

中国家庭追踪调查： 理念与实践

社会
2014·2
CJS
第34卷

谢宇 胡婧炜 张春泥

摘要：为了让研究者对中国家庭追踪调查(CFPS)有更多了解，本文从四个方面介绍其背景和特色。在设计理念上，CFPS以个人和家庭为研究主体，以多层次问卷和跟踪性调查为研究视角，充分考虑了社会现象的异质性、复杂性和时间性；在调查实施上，CFPS采用城乡一体的多阶段、内隐分层和与人口规模成比例的抽样方法保证样本的代表性，借助先进的调查技术提高访问质量，设计了兼顾研究需要和可操作性的追访原则；在问卷内容上，CFPS不仅涵盖内容广，而且在整合城乡问卷，收集家庭关系、人口流动、事件史、认知能力和少儿发展等相关数据上也均有长处；在数据上，本文以经济不平等、婚姻变迁、认知能力三个方面的数据研究为例，展示了CFPS的追踪设计和特色内容为社会科学研究带来的潜力。

关键词：中国家庭追踪调查 社会调查 CFPS

The China Family Panel Studies: Design and Practice

XIE Yu HU Jingwei ZHANG Chunni

Abstract: The China Family Panel Studies (CFPS), launched by Peking University, is a nationwide, comprehensive, longitudinal social survey. The project aims to document historically unprecedented social changes that are currently taking place in China in different domains by repeatedly collecting information from a sample of individuals, households, and communities over an extended period. In order to help researchers better understand the CFPS project

* 作者1:谢宇 CFPS项目负责人,美国密歇根大学社会学系,北京大学社会研究中心(Author 1: XIE Yu, PI of CFPS; Department of Sociology, University of Michigan; Center for Social Research, Peking University) E-mail: yuxie@umich.edu;作者2:胡婧炜 北京大学社会研究中心(Author 2: HU Jingwei, Center for Social Research, Peking University);作者3:张春泥 北京大学社会学系(Author 3: ZHANG Chunni, Department of Sociology, Peking University)

本文在数据统计过程中得到了北京大学中国社会科学调查中心孙妍、戴利红和吴琼的帮助,特此感谢。文责自负。

and its data, this article describes the background and characteristics of the CFPS in four aspects. In research design, the CFPS adopts multiple-level questionnaires and a panel design to track changes in individuals and households so as to allow researchers to study heterogeneity, embeddedness, complexity, and time-dependency of social phenomena. In implementation, it uses multi-stage, implicit stratification, and probability proportion to size sampling methods with a sampling frame that integrates rural and urban populations to obtain a nationally representative sample. To assure data quality, the CFPS uses advanced computer-assisted personal interviewing (CAPI) techniques in its fieldwork. By now, the 2010 baseline survey, the 2011 small-scale followup survey for maintenance, and the 2012 full-scale followup survey have been completed. All follow-up strategies have met many research needs but remained practical. In contents, the CFPS learned from the methods and experiences from the most influential survey projects in the world. The questionnaires not only cover a wide range of topics but also consist of intergraded modules for rural and urban interviews and gathering information of family structure and family members, migrant mobility, event history (e. g., history of marriage, education, and employment), cognitive ability, and child development. Finally, we present preliminary findings about income inequality and poverty, marital events and cohabitation, and cognitive ability based on the 2010 and 2012 CFPS data, as demonstrations of the CFPS's potentials for social science, owing to its strengths in research design and topical contents.

Keywords: China Family Panel Studies, social survey, CFPS

“社会学的重要贡献之一，就在于它提供信息的能力。”

——Stanley Lieberman(1992:2)

从 20 世纪 80 年代至今,中国社会发生了翻天覆地的变化。一方面,经济增长、教育扩张与人口转型等成就斐然(Xie,2011);另一方面,贫富差距扩大、教育不平等加剧、传统婚姻家庭观念日渐瓦解等社会问题也接踵而至(谢宇、胡婧炜,2013)。中国正经历的这场变革根植于其独有的政治、经济、文化与社会环境,过去没有,将来也不可能在其他社会情境中重现,其意义堪比 14 世纪的意大利文艺复兴、16 世纪的德国宗教改革和 18 世纪的英国工业革命等重大历史转折性事件(Xie,

2011)。它为处于这个时代的社会科学研究者提供了一次千载难逢的观察、记录、解读社会发展和历史演进的机遇。然而,面对多元、复杂和充满异质性的中国社会,即使是经验丰富的社会科学工作者,他们的了解和认识可能也只一知半解。而且,个体认知不可避免地受到个人经历与价值观念的影响,即便是学者也可能对社会现象轻易做出带主观偏见的评判或受意识形态影响的推断。那么,人们何以全面、客观地了解身边所发生的变化?何以清晰、准确地分析和解释引起这些变化的机制?构筑在经验证据基础上的实证研究变得尤为重要。

借助调查数据研究社会现象的做法在中国已被越来越多的社会科学研究者认同和采纳。自 20 世纪 80 年代末起,许多旨在了解中国社会各方各面的调查项目应运而生,如中国家庭收入项目(CHIP)¹、中国综合社会调查(CGSS)²等。这些调查项目为研究特定时期的中国社会提供了重要依据。然而,由于局限于单一时点的截面(cross-sectional)设计、调查对象层次的单一化和调查内容的有限性,这些调查项目不能满足对变迁本身和变迁机制研究的需要。高水准的实证研究更需要高质量的跟踪数据支持(任强、谢宇,2011)。

北京大学的中国家庭追踪调查(China Family Panel Studies,以下简称 CFPS)³作为一项全国性、综合性的社会跟踪调查项目,它的开展正是以采集动态经验数据、记录当前中国的社会变化为目的,意在为现在及将来的研究者提供了解和研究当今中国最为翔实、客观的一手数据。CFPS 长期跟踪收集个体、家庭、社区三个层次的数据,其调查内容包括村/居概况、家庭关系、人口流动、家庭经济、居住与设施、工作与收入、教育、婚姻、健康、态度观念、认知能力和社会交往等诸多主题。自 2010 年正式实施基线调查起,迄今为止 CFPS 已采集三期数据,具备了开展跟踪分析的数据基础,并开始为学术研究和公共政策提供数据支持。⁴

为了让研究者对该项目的背景和特色有更多了解,本文将从设计理念、调查实施过程、内容特色、数据结果四个方面对 CFPS 展开全面介绍。

1. CHIP 始于 1988 年,最近两期分别在 2002 年和 2007 年实施,详尽介绍参见:李实等(2013)。

2. CGSS 始于 2003 年,该项目的详尽介绍参见:Bian and Li(2012)。

3. CFPS 的中文曾用名“中国家庭动态跟踪调查”。该项目由北京大学 985 项目资助,北京大学研究团队设计,北京大学中国社会科学调查中心实施。

4. 项目介绍与数据下载网站:<http://www.iss.edu.cn/cfps/>。

一、设计理念

CFPS的设计理念主要是:同时将个人和家庭作为研究主体;以多层次的、动态的视角收集数据。前者依据的是个体变异性在社会科学研究中的重要性,以及家庭在中国社会结构中的基础性地位;后者依据的是社会系统的复杂性和社会现象的时间性。

(一) 研究主体:个人与家庭

社会科学研究的本质内容是变异性(谢宇,2012)。同一总体中的不同个体虽然具有一定共性,但他们在具体特征上仍存在实质性差异。对这种差异的关注成为社会科学有别于自然科学的独特之处。个体变异性不仅表现在那些由基因决定的、与生俱来的差异上,同时也表现为个体在生命历程中受社会环境和特定个人经历等因素影响而产生的变化。变异性给研究人类社会的因果关系带来了复杂性和不确定性:个体在受到特定社会影响前具有差异性,对接受社会影响具有选择性,对社会影响的反应同样有差异性(Xie,2013)。这些差异随时间发生的变化,均会对个体的社会结果产生影响。正因为这些个体差异的存在,我们只能得到对总体平均有意义的统计规律,但这些规律对具体个体并不一定适用(谢宇,2012),因此,即便是对宏观现象和社会总体的研究也不能不关注微观个体。也正因为个体差异的存在,我们不能把不同的个体单元等同看待,更不能忽视个体单元随时间发生的不同变化。

尽管CFPS以社会变迁这一宏大议题为出发点和研究目的,其研究设计和采集数据的落脚点仍是具体的个人。如前所述,个人是反映人类社会变异性的最基本单元,许多社会现象,如社会不平等、人口素质、生活质量、社会心理和文化等,最终都会具体化为个人层次的收入、职业、教育、健康、幸福感及态度等;社会的变迁,如代际流动、人口结构的改变、家庭制度的演变等,最终也可以通过个人层次的变化体现。所以,要了解社会,就必须先了解社会中不同个人在角色、地位、行为和态度等方面的差异。正因如此,CFPS最基本、最重要的调查和研究单元是社会中的个人,其长期跟踪的对象也是个人。

中国社会具有多层次结构,这体现为个人嵌套在层层的社会组织(如工作单位、村庄、家庭等)中。在这一嵌套结构中,家庭是最基础和最直接影响个人角色、地位、行为和态度等的社会组织。

首先,家庭构成了个人社会化最初始也是最重要的环境。家庭赋予个人最初始的社会地位;个人从出生起在相当长的一段成长过程中要通过家庭来接触、学习社会规则(Duncan, *et al.*, 1998; McLanahan and Percheski, 2008);家庭环境对个人的态度、行为、期望也有持续性的影响(Lareau, 2011; Thornton and Camburn, 1987)。因此,要研究个人,必须要了解个人所处的家庭。

其次,家庭是代际关系的重要桥梁(Blau and Duncan, 1967; Sewell and Hauser, 1975)。对代际流动和代际关系的研究,例如父母社会经济地位对子女地位获得的影响、家庭资源在不同子女间的分配方式、成年子女和父母之间资源的交换与流向等,均需要建立在在对家庭结构和代际关系清晰、全面认识的基础上,也需要了解家庭成员的详细信息。

第三,家庭是研究婚姻与性别机制的平台。中国绝大多数成年男女都会结婚(Yu and Xie, 2013)。来自不同家庭的男女通过婚姻形式组成新的家庭,资源在通过婚姻形式结成家庭的过程中被重新分配和组合(Kalmijn, 1998)。男女在劳动分工和社会经济地位等方面的性别差异也会在婚姻和家庭中得到体现(Yu and Xie, 2012)。

