

中国民生发展报告

2018~2019

谢宇 张晓波 涂平 任强 黄国英 / 著

CHINA FAMILY PANEL STUDIES 2018-2019

中国民生发展报告

2018~2019

谢宇 张晓波 涂平 任强 黄国英 / 著

CHINA FAMILY PANEL STUDIES 2018-2019

目 录

致 谢	1
导 论	谢 宇 / 1
一 大数据与社会调查	2
二 CFPS 数据的价值	5
三 本书章节介绍	6

· 第一部分 综述报告 ·

第一章 收入分配	赵晓航 黄国英 谢 宇 / 11
一 2016 年家庭纯收入的水平和结构	12
二 2014 年、2016 年家庭收入水平的变化	15
第二章 家庭支出与消费	涂 平 张春泥 / 21
一 家庭支出水平、构成及分层概况	21
二 家庭支出构成类型	26
三 家庭支出类型的相关因素	29
四 本章小结	31
第三章 迁移	王 非 / 35
一 人口迁移的整体情况	36

二 乡城迁移与城镇本地劳动人口对比分析	47
三 本章小结	50
第四章 性别差异	於 嘉 / 54
一 教育	55
二 工作与社会参与	61
三 家庭	73
四 本章小结	75
第五章 婚姻与家庭	陈 沁 张冰莹 周 羿 / 80
一 初婚	81
二 同居	88
三 离婚	91
四 婚姻与劳动力市场	94
五 本章小结	98

· 第二部分 专题研究 ·

第六章 少儿的身体发展	黄国英 / 103
一 少儿的身高、体重发育	104
二 少儿的健康发育状况	108
三 家庭收入的影响	112
四 本章小结	113
第七章 农村家庭子女性别与医疗照料	阎 淑 任 强 / 115
一 子女性别、年龄与疾病报告	117
二 子女性别与就医、住院治疗及医药花费	118
三 年龄与性别差异	122
四 同胞竞争与性别差异	125
五 本章小结	129

第八章 家庭背景对子女教育成就的影响	柳皓然 李汪洋 谢 宇 / 133
一 文献与研究设计	134
二 数据与变量	136
三 分析与结果	137
四 本章小结	145
第九章 课外辅导	张月云 / 147
一 课外辅导的参与率	149
二 课外辅导的经济支出	153
三 课外辅导的具体内容	157
四 本章小结	161
第十章 家庭养育观念及其影响因素	李汪洋 柳皓然 / 164
一 家庭养育观念：数据与测量	166
二 养育子女的观念	167
三 生养子女的理由	171
四 本章小结	174
第十一章 离异家庭中的少儿	张春泥 / 177
一 离异单亲家庭的分布与特征	181
二 离异家庭的养育投入	185
三 离异家庭的少儿表现	191
四 父母争吵与少儿表现	200
五 本章小结	204
· 第三部分 方法介绍 ·	
第十二章 中国民营企业家特征分析	张晓波 李德洗 王雨濛 / 213
一 中国企业创新创业调查 (ESIEC)	215

二	民营企业家基本特征	217
三	民营企业家的价值观特征	222
四	中国河南省民营企业的经营情况	224
五	本章小结	227
第十三章	CFPS 方言编码报告	黄国英 / 231
一	CFPS 方言信息和地址信息	232
二	地址信息匹配方言编码过程	234
第十四章	问卷调查中的访员观察	吴琼 甄祺 / 239
一	访员观察：用途及采集方法	239
二	CFPS 个人自答问卷的访员观察	242
三	人群分布特征	249
四	本章小结	251
第十五章	CFPS 样本追踪设计及执行效果	孙妍 管亚男 / 254
一	CFPS 样本追踪设计	255
二	CFPS 样本追踪情况	259
三	本章小结	262
结论	从早期到将来	谢宇 张春泥 / 265

