新在于:一是把滚雪球抽样与一套特别的程序结合起来,二是利用一个数理统计模型修正某些个体拥有更大入样概率的问题(Gile & Handcock, 2010)。

在RDS中,样本通过被访者推荐的方式进行收集。第一个接受访问的对象被称为“种子”(seed),他/她在接受访问的同时,我们以物质奖励

^① 赫克索恩专门建立了一个网站来介绍被访者驱动抽样,网址为 <http://www.respondentdrivensampling.org>。

的方式激励其邀请目标群体中的其他成员参与调查;以此类推,后续的被访者也被激励推荐其他成员参与调查,这个过程一直持续到样本实现“均衡”。RDS的这一推荐过程可以被认为是马尔科夫过程。^① 根据马尔科夫大数定律,^②样本在招募被访者的过程中会实现均衡,即最终样本的构成变得稳定,并与初始的种子相互独立(Heckathorn,1997)。赫克索恩在2002年发表文章,从马尔科夫大数定律的理论出发,对样本的均衡性做了详细的论证,指出理论上的均衡样本构成可以通过方程计算出来(Heckathorn,2002)。此外,萨尔加尼克和赫克索恩等人努力改进了RDS的抽样概率估计方法,打破了非随机样本不能对总体情况进行无偏估计的困局(Salganik & Heckathorn,2004;Volz & Heckathorn,2008)。

在国外,RDS主要应用于诸如艾滋病患者(King et al.,2009)、吸毒人员(Abdul-Quader & Heckathorn,2006)和性工作者(Johnston et al.,2006)等隐藏群体^③的疾病防控调查中,其中有些研究涉及到流动人口,如对泰国—柬埔寨边境上的流动移民工人的疟疾防治问题调查(Khamsiriwatchara et al.,2011)以及中东欧男性移民的男男性行为调查(Evans et al.,2011)。

但RDS不仅限于隐藏群体的抽样调查,赫克索恩和其他人也在非隐藏群体中尝试使用过,如对爵士乐手(Heckathorn & Jeffri,2001)、工人(Bernhardt et al.,2013)以及美国本土亚文化群体的调查(Walters et al.,2001)。2007年,赫克索恩还曾尝试在互联网上对在校大学生进行RDS抽样,并与常规抽样方法做了比较分析,他认为在确保RDS随机推荐假设和推荐链条足够长的前提下,RDS估计值不仅合理,还可以进一步提高精确性(Wejnert & Heckathorn,2008)。

目前,国内RDS主要应用于艾滋病患者、男男性行为、吸毒人员等

① 马尔科夫过程(Markov process)是一类随机过程。它的原始模型马尔科夫链是俄国数学家A. A. 马尔科夫于1907年提出的。这类随机过程具有“无后效性性质”,即当随机过程在某一时刻 t_0 所处的状态已知的条件下,过程在时刻 $t > t_0$ 时所处的状态只和 t_0 时刻有关,而与 t_0 以前的状态无关,则这种随机过程称为马尔科夫过程(Suhov & Kelbert,2008)。

② 马尔科夫大数定律(Suhov & Kelbert,2008),指对于随机变量序列 $\xi_1, \xi_2, \dots, \xi_n$,若有 $\frac{1}{n^2}D(\sum_{i=1}^n \varepsilon_i) \rightarrow 0, n \rightarrow \infty$,则有 $\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{i=1}^n \varepsilon_i - \frac{1}{n} \sum_{i=1}^n E\varepsilon_i\right| < \varepsilon\right) = 1$

③ 所谓“隐藏群体”(hidden populations)有两个特征:第一,无法知道他们的人口规模和群体边界,难以建立抽样框;第二,由于这些群体常常被污名化或者存在违法行为,他们往往因为不愿暴露身份而拒绝接受调查,或者出于对隐私的保护而在调查中提供虚假信息(Heckathorn,1997)。艾滋病患者、吸毒者、同性恋及性工作者等等,都是典型的隐藏群体。

疾病预防控制方面的调查(许娟等,2007;刘石柱等,2005)。在社会科学研究中,赵延东(2007)较早注意到 RDS,并专门撰文对其做了较为全面的介绍;张文宏和雷开春(2008,2009)于 2007 年应用 RDS 调查了上海城市新移民;梁玉成(2013)于 2011 年使用该方法调查了非洲裔外国人。值得注意的是,2012 年邱培媛等人发表文章,介绍了 RDS 在四川省成都市采集外来工样本中的应用情况,结果表明 RDS 在外来工调查中是一种稳定的抽样方法。这些研究的共同点是把 RDS 应用到了非隐藏群体的调查中(Qiu et al.,2012)。

国内对 RDS 的应用总结了不少经验,也存在明显的局限。首先,应用领域主要集中于疾病预防控制方面的调查,这些调查采集的数据仅限于少数人口学和医学变量,对于复杂的社会科学调查借鉴意义有限。虽然邱培媛等人首次在外来工群体中采用 RDS 进行抽样调查并对结果进行了详细分析,但他们的研究仍属于疾病预防控制领域(Qiu et al.,2012)。其次,赵延东(2007)对移民儿童生活状况,张文宏、雷开春(2008,2009)对上海新移民和梁玉成(2013)对非洲裔外国人的调查总结了一些操作经验,但缺乏详细的效果评估,而且调查仅限于一地,样本规模较小。

2014 年 6 月 27 日到 8 月 4 日,为完成教育部哲学社会科学研究重大课题攻关项目“户籍限制放开背景下促进农民工中小城市社会融合的社会管理和服务研究”(项目号:13JZD018)的任务,本课题组首次在全国 7 个省 13 个城市(城区)中采用 RDS 进行了较大规模的农民工抽样。本文拟在此次调查基础上,以详细的数据分析评估 RDS 在农民工调查中的应用效果,并对相关问题进行讨论。

二、RDS 理论

以往的滚雪球抽样被认为是方便抽样,因为所获得的样本不可能产生无偏估计(Berg,1988)。赫克索恩等人挑战了这一观点,在滚雪球抽样的基础上,根据社会网络理论的基本原理建立了 RDS。通过 RDS 收集真实数据,他们对抽样过程中的非随机社会网络结构做了解释,同时计算了无偏的总体估计效果(Salganik & Heckathorn,2004; Volz & Heckathorn,2008)。

(一)基本原理

RDS 用抽样过程中获得的目标群体社会网络关系数据来计算总体估计值,其原理源自于抽样过程中以下两个观察发现:

第一,如果推荐链足够长,抽样结果会实现均衡,亦即样本构成比例会保持稳定,所有被抽中的目标群体成员与种子相互独立(Heckathorn,2002)。推荐链的长度以推荐批次计算,每个批次代表链条的一单位长度。由于种子没有相应的推荐人,他们被记为第 0 批,种子直接推荐的被访者构成第 1 批,而第 1 批被访者推荐的目标群体成员构成了第 2 批,如此类推,依次向下推荐后续批次。经过数个批次后,样本构成变得稳定,也即实现了样本均衡。达到样本构成均衡所需的批次数是通过马尔科夫链模型对每个指标进行模拟计算得到的。在马尔科夫链模型中,根据前一批次的样本构成,样本转移概率被用来计算每一批样本构成。当前一批样本构成与后一批样本构成之间的比例差异小于 2% 时,样本构成即实现均衡。