第四,中国人常以家庭为单位从事经济活动和进行社会交往。中国人生活中的一些重要方面,如经济生活、居住、抚育子女和赡养老人等,均在家庭中进行。要关注中国社会,我们就必须从家庭层面对这些相关的主题开展调查。

最后,家庭在中国文化中具有非同寻常的意义。中国人崇敬祖先,重视孝道,婚姻讲究门当户对和传宗接代,事业追求光宗耀祖,这些传统价值观念都体现出家庭及家族在人们生活中的重要地位(Chu and Yu, 2010; Whyte, 2004)。虽然传统的家庭观念现在正逐渐被侵蚀,但从中国父母对子女的投入、家庭或家族关系网络对家庭成员的影响,以及家庭内部资源的转移等方面仍可看到家庭在中国文化中的重要性。

综上所述,对中国社会的了解与研究不可能脱离对家庭这一社会细胞的研究,而家庭也正是 CFPS 一个重要的研究和调查单元。CFPS 采用了可明确定位家庭成员间关系、描述家庭结构网络的方法,对家庭关系和家庭成员信息展开全方位的深度调查;同时,它也详细采集了家庭层次经济社会生活的各方面数据。家庭数据的收集丰富了经验资料的层次,为研究者分析中国社会提供了更多有价值的信息。

（二）研究视角：社会系统的复杂性与社会现象的时间性

社会是一个庞大而复杂的动态系统，它的复杂性首先体现为社会现象的多层次性，且不同层次紧密相扣，不可割裂。例如，宏观的国家政策和经济环境会影响微观的个人收入，而微观的个人行为的集合也会引起宏观变化。社会系统的复杂性还体现为社会现象的多维度性，不同维度紧密联系和互相影响。例如，个人的职业会影响到其收入和社会关系网络，进而又会影响到其健康和生活满意度等。

针对社会系统的复杂性，CFPS 开展了多层次和多维度的全方位调查。社区、家庭和個人是 CFPS 调查的三个重要层次。在每个层次上，CFPS 均从多个维度采集信息：在社区层次调查社区的政治环境、村/居面貌、基础设施、人口、资源、交通、医疗卫生和财政收支等；在家庭层次调查家庭的结构与成员关系、生活条件、社会交往、收入支出和资产状况等；在个人层次调查个人的教育、职业、收入、婚姻、心理与生理状况、观念与态度等。通过这一设计，调查对象不再是孤立的，个人、家庭和社区三者可以在数据上关联起来。

社会现象具有时间惯性、时间持续性和时间不对称性。过去的事件影响现在的行为，现在的经历影响将来的社会结果。许多社会现象都是时间积累的结果。社会现象在时间上的这些特性加深了社会系统的复杂性，提高了社会研究的难度。时间性是研究社会现象与社会变化的一个重要概念。从方法论角度讲，时间是一种信息，社会科学研究需要了解什么事情先发生和先被观察到，什么事情后发生和后被观察到（谢宇，2012）。跟踪调查关注社会现象的动态趋势与个体随时间的变化，是研究社会现象时间性的最有效途径。它通过在不同时点上对同一人群（同一样本）的重复观察，能够掌握具体个体在不同时点的状态，对研究总体异质性、因果机制以及状态变化等社会科学研究的重点课题有极高的价值（任强、谢宇，2011）。跟踪调查虽然成本昂贵、设计复杂和操作困难，但与截面调查和趋势调查相比，它能够采集更多有价值的信息，为科学研究带来更大回报。正因为如此，CFPS 在设计之初便决定以跟踪调查的方式采集固定调查对象在不同时点上的数据，对全部基因成员⁵展开长期跟踪调查。

5. CFPS 将 2010 年基线调查界定出来的所有家庭成员及其今后的新生血缘/领养子女视为基因成员。

二、调查实施

CFPS项目由北京大学中国社会科学调查中心(以下简称“调查中心”)实施。基线调查于2010年4月开始,2011年2月结束。2011年7月到2012年2月,调查中心进行了以样本维护和测试调查为目的的小规模追访。首轮全部样本的追踪调查在2012年7月到2013年3月进行。CFPS调查的全部实施过程,从抽样到实地访问,从初访到追访,均由调查中心团队独立完成。下面我们将介绍数据使用者最关心的抽样设计、历次实地访问情况和追访原则。

(一) 抽样设计

CFPS调查的目标总体为中国25个省、直辖市、自治区(不含香港、澳门、台湾、新疆、西藏、青海、内蒙古、宁夏、海南)中的家庭户和家庭户中的所有家庭成员。居住在传统居民住宅内,且家中至少有一人拥有中国国籍的独立经济单元,便可视为一个满足项目访问条件的家庭户。⁶CFPS定义的家庭成员包含家庭户中经济上联系在一起的全部直系亲属,⁷和经济上联系在一起且连续居住时间满3个月的全部非直系亲属。由于CFPS样本覆盖的25个省、直辖市、自治区的人口约占全国总人口(不含港澳台地区)的95%,因此,可被视为一个全国代表性样本。

CFPS分6个独立子抽样框抽取样本,相应得到6个子总体。其中,上海、辽宁、河南、甘肃和广东5省市构成了5个独立子样本框(称为“大省”)。CFPS对“大省”均以过度抽样的方式抽取样本,因此其样本具有地区的独立代表性,可以用于省级推断和地区比较。每个“大省”目标样本规模为1600户。另一个独立子样本框由其他20个省、直辖市、自治区共同构成(称为“小省”),目标样本规模为8000户。因而CFPS总的样本规模为16000户。5个“大省”经二次抽样后获取的样本与“小省”样本框中的样本共同构成具全国代表性的总样本。⁸

在抽样方法上,考虑到中国巨大的地区差异,为节省调查成本,提

6. 最初我们还要求受访家庭户中至少有一名成员在抽样社区居住时间满6个月,但在执行过程中,这一条件被取消,实际被这一条件过滤掉的仅有极少数家户。

7. 关于直系亲属的界定参见:孙妍等(2011)。

8. 该样本也称再抽样样本或整合样本。5个“大省”未经二次抽样的样本与“小省”样本合并后的总样本称为完全样本。CFPS数据提供了识别再抽样样本的变量。

高抽样的代表性, CFPS 采用多阶段、内隐分层(implicit stratification)和与人口规模成比例(probability proportional to size, PPS)的系统概率抽样方式。抽样分三个阶段进行:抽取行政性区/县、抽取行政性村/居委会和抽取家庭户。⁹前两个阶段的抽样框使用了根据官方行政区划资料制成的区/县名录和村/居名录,第三个阶段的末端抽样框使用了通过实地绘制地图获得的家户住址。CFPS 分层抽样的主要变量是行政区划和社会经济水平(socioeconomic status, SES)。在同级行政层,地方人均 GDP 是社会经济水平的主要排序指标;在无法获得 GDP 指标的条件下,则采用非农人口比例或人口密度作为替代指标。在前两个阶段的每一抽样框中,根据这些指标对行政区/县或村/居排序后,从随机起点以等距方式抽取区/县或村/居样本,按与人口规模成比例的概率进行内隐分层抽样。例如,在“小省”子总体样本框中,将同省的区/县排列在一起能保证每个省内区/县被抽中的概率与该省人口规模大约成比例。同时,我们对省又进行了社会经济水平的排序,使得在两省交界处的样本单元具有相对相似性,因此,通过随机方式多抽或少抽一个样本单元(即区/县)都不会影响到抽样的效率。也就是说,在随机抽样已经能保证总体代表性的基础上,对社会经济水平这一辅助信息的利用又提高了样本代表总体的效率。在末端抽样时,由于事先没有家户社会经济水平指标的数据,我们就借助地理信息,按照社区内的行走路线对家庭户进行排序。由于住宅的地理位置通常含有社会经济水平的信息,这种排序方式对提高样本效率依然有效。

CFPS 抽样设计的另一特点是打破了农村与城市分开抽样的传统,采用城乡一体化的抽样方式。这样设计主要是因为官方对农村与城市的行政划分已难以及时反映中国快速城市化的现实。因此,CFPS 样本的城乡属性不再是依据事先由城乡行政划分的抽样框来鉴别,而是依据社区、家庭以及个人问卷中关于村居属性、农业与非农活动、户口等多个变量来判断。

为了评估 CFPS 样本对总体人口的代表性,我们将基线调查中基

9. 上海因不同于其他“大省”,样本的抽取略有不同。关于抽样的具体设计可参见:谢宇、邱泽奇、吕萍. 2012. 中国家庭追踪调查抽样设计,北京大学中国社会科学调查中心《中国家庭追踪调查技术报告系列(CFPS-1)》,网址:<http://www.issu.edu.cn/cfps/d/file/wd/jsbg/2010jsbg/c70c703752f7f6f24e27b5a8a85c1cba.pdf>。

因成员的性别—年龄加权前分布与全国第六次人口普查(以下简称“六普”)的性别—年龄分布作比较。图 1 中的性别—年龄金字塔从 0 到 100 岁,每 5 岁为一组,分别统计了男性和女性人数占总人数的比例。其中,a 是基于 CFPS 2010 年家庭关系数据库的再抽样数据,b 是基于“六普”短表的汇总数据。我们发现,CFPS 家庭成员的性别—年龄结构与“六普”的人口结构分布直观上相当吻合。¹⁰

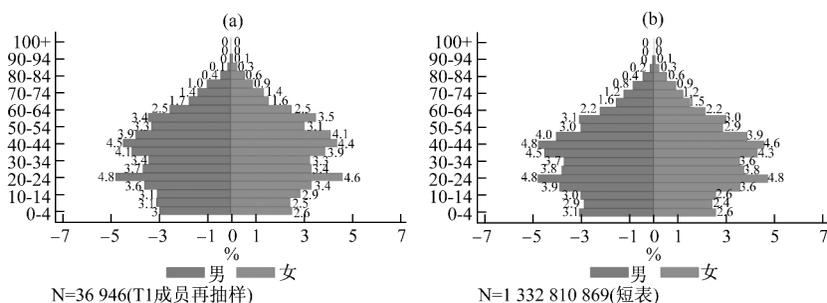


图 1:CFPS 基线调查和第六次人口普查的年龄—性别结构

CFPS 2010 对全国完全样本和全国再抽样样本分别计算了家庭问卷、成人问卷和少儿问卷三个数据库的权数。¹¹权数的计算包括抽样设计权数、无应答调整权数、事后分层调整权数的计算以及对权数的极值调整。其中,抽样设计权数为三个阶段抽样概率的乘积的倒数,并对再抽样样本考虑了从第一阶段所抽取的样本区/县中再抽取再抽样样本区/县的概率。无应答调整权数的计算在家庭层面上采用了加权组调整的方法,综合考虑了家庭成员问卷完成的数量占村/居样本中所有家庭样本数量的比例,以及完成家庭成员问卷的家户中完成家庭问卷的家户数与需作答家庭问卷的家户数的比例;在个人层面上则采用了基于 Logistic 模型的联系层次上(相对于无联系样本)的应答倾向概率和拒访层次上(相对于非拒访样本)的应答倾向概率作为调整系数。事后分层调整主要在个人问卷层面进行,以性别、年龄和城乡作为主要分层