致 谢

“中国家庭追踪调查”项目资助：北京大学、国家自然科学基金

北京大学中国社会科学调查中心理事会

高松、王博、陈宝剑、张平文、周辉、刘波、张静、李强

北京大学中国社会科学调查中心

主任：李强 副主任：任强、赵耀辉

北京大学中国社会科学调查中心学术委员会

谢宇（主任）、白建军、陈玉宇、丁小浩、国光、郭志刚、刘爱玉、刘德寰、乔晓春、姚洋、张拓红、章政、曾毅、朱天飏、周雪光、周晓林

调查技术咨询：William Axinn、Beth-Ellen Pennell、杨倩

“中国家庭追踪调查”国际学术咨询委员会

谢宇（主任）、陈松蹊、国光、唐文方、吴晓刚、Robert Hauser、James Heckman、James Lepkowski、Narayan Sastry、Duncan Thomas、Donald Treiman、Robert Willis、Jean Yeung

“中国家庭追踪调查”项目负责人

谢宇、张晓波、涂平、任强

“中国家庭追踪调查”2016年调查问卷设计

总负责人：谢宇

家庭经济问卷：涂平、张晓波

成人问卷：谢宇、任强、孙妍

少儿问卷：谢宇、吴琼、孙妍

家庭成员问卷：孙妍

社区问卷：涂平

问卷修订工作：孙妍、吴琼、张春泥、李汪洋、吕梦捷

“中国家庭追踪调查”项目 2016 年调查执行

调查执行协调：涂平、孙妍

调查技术合作机构：北京大学图书馆

实地执行配合机构：民政部政策研究中心

访问管理：丁华（负责人）、陈敏燕、叶雪、张雅欣、葛新兴、王京、周滢滢、魏晓雯、徐丽波、仇新晨、孙彩琴

调查质量支持：丁华（负责人）、滕学亮、王琪尧

调查信息管理系统：姚佳慧（负责人）、孙帅、于双、张永建、洪洋、刘昊、王冰

调查数据支持：任莉颖（负责人）、吕萍、王雪音、王隆玉、王堃、马文婷

调查后勤支持：陈敏燕（负责人）、李国华、贾丹丽、马耘、邵娜

调查数据整理：吴琼（负责人）、吕萍、戴利红、张聪、张婧申、甄祺、阎淑、谷丽萍。

在此我们也向其他为“中国家庭追踪调查”项目做出贡献的学者表示感谢，他们是：邱泽奇、刘世定、李建新、沈明明、陈育德、黄桂田、雷晓燕、李国平、宋新明、萧群、徐湘林、杨伯淑、袁瑞军、岳昌君、张千帆、邹艳辉、于学军（原卫生和计划生育委员会）、郭震威（原卫生和计划生育委员会）、韩俊丽（原卫生和计划生育委员会）、李培林（中国社会科学院）、李友梅（上海大学）、李路路（中国人民大学）、蔡禾（中山大学）、郭有德（复旦大学）、雷洪（华中科技大学）、李实（北京师范大学）、罗楚亮（北京师范大学）、刘精明（清华大学）、刘玉照（上海大学）、仇立平（上海大学）、汪华（华东理工大学）、王广州（中国社会科学院）、翁定军（上海大学）、张燕（华东理工大学）、祝建华（香港城市大学）、Colette Browning（Monash University）、Michael Carter（University of Wisconsin-Madison）、David Lam（University of Michigan）、Arland Thornton（University of Michigan）、Nora Schaeffer（University of Wisconsin-Madison）。

此外，我们还要感谢聂雨琪、徐嘉树、黄国英、张春泥、张春雪、刘雯、谭晨昕为《中国民生发展报告 2018 ~ 2019》所做的文字编辑及校对工作。

导 论

谢 宇*

近些年来，随着“大数据”这一概念在国内的普及，越来越多的来自各学科领域的学者意识到调查与量化科学研究对中国研究的重要性，这种意识超越学术界，影响了整个社会。很多学者、学术机构或社会企业都进行过采集数据等一系列尝试，其中不乏值得肯定的工作成果，但绝大部分调查通常是区域性或专题性的，样本量有限，样本代表性有限，无法满足多样化的研究需要。早在2006年，北京大学就成立了中国社会科学调查中心（以下简称“调查中心”），致力于社会科学领域的调查数据采集和对中国社会问题的实证研究。依托调查中心，中国家庭追踪调查（China Family Panel Studies, CFPS）项目得以立项^①，该项目力图为学术界、为所有对中国社会感兴趣的学者，乃至社会各界人士提供公开、免费、高质量的数据。以CFPS数据为基础，自2009年起，调查中心推出“中国民生发展报告”（以下简称“报告”）系列^②，对与中国民生相关的主要议题及热点话题进行全面描述和分析性解读。其中，我参与写作并负责主编了2013年、2014年、2016年的报告，并有幸于2018年继续主持报告的编写工作。