第二,在 RDS 抽样过程中,通过收集两种信息,可以计算目标群体的抽样框(Heckathorn,2002)。一种信息是推荐记录。RDS 要特制一个招募卡片(coupon),用以记录被推荐者的编号和推荐者与被推荐者的关系。这些信息可以控制和解释非随机社会网络结构引起的偏差。另一种信息是被访者所报告的他/她所知道的其他目标群体成员数量。在一个通过网络获得的样本中,一个个体的内含概率(inclusion probability)是指他/她同所联系目标群体成员数量的比例,被称为他/她的度(Volz & Heckathorn,2008)。^① 萨尔加尼克和赫克索恩指出,一旦样本达到均衡,目标群体中所有关系(即两个有关联的成员之间的联系)都将有相同的被推荐概率,这样一个拥有 50 度的个体被推荐的概率就是仅有 10 度的个体的 5 倍(Salganik & Heckathorn,2004)。但那些具有高度数的个体可能会导致 RDS 样本出现偏差,萨尔加尼克和赫克索恩通过导出一个子群平均度数估计值来予以修正(Salganik & Heckathorn,2004)。这个值是两个汉森-赫维茨(Hansen-Hurwitz)估计值的比值。我们知道,汉森-赫维茨估计是无偏估计(Brewer & Hanif,1983),两个无偏估计值的比率在 n^{-1} 序列基础上是渐进无偏的,

^① 度(degree)即社会网络理论中对两个个体之间的连接的描述,如六度空间理论,英文名称是“six degrees of separation”。

这里的 n 是指样本规模 (Cochran, 1977; Salganik & Heckathorn, 2004)。这个估计值可以用来修正 RDS 中变量估计的度数偏差 (Wejnert & Heckathorn, 2008)。

(二) RDS 估计

与传统抽样设计不同, RDS 采用的是适应性抽样设计, 抽样程序受到目标群体已知网络的影响, 因此必须使用一个特别的估计程序。

传统抽样估计是先从总体中随机抽取样本, 再以样本直接估计总体, 如图 1(a) 所示。被访者驱动抽样与此不同, 采取了两步间接估计方法, 如图 1(b) 所示 (Salganik & Heckathorn, 2004)。首先用样本估计目标群体的社会网络结构, 然后以社会网络信息估计不同子群体在总体中的比例。这样就可以尽量避免其他链式推荐抽样中已知存在的许多问题。

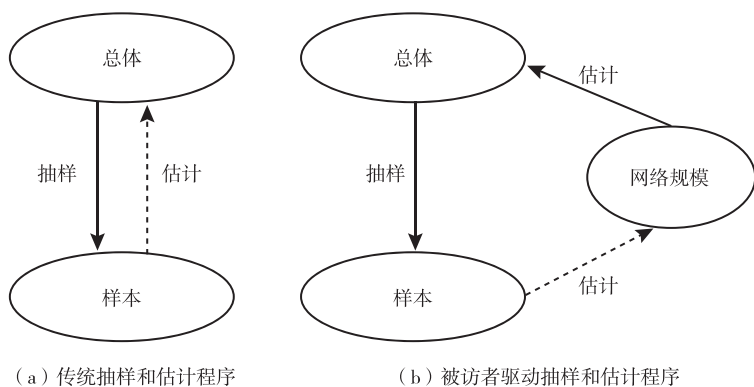


图 1 传统抽样与被访者驱动抽样差异

具体的估计程序是, 用度估计值结合招募数据计算出子群规模比例的 RDS 估计值 (Salganik & Heckathorn, 2004; Heckathorn, 2002):

$$\hat{P}_x = \frac{\hat{S}_{yx} \hat{D}_y}{\hat{S}_{yx} \hat{D}_y + \hat{S}_{xy} \hat{D}_x} \quad (1)$$

公式(1)中, S_{xy} 是指子群 X 成员推荐子群 Y 成员的比例 (例如, 男性个体推荐女性个体的比例), 与之相对, S_{yx} 是指子群 Y 成员推荐子群

X 成员的比例; D_x 是子群 X 个体的平均度数估计, D_y 是子群 Y 成员的平均度数估计。相对内含概率(relative inclusion probabilities)即样本权重,可以用 RDS 估计值和样本构成比例计算出来:

$$W_x = \frac{\hat{P}_x}{C_x} \quad (2)$$

公式(2)中, W_x 指子群 X 在总体中的比重, C_x 指子群 X 在样本中的比重, \hat{P}_x 即子群规模的 RDS 估计值。这里需要注意的是,韦吉内特和赫克索恩明确指出, RDS 估计值的渐进无偏性取决于以下五个假设(Wejnert & Heckathorn, 2008):

- (1) 被访者之间维持双向关系,即他们相互知道对方为目标群体成员;
- (2) 被访者在一个网络中都是以单一要素联系在一起的;
- (3) 抽样过程伴随着替换;
- (4) 被访者能够准确报告他/她个人的网络规模或他/她的度;
- (5) 被访者推荐目标群体成员是随机选择。

前三个假设指明了可以实施 RDS 的总体必须满足的条件(Wejnert & Heckathorn, 2008)。首先,推荐过程可以实现,被访者必须可以接触到并识别出满足条件的目标群体成员。RDS 特别适合那些内部具有明显结构和互动的群体,在这类群体中,成员间的互动保证了成员之间关系的维持。进一步而言, RDS 估计的基础是双向关系的网络结构,即被访者 A 推荐了被访者 B,那么被访者 B 也有可能推荐被访者 A,这样在 RDS 实施过程中,就能确保被访者去招募他们的熟人或朋友,而不是一个陌生人。第二,目标群体被假定为构成单一的群体,也即所有目标群体成员必须能够从任意一个被访者接触到。如果允许被访者不仅可以推荐那些具有特定关系的个体,还可以推荐其他朋友或熟人,则个体的群体网络度就会大大超出单一要素构成的群体网络度(例如,一个农民工被访者只能推荐与之有关联的其他农民工亲戚、老乡或者工友,而不能推荐其他诸如当公务员的亲友,否则他所报告的网络规模就会大大超出他在农民工群体中的网络度数)。需要注意的是,由于

实际的社会网络并非完全随机,肯定不存在任何可以完全把一个子群同群体中其他子群隔离开来的社会或结构性障碍,因此 RDS 不能用于跨界抽样,同样,一个种子不能推荐在不同城市中的同一属性成员。第三,抽样过程被假定为伴随着替换,以便招募一组有效的目标群体成员,避免出现浪费。其结果是,抽样率应该保持足够小,以确保这个模型是恰当的。

第四个假设是被访者可以准确提供个体网络规模的信息。赫克索恩认为,网络指标的稳定性研究表明 RDS 式的指标属于最为可靠的指标(Wejnert & Heckathorn, 2008)。第五个假设假定被访者的推荐是从他/她的个人网络中随机选取的。赫克索恩认为这个随机性假设可以通过仔细的抽样设计来实现(Wejnert & Heckathorn, 2008)。也就是说,后面两个假设主要是针对抽样过程中的要求。韦吉内特从经验上证明了是可操作的(Wejnert, 2009)。

总的来看,RDS 估计的根据是群体成员之间相互熟识,以其社会联系可以用来获取该群体的样本。众所周知,“滚雪球”抽样不能产生具有代表性的数据,因为使用社会联系招募新的被访者不是在总体中随机产生的;但通过在抽样过程中收集被访者及其目标群体的社会网络规模信息,可以计算由非随机网络结构所造成的偏差,RDS 研究者可以由此计算出无偏总体估计,并对总体做出推断(Heckathorn, 2002)。