10. 我们计算了 CFPS 各性别年龄组与“六普”数据相比的抽样偏误率,详细内容参见:谢宇, 2012. 中国家庭追踪调查(2010)用户手册(第二版),北京大学中国社会科学调查中心 (<http://www.iss.edu.cn/cfps/wd/jsbg/2010jsbg>)。

11. 全国完全样本的权数为 5 个“大省”和 1 个“小省”共 6 个子总体的全部样本的权数的合并,全国再抽样样本的权数为 5 个“大省”经再抽样后的样本的权数与“小省”样本的权数的合并。

变量。极值调整主要是将权数控制在一定的范围内,其目的是控制权数方差,保证估计效率。¹² 追访数据将同样提供家庭和个人问卷的权数,但个人问卷的权数仅针对基因成员,在计算时将在2010年个人权数的基础上考虑其在调查当年被成功追访到的概率。

(二) 实地访问

CFPS 实地访问借助了 CAPI(计算机辅助面访调查)、CATI(计算机辅助电访调查)和 CAWI(计算机辅助网访调查)三种技术工具,以面访为主,网络和电话访问为辅。CFPS 共有社区问卷、家庭成员问卷、家庭问卷、成人问卷和少儿问卷五种主体问卷类型,受访人分别为最熟悉情况的社区干部、最熟悉情况的家庭成员、成年和少儿家庭成员。

2010年CFPS基线调查共发放样本19986户,¹³涉及25个省、直辖市、自治区的162个区/县的649个村/居,最终完成了635个村/居的14960户的访问,共界定出基线基因成员57155位。其中,42590位基因成员完成了2010年的个人访问(成人33600位、少儿8990位)。此次调查在家庭层面的累积应答率为81.3%,在个人层面应答率为84.1%。¹⁴

CFPS 2011年的调查规模较小。出于样本维护的目的,同时作为对2012年全部样本追踪调查前的一次预调查,2011年CFPS仅计划访问基线调查中的部分样本。考虑到青少年在身心发育、学校教育和行为态度等方面变化较快,相对密集的数据采集将对青少年开展研究更有利,2011年CFPS将调查对象锁定为基线调查中完访的青少年(至2011年年龄为18岁及以下)基因成员,同时对基线调查中完访的家庭进行了家庭层面的追访。对18岁以上的成年基因成员、新出现的家庭成员和新组成的家庭没有采集信息。在家庭层面,2011年共完成访问

12. 关于权数计算的详细方法参见:吕萍、谢宇. 2013. 中国家庭追踪调查2010年基线调查权数计算(第二版),北京大学中国社会科学调查中心《中国家庭追踪调查技术报告系列(CFPS-17)》,网址:<http://www.iss.edu.cn/cfps/d/file/p/a88086f58ae9d67eb1a8a1a5bf24563c.pdf>。

13. CFPS2010年基线调查实际样本的规模参考了2008年和2009年预调查所得的预估应答率,采用按应答率比例扩大样本规模的方法,以保证获得预计的有效样本家户数量。

14. 此个人层面的应答率是完访家庭内部的个人应答率。此外,在家庭层面,合作率为96.6%,联系率为84.1%,拒绝率为2.7%;在个人层面,合作率为87.0%,联系率为96.7%,拒绝率为8.5%。所有结果按照AAPOR标准进行计算,具体方法参见:孙妍. 2012. 中国家庭追踪调查2010年基线调查样本联系情况,北京大学中国社会科学调查中心《中国家庭追踪调查技术报告系列(CFPS-5)》,网址:<http://www.iss.edu.cn/cfps/d/file/wd/jsbg/2010jsbg/06068e1503be9b60e67d3f4cfe775fb3.pdf>。

13 130 户家庭,追踪成功率为 89.1%。¹⁵ 在个人层面,共完成访问 8 803 位基因成员(成人 1 279 位,少儿 7 524 位)。¹⁶ 其中,在 2010 年完访且满足 2011 年访问年龄条件的 9 701 位基因成员中,共有 7 696 位基因成员在 2011 年成功访问,追踪成功率为 88.8%。¹⁷

2012 年,CFPS 对全部个人样本及所在家庭开展了追踪调查。其中,个人样本包含 2010 年的所有基因成员,以及 2010 年后出生或领养的新基因成员;家庭既可能是 2010 年的原家庭,也可能是由于基因成员结婚、分家等原因而派生出来的新家庭。¹⁸ 此外,CFPS 同样采集了基因成员所在家庭中其他非基因家庭成员的基本信息。在 2010 年完访的 14 960 户家庭中,有 12 725 户在 2012 年成功完成家庭层面的访问,家户层面追踪成功率为 85.1%。在这完访的 12 725 户家庭及其派生出的 728 个新家庭中,共界定出基因成员 52 336 人,¹⁹ 非基因成员 2 737 人。最终完成个人访问的基因成员为 42 970 人,非基因成员为 1 714 人。其中,2010 年完访的 42 590 位基因成员中,共有 33 956 人在 2012 年被成功追踪,个人层面的追踪成功率为 80.6%。²⁰ 在 2012 年界定出的 52 336 位基因成员中,共有 8 477 人由于外出或新组家庭的原因离开原家庭,其中 8 341 人需要进行异地追访,共访问成功 5 756 人,异地追踪成功率为 69.0%。表 1 对三年访问的基本情况进行了汇总。

(三) 追访原则²¹

跟踪调查重复观察同一样本在不同时点的状况及其历时变化。由

15. 计算中排除了 219 户不需访问家庭的影响。其中,212 户为 2011 年不要求访问的搬迁或疑似搬迁家户,7 户为全家去世家户。

16. 2011 年对访问时不在家青少年基因成员尝试性地使用电访、网访与邮寄问卷相结合的方式进行了异地追访。但由于此部分的数据清理工作尚未完成,本文中关于 2011 年的联系结果统计均未考虑此部分数据。

17. 此处仅计算了面访的追踪成功率(原因参见上文)。计算中排除了 1 036 位不需面访的基因成员的影响。其中,24 人为死亡以及调查当年无需访问的出境、出家、入狱、参军的基因成员,1 012 人为需要异地追访的基因成员。

18. 读者需要注意,CFPS 追访调查中的“新家庭”(也称新组家庭或另组家庭)是相对于 CFPS 原有家庭样本而言的。它并不一定是新建家庭,而可能是本来就存在的家庭,由于第一次进入 CFPS 调查,因而也称为“新家庭”。

19. 其中,50 665 人为原有基因成员(含 639 位去世人员),1 671 人为新生/领养基因成员。

20. 计算中排除了死亡人员(408 人)和出境、出家、入狱、参军四类在调查当年不需要访问人员(40 人)的影响。

21. 此处为 CFPS 通用的追访原则,但在具体的年份,追访会有不同的操作策略。具体可参考 CFPS 相关年份用户手册与技术报告。关于 2012 年的追访操作策略还可参考谢宇、胡婧炜(2013)。

表 1:CFPS2010、2011、2012 年调查基本情况汇总

	2010	2011	2012
调查对象			
家庭层面	随机抽样的 19 986 户家庭	2010 年全部完访家庭	基因成员所在的家庭
个人层面	样本家庭中的全部家庭成员	2010 年完访的 18 岁及以下基因成员	基因成员及其所在家庭的其他家庭成员
社区层面	样本家庭所在 649 个村居	—	—
样本量			
社区	635	—	—
家庭	14 960	13 130	13 453
基因个人	42 590	8 803	42 970
成人	33 600	1 279	34 447
少儿	8 990	7 524	8 523
访问方式	CAPI	CAPI + CATI + CAWI + 邮寄问卷	CAPI + CATI
联系结果			
家庭层面 (%)	81.3(应答率)	89.1(追踪成功率)	85.1(追踪成功率)
个人层面 (%)	84.1(应答率)	88.8(追踪成功率)	80.6(追踪成功率)
异地 (%)	—	—	69.0(追踪成功率)

于 CFPS 的研究主体是家庭和个人,所以家庭的分裂组合,以及家庭人口的出生、死亡、迁移等变化均会对选定谁为重复观察的样本提出挑战。接下来,我们简要介绍 CFPS 确立追访样本的基本原则,该原则主要依研究需求而定,但同时也考虑了时间、经费和技术等客观条件的限制。

CFPS 每一年或两年进行一次全部样本的调查。图 2 展示了 CFPS 通用的追访流程。2010 年基线调查界定出来的所有家庭成员及其今后的新生血缘/领养子女为 CFPS 的基因成员,他们是 CFPS 的永久追踪对象,直至死亡。CFPS 基因成员调查时所在的家庭也是 CFPS 当年的调查对象,但这些家庭今后一旦没有基因成员存在(如基因成员从属新的家庭或死亡),则终止调查。因此,在 CFPS 样本框中,随着时间的变迁,新的基因成员会出生,已有基因成员会死亡;新家庭会因为婚姻、分家等原因而不断产生,旧家庭会因为基因成员死亡或另组新的家庭而不断分化和消失。在无样本流失的情况下,CFPS 样本正好可以反映中国人口与家庭的自然更替与变化,具有可持续自我更新的特性。