目前，CFPS已成功实施了四期全样本追访工作，前三期追访调查分别

* 谢宇，北京大学讲席教授、社会研究中心主任，美国普林斯顿大学社会学教授、当代中国中心主任。

① CFPS项目自2006年开始前期筹划与准备工作，并于2008年、2009年进行了两轮预调查，2010年正式开始实施并完成了基线调查的执行工作。

② “中国民生发展报告”系列丛书在2009~2016年间共出版8辑（每年一辑），前三辑报告名称为《中国报告·民生》，从2011年起，该系列丛书成为教育部哲学社会科学系列发展报告资助项目，自此更名为“中国民生发展报告”。

于 2012 年、2014 年、2016 年实施，最新一期 2018 年的数据尚未发布。^①因此，本书的各章节内容主要基于 2016 年及此前的数据，对中国社会近几年的民生热点进行全面、客观的分析。此外，本书也将为各位读者介绍和分享 CFPS 的调查方法与经验。

一 大数据与社会调查

近年来，“大数据”概念的兴起和传播一定程度上让一些人担忧“传统”社会调查在社会研究中的地位会有所动摇，甚至盲目地认为社会调查已经过时，在研究价值上比不上大数据。诚然，大数据与计算科学的发展的确带来了研究方法的创新，为一些议题的研究开拓了新局面，也引领了一些过去社会科学难以涉足的新领域的发展。但就大数据的特征和现阶段的发展而言，将其运用于社会研究仍存在不少局限。

很多人误以为大数据就是大样本乃至海量样本的数据，因此会将许多大规模问卷调查也归于此类，这种看法显然没有真正理解大数据的含义。所谓“大数据”，是我们现代生活或工作留下来的数字脚印（footprint）。现代化生活中诸多电子化的生活和工作方式（比如打电话、网购等）为我们提供了可以保存和利用电子化数据的机会。从这个角度说，“大数据”其实并不是为研究而产生，而是我们日常生活留下的痕迹。以此为前提，我们不难发现，大数据有以下几个典型的特征：首先，它并不是一个新生事物，数据本身在很早以前就已经存在了。比如美国早在 18 世纪就有人口普查，有各种档案和文本，虽然这些信息并没有完全电子化，但是它们已经被记录并保存下来，是可以实现电子化的。

其次，大数据的重点并不在于“大”。虽然我们的生活每时每刻都在不断地产生数据，积累了海量的样本，但这个意义上的“大”并不是它的主要特征，大数据的主要特征在于其产生方式不同，它是随着我们的日常生活和工作自然而然地产生，并不是专门为了研究而产生的。

最后，大数据具有“被动性”。有时候我会将大数据称作被动数据，因为它不是为研究而有目的地收集的，而是被动产生的。比如，打电话的通话记录会被保留，在这一情境里，人们并不是为了研究而去记录通话数

^① 2018 年下旬，CFPS 发布了跨年核心变量库（2010 ~ 2016 年）。

据，打电话只是日常工作和生活的一部分，随之产生的数据是被动记录的。

在中国，长期以来，由于各种各样的原因，民众很难了解到真实的数据。在没有其他数据来源的情况下，大数据的出现和普及使大家有了一些了解现实情况的渠道。它打破了过去数据严重缺乏或被垄断的局面，在采集方式上使从前一些难以收集和掌握的数据（如交通、社交网络数据）的获取变得更容易，给人们提供了过去没有的数据、新的知识和观察视角，这种从无到有的变化，是它的一个重要贡献。但尽管如此，从研究的角度来说，大数据也有缺陷。