(三)对农民工调查的适用性

根据 RDS 的前三个基本假设,可以判断农民工符合 RDS 对目标群体的要求。首先,众所周知,我国农民工是一个网络化的群体。农民工在求职过程中往往会借助亲戚朋友等社会关系网络(张春泥、刘林平, 2008),并在打工地形成与不同类型个体相互交织的社会网络(李树苗等, 2006)。显然,这些网络能满足 RDS 所要求的被访者之间维持双向关系的要求。第二,农民工与城市工人有着清晰的群体边界,城乡二元结构使两种工人群体在特征上有明显的区别,其中户籍身份是这些特征差异的核心,因此农民工群体可以根据户籍清晰地识别出来。第三,由于我国的农民工群体规模庞大,而常规的抽样调查一般仅抽取几百到数万之间的样本,所以抽样过程中是否放回可以忽略。

就第四条假设要求准确提供个体网络规模的信息来说,农民工的

社会关系往往由亲人、老乡和工友等构成,交往的圈子较为明确、单一和狭窄,他们互相之间也较为熟悉,所以在实际调查中准确汇报网络规模的可能性很大。对于假设五中要求被访者在推荐目标群体成员时进行随机选择,由于农民工的职业相对于其他隐藏群体更具有公开性、合法性,较少隐私性,推荐者在推荐时更少顾虑,而且被推荐者拒访的可能性也更小。至于交通和时间安排等影响被访者拒绝访问的因素,则只是技术层面的问题,可以克服。

当然,RDS发端于小规模隐藏群体的抽样调查,在农民工调查中使用RDS,要考虑农民工是大规模非隐藏群体的现实。尤其是在大城市中,农民工群体规模庞大,分布于各行各业,如何才能获得一个能够覆盖整体的样本,建立合理的样本代表性评估指标,这些问题都可以在调查和研究中进行探索。

三、抽样和分析方法

(一)调查对象

本次问卷调查的对象是中小城市的农民工,为明确调查对象识别的可操作性,农民工被界定为农业户籍,跨县区(乡镇)流动,非正规大专(非正规指电大、函授、自考等教育方式)及以下学历,正式在企业或单位上班的就业者。也就是说,调查对象排除了城市户籍移民工人、在本乡镇就业的农民工、正规大专以上学历者,以及无就业单位的散工和在校上学的实习生、兼职生等。调查采用面对面访谈的方式进行。

(二)RDS实施

根据RDS规定,本次农民工被访者驱动抽样的具体实施步骤如下:

(1)每个城市选取2-3个农民工做种子,选种子时注意行业、城市区位,种子记为第0批被访者。

(2)种子接受调查后给予一定现金或礼品作为接受访谈的报酬,同时每个种子要推荐2-3个自己熟识的农民工,作为第1批被访者。推荐时,作为推荐人的种子要报告被推荐人的姓名、性别、年龄、行业、联系方式以及相互间的关系,确定被推荐人适合接受调查后,推荐人与被推荐人联系,确定接受调查的时间和地点。推荐人每成功推荐1个

农民工,又给予一定现金或礼品作为报酬。

(3)后续招募依第2步类推,形成第2批、第3批、第4批等后续批次的被访者,为了确保推荐链条足够长,访问员要努力使每个种子的后续链条至少推荐4层以上。

(4)每个接受调查的被访者记录一个RDS卡片,卡片内容包括RDS编号和与推荐人的关系。

(5)当样本量达到规定的样本规模时,抽样结束。本次研究计划每个省调查280个农民工,其中地级市调查160人,县级市调查120人,广东省仅广州市,调查农民工280人。^①

本次农民工RDS实施与以往RDS操作最大的不同在于没有固定的调查点。隐藏群体调查出于保护隐藏群体人员个人隐私的需要,一般会采取固定点调查,非隐藏群体则没有这一限制。在抽样调查过程中,访问员在获得被访者的推荐信息后及时联系下一个被访者,在确定接受调查的地点和时间后及时找到这个被访者,对其展开访谈。这种非固定点调查有如下好处:

(1)有助于提高样本的代表性。前文中,RDS第五条假设要求被访者推荐目标群体成员是随机选择,韦吉内特(Wejnert,2009)曾对这个假设的可操作性提出过质疑,即如何能够保证推荐是随机选择。例如,如果被访者主观上随机推荐了一个目标群体成员,但该成员因为客观上的原因,如距离调查地太远或因为工作时间的安排不能前来接受调查,那么第五条随机选择假设就大打折扣。邱培媛等人在成都运用RDS调查农民工的结果分析表明,固定调查地点、等待被访者上门接受调查会导致一些距离调查地点远的调查对象出于成本(交通、误工等)收益(接受访谈和推荐下一批次被访者的报酬)的考虑拒绝或放弃接受调查,进而影响每个目标群体成员的入样概率(Qiu et al.,2012)。非固定点调查可以很好地处理这个问题,即通过访问员前去寻找被推荐者,可以大大降低因时间、交通等问题导致的拒访率,进而确保被访者推荐的随机性,提高样本代表性。

(2)有助于防止重复调查和识别调查对象身份的真实性。赫克索

① 根据RDS的要求,样本规模取决于样本构成是否达到均衡,只有当样本构成均衡时,抽样才能结束。由于我们调查的省市较多,而调查经费又较有限,因此不得不对每个城市的样本规模做出限定,但根据我们调查前的计算,如果每个城市分配两个种子,每个被访者成功推荐两个农民工,120-160个样本量基本能确保每个种子推荐6层以上。

恩强调,在 RDS 实施中,需要建立一个被访者身份数据库来记录被访者的身份特征,从而避免发生重复调查;同时避免被访者因推荐奖励把非群体成员纳入推荐行列(Heckathorn, 1997)。在本次调查中,小组督导带领访问员到市区的不同位置进行调查,在抽样过程中绘制推荐网络图并登记相应的人口学信息,每天各小组督导向总督导汇总信息,总督导对这些信息再进行汇总和识别。这一方面可以防止对某个个体重复调查,另一方面可以根据推荐信息、工作场所和当面询问的方式避免非目标群体成员被纳入调查。

(3)有助于缩短调查时间,节约调查费用,提高调查效率。以往固定点 RDS 调查往往需要数月,甚至耗费一年以上时间,且要吸引目标群体成员前来接受调查往往也需要付较高的报酬。例如邱培媛等人在成都的调查从 2008 年 9 月开始到 2009 年 6 月结束,用了 10 个月时间;费用上,每个被访者接受访谈的报酬是 40 元,推荐费是阶梯式给付,即推荐 1 人 10 元,2 人 25 元,3 人 45 元(Qiu et al., 2012)。而非固定点调查可以取消等待被访者前来接受调查的时间,同时大大降低被访者在交通和时间上的成本,节约调查费用。如本次农民工调查从 2014 年 6 月底开始至 8 月初结束,整个调查前后历时 1 个月左右,各地调查用时 15 至 25 天不等(见表 1),费用也较为节省。

表 1 调查地点与时间

省份	城市		起止时间
	地级市	县级市	
江苏	常州	武进	7 月 1 日 - 7 月 15 日
浙江	金华	义乌	6 月 27 日 - 7 月 9 日
山东	泰安	肥城	7 月 10 日 - 7 月 26 日
陕西	咸阳	兴平	7 月 5 日 - 8 月 2 日
湖南	岳阳	汨罗	7 月 18 日 - 8 月 3 日
贵州	遵义	凯里	7 月 8 日 - 7 月 20 日
广东	广州(副省级)		7 月 10 日 - 8 月 4 日

当然,非固定点调查也存在不足,主要是增加访问员工作难度及对其监督的难度,对访问员的数量和积极性要求较高。

(三) 分析方法

RDS 有特定的数据整理、处理方法及专门的分析软件 RDSAT。主要分析步骤有二:一是描述分析,二是均衡分析。描述报告的主要内容是每个城市和整个样本的种子及其每一批次的被访者,包括种子的数量、招募人数和招募批次,以及总体的招募情况。