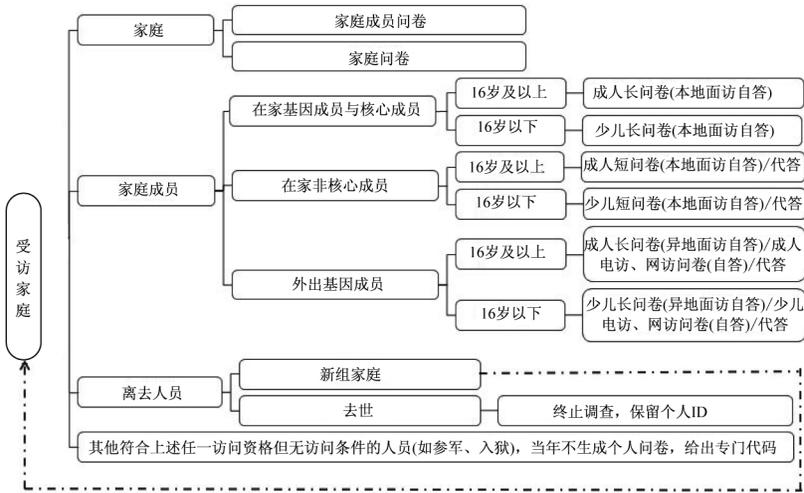


图 2:CFPS 追访流程与问卷生成规则

为了保证研究需求的前提下尽可能降低成本和提高访问效率, CFPS 在 2012 年的调查中对家庭成员性质进行了分类:将基因成员所在家庭中的所有与基因成员有父母、子女、配偶关系的非基因成员视为调查当年该家庭的核心成员,将基因成员与核心成员以外的家庭成员视为非核心成员。其中,基因成员与核心成员需要回答完整的个人问卷(也称“长问卷”),非核心成员则只需回答简短的个人问卷(也称“短问卷”),或者仅通过他人代答的方式采集少量的关键变量信息。通过这样一种设计,我们既可以收集到最重要的一部分家庭成员的详细资料,又可以了解到相对次要的家庭成员的基本背景,从而帮助研究者了解基因成员所处的家庭环境及成员关系。不过,与基因成员不同,核心成员与非核心成员不是 CFPS 长期追踪的对象,他们与基因成员的家庭关系一旦断裂(即不属于基因成员所在家庭的成员),对他们的调查将终止。此外,CFPS 规定,调查时出境、出家、入狱或参军的基因成员在调查当年无需进行个人访问。

样本流失是所有追踪调查都面临的问题。为了尽可能采集到受访者信息,CFPS 采取了两方面措施:一是对不能成功面访的个人或者家庭,通过电话访问、网络访问等辅助方式,尽可能实现对其本人的访问;二是对经各种努力未能访问到本人的个别基因成员,通过他人代答的

方式收集其一些重要的客观指标,避免其信息完全缺失。

三、内容特色

CFPS在设计之初借鉴了世界上一些富有影响力的调查项目的方法与经验。这些项目主要有PSID(Panel Study of Income Dynamics)、NLSY(National Longitudinal Survey of Youth)和HRS(Health and Retirement Study)等。与这些调查项目一样,CFPS的调查内容丰富而广泛(见表2),可供多学科、多领域各类主题的研究。CFPS的调查内容不仅涵盖了社会科学研究中被普遍关注的主题,同时也具有自身的特色,以下我们将介绍CFPS问卷设计的一些独特之处。

(一)完整的家庭结构网络与家庭成员信息

家庭、婚姻、代际关系及流动既是社会学的传统议题,也是当今社会科学研究的一个热门方向。这一领域的研究潜力取决于数据在多大程度上能够将家庭及重要亲属的关系网络数据化,以及在多大程度上能详尽收集家庭成员的信息。目前的绝大多数国内调查只抽取家庭中的单个个人或少数成员收集数据,所调查的家庭关系也是以具体受访人为核心延展出的家庭关系,最多只能掌握受访者与部分未受访家庭成员的关系,却无法获知未受访成员之间的关系,更无法获得跨代或多代研究所需要的家庭关系。此外,以单个个人为核心收集的其他家庭成员的信息往往不完整,限制了相关领域更深入的研究。例如,研究婚姻关系受限于只知道夫妻单方的数据;研究代际关系或代际流动受限于只知道父母和子女其中一方的信息,或只知道同住的父母、子女的信息而不知道不同住的其他子女的信息。一般的调查以单个个人或少数成员为核心采集数据,那么,受访者以外的其他大部分人的信息,即使是父母、子女、配偶这些重要的家庭成员,也只能通过代答的方式采集,这不仅影响数据的准确性,也极大地限制了可调查的内容。

为了描述出一个完整、精确的家庭结构网络,经过2008年和2009年预调查的尝试后,CFPS在2010年的基线调查中创造性地采用了一套T表系统来采集家庭关系与家庭成员信息。

T表位于2010年家庭成员问卷的起始部分,由T1、T2、T3三张表构成(图3)。其中,T1表记录每一位家庭成员的基本社会人口特

表 2:2010 年、2011 年和 2012 年 CFPS 问卷内容模块对比表

	2010	2011	2012		2010	2011	2012
	社区问卷				家庭问卷		
	√			地理交通	√	√	
家庭成员问卷	√			家庭生活条件	√		√
	√	√	√	社会交往	√		
成人问卷				农业生产		√	√
基本信息	√	√	√	非农经营	√	√	√
流动迁移	√	√	√	家庭成员收入	√	√	
兄弟姐妹	√			外出务工	√		
特殊经历	√		仅参军	转移支付	√	√	√
学历教育	√	√	√	拆迁/征地补偿	√	√	√
非学历教育			√	家庭支出	√	√	√
语言	√	√	√	房屋	√	√	√
婚姻	√	√	√	土地	√	√	√
与子女关系	√	√	√	其他家庭资产、	√	√	√
工作与收入	√	√	√	财物与消费品	√	√	√
退休养老	√		√	金融资产	√	√	√
时间利用	√	√		债权债务	√	√	√
休闲与出行	√	√		少儿问卷(家长代答部分)			
手机和网络	√	√		基本信息	√	√	√
社会关系	√	√		流动迁移	√		
宗教			√	家长照管	√	√	√
政治	√	√	√	个体成长	√		√
主观态度	√	√	√	健康	√	√	√
健康	√	√	√	教育花费	√	√	√
日常行为	√	√	√	父母养育观念	√	√	√
精神测量	√	√	√	上学	√	√	√
认知测试	√	√	√	课外辅导	√	√	√
特殊身体活动测试	√	√	√	少儿问卷(少儿自答部分)			
父母信息			√	教育史与上学	√	√	√
				工作	√	√	√
				时间利用	√	√	
				语言	√	√	√
				社会交往	√	√	√
				手机网络	√	√	
				健康	√	√	√
				主观态度	√	√	√
				行为与精神测量	√	√	√
				认知测试	√	√	√

征, T2 表采集 T1 表中所有家庭成员的父母、配偶、子女的姓名信息。T2 表中填写的既可能是 T1 表成员,也可能是 T1 表成员的非同住的父母、配偶、子女,后者的基本社会人口特征通过 T3 表采集。²²通过 T2 表,研究者可以得到一个全面的家庭及亲属关系网络。通过 T1 和 T3

22. 关于 T 表的具体操作方案参见:谢宇. 2012. 中国家庭追踪调查(2010)用户手册(第二版), 北京大学中国社会科学调查中心(<http://www.iss.edu.cn/cfps/wd/jsbg/2010jsbg>)。

表,研究者可以获取每一位家庭成员及其不同住父母、子女和配偶的基本社会人口信息。此外,研究者通过家庭关系的匹配²³还可以将 T1 表中不同成员的个人问卷相互联系,得到完整、对称的父母—子女数据、夫妻数据、同住兄弟姐妹数据,在一些家庭中还可获得跨代、多代成员间的完整个人数据。

T1 表:同住家庭成员表								
个人编码	姓名	出生日期/ 属相年龄	性别	婚姻 状况	最高 学历	主要 工作	行政管理 职务	外出人员 信息
101								
102								
...								
301								
302								
...								

T2 表:家庭成员父母、配偶、子女关系表								
个人编码	姓名	父亲	母亲	配偶	孩 1	孩 2	...	孩 10
101								
102								
...								

T3 表:家庭成员不同住父母、配偶、子女列表								
个人编码	姓名	出生日期/ 属相年龄	性别	婚姻 状况	最高 学历	主要 工作	行政管理 职务	居住地与 户口信息
301								
302								
...								

图 3:T 表设计问卷截图

与传统社会调查相比,使用 T 表采集家庭关系的优势总结为三点:第一,传统社会调查通常以家庭中的某一位受访者为核采集其他家庭成员的基本信息及其与受访者的关系。这种做法假定家庭关系中只有受访者一个核心,该核心通常是采用随机抽取的方式或以人为指定“户主”的方式来确定。但无论采取哪种方法,关于核心(尤其是单核心)的假定都把家庭关系理解得过于简单,因为逻辑上家庭中的每一个人都可以作为家庭关系的核心,家庭关系应该是由多个核心连接起来的树状网络(family tree),而从单一核心出发(如户主或随机抽取的受访者)的辐射状结构仅仅是树状家庭网络中的一小部分,因而只能采集到每个(或几个)家庭成员与受访者之间的关系,却无法获知除受访者

23. 关于 T 表的使用及家庭关系的匹配参见:孙玉环、谢宇、胡婧炜、张春泥、许琪、黄国英. 2012. 中国家庭追踪调查 2010 年家庭关系原始数据库的分解与匹配,北京大学中国社会科学调查中心《中国家庭追踪调查技术报告系列(CFPS-6)》,网址:<http://www.iss.edu.cn/cfps/d/file/wd/jsbg/2010jsbg/771ddeed0dc54dd0e20ac0aae9504c3c.pdf>.