1. 代表性问题

大数据的代表性是有缺陷的，在研究中，我们难以得知大数据所代表的总体是什么以及如何分布。比如，有的人用手机，有的人不用；有的人一天发很多条微博，也有人一条都不发；有的人有几个银行账户，有的人没有账户……如果产生数据的人和不产生数据的人之间不存在差异，我们尚可以忽略代表性的问题，但如果产生数据的人和不产生数据的人之间存在系统性的差异，就会产生选择性问题，这是海量的样本难以解决的。因为这些行为数据是被动记录的，很难进行人为控制，只能用统计的方法尽量减少偏差。我们所获取的大数据的总体情况究竟如何、总体是否随时间的变化而发生变化、能不能进行推论、推论的对象是谁，这些异质性问题和选择性问题都很难解决。

2. 可靠性和质量问题

大数据的变量可靠性和数据质量也是不确定的。比如，在机器学习的文本分析中，表达正面情绪的一个形容词和一个主语放在一起，机器认为它是正面的，但它实际上可能表达的是讽刺或者其他的含义。人只有完整读完这段话才能理解它背后的情绪，但机器不一定能读出来，诸如这样的干扰（noise）还有很多例子。此外，海量的数据还意味着数据清理的代价也很大，也许很多人不做数据清理，可能也不做数据验证（verification）。但若不做这些最基本的工作，数据质量很可能存在问题。

3. 测量（measurement）问题

在传统调查搜集数据的实际操作中，一些问题的设计与提问本身就存在一定的难度，比如提问与情感、态度、观念有关的问题等。而在被动数据中，只有表达了强烈的情绪，才能通过数据读出这个人的情感。可是一般情况下人的情感未必有那么强烈，也未必会直接进行表达，而且很难用设计来把内在（隐私）的东西挖掘出来（这也是不合法的）。大数据作为

被动数据，难以对上述问题进行细致的处理，也不能通过设计来改进，所以它的测量是有限的。

4. 数据整合（data integration）和数据厚度问题

大数据的分散性造成了数据整合的困难。假设我们拥有电话数据和银行账户数据，将这两类数据整合起来看似可以获取用户的很多信息，为服务用户和分析研究提供支持。但事实上，有些用户的钱未必都存在自己的账户里，或者电话号码并不是用本人的身份证注册的，诸如此类的情况使整合后的数据存在很多问题，难以用于分析和研究。还有很多类似的情况，比如学历、病历等。合并这些数据是非常困难的。首先，电子数据的产生时间并不长，可以使用的信息有限；其次，大数据的变量非常少，不足以用来全面描述和分析人们长期的生活轨迹、生活事件和后果之间的关联。从这个角度来说，大数据虽然“大”，但信息仍太“单薄”，缺乏厚度。

就上述方面，传统的社会调查数据较之大数据则拥有许多优势。调查数据是依据一定的研究目的设计并有针对性地采集而获得的。在概率抽样调查中，总体和样本框是明确的，样本选取的过程是已知的，即便是有失访或拒访的人群，这些人群的一些基本特征也是能够被了解并且被纳入对数据质量的评价中。社会调查中的变量测量方案通常是严格根据研究需求而设计的，有些调查可能有几千个变量，涵盖研究对象的各个方面，甚至可以根据调查需要，从不同受访者的角度来采集更可靠或更相关的信息。相比之下，大数据大多记录的是单一方面的行为痕迹（比如微信、微博），研究者对个体的其他特征了解得非常少，又缺乏长期跟踪，很难深入探索异质性、因果性等问题，在数据厚度上难以与传统社会调查媲美。

不过，所有社会现象的数据都是有局限性的，这也是为什么人们有必要采集不同类型的数据，从不同方面、不同角度看待同一个问题（谢宇，2018）。特定的数据服务于特定的目的，如果使用数据的目的是为了推断个体层次的行为，则大数据的价值在于能够提供个体某些方面的行为细节，因此它在诸如个性化的广告推送、个性化的政策实施等领域大有作为。但如果使用数据的目的是为了推断总体特征，样本代表性还是优先于细节上的准确性。因此，谈到数据的利与弊，我们应该时刻提醒自己使用数据是为了服务于研究哪些问题。在社会科学领域的大多数研究中，最重要的仍是把握个体的异质性（谢宇，2006）。由于个体的异质性很强，若要探究个体特征，需要不同的测量，包括对基因、出生状况、家庭背景、