均衡分析主要是通过分析样本中的几个指标及其构成,来探讨样本的均衡性和代表性。邱培媛等人选择的农民工群体特征指标包括性别、年龄和行业(Qiu et al., 2012)。其中,性别和年龄是一般性指标,任何群体都会采用这两个指标,但行业作为分析外来农民工的群体特征指标有值得商榷的地方。行业分布具有集聚效应,具体表现为:(1)城市一般都有其主导产业,行业门类并不齐全,尤其中小城市更是如此。(2)某个行业在城市中的布局也相对集中,例如工业区、商业区和科教文卫区等。因此,以行业划分农民工的子群体会导致大量的群内推荐,引发同质性问题,进而影响样本均衡。(3)行业并不是群体的本质属性,因为一个个体可以在不同时段内变换多种行业,农民工更是如此,他们跨行业换工作的现象十分常见。我们认为以工种作为农民工群体特征指标较为合适。一是工种在不同类型企业中的分布较为均衡,无论在哪个单位都存在相应比例的工种或岗位,个体跨工种交往的可能性很大;二是工种稳定地反映了个体的工作能力和工作性质。具体的分析变量见表2。

表2 变量及其说明

变量名	变量说明
id	样本编号
netz	个体在目标群体中所拥有的网络规模,即“度”
cpr	接收的卡片编号
cpg	给出的卡片编号
wave	所属推荐批次
seed	所属种子,以种子编号标记
gender	性别:1 = 男, 2 = 女
ageg	年龄组:1 = age < 26, 2 = age 26 - 35, 3 = age 36 - 45, 4 = age > 45
workt	工种分类:1 = 普工(包括流水线生产工、其他生产工、后勤服务人员、服务员、保安、清洁工、建筑工、农林牧渔工人等), 2 = 技工(包括技术员、质检员、文员、司机和销售业务员等), 3 = 中低层管理者(包括班组长、领班、中低层管理人员), 4 = 其他(以上未包含的职业或岗位)

四、数据分析结果

(一)描述分析

表3显示了本次RDS抽样使用的种子数、种子招募人数和招募批次的情况。此次7省13市(区)抽样总共使用了75个种子,平均每个城市使用5.77个种子;这75个种子招募了1945人,平均每个种子招募约26人。

表3 种子及其招募情况

城市	种子数	种子招募(不含种子)					招募批次(不含种子)	
		招募数	均值	最大值	最小值	招募最多的两个种子占比(%)	均值	最大值
常州	4	162	40.50	59	28	61.11	4.00	8
武进	3	119	39.67	49	31	73.95	4.55	11
金华	7	125	17.86	29	5	41.60	3.54	7
义乌	8	155	19.36	35	7	39.35	3.14	6
泰安	11	101	9.18	35	0	46.53	3.20	9
肥城	2	174	87.00	98	76	100.00	5.66	12
咸阳	8	182	22.75	77	2	70.33	3.43	8
兴平	4	93	23.25	33	6	68.82	3.28	7
岳阳	5	161	32.20	42	22	49.07	4.70	12
汨罗	5	119	23.80	40	8	57.98	4.84	11
遵义	5	159	31.80	74	3	72.96	5.00	11
凯里	7	118	16.86	28	6	41.53	2.90	6
广州	6	277	46.17	202	1	89.17	11.19	22
全部样本	75	1945	25.93	202	0	—	5.11	22

由于样本规模的限定,种子越多,每个种子招募的人数就越少。但反映招募效果较好的统计值是招募人数的累积百分比。从表3的“招募最多的两个种子占比”可知,有8个城市招募人数最多的两个种子累积招募比例超过了50%。

在整个样本中,广州有一个种子(编号为MA)招募最多,招募了202人,韦吉内特和赫克索恩把这类招募了很多被访者的种子定义为

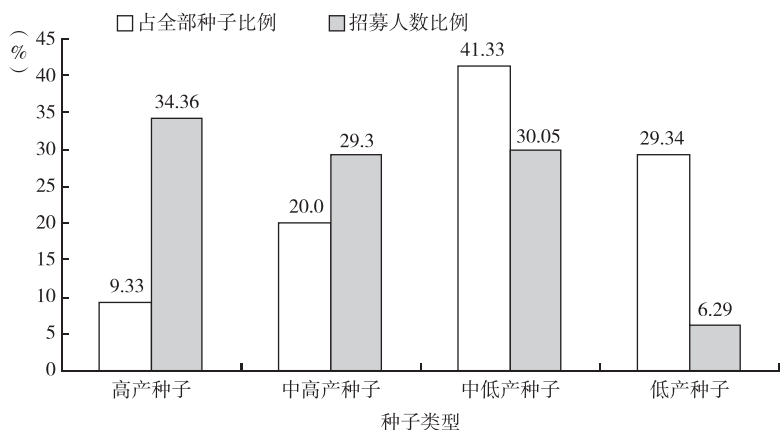


图2 种子类型及其招募比例

“超级种”(super seed)(Wejnert & Heckathorn, 2008);泰安有一个种子(编号为EK)未招募到任何被访者,邱培媛等人把这类种子定义为“不育种”(infertile seed)(Qiu et al., 2012)。从图2可知,有9.33%的高产种子招募人数在50人及以上,20%的中高产种子招募人数在30-49人,41.33%的中低产种子在10-29人,29.34%的低产种子招募人数在10人以下;其中,高产种子招募了总人数的34.36%,数量众多的低产种子仅招募了6.29%。

表3所显示的招募批次中,广州的平均招募批次最多,达到11批,其最高批次达到了22批;凯里的平均招募批次最少,约3批,最高招募批次6批;有7个城市的平均招募批次在4层及以上。整个样本的平均招募批次为5批,各批次招募人数增长趋势见图3。从图3中可以看出,实际招募人数的增长趋势与计划招募人数(每个城市两个种子,共26个种子,每个被访者招募2人)的指数增长趋势有区别。按照理想的指数增长,6至7个批次就可以完成2000多个样本的收集,但实际上多数种子招到第3批达到高峰,到第4批招募人数开始下降,至第11批后招募人数进入平缓增长阶段。但与以往的RDS实践(Wang et al., 2005)相比,本次抽样的招募趋势表现明显要好一些。