以外的那些家庭成员之间的关系。CFPS的T表设计突破了单一核心的限定,允许每一名家庭成员轮流成为家庭关系的核心。因此,研究者不仅可以知道家庭成员之间直接的父母、配偶、子女关系,还可以推断出一些间接亲属关系,比如,继父母—子女关系、跨代的祖父母—孙子女关系、同辈的兄弟姐妹关系等。

第二,针对传统社会调查中同一家庭不同成员信息收集不完整、不对称的问题,T1表用代答的方式采集了无论老幼所有家庭成员的基本信息,这一信息采集不受限于具体家庭成员是否接受个人访问的影响。不仅如此,通过T3表,研究者还可以得到与受访者不同住的父母、配偶、子女的一些基本社会人口信息,这弥补了传统调查中由于分家导致重要亲属信息缺失的问题。²⁴

第三,T表建立的个人编码体系可以对家庭成员进行准确定位,解决了传统社会调查无法将受访者填写的家庭关系与具体的个人联系起来的问题。唯一的个人编码为后续的追踪访问奠定了基础,也让研究家庭成员边界的持续变动(如子女离家、父母迁入子女家庭养老等)成为可能。

综上所述,T表的设计克服了传统社会调查收集的家庭关系不明确、信息不完整的问题,为研究者掌握家庭关系全貌、了解全部家庭成员信息、匹配家庭成员数据等前沿研究的需求提供了难得的资料。

(二) 整体抽样下的城乡问卷一体化

前文提到,CFPS采用了城乡一体的抽样方式。反映在问卷设计上,CFPS采用了城乡整合的问卷,不再像传统调查一样分城市和农村两套问卷进行调查。

传统社会调查分城乡两套问卷的做法主要是鉴于中国过去城乡间制度上的巨大差别。但是,由于近年来中国快速的城市化进程,对农村与城市的行政划分已难以及时反映城乡边界的实际变化。而且,随着农村的现代化发展和城乡人口、资源、信息流动的日益频繁,城乡家庭和个人在许多重要特征上的界限也在不断模糊,而城市内部和农村内部个体间的异质性却在扩大。因此,城乡使用不同问卷经常会出现问

24. 但是,在2010年基线调查的实地访问中,由于没有完全按照原设计执行,T3表中的部分去世人员(尤其是父母)的信息没有被采集。为弥补这一缺陷,我们在2012年的调查中对所有个人问卷的成人受访者补问了其父母的出生年月、教育、职业和政治面貌。

卷对于地区、家庭或个人不适用的情况,整合城乡问卷变得很有必要。

与传统社会调查相比,一方面,CFPS 能够从多角度判断受访者的城乡属性:社区问卷的村/居属性(村委会/居委会)反映了家庭居住地的城乡行政属性;家庭问卷采集的家庭从事农业生产与非农经营的相关信息反映了家庭经济活动上的农业和非农属性;个人问卷采集的受访者户籍状况、从事农业与非农工作等信息可用于判断户口上的城乡人口划分和个人职业的农业或非农属性。

另一方面,城乡问卷一体化解决了问卷适用性问题。借助 CAPI/CATI/CAWI 系统的帮助,CFPS 可以在访问中即时生成适用于各类城乡人群的个性化问卷。以工作模块为例,无论受访人是农业户口或非农业户口,无论其居住在村委会或居委会,只要其从事了农业工作,访问系统可以立即调用农业工作模块,采集其农业工作信息;如果其从事了非农工作,也可以立即调用非农工作模块,采集其非农工作相关信息。

(三) 流动人口

人口迁移与流动是自 20 世纪 90 年代以来中国社会的一个重要现象,“六普”数据显示,2010 年中国流动人口已达到 22 143 万人。²⁵人口迁移与流动不仅改变了城乡、地区间的人口结构与分布,还改变了劳动力市场的结构和社会分层。一些新出现的社会问题也与人口流动有关,例如流动人口犯罪、留守儿童问题、农村离婚率上升、养老问题等(Davin,1996;Silverstein, *et al.*, 2006;Lu,2012;吴要武,2013)。

对流动人口的研究和政策制定迫切需要可靠的数据。目前对流动人口的调查主要存在三方面的困难:首先,一般对流动人口(如农民工)的调查是对流入地人口进行抽样,因而只能访问到当前的流动人口,无法访问曾经有流动经历但目前已停止流动的人(如回流者)和留在流出地的潜在流动者。流动者在许多特征上具有选择性,研究数据如果缺乏回流者或留守者这些比较群体,其结论可能会存在选择性偏误,也无法研究流动的选择性。其次,截面调查通常只采集到流动人口调查当时的情况,却没有采集其流出之前的情况。由于缺少同一个体流动前后状态的对比,因而难以判断流动与状态改变之间的因果关系。即使

25. 参见:马建堂. 2011. 第六次全国人口普查主要数据发布,国家统计局网站(http://www.stats.gov.cn/tjfx/jd/fx/t20110428_402722238.htm)。

有的截面调查收集了历次迁移或流动的回顾性数据,但这种做法仍有缺陷:受访者的回忆难免有误差,而且由于不同流动者流动经历的时间起点不同,这种做法会降低流动经历的可比性,难以据此准确描述人口流动随时间变动的趋势。最后,流动人口的高度流动性给抽样和访问均造成了困难,影响了数据的代表性与准确性。

为避免上述问题,CFPS选择的调查对象既包括流出地家庭外出但尚未在外安家的个人,也包括流动到外地安家的流动者,这为研究流动的选择性、比较流动者与非流动者间的差异提供了条件。通过对同样本的长期追踪,CFPS收集到受访者一生的居住地址的变化情况,这让流迁经历的数据更完整、更准确。而且,将流迁信息与历次追踪收集的其他数据相联系,研究者还可以分析流动前后的变化,从而更好地研究流动与其他事件或状态之间的因果性。此外,CFPS每年均从上次调查时的家庭入手来调查家庭内部成员的出入状况,因而能够更有效地捕捉到个人的去向和联系信息。最后,电访、网访有效提高了流动样本的应答率。即使在本人面访、电访或网访都不成功的情况下,其他家庭成员为其完成的代答信息也可以一定程度上弥补缺失。

(四) 认知测试

经济学家和社会学家在研究影响人们获得社会经济地位(如收入、职业)的因素时,人力资本都是最常用于解释个体间收入(或其他劳动力市场结果)差异的重要变量(Mincer,1974;Sewell and Hauser,1975;Xie and Hannum,1996)。人力资本是指一系列能够提高劳动者生产力的技能或特征,对这一概念最常见的操作化是Mincer模型用学校教育和工作经验来测量(Mincer,1974),许多有关现代社会经济地位获得的研究均以Mincer模型为基础。但是,学校教育、培训和工作经验仅是人力资本中的一部分。人的能力,尤其是认知能力,会直接影响劳动者生产力。认知能力既体现了人的智力因素的影响,也体现了学校教育、培训对能力的培养。忽略认知能力很可能导致高估正式教育对收入的影响(Griliches,1977)。认知能力测试在国外已广泛运用于社会调查、企业管理、军队选拔等领域,并为大量经验研究提供解释变量。但受调查技术和成本的限制,国内社会调查极少提供认知能力的测试。

CFPS调查的一个特色内容是对所有10岁及以上的少儿和全部成人的认知能力发展状况进行长期的测试与评估。目前CFPS共拥有

两套测试题,其中一套由识字题和数学题组成,测试受访者的识字水平与数学计算水平,使用于2010年和2011年;另一套由记忆题和数列题组成,测试受访者的长短期记忆能力和算术推理能力,使用于2012年。

CFPS准备长期轮换使用上述两套测试题。研究者利用CFPS的认知测试数据可以了解受访者认知能力的长期发展情况。借助CFPS其他方面的详细信息,研究者也可以研究影响受访者认知能力的因素、认知能力对个人成就的影响,以及认知能力与行为、态度之间的关系。

(五) 少儿的成长与发展

少儿的成长与发展是一个重要的研究主题。少儿时期的成长和发展对入成年以后的社会经济成就、行为、态度均有持续性影响。但是,国内大多数关于家庭的社会调查或综合性社会调查仅以成人为受访对象,缺少针对少儿成长与发展的调查内容,更没有可以从总体上反映少儿从出生起的整个成长过程的追踪数据,而关于少儿的专题调查又缺少完整的家庭背景信息。相比之下,对少儿建立专门的数据库,同时又有翔实的家庭背景信息可供综合分析,这是CFPS独有的优势。

CFPS采用独立问卷长期跟踪少儿的成长与发展状况。在调查对象上,基因成员与核心成员中所有少儿(16岁以下)都将接受调查。其中,10岁以下的少儿问卷全部由父母(或监护人)代答;10岁及以上少儿问卷既有父母(或监护人)代答部分,也有少儿自答部分。对于历次调查中的新生儿,研究者可以得到其从出生开始的完整成长信息。

在调查内容上,少儿问卷收集了少儿成长过程中各个方面的信息,包括早期发育、学校教育、认知能力、个性特征、心理素质、生理健康、学习与日常行为习惯等,这些对研究少儿的成长与发展、少儿阶段特征的持续影响具有非常重要的价值。

(六) 完整的事件史信息

社会变迁反映到个人层面上就是个体的生命历程(Elder,1985)。通过比较不同时代经历者的生活轨迹和人生经历,研究者能够对社会变迁有更细致、深刻的理解。生命历程的研究需要时间性的数据。研究者不仅需要知道事件或经历发生与否,还需要知道事件或经历何时发生、持续多久,以及不同事件或经历的先后次序。

在2010年的基线调查中,CFPS对所有基因成员通过回顾(retrospective)的方式收集了教育史和婚姻史的详细信息,并在之后的

调查中持续跟踪教育、婚姻的变化。在教育史方面,CFPS 调查了受访者一生的教育经历,包括各个阶段的就读时间、专业、学校性质和学业完成情况等;在婚姻史方面,CFPS 调查了受访者的初次婚姻和自初访起一生中历次婚姻的起止时间和变化情况,以及不同婚姻阶段中包括配偶年龄、双方如何认识等在内的具体信息。此外,CFPS 从 2012 年起增加了对个人职业变化的跟踪调查。

CFPS 对事件史的收集还关注一些生命历程中的重要转折性事件,例如参军、上山下乡、饥荒等,也调查一些特殊的阶段,如同居。近年来,同居现象日益增加,这一变化趋势对婚姻与性别研究具有极大的价值。但在过去由于同居一直被认为是敏感问题,很少有调查收集相关的数据。CFPS 极具前瞻性地有这样一些社会现象出现的早期便将其纳入了调查。

(七) 访员观察数据

作为长期追踪的调查项目,CFPS 尤其重视对访问质量的评估,这需要收集访问过程本身的数据。CFPS 在每一套问卷末尾均收集了一组访员的观察数据(图 4),主要包括访问的环境、受访者在访问中的行为态度、受访者个人特征等。这部分数据的采集由访员直接通过观察填写,无须向受访者提问。所获得的数据不仅可以作为研究中的分析变量使用,也可以用来评估受访者回答的可靠性,还可以用于对调查执行过程本身的研究,为改进访问工作、提高数据质量提供科学依据。

四、初步发现

CFPS 的追踪设计和内容特色最终会体现为引导有潜力的经验研究及获得有价值的发现。虽然 CFPS 目前只积累了三期调查数据,数据的开发和使用仍处在初期阶段,但我们的团队已经从初步的使用中得到有意义的发现。我们将以基于 CFPS 数据的几个研究为例,²⁶向读者展示 CFPS 数据的开发潜力。

(一) 收入不平等和贫困

贫富差距的扩大是伴随中国经济增长出现的重大民生问题之一。CFPS 的 2012 年调查显示,在全国居民对几大民生问题严重性的评价

26. 此处所列研究发现的详尽分析和解释可参见《中国民生发展报告 2013》相应章节。

中,贫富差距位居榜首,其严重程度被认为超过了腐败、就业、环境污染、住房、教育和医疗等问题(吴琼、谢宇,2013)。那么,目前中国的收入不平等和贫困程度到底如何?其变动趋势是上升还是下降?