教育经历、职业生涯、中老年生活等整个生命历程 (life course) 的各个方面数据的采集, 并且是长期的、追踪性的调查, 这种数据采集工作是基础性的, 需要长时间的积累, 目前中国仍缺乏这种持续多年的追踪调查 (panel survey)。

二 CFPS 数据的价值

作为一项全国性追踪调查, CFPS 长期跟踪收集个体、家庭、社区三个层次的数据, 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、认知能力和社交交往等诸多主题 (谢宇、胡婧炜、张春泥, 2014)。

正如前文所述, 追踪调查最重要的价值之一是可以追踪调查对象完整的生命历程, 了解调查对象的特征、行为、经历是如何随着时间的推移而发展变化的 (谢宇等, 2014)。作为追踪调查的个体是反映人类社会变异性的最基本的单元, 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体层次的收入、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) 最终也可以通过个体层次的变化体现出来。所以, 要了解社会, 就必须先了解社会中不同的个体长期以来在角色、地位、行为和态度等方面的差异和变化。CFPS 的首要分析单元即是个体——基因成员及其核心亲属, 它针对这些个体采集了全方位、多角度的回溯性和追踪性的数据。

在追踪个体生命历程发展变化情况的同时, CFPS 也追踪了家庭和社区的变化。中国社会的结构具有多层次性, 这体现为个体嵌套在层层的社会组织 (如工作单位、村庄、家庭等) 中。在这种嵌套结构中, 家庭是最基础也是对个体产生最直接影响的社会组织, 我们对中国社会的认识与研究往往不能将研究对象置于家庭之外。因此, 家庭也是 CFPS 中一个重要的研究对象和调查单元。CFPS 采用了可明确定位亲属关系、描述家庭结构的方法, 对家庭关系和家庭成员展开了全方位的深度调查。同时, 它也详细采集了家庭层次上经济社会生活各方面的数据。CFPS 在家庭数据采集上的优势很大程度上丰富了该调查经验资料的层次, 为研究者分析中国社会提供了更多有价值的信息 (谢宇、胡婧炜、张春泥, 2014)。除家庭外, 个

体的生命历程也与其所生活的社区环境密切相关。相比于个体和家庭，社区的变化往往更难以观测。首先，一个社区的形成是由一个个单独的个体或家庭彼此在自觉或不自觉的居住选择基础上互动形成的结果，每个社区都是在这种动态过程中形成和发展的；其次，在一个已经成型的社区内，个体或家庭的行为活动、发展轨迹并非仅单向地受社区环境的结构性制约，个体或家庭的行为及其互动也会直接或间接地引发社区环境的变化。CFPS 力图借鉴西方学者发展的调查方法 (Axinn et al., 1997)，尽可能真实、完整地记录中国社区的变迁，为研究个体、家庭、社区的多层互动及其影响留下宝贵的资料 (谢宇、张晓波、徐宏伟、靳永爱、胡婧炜, 2014)。

三 本书章节介绍

本次出版的《中国民生发展报告 2018 ~ 2019》正是依托 CFPS 数据来呈现中国社会在多个重要领域的面貌与变化。本书共分为 15 章。考虑到读者的不同背景，我们将本书分为综述报告、专题研究、方法介绍三个部分，以满足不同读者的阅读需求。

综述报告 (第一至五章) 涵盖了收入与支出、人口迁移、性别、婚姻等与民生发展息息相关的议题。第一章“收入分配”描述了 2016 年中国家庭收入的水平与构成，以及从 2014 年至 2016 年收入的变化。第二章“家庭支出与消费”主要根据 CFPS 2016 的家庭支出数据对家庭进行分类，描述及概括了家庭支出的几种主要模式及其特征。第三章“迁移”聚焦乡城迁移人口，系统描述了人口迁移的整体情况，在与城镇本地人口的对比中呈现了处于劳动年龄的乡城迁移人口的特征。第四章“性别差异”主要从教育、工作与社会参与、家庭等方面对我国男女两性的差异进行了描述和分析。第五章“婚姻与家庭”则基于 CFPS 2016 描述了中国人婚姻行为的几个典型特征，并着重探讨了婚姻和生育行为与女性劳动参与的关系。

