

中国健康与养老追踪调查  
– 2011-2012 年全国基线调查用户手册

赵耀辉

**John Strauss**

杨功焕

**John Giles**

**Peifeng (Perry) Hu**

胡以松

雷晓燕

刘嫚

**Albert Park**

**James P. Smith**

王亚峰

2013 年 4 月

## 前言

这本数据使用手册描述了中国健康与养老追踪调查 2011-2012 年全国基线调查的全程进展，从问卷设计、调查实施到数据发布都囊括在内。我们编写这本手册的主要目的是帮助数据使用者更好地理解 and 运用这套数据。

中国健康与养老追踪调查(China Health and Retirement Longitudinal Study, 简称 CHARLS)是对中国中老年人进行的一项调查，调查对象是随机抽取的家庭中 45 岁及以上的人。调查旨在建立一套高质量的公开的微观数据库，数据包括从广泛的社会经济状况到个人健康状况方面的信息，满足老龄研究的需要。

CHARLS 是在健康与养老调查(Health and Retirement Study, 简称 HRS)的基础上进行的调查，与之相关的类似调查有英国老年人追踪调查 (English Longitudinal Study of Aging, 简称 ELSA)，欧洲大陆 19 个国家的欧洲健康、养老和退休调查 (Survey of Health, Aging and Retirement in Europe, 简称 SHARE)。考虑到在全国做调查的复杂性，我们先选取两个省做预调查。这两个省分别是内陆较穷的省份甘肃省和沿海富裕的省份浙江。预调查调查样本来自 32 个县/区的 95 个社区/村庄，共 1570 户家庭中的 2685 个人。预调查最终产出了一套高质量的调查数据，证明了在中国进行健康与养老类型的调查是可行的。在预调查的经验基础上，CHARLS 项目于 2011-2012 年进行了全国基线调查。为了保证样本的代表性，CHARLS 基线调查覆盖了全国 150 个县、区的 450 个村、居。我们成功访问了 10257 户家庭的 17708 个人，总体上代表中国中老年人群。

## 致谢

中国健康与养老追踪调查(China Health and Retirement Longitudinal Study, 简称 CHARLS)是一项巨大的工程, 需要很多人投入很大的精力。在此, 我们要感谢 CHARLS 研究团队的所有成员、实地调查的所有人员和每一位受访者。感谢您们为 CHARLS 项目投入了大量的时间、精力和热情, 感谢您们对 CHARLS 项目的理解与支持!

CHARLS 项目汇集了国内外许多知名学者的智慧、精力和汗水。该项目负责人是北京大学国家发展研究院(中国经济研究中心)赵耀辉教授, 另外两位负责人是南加州大学 John Strauss 教授和中国疾控中心的杨功焕教授。世界银行的 John Giles 博士、洛杉矶加