Z1	这份问卷最主要是由谁完成的?	
Z101	访问过程有哪些家庭成员参与了回答?	
Z102	访问时,除了家庭成员外,还有谁在场?【可多选】	
	1. 亲友 2. 督导 3. 邻居 4. 村居干部 77. 其他【请注明】_____ 78. 没有其他人在场	
Z103	访问过程中主要使用以下哪种语言?【单选】	
	1. 普通话(跳至 Z201) 5. 方言(继续回答 Z104)	
Z104	具体是什么方言_____	
Z201	受访者的理解能力:	很差—1—2—3—4—5—6—7—) 很好
Z202	受访者的健康状况:	很差—1—2—3—4—5—6—7—) 很好
Z203	受访者的衣装整洁程度:	很差—1—2—3—4—5—6—7—) 很好
Z204	受访者的外貌:	很差—1—2—3—4—5—6—7—) 很好
Z205	受访者的普通话熟练程度:	很差—1—2—3—4—5—6—7—) 很好
Z206	受访者对调查的配合程度:	很差—1—2—3—4—5—6—7—) 很好
Z207	受访者智力水平:	很低—1—2—3—4—5—6—7—) 很高
Z208	受访者的待人接物水平:	很差—1—2—3—4—5—6—7—) 很好
Z209	受访者对调查的兴趣:	很低—1—2—3—4—5—6—7—) 很高
Z210	受访者对调查的疑虑:	很低—1—2—3—4—5—6—7—) 很高
Z211	受访者回答的可信程度:	很低—1—2—3—4—5—6—7—) 很高
Z212	受访者的语言表达能力:	很差—1—2—3—4—5—6—7—) 很好
Z5	访问过程中受访者在多大程度上表现出急于结束调查:	很低—1—2—3—4—5—6—7—) 很高

图 4:CFPS 访员观察题问卷

对中国收入不平等和贫困的水平和趋势的估计与预测在经济学界一直广受重视。许多研究用不同来源的截面数据描述过收入不平等程度和贫困水平的历时变化(李实、赵人伟,1999;李实等,2013;胡鞍钢等,2006;胡兵等,2007;Ravallion and Chen,2007),但由于抽样方案和测量上的差异,具体的研究发现仍存争议。最近一次关于收入不平等程度的争议发端于西南财经大学中国家庭金融调查(CHFS)²⁷公布的高达 0.61 的基尼系数(Orlik and Davis,2013;Fisher,2012),以及国家

27. CHFS 开始于 2011 年,该项目的详尽介绍可参见:Gan, *et al.* (2013)。

统计局数据显示的自 2008 年起基尼系数逐年下降的发现。²⁸争议的焦点仍是收入不平等的变动趋势。但单纯从全国基尼系数在不同截面时点上的变化来看不足以得出可靠的结论,因为无论是基尼系数,还是贫困率,其数值在两个时点间的整体变化可能体现为多种形式。例如,收入不平等的上升既可能是穷人更穷、富人更富;也有可能是穷人和富人的收入都提升,但富人的收入上升得更快;还有可能是穷人和富人的收入均下降,而穷人的收入损失更严重。又例如,贫困率保持不变,这既可能表现为既有贫困家庭持续贫困,也可能表现为脱贫与新增贫困家庭在数量上相互抵消。这些变化形式的性质是不同的,对政策的意义也不同。所以,截面数据难以反映收入或贫困流动的真实过程。

CFPS 的 2010 年和 2012 年追踪数据有助于分析收入不平等和贫困的变动过程。图 5 总结了不同来源数据估计的基尼系数,其中也包括来源于两期 CFPS 数据的估计。首先,我们看到 CFPS 估计的基尼系数水平介于 CHFS 和国家统计局的估计之间,但更接近于国家统计局的估计。相比之下,CHFS 的估计明显高于 CFPS,也高于 CGSS 和 CHIP 这些学术调查数据的估计。其次,从 2010—2012 年基尼系数的变化看,CFPS 几乎相同的家庭样本也反映出全国基尼系数确实略微下降,这一点似乎与国家统计局的数据相符。但是,如果分别对农村和城镇的 CFPS 样本数据进行计算就会发现,农村和城镇内部的基尼系数几乎没有变化(谢宇等,2013)。图 6 描述的是 CFPS 同一批家庭在 2010 年的收入组别和 2010—2012 年间的平均收入增长率。该图在一定程度上解释了为什么在农村和城镇基尼系数几乎不变的情况下,全国基尼系数会降低:农村家庭的收入增长比率高于城镇家庭,故城乡收入差距正在缩小;中等和中上收入家庭的收入增长比率高于高收入家庭,这也有助于缩小中等、中上收入家庭与高收入家庭间的贫富差距。全国基尼系数水平下降并不意味着收入不平等得到了整体改善。图 6 还显示低收入家庭的收入增长比率明显低于中等和高收入家庭。此外,从 90/10 比率上看,贫富

28. 参见:网易财经(意见中国:网易经济学家访谈录). 2013. 甘犁:中国基尼系数高不是坏事,网址:<http://money.163.com/13/0515/14/8UU3AB7H00254TV5.html>;岳希明、李实. 2013. 如何看待不同的基尼系数,网址:<http://www.ciidbnu.org/news/201302/20130207203015706.html>;财新网. 2013. 李实:难以判断基尼系数已呈下降趋势,网址:<http://china.caixin.com/2013-01-21/100484878.html>。

两极家庭的收入差距在两年间略有扩大(谢宇等,2013)。这些均意味着收入不平等在局部人群中甚至有所上升。

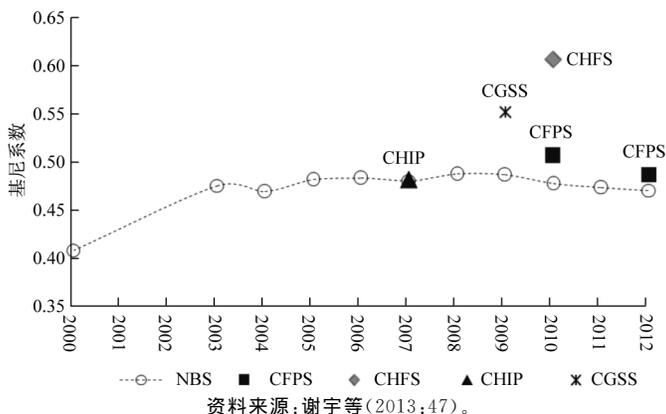


图 5: 2000—2012 年不同数据来源的基尼系数估计

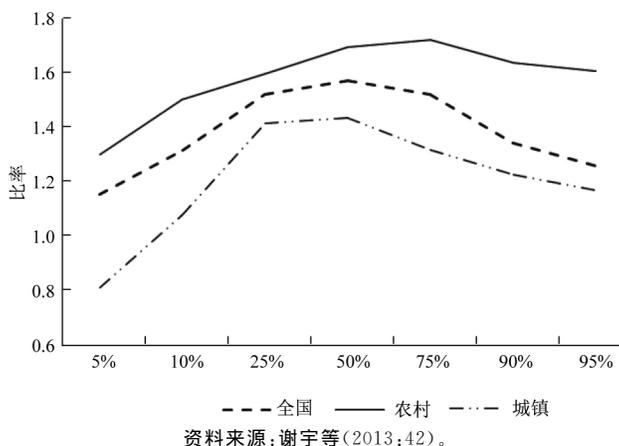
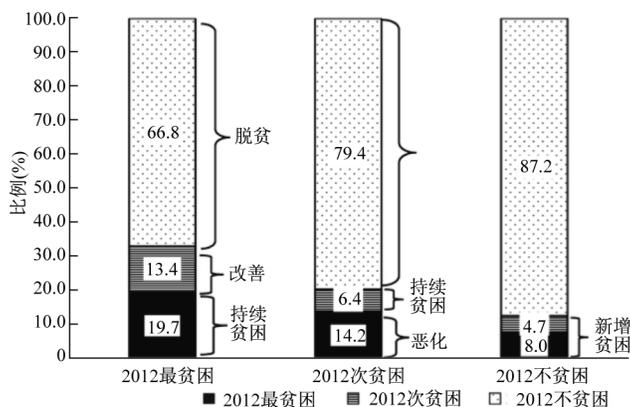


图 6: 2010—2012 年人均家庭收入分位数的增长比率

在贫困方面,CFPS 通过对同一批家庭的追访首次展现了家庭贫困状态的动态变化。图 7 中,2010 年和 2012 年农村家庭贫困状态均分为三类:最贫困(人均年收入在 1 美元/天以下)、次贫困(人均年收入为 1—1.5 美元/天)、不贫困(人均年收入高于 1.5 美元/天)。我们看到 2010 年最贫困和次贫困的农村家庭分别有 67%和 79%在 2012 年脱贫,有 13%的农村家庭虽然没脱贫,但贫困状况有所改变。但是仍

有 20% 的农村最贫困家庭和 6% 的农村次贫困家庭的贫困状态持续, 14% 的次贫困家庭贫困程度恶化, 以及有 13% 的新增贫困。以上发现说明相当一部分低收入家庭的经济水平是在贫困线上下浮动, 大多数贫困现象是暂时的, 构成贫困人口的家庭是不断变动的。



资料来源: 张晓波等(2013: 62)。

图 7: 农村家庭贫困状况变动

经济不平等是经济学和社会学长久讨论的议题, 以往许多官方和学术机构的数据已为收入不平等和贫困研究提供了丰富的经验证据, 但 CFPS 作为新的追踪数据为这一旧议题的讨论提供了新视角和新发现。研究者不仅可以利用 CFPS 数据检验以往数据发现的可信度, 还可以通过追踪数据观察到收入流动、贫困流动等这些在截面数据中无法观察到的动态过程。