图4显示了肥城种子FA(id=1643)的推荐链效果图,该种子招募了12批,共98个被访者。

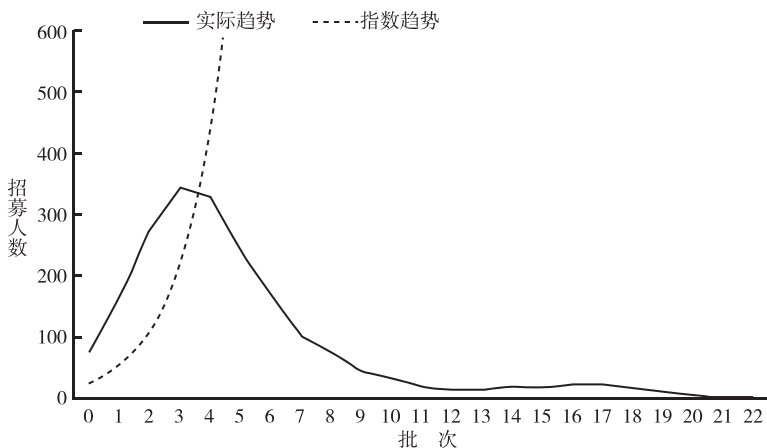


图3 RDS招募趋势

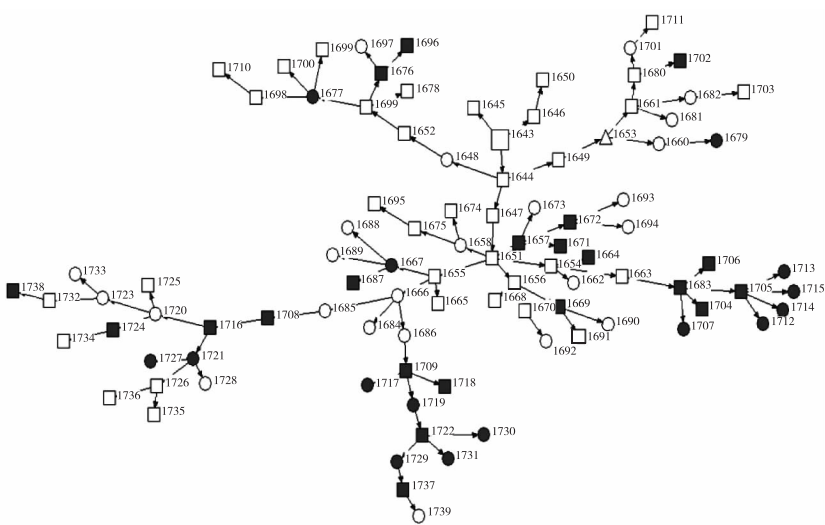


图4 肥城种子FA推荐链

(二) 均衡分析

表4展示的是性别、年龄和工种三个指标构成随着招募批次的逐渐增加发生变化的趋势。在初期招募的75个种子中,男性约占70%,36岁以下青年人约占65%,普工居多,但中低层管理者也占16%。由于初期招募的种子不是随机抽取的,因此肯定存在偏差,但随着招募批次的增加,性别、年龄和工种的构成明显发生变化。在最后的样本构成中(第1-22批),男性比例下降到53.52%;36岁以下青年人下降到47.51%,45岁以上中老年人增加了10%,占24.88%;普工占比先降后升,维持在60%多,中低层管理者的比例下降至5.19%。总的来看,各个指标构成在1-5批次内发生了较大变化,随后便逐渐稳定下来。

赫克索恩(Heckathorn,1997)明确指出,“由RDS产生的样本可以以很快的速率(如几何级)接近均衡”,表5显示了各地抽样均衡收敛所需招募批次和构成比例。从表5中可知,性别构成均衡所需招募批次在2-6批,年龄构成均衡多数在3-5批,泰安、肥城、岳阳和遵义需要8批次以上;工种构成均衡在1-5批。整个样本要达到均衡收敛,性别仅需4批,年龄仅需5批,工种仅需3批。

引起RDS样本偏差,威胁样本代表性的另一个因素是“同质性”或“近亲繁殖”(Heckathorn,1997,2002),即子群内部相互招募。赫克索恩指出,当近亲繁殖普遍存在时,接近均衡的速度会很慢,所以在有限批次内,均衡只有在近亲繁殖不会很极端的时候才能达到;而当子群边界不能打破时,就要在相应的其他子群(如与男性相对的女性)中进行类似的抽样,在不跨越抽样群体边界的情况下,确保这些近亲繁殖是对等的,即确保所有子群的同质性水平一样,RDS才能获得一个无偏样本(Heckathorn,1997)。图5显示了本次RDS抽样的同质性水平。从图5中可知,性别、年龄和工种均存在不同程度的同质性,但未达到完全同质的水平。其中,男性与女性的同质性水平差不多;年龄的同质性水平差异较大,这在一定程度上影响了年龄的均衡速度,这从表5中年龄组均衡所需批次中可以看出;工种的同质性较低。总的来看,样本在较短批次内就达到了收敛,同质性未对均衡造成太大影响,属于可接受的水平。

RDSAT软件还提供了样本构成估计的结果,包括招募样本分布、招募样本构成、均衡样本构成和估计总体构成,见表6。为了进一步明确样本的代表性,本文分别计算了招募样本构成同均衡样本构成和估

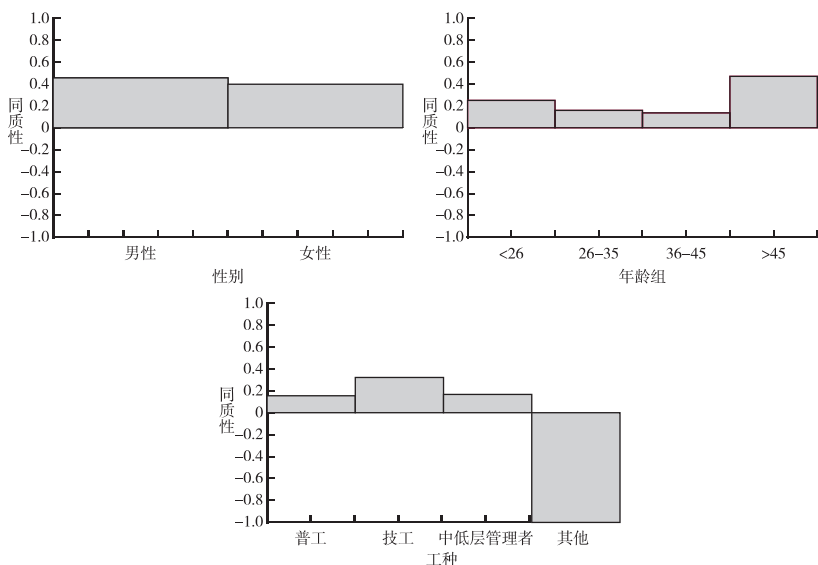
表 4 RDS 样本量和样本构成比例随招募批次变化的趋势

招募		累计样本构成										
累积批次	招募人数	累积样本数	性别(%)		年龄组(%)				工种(%)			
			男	女	<26	26-35	36-45	>45	普工	技工	中低层管理者	其他
0	—	75	69.33	30.67	30.67	34.67	20.00	14.66	64.00	20.00	16.00	.00
1	163	163	66.87	33.13	28.83	30.06	22.09	19.02	55.21	35.58	7.98	1.23
1-2	272	435	64.83	35.17	26.67	27.59	23.44	22.3	59.31	33.33	6.67	.69
1-3	342	777	61.78	38.22	28.31	25.61	23.94	22.14	62.16	31.27	5.79	.78
1-4	329	1106	59.22	40.78	28.48	24.32	24.96	22.24	63.02	30.92	5.42	.64
1-5	244	1350	57.26	42.74	27.78	24.15	25.7	22.37	63.48	30.52	5.26	.74
1-6	167	1517	56.16	43.84	26.76	23.47	25.58	24.19	63.68	30.32	5.34	.66
1-7	104	1621	55.58	44.42	26.16	23.07	25.97	24.8	63.91	29.92	5.49	.68
1-8	76	1697	55.10	44.90	25.87	23.04	26.16	24.93	64.05	29.76	5.48	.71
1-9	45	1742	55.17	44.83	25.72	22.85	26.34	25.09	64.18	29.45	5.57	.80
1-10	33	1775	54.99	45.01	25.8	22.82	26.48	24.90	64.11	29.58	5.52	.79
1-11	21	1796	54.57	45.43	25.72	22.77	26.68	24.83	64.03	29.73	5.46	.78
1-12	15	1811	54.50	45.50	25.57	22.86	26.72	24.85	63.94	29.82	5.47	.77
1-22	134	1945	53.52	46.48	25.40	22.11	27.61	24.88	65.09	29.00	5.19	.72
全样本	1945	2020	54.11	45.89	25.59	22.57	27.33	24.51	65.05	28.67	5.59	.69