专题研究 (第六至十一章) 聚焦少儿发展,^① 探讨了家庭对少儿身心发展、教育与行为的影响。其中，第六章“少儿的身体发展”报告了少儿

^① 自 2019 年起，“中国民生发展报告”系列丛书将以“综述报告”、“专题研究”、“方法介绍”这三个模块为固定结构，每年的“专题研究”部分都将更换不同的专题进行重点探讨，今年的主题为“少儿发展”。

身体健康的不同方面，描述了少儿在身高、体重、生长发育迟缓率、体重过高或过低等指标上的发展状况，及其在性别、城乡和家庭收入上的分化。第七章“农村家庭子女性别与医疗照料”主要利用 CFPS 2014 农村多子女家庭的数据，探究了中国农村家庭对患病少儿的医疗决策与照料行为上的性别偏好，以及年龄和同胞竞争因素对少儿医疗照料的性别差异的影响。第八章“家庭背景对子女教育成就的影响”以家庭与少儿教育发展为主题，分析了家庭背景与少儿学业成就之间的关系。第九章“课外辅导”探讨了少儿的课外辅导参与率、课外辅导支出及其在城乡与家庭之间的差异。第十章“家庭养育观念及其影响因素”重点探讨了中国家庭养育观念以及家庭社会经济地位对其的影响。第十一章“离异家庭中的少儿”较为全面地评估了当今中国父母离婚对孩子发展的影响，重点探究了中国离异家庭的特征，父母离异对家庭教育和养育投入的影响，并评估和考察了离异单亲家庭和双亲家庭少儿在学业、在校表现、心理特征、越轨行为等方面的表现，以及父母婚内争吵对少儿各方面发展的影响。

方法介绍（第十二至十五章）部分主要向读者介绍和分享了 CFPS 使用的一些重要技术和变量处理方案，同时也介绍了 CFPS 与其他一些数据的整合。其中，第十二章“中国民营企业特征分析”主要介绍了中国企业创新创业调查数据如何与 CFPS 2016 数据相整合来刻画中国民营企业和民营企业家的特征。第十三章“CFPS 方言编码报告”提供了基于汉语方言分类谱系构建的方言类型的标准编码，以此作为对原编码体系的补充，弥补了原有方案的缺失信息。第十四章“问卷调查中的访员观察”介绍了大型问卷调查中访员观察数据的用途、采集方法、质量，并以 CFPS 为例，报告了个人问卷中访员观察数据的采集、用途和分布特征，为数据用户使用该数据提供了参考建议。第十五章“CFPS 样本追踪设计及执行效果”从设计与执行两个方面出发，具体介绍了 CFPS 在样本追踪上尝试过的方案及效果，并与国际同期同类型调查相比较。

尽管我们在本书中尽可能多地向读者呈现了有关中国社会变迁的重要议题和数据发现，但这些分析仍然是初步的。要认识和理解当代中国社会，仍然有待更深入、细致的研究。希望各位读者能够从本书中获得一些启发，也希望 CFPS 数据能够为更多学者的实证研究提供支持，产生更多优秀的成果。

参考文献

Axinn, William G. , Jennifer S. Barber, & Dirgha J. Ghimire. 1997. "The Neighborhood History Calendar: A Data Collection Method Designed for Dynamic Multilevel Modeling." *Sociological Methodology Sociological Methodology* 27 (1): 355 - 392.

谢宇, 2006, 《社会学方法与定量研究》, 北京: 社会科学文献出版社。

谢宇, 2018, 《专访谢宇教授(上): 大数据的重要价值不是“大”》, <https://mp.weixin.qq.com/s/qSYuKZA6nbVuT9WE4e4pdw>, 2018 - 08 - 24/2019 - 02 - 22。

谢宇、胡婧炜、张春泥, 2014, 《中国家庭追踪调查: 理念与实践》, 《社会》第 2 期。

谢宇、张晓波、徐宏伟、靳永爱、胡婧炜, 2014, “导论”, 载谢宇、张晓波、李建新、于学军、任强著《中国民生发展报告 2014》, 北京: 北京大学出版社。