(二) 婚姻稳定性

婚姻是家庭研究的主要议题之一。随着第二次人口转型(second demographic transition)从西方发达国家向全球逐渐扩散, 人口学研究的一个重要转向是从对生育率的单一关注转变为对与生育相关的婚姻、家庭制度的多元关注。第二次人口转型始于 20 世纪 50 年代, 以离婚率上升、同居日益普遍、婚外生育增多等趋势为特点(Lesthaeghe, 2010)。但这些趋势在全球范围内的扩散并不一致, 不同国家婚姻、家庭领域的变迁具有差异性。这些差异也是人口学、社会学关注的问题。

在计划生育政策和经济发展的双重推动下, 到 20 世纪 90 年代, 中国已经完成了从高出生率、高死亡率到低出生率、低死亡率转变的第一次人

口转型(Xie,2011)。虽然中国社会婚姻、家庭的变化已显现出第二次人口转型的一些特点,如离婚、同居现象的增多,但中国在多大程度上进入了第二次人口转型?离婚、同居等现象在中国是否真的日益普遍?局部的、零星的观察不能代替对总体趋势的描述,研究者需要更好的数据。

CFPS 以回顾性问卷和追踪调查的方式收集的婚姻史数据为描述中国社会婚姻的变迁提供了难得的资料,下面我们将以结婚、离婚和同居三个例子来说明这些事件史数据的价值。

图 8 是根据 2010 年 CFPS 婚姻模块的回顾性信息统计不同出生年代人群 18—40 岁进入初婚的累计比例。我们可以看到,婚姻在中国仍然是一个普遍现象:无论出生于哪个年代,在 40 岁以前,超过 80% 的人都至少经历了一次婚姻。相比于出生在 20 世纪 40、50 年代的人,生于 60、70 年代的人在 40 岁以前结婚的比例不但没有降低,反而有所升高。在初婚时间上,生于 40 年代的人明显结婚较早,50% 的人的初婚发生在 22 岁左右。60—80 年代,出生年代越晚,进入婚姻越迟。在较晚出生队列中初婚时间的推迟一定程度上与年轻人面临的经济压力上升有关,他们需要更多时间来实现组建家庭所需的经济积累(Mu and Xie,2014; Yu and Xie,2013)。但生于 50 年代的人例外,他们结婚普遍比生于 60、70 年代者更晚,这一推迟可能是受 70 年代“晚婚”政策的影响。

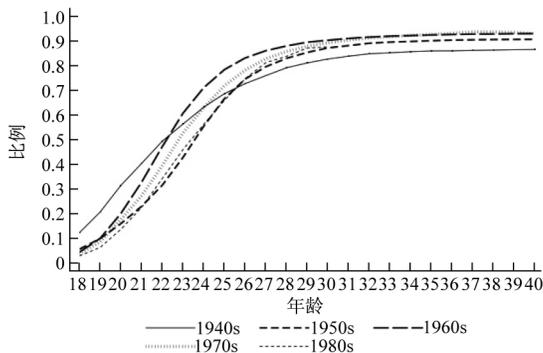


图 8: 18—40 岁各年龄组进入初婚的累计比例

官方数据表明,中国的离婚率在过去的 30 年有所上升。²⁹但离

29. 参见:法制晚报. 2010. 民政部:中国离婚率连续 30 年上升,网址:<http://news.163.com/10/1003/13/612T8NPB0001124j.html>。

婚率的计算通常是用当年的离婚次数除以人口总数,这一比率受人口结构的影响,难以衡量婚姻的稳定性。利用 CFPS 的婚姻史数据,我们能够观察到同一段婚姻随时间的变化过程。图 9 分不同初婚结婚年代描述了初婚婚姻持续时间与初婚离婚发生的关系。我们使用了 Kaplan-Meier 生存曲线来描述结合于同一年代的婚姻随着婚姻年份的变化能够维持下来的比例。首先,我们观察到中国的离婚比例总体上看仍不高,无论是 20 世纪 70 年代以前的婚姻,还是到 90 年代的婚姻,婚姻维持 20 年仍没有以离婚告终的比例均超过 90%。但我们还是可以看到随着时代推移,婚姻的不稳定性增加:80 年代、90 年代、2000 年以后的婚姻在前十年发生离婚的比例高于 70 年代或以前的婚姻。

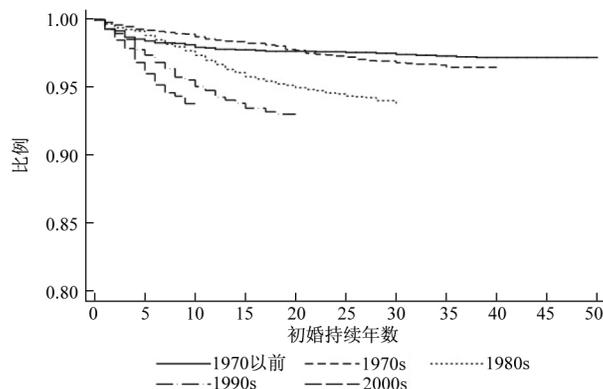


图 9:分初婚结婚年代随婚姻持续时间的初婚 Kaplan-Meier 生存曲线

表 3 描述了婚前同居比例随初婚年代的变化趋势。在 20 世纪 80 年代以前结婚的夫妻中,婚前同居的比例低于 2.0%,未婚同居的现象十分少见。但自 1980 年以来,婚前同居的比例开始迅速上升,80 年代为 5.0%,90 年代升至 12.1%,2000 年后已上升到 32.6%。换言之,在 2000 年以后结婚的夫妻中约有三分之一在婚前同居过。数据同时显示,发达地区(如上海、广东)婚前同居的比例高于欠发达地区(如甘肃),高教育水平者婚前同居的比率高于低教育水平者(许琪等,2013)。

CFPS 的数据显示,中国的婚姻制度总体来说仍然比较稳固,主要表现为初婚的普遍性和较低的离婚比例。尽管同居现象日益增多,但

并不意味着婚姻制度的松动。因为同居很可能是婚前的过渡状态或准备阶段,而不是替代婚姻的生活选择(Thornton, *et al.*, 2007)。

表 3:分初婚年代的初婚同居状况(%)

初婚年代	未同居	同居	合计(样本量)
1970 以前	98.2	1.8	100.0 (1 876)
1970~1979	98.0	2.0	100.0 (1 647)
1980~1989	95.0	5.0	100.0 (2 763)
1990~1999	87.9	12.1	100.0 (2 283)
2000~2012	67.4	32.6	100.0 (2 765)
合计	87.8	12.2	100.0 (11 334)

资料来源:许琪等(2013:312)。

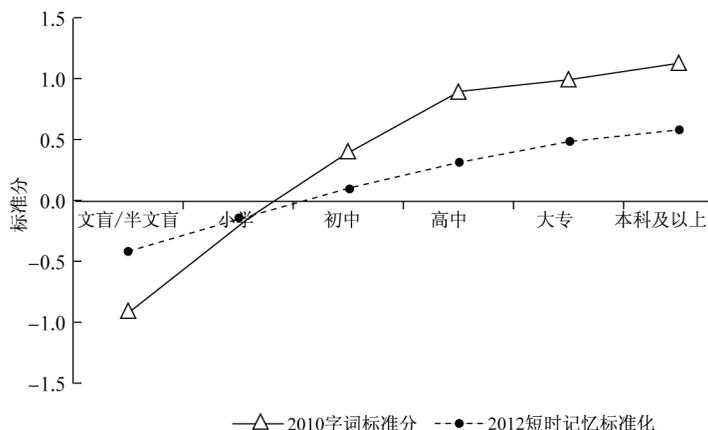
婚姻变迁是目前人口学、社会学分层领域和女性学领域学方兴未艾的研究方向。中国的婚姻变迁研究以前主要受到数据的限制,尤其缺乏反映整个婚姻历程及其细节的数据。CFPS 的婚姻史模块将回顾性和追踪性设计相结合,全面反映每段婚姻的去、现在和将来。其对同居、相识方式、配偶信息等数据的收集也为婚姻研究提供了更多可供分析的细节。

(三) 认知能力

在调查中引入认知能力的测试是 CFPS 的特色之一。目前国内的研究很少使用认知能力这一变量。此处,我们对认知能力测试的内容进行评估,并简要报告使用认知能力变量的初步发现。

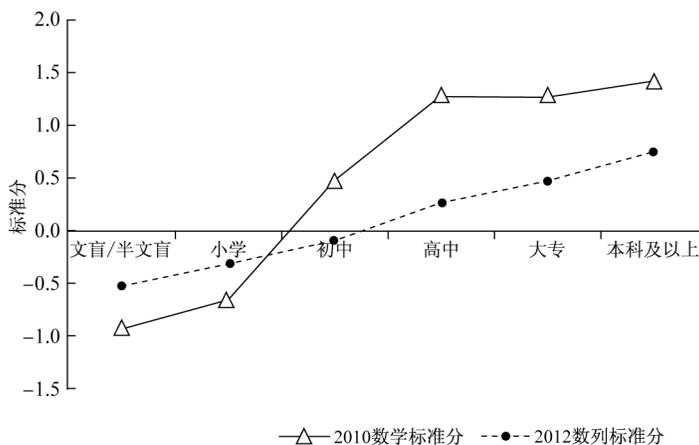
图 10 是分受教育程度的 2010 年字词测试和 2012 年短时记忆测试的标准化得分。图 11 是分受教育程度的 2010 年数学测试和 2012 年数列测试的标准化得分。2010 年与 2012 年测试的区别在于,2010 年的题目是根据受访者的受教育程度来选择相应的答题起点,测试题目也是基于中小学课本,因此其主要是反映受访者经学校教育获得的字词和数学计算能力;2012 年的题目则侧重测量受访者的记忆和算术推理这两项更为“内在”的认知能力。基于以上设计,如果 CFPS 认知测试得到的数据可靠,那么,我们应该观察到两套认知能力测试的分值都将与受教育程度正相关,因为认知能力强的人通常能够在学习中表现更好,也更可能取得较高的教育成就;此外,我们还应该观察到 2010 年的认知测试得分与受教育程度的关联性比 2012 年强,因为前者反映的是学校教育的影响。图 10 和图 11 均显示 CFPS 对认知能力的测量具有效度:在用年龄标准化后,两套测试的标准分均与受教育程度正相