表 5 样本均衡收敛所需招募批次和构成比例^a

城市	批次	性别		批次	年龄组				工种				
		男 (%)	女 (%)		<26 (%)	26-35 (%)	36-45 (%)	>45 (%)	普工 (%)	技工 (%)	中低层 管理者 (%)	其他 (%)	
													批次
常州	3	43.2	56.8	5	30.9	18.6	31.6	18.9	1	81.5	16.9	1.5	0
武进	5	57.2	42.8	5	31.0	14.0	15.7	39.3	3	81.7	13.8	4.5	0
金华	2	75.0	25.0	3	34.0	28.6	16.2	21.3	2	64.9	26.7	7.5	.9
义乌	4	47.3	52.7	4	43.8	27.0	21.7	7.5	3	74.4	17.9	7.2	.5
泰安	6	65.4	34.6	17	30.1	22.3	27.1	20.5	5	51.9	44.5	3.6	0
肥城	6	43.4	56.6	8	8.1	26.2	31.5	34.3	— ^b				
咸阳	6	69.2	30.8	5	15.5	18.1	29.8	36.7	2	74.3	24.1	1.1	.5
兴平	6	55.8	44.2	4	19.3	26.3	27.5	26.9	3	74.2	20.8	5.0	0
岳阳	4	48.8	51.2	8	28.3	18.6	25.9	27.2	6	37.7	51.7	7.8	2.7
汨罗	4	59.9	40.1	4	14.7	25.4	37.6	22.4	2	52.9	34.2	12.5	.4
遵义	3	38.6	61.4	10	13.7	17.4	37.1	31.8	2	79.8	16.4	3.2	.6
凯里	4	55.2	44.8	3	48.2	30.5	15.0	6.3	2	57.5	38.2	4.3	0
广州	2	50.5	49.5	4	28.1	22.4	29.8	19.7	4	65.6	26.8	6.4	1.1
全样本	4	53.2	46.8	5	27.0	22.3	26.7	24.0	3	66.4	28.1	4.8	.7

注：a. 本表数据由 RDSAT 软件估计得来，收敛值为 2%，即构成比例随着批次增加，其变化范围在 2% 以内。b. 在肥城样本的“工种”特征中，“中低层管理者”子群只在自己群体内推荐被访者，所以无法使用这组数据对工种的均衡收敛所需批次进行估计。



注: +1 为完全同质, -1 为完全异质。

图 5 同质性指数

计总体构成的差异。招募样本构成与均衡样本构成的差异以加权平均绝对差异来反映,从表 6 中可知,性别、年龄和工种的加权平均绝对差异分别为 1.8%、0.3% 和 0.3%,三者均小于 2% 的收敛值。样本代表性检验显示,性别、年龄和工种三个指标中除了大于 45 岁组与普工组具有显著差异外($p < 0.001$),其他各组均没有显著差异($p > 0.05$)。也就是说,招募样本中大于 45 岁组和普工组的比例低于估计总体中的比例,招募人数过少,需要多招募一些被访者。

表 6 招募样本构成的 RDS 估计 N = 1945(不包含种子)

统计项	性别	
	男	女
招募样本分布	1041	904
招募样本构成(P_s)(%)	53.5	46.5
均衡样本构成(P_e)(%)	51.7	48.3
估计总体构成(\hat{P})(95% 置信区间)(%)	54.2 (45.3, 54.3)	45.8 (45.7, 54.7)

续表 1

统计项	性别			
	男		女	
P_s 与 P_e 的加权平均绝对差异 ^a (%)	1.8			
t 检验 ^b ($P_s - \hat{P}$)	t = -.452, p > .5		t = .422, p > .5	
统计项	年龄组			
	<26	26-35	36-45	>45
招募样本分布	494	430	537	484
招募样本构成(P_s)(%)	25.4	22.1	27.6	24.9
均衡样本构成(P_e)(%)	25.9	22.1	27.2	24.8
估计总体构成(\hat{P}) (95% 置信区间)(%)	23.0 (17.0, 23.2)	18.6 (13.8, 18.6)	26.3 (24.5, 32.1)	32.1 (31.1, 40.1)
P_s 与 P_e 的加权平均绝对差异(%)	.3			
t 检验($P_s - \hat{P}$)	t = 1.225, p > .2	t = 1.749, p > .05	t = .674, p > .5	t = -3.663, p < .001
统计项	工种			
	普工	技工	中低层管理者	其他
招募样本分布	1266	564	101	14
招募样本构成(P_s)(%)	65.1	29.0	5.2	.7
均衡样本构成(P_e)(%)	65.5	28.9	4.9	.7
估计总体构成(\hat{P}) (95% 置信区间)(%)	70.8 (67.9, 75.1)	25.2 (21.4, 28.3)	3.3 (2.1, 4.3)	.7 (.2, 1.1)
P_s 与 P_e 的加权平均绝对差异(%)	.3			
t 检验($P_s - \hat{P}$)	t = -4.255, p < .001	t = 1.989, p > .05	t = .860, p > .2	t = 0, p > .5

注:a. 加权平均绝对差异是招募的实际样本与均衡样本之间的差异,以各子群体的样本频次为权重。b. 该 t 检验的具体方法参见 Wang et al, 2005。

五、结论与讨论

综合以上分析,可以得出如下结论:

第一,本次 RDS 实施中,针对中国农民工的实际情况,我们采用了与以往不同的非固定点调查法。从实施的过程和结果看,这个方法有助于提高样本的代表性,识别调查对象身份的真实性,防止重复调查,同时缩短调查时间,提高调查效率。

第二,本文认为,对农民工来说,以工种取代行业作为评估指标较为恰当,一方面工种在不同类型企业中的分布较为均衡,另一方面工种反映了个体的工作能力和工作性质,具有较高的稳定性。

第三,此次 RDS 抽样共使用 75 个种子,平均每个城市 5.77 个,多数城市仅用 2-3 个种子就招募了大多数被访者,平均每个种子招募约 26 人,有近 1/3 的高产和中高产种子招募了 64% 的后续被访者,这些数据说明本次调查招募人数的集中度较高。整个样本的平均招募批次为 5 批,13 个城市中有 7 个城市的平均招募批次在 4 层及以上,实际招募人数的增长趋势较以往其他 RDS 抽样要好。

第四,从性别、年龄和工种三个指标构成随着招募批次逐渐增加发生变化的趋势看,各个指标构成在 1-5 批次内都发生了较大变化,随后便逐渐稳定下来。均衡所需批次估计表明,整个样本要达到收敛,性别、年龄和工种分别仅需 4 批、5 批和 3 批。同质性指数显示,性别、年龄和工种存在不同程度的同质性,但没有影响样本的均衡收敛。

第五,性别、年龄和工种的加权平均绝对差异值分别为 1.8%、0.3% 和 0.3%,三者均小于 2% 的收敛值。样本代表性检验显示,招募样本与估计总体之间在性别构成上没有显著差异;在年龄和工种中除大于 45 岁组与普工组具有显著差异外,其他各组均没有显著差异。

以上分析结果表明,本次对农民工的 RDS 抽样所获样本达到了均衡,且在较短批次内快速收敛,所获样本具有较好的代表性。此外,与 2013 年农民工监测报告中的部分指标相比较,在年龄、受教育程度、从事产业、技能培训和合同签订等方面具有较高的一致性,这也在很大程度上说明了 RDS 在农民工抽样中的有效性(指标对比参见附录)。