关,且 2010 年的字词、数学测试得分与教育程度的关联更强,而 2012 年的记忆与数列测试得分随教育程度的变化则相对平缓。



资料来源:黄国英、谢宇(2013:127)。

图 10:受教育程度与 2010 字词测试、2012 短时记忆测试标准分分布



资料来源:黄国英、谢宇(2013:128)。

图 11:受教育程度与 2010 数学测试、2012 数列测试标准分分布

对 CFPS 认知能力变量的初步使用发现,在控制了少儿的性别、年龄、在学年级和户籍后,认知能力越强的少儿英语水平越高,其越轨行为(如抽烟、喝酒)和早恋行为的发生概率越低。对成人来说,在受教育程度、工作经验、户籍、地区等条件相同的情况下,认知能力越强者收入

越高,入党可能性越大。在60岁以上的老年人中,认知能力越好者,生活自理情况越好(黄国英、谢宇,2013)。

认知能力的社会科学研究在中国还在起步阶段。CFPS的认知测试填补了国内相关调查数据收集的空白。CFPS对认知能力的测量具有可靠性,为进一步探索认知能力与个体发展、地位获得、健康等社会结果之间的因果关系提供了可能。

以上对经济不平等、婚姻变迁与认知能力三个主题的介绍很好地反映了CFPS数据的研究潜力。更多的数据发现不在此处一一展示,我们希望能有更多的研究者能够使用和开发CFPS数据,并据此作出更有意义的研究,得到更有价值的结论。

五、总结

CFPS是目前中国规模最大和内容最全面的社会跟踪调查项目。为了让广大研究者对该项目及数据有更多了解,本文介绍了该项目的设计理念、实施过程、内容特色,并以个别领域的数据发现为例展现了该数据运用于社会科学研究的潜力。

在设计理念上,CFPS以个体和家庭为研究主体,以多层次问卷和追踪调查为视角,充分考虑了社会现象的差异性、嵌套性、复杂性和时间性。在调查实施上,CFPS采用城乡一体化抽样方式获得了具有全国代表性的样本;它借助先进的调查技术提高了面访的质量,并结合电访、网访技术有效降低了样本拒答和流失;它确立了合理的追访原则,提供了追踪调查的可操作性。在调查内容的设计上,CFPS不仅吸收了一般社会科学领域调查的大多数内容,还具有自身的特色,例如创新性的T表设计、与时俱进的城乡一体化问卷设计以及事件史、认知能力、少儿发展、访员观察等专题内容。最后,我们展示了使用CFPS数据研究经济不平等、婚姻变迁和认知能力的初步结果,这些例子反映了CFPS的追踪设计和特色内容在解答既有研究问题和拓展新议题上的潜力。

本文开篇时提到,中国目前经历的这场社会变迁史无前例。社会科学研究者应该客观、全面地记录这场变迁的过程,深入细致地研究变迁的机制和后果,为学术知识的积累和社会的改革与实践作出贡献——这是学者的使命也是时代的使命。收集高质量的社会变迁数据仅是一个开始,而更多的学者能够利用这些数据作出有意义的研究和

得到有用的结论才是 CFPS 的真正价值所在。

参考文献 (References)

- 胡鞍钢、胡琳琳、常志霄. 2006. 中国经济增长与减少贫困[J]. 清华大学学报(哲学社会科学版)(5):105-115.
- 胡兵、赖景生、胡宝娣. 2007. 经济增长、收入分配与贫困缓解[J]. 数量经济技术经济研究(5):33-42.
- 黄国英、谢宇. 2013. 认知能力[G]//中国民生发展报告 2013. 谢宇、张晓波、李建新、于学军、任强. 北京大学出版社:118-147.
- 李实、佐藤宏、史泰丽. 2013. 中国收入差距变动分析:中国居民收入分配研究 IV[M]. 北京:人民出版社.
- 李实、赵人伟. 1999. 中国居民收入分配再研究[J]. 经济研究(4):3-17.
- 任强、谢宇. 2011. 对纵贯数据统计分析的认识[J]. 人口研究(6):3-35.
- 孙妍、严洁、丁华、顾佳峰、刘月、姚佳慧、邹艳辉. 2011. 中国家庭动态跟踪调查(2010)访员培训手册[M]. 北京大学出版社.
- 吴琼、谢宇. 2013. 态度与观念[G]//中国民生发展报告 2013. 谢宇、张晓波、李建新、于学军、任强. 北京大学出版社:344-367.
- 吴要武. 2013. 独生子女政策与老年人迁移[J]. 社会学研究(4):1-25.
- 谢宇. 2012. 社会学方法与定量研究(第二版)[M]. 北京:社会科学文献出版社.
- 谢宇、胡婧炜. 2013. 导论[G]//中国民生发展报告 2013. 谢宇、张晓波、李建新、于学军、任强. 北京大学出版社:1-26.
- 谢宇、张晓波、许琪、张春泥. 2013. 收入分配[G]//中国民生发展报告 2013. 谢宇、张晓波、李建新、于学军、任强. 北京大学出版社:27-53.
- 许琪、李建新、于学军. 2013. 婚姻与家庭[G]//中国民生发展报告 2013. 谢宇、张晓波、李建新、于学军、任强. 北京大学出版社:305-343.
- 张晓波、张春泥、谢宇. 2013. 贫困与失业[G]//中国民生发展报告 2013. 谢宇、张晓波、李建新、于学军、任强. 北京大学出版社:54-91.
- Bian, Yanjie and Lulu Li. 2012. "The Chinese General Social Survey (2003-2008): Sample Designs and Data Evaluation." *Chinese Sociological Review* (45):70-97.
- Blau, Peter and Otis Dudley Duncan. 1967. *The American Occupational Structure*. New York: John Wiley & Sons.
- Chu, C. Y. Cyrus and Ruoh-Rong Yu. 2010. *Understanding Chinese Families: A Comparative Study of Taiwan and Southeast China*. Oxford, UK: Oxford University Press.
- Davin, Delia. 1996. "Gender and Rural-Urban Migration in China." *Gender and Development* 4(1):24-30.
- Duncan, Greg J., W. Jean Yeung, Jeanne Brooks-Gunn, and Judith R. Smith. 1998. "How Much does Childhood Poverty Affect the Life Chances of Children?" *American Sociological Review* 63(3):406-423.
- Elder, Glen H., Jr. 1985. "Perspectives on the Life Course." In *Life Course Dynamics: Trajectories and Transitions, 1968-1980*, edited by Glen H. Elder, Jr. Ithaca, NY: Cornell University Press:23-49.
- Fisher, Max. 2012. "Study: Income Inequality Skyrockets in China, Now among World's Highest." *The Washington Post*, Dec. 11 (<http://www.washingtonpost.com/blogs/worldviews/wp/2012/12/11/study-income-inequality-skyrockets-in-china-now-among-worlds-highest/>).
- Gan, Li, Zhichao Yin, Nan Jia, Shu Xu, Shuang Ma, and Lu Zheng. 2013. *Data You Need to*

- Know about China: Research Report of China Household Finance Survey · 2012*. London: Springer.
- Griliches, Zvi. 1977. "Estimating the Returns to Schooling: Some Econometric Problems." *Econometrica* 45(1):1-22.
- Kalmijn, Matthijs. 1998. "Intermarriage and Homogamy: Causes, Patterns, Trends." *Annual Review of Sociology*(24):395-421.
- Lareau, Annette. 2011. *Unequal Childhoods: Class, Race, and Family Life, with an Update a Decade Later*. Berkeley, CA: University of California Press.
- Lesthaeghe, Ron J. 2010. "The Unfolding Story of the Second Demographic Transition." *Population and Development Review* 36(2):211-251.
- Lieberman, Stanley. 1992. "Einstein, Renoir, and Greeley: Some Thoughts about Evidence in Sociology: 1991 Presidential Address." *American Sociological Review* 57(1):1-15.
- Lu, Yao. 2012. "Education of Children Left Behind in Rural China." *Journal of Marriage and Family* 74(2):328-341.
- McLanahan, Sara and Christine Percheski. 2008. "Family Structure and the Reproduction of Inequalities." *Annual Review of Sociology* (34):257-276.
- Mu, Zheng and Yu Xie. 2014. "Marital Age Homogamy in China: A Reversal of Trend in the Reform Era?" *Social Science Research* (44):141-157.
- Mincer, Jacob. 1974. *Schooling, Experience, and Earnings*. New York: Columbia University Press.
- Orlik, Tom and Bob Davis. 2013. "Relief on China Growth Comes with Caveats." *The Wall Street Journal*, Jan. 18. (<http://online.wsj.com/article/SB10001424127887323783704578249371234336736.html>)
- Ravallion, Martin and Shaohua Chen. 2007. "China's (Uneven) Progress against Poverty." *Journal of Development Economics* 82(1):1-42.
- Sewell, William H. and Robert M. Hauser. 1975. *Education, Occupation, and Earnings: Achievement in the Early Career*. New York: Academic Press.
- Silverstein, Merrill, Zhen Cong, and Shuzhuo Li. 2006. "Intergenerational Transfers and Living Arrangements of Older People in Rural China: Consequences for Psychological Well-Being." *Journal of Gerontology: Social Sciences* 61B(5):S256-S266.
- Thornton, Arland and Donald Camburn. 1987. "The Influence of the Family on Premarital Sexual Attitudes and Behavior." *Demography* 24(3):323-340.
- Thornton, Arland, William Axinn, and Yu Xie. 2007. *Marriage and Cohabitation*. Chicago: University of Chicago Press.
- Yu, Jia and Yu Xie. 2012. "The Varying Display of 'Gender Display'." *Chinese Sociological Review* 44(2):5-30.
- Yu, Jia and Yu Xie. 2013. "Changes in the Determinants of Marriage Entry in Post-Reform Urban China." *PSC Research Report* 13-802 (September), University of Michigan.
- Whyte, Martin K. 2004. "Filial Obligations in Chinese Families: Paradoxes of Modernization." In *Filial Piety: Practice and Discourse in Contemporary East Asia*, edited by Charloittelkels. Stanford, CA: Stanford University Press.
- Xie, Yu. 2011. "Evidence-Based Research on China: A Historical Imperative." *Chinese Sociological Review* 44(1):14-25.
- Xie, Yu. 2013. "Population Heterogeneity and Causal Inference." *Proceedings of the National Academy of Sciences (PNAS)* 110(16):6262-6268.
- Xie, Yu and Emily Hannum. 1996. "Regional Variation in Earnings Inequality in Reform-Era Urban China." *American Journal of Sociology* 101(4):950-992.

责任编辑:张 军