当然,本次调查也存在一些问题,如个别城市种子偏多、推荐链条

不够长等。究其原因,RDS 抽样对调查员的责任感、人际交往经验、语言能力和时间投入都要求较高,因而对调查员采取何种激励方式需要进一步探索,对调查实施和监督都需要进行有效管理。

长期以来农民工调查始终苦于没有一个完整的抽样框。风笑天(2003)对发表于《中国社会科学》和《社会学研究》上的 141 篇实证研究报告的解析和朱磊(2014)对近年几项较有代表性的农民工抽样调查的分析都指出,多数研究在抽样方法上均存在这样那样的问题。这在很大程度上说明,要获取一个完整的抽样框,给每一个目标群体成员均等的入样概率非常困难,正如莱文和福克斯所言:“随机抽样的这个特性要求总体的每一个成员在抽样前必须被确定,通常是通过获取一个列出了总体的每一个成员的清单来满足这个要求。我们稍稍一想就会发现获得这样一个总体的完整清单不会是一个容易的任务,尤其是在研究一个巨大而复杂的总体时”(莱文、福克斯,2008:168)。很明显,在社会调查研究中,能获得完整的抽样框并不多见,缺乏完整的抽样框,甚至没有抽样框则较为普遍。进一步说,从事过抽样调查的研究人员都清楚,即使有完整的抽样框并进行随机抽样,在实际执行中也会因为空户、拒访和调查员作弊等各种原因使随机抽样在操作中变形并引起抽样偏差。由于对高流动性的农民工进行严格的随机抽样难度太大,许多学者会在农民工调查中采取偶遇、滚雪球、配额和判断等非随机抽样方法不难理解。因此,在缺乏抽样框的条件下,寻找切实可行又能减少抽样偏差的方法才是科学务实的做法。自 1997 年以来,赫克索恩等人不仅从数学理论上证明了 RDS 的合理性,在实际应用中不断评估和完善 RDS,而且把 RDS 的应用从隐藏群体扩展到了非隐藏群体。本研究表明,RDS 为我们走出农民工问卷调查抽样缺乏抽样框的困局开辟了新的路径。

“不管抽样计划设计得多么完美,研究者有多么认真仔细,不发生任何作弊或是错误,抽样误差总会存在”(莱文、福克斯,2008:171)。RDS 抽样虽然不是对目标群体的完全随机抽样,但可以在努力改进和严格执行抽样程序的基础上获得一个渐进无偏的样本,并对样本偏差做出估计。无疑,RDS 为农民工及其他缺乏抽样框的群体抽样提供了一个获得更具代表性的样本的替代性方法。限于篇幅,我们将在后续研究中进一步总结本次调查中的经验和不足,为以后的 RDS 抽样实践提供借鉴。

附录:本次调查结果与 2013 年农民工监测报告部分指标对比

	2013 年报告	本次调查	2013 年报告		本次调查
80 后 (%)	46.6	46.31	新生代农民工受教育程度 (%)		
从事第二产业 (%)	56.80	55.93	初中以下	6.10	6.75
接受过技能培训 (%)	32.70	31.94	初中	60.60	50.96
劳动合同签订率 (%)	41.30	44.44	高中	20.50	20.34
日工作超 8 小时 (%)	41	55.07	职高	—	2.57
被拖欠工资 (%)	.80	7.65	中专	—	11.46
养老保险 (%)	15.7	24.35	技校	—	3.00
工伤保险 (%)	28.5	31.1	大专及以上学历	12.80	4.93
医疗保险 (%)	17.6	26.36	老一代农民工受教育程度 (%)		
失业保险 (%)	9.1	17.14	初中以下	24.70	32.59
生育保险 (%)	6.6	14.76	初中	61.20	50.88
人均月收入 (元)	2609	2864.52	高中	12.30	12.83
2014 年人均月收入 (元)	2864	2864.52	职高	—	.65
人均月生活消费支出 (元)	892	1190.86	中专	—	1.66
租房居住比重 (%)	36.70	41.84	技校	—	.28
平均每月工作时间 (天)	25.2	27.1	大专及以上学历	1.80	1.11
平均每天工作时间 (小时)	8.8	9.29	—	—	—

注:本表数据除“2014 年人均月收入 (元)”一项外,均来源于国家统计局 (2014),“2014 年人均月收入 (元)”的数据来自国家统计局 (2015)。

参考文献:

- 艾尔·巴比,2005,《社会研究方法》,邱泽奇译,北京:华夏出版社。
- 风笑天,2003,《结果呈现与方法运用——141 项调查研究报告的解析》,《社会学研究》第 2 期。
- 国家统计局,2014,《2013 年全国农民工监测调查报告》(http://www.stats.gov.cn/tjsj/zxfb/201405/t20140512_551585.html)。
- ,2015,《2014 年国民经济在新常态下平稳运行》(http://www.stats.gov.cn/tjsj/zxfb/201501/t20150120_671037.html)。
- 莱文,杰克·詹姆斯·艾伦·福克斯,2008,《社会研究中的基础统计学 (第九版)》,王卫东译,北京:中国人民大学出版社。
- 李树茁,任义科、杨绪松,2006,《中国农民工的整体社会网络特征分析》,《中国人口科学》第 3 期。

- 梁玉成,2013,《在广州的非洲裔移民行为的因果机制——累积因果视野下的移民行为研究》,《社会学研究》第1期。
- 刘石柱、冯成玉、阮玉华、巫俊林、周枫、陈康林、何益新、邵一鸣,2005,《应答驱动抽样方法在吸毒人群招募中的应用》,《中国公共卫生》第11期。
- 唐纳德·特雷曼、陆瑶、齐亚强,2012,《人口数据收集的新方法》,梁在主编《人口学》,北京:中国人民大学出版社。
- 许娟、张洪波、郑迎军、王君、朱军礼、李照荣、朱义彬、胡中旺、张晓鹏,2007,《男男性行为者HIV自愿咨询检测需求与利用》,《中国公共卫生》第9期。
- 张春泥、刘林平,2008,《网络的差异性和求职效果》,《社会学研究》第4期。
- 张文宏、雷开春,2008,《城市新移民社会融合的结构、现状与影响因素分析》,《社会学研究》第5期。
- ,2009,《城市新移民社会认同的结构模型》,《社会学研究》第4期。
- 赵延东,2007,《受访者推动抽样:研究隐藏人口的方法与实践》,《社会》第2期。
- 朱磊,2014,《流入地抽样抑或流出地抽样?——对当前农民工研究中抽样方法的评析》,《青年研究》第1期。
- Abdul-Quader, Abu S. & Douglas D. Heckathorn 2006, "Effectiveness of Respondent-Driven Sampling for Recruiting Drug Users in New York City: Findings from a Pilot Study." *Journal of Urban Health: Bulletin of the New York Academy of Medicine* 83(3).
- Bernhardt A., Spiller M. W. & D. Polson 2013, "All Work and No Pay: Violations of Employment and Labor Laws in Chicago, Los Angeles and New York City." *Social Forces* 91(3).
- Berg, S. 1988, "Snowball Sampling." In S. Kotz & N. L. Johnson (eds.), *Encyclopedia of Statistical Sciences*. New York: Wiley.
- Brewer, K. R. W. & Muhammad Hanif 1983, *Sampling with Unequal Probability*. New York: Springer-Verlag.
- Cochran, W. G. 1977, *Sampling Techniques*. New York: Wiley.
- Evans, A. R., G. J. Hart, R. Mole, C. H. Mercer, V. Parutis, C. J. Gerry, J. Imrie & F. M. Burns 2011, "Central and East European Migrant Men Who Have Sex with Men in London: A Comparison of Recruitment Methods." *Sexually Transmitted Infections* 87(4).
- Gile, Krista J. & Mark S. Handcock 2010, "Respondent-Driven Sampling: An Assessment of Current Methodology." *Sociol Methodol* 40(1).
- Heckathorn, Douglas D. 1997, "Respondent-driven Sampling: A New Approach to the Study of Hidden Populations." *Social Problems* 44(2).
- 2002, "Respondent-driven Sampling II: Deriving Valid Population Estimates from Chain-referral Samples of Hidden Populations." *Social Problems* 49(1).
- Heckathorn, Douglas D. & Joan Jeffri 2001, "Finding the Beat: Using Respondent-driven Sampling to Study Jazz Musicians." *Poetics* 28(4).
- Johnston, L. G., K. Sabin, M. T. Hien & P. T. Huong 2006, "Assessment of Respondent Driven Sampling for Recruiting Female Sex Workers in Two Vietnamese Cities: Reaching the Unseen Sex Worker." *Journal of Urban Health* 83(1).

- Khamsiriwatchara, A. , P. Wangroongsarb, J. Thwing, J. Eliades, W. Satimai, C. Delacollette & J. Kaewkungwal 2011, "Respondent-driven Sampling on the Thailand-Cambodia Border. I. Can Malaria Cases be Contained in Mobile Migrant Workers?" *Malaria Journal* 10(1).
- King, W. D. , S. Larkins, C. Hucks-Ortiz, P. C. Wang, P. M. Gorbach, R. Veniegas & S. Shoptaw 2009, "Factors Associated with HIV Viral Load in a Respondent Driven Sample in Los Angeles." *AIDS and Behavior* 13(1).
- Qiu, Peiyuan, Yang Yang, Xiao Ma, Fang Wu, Ping Yuan, Qiaolan Liu & Eric Caine 2012, "Respondent-driven Sampling to Recruit in-country Migrant Workers in China: A Methodological Assessment." *Scandinavian Journal of Public Health* 40(1).
- Salganik, Matthew J. & Douglas D. Heckathorn 2004, "Sampling and Estimation in Hidden Populations Using Respondent-Driven Sampling." *Sociological Methodology* 34(1).
- Suhov, Y. & M. Kelbert 2008, *Probability and Statistics by Example; Volume 2, Markov Chains: A Primer in Random Processes and Their Applications*. New York: Cambridge University Press.
- Volz, Erik & Douglas D. Heckathorn 2008, "Probability Based Estimation Theory for Respondent Driven Sampling." *Journal of Official Statistics* 24(1).
- Walters, K. L. , P. F. Horwath & J. M. Simoni 2001, "Sexual Orientation Bias Experiences and Service Needs of Gay, Lesbian, Bisexual, Transgendered, and Two-spirited American Indians." *Journal of Gay & Lesbian Social Services* 13(1-2).
- Wang, Jichuan, Robert G. Carlson, Russel S. Falck, Harvey A. Siegal, Ahmed Rahman & Linna Li 2005, "Respondent-driven Sampling to Recruit MDMA Users: A Methodological Assessment." *Drug and Alcohol Dependence* 78 (2).
- Wejnert, Cyprian 2009, "An Empirical Test Of Respondent-Driven Sampling: Point Estimates, Variance, Degree Measures, and Out-Of-Equilibrium Data." *Sociological methodology* 39(1).
- Wejnert, Cyprian & Douglas D. Heckathorn 2008, "Web-Based Network Sampling: Efficiency and Efficacy of Respondent-Driven Sampling for Online Research." *Sociological Methods and Research* 37(1).

作者单位:南京大学社会学院
责任编辑:杨 可

regions in China have transformed from drawing-type government to suspended-type government. Meanwhile, a complicated fiscal transfer payment system has been set up to keep local government running and provide public goods and services. With the downward flowing of transfer payments from central government to grass-roots governments, program funds are becoming increasingly important in the transfers, which is called as “governing the state through programs”. In the era of “governing the state through programs”, township government faces a dilemma: With the upward centralization of resources and power, township government plays an indispensable role in implementing the programs in the villages. Thus township-government has changed from suspended-type to assisted-type. However, with lots of affairs but short of funds and power, the township government changes from technical governing to non-technical governing.

“Transient Association”: Africans’ social organizations in Guangzhou
..... *Niu Dong* 124

Abstract: Africans in Guangzhou have formed a kind of social organizations, which usually bases on common nationality, centers on leaders, provides internal mutual aid and guidance, and represents their people in interacting with other organizations or groups. The emergence of those associations is actually a response to Africans’ general sense of sharing, national identity strengthened when they came into Guangzhou and the stressful situation Guangzhou authority has constructed. Such associations are characterized by its single-level development, informal operation, imperfect functions and unstable membership, which make them differ from the ones Africans formed in immigrant countries. In order to reveal the temporality, mobility and uncertainty of Africans’ existence in Guangzhou, this paper argues that “transient association” can be used to describe the social organizations. Moreover, the paper further discusses the tension brought by the “transient” attribute of Africans and their associations in Guangzhou to both the popular analytical paradigm of migration studies and present governance of foreigners in China.

Application of Respondent-Driven Sampling in the Survey of Chinese Rural-to-urban Migrant Workers: Practice and evaluation
..... *Liu Linping, Fan Changyu & Wang Ya* 149

Abstract:As there has been a long-term difficulty that researchers do not have proper sampling frame in the survey of Chinese rural-to-urban migrant workers, this study adopted the Respondent-Driven Sampling (RDS) and conducted a large-scale questionnaire survey in 13 cities (districts), 7 provinces in China. Based on the

Chinese rural-to-urban migrant workers' real situation, this study made use of the unfixed-sites survey and improved the original method. This study is also an evaluation of the application of RDS used in the survey of migrant workers. The results show that samples obtained by the RDS in this study reached equilibrium and converged rapidly in a few of recruiting batches. Thus, samples in this study have good representativeness. This paper argues that although RDS is not a completely random sampling method for the target groups, researchers can obtain an asymptotic unbiased sample through improvement and strict sampling procedure, and the bias of the sample can also be estimated. RDS can be used as an effective alternative approach for the survey of migrant workers and other groups which do not have a clear sampling frame.

Patterns of Divinity-Humanity Relations and Construction of Believing Model;Based on survey data of Yangtze River Delta in China

..... *Li Xiangping & Li Feng* 174

Abstract:There are two opposing perspectives in sociology of religion in China; The one follows the pattern of western traditional structuralism, the other refuses it and emphasizes the particularity of Chinese belief. This leads to the binary opposition between Chinese belief and western religion, diffused religion and institutionalized religion. Drawing on data from the 2012 survey “Belief and Religion in the Yangtze River Delta”, this paper represents the status and role of the god-person relationship in shaping belief patterns in China from the perspective of relationalism methodologically. The results show that; (1) the polytheists are more identified with the reciprocal exchange relationship between god and man than the monotheists; (2) the reciprocal exchange relationship between god and man forms a private religious experience; (3) the monotheists' higher tendency of group religious experience is due to their lower inclination of reciprocal exchange relationship between god and man. The Chinese belief models and western religion structures are not fundamentally different. The construction of the relationship between god and man is the core of them, which is also the common platform to study religion sociologically both in China and western society.

The Relative Risk Aversion or the Credential Decline Aversion? A Reconsideration of the Generation Mechanism of China's Educational Inequality (1978 – 2006)

..... *Hou Liming* 192

Abstract:During the social transition period of China, how do the family educational strategies affect children's educational attainment? Based on the rational choice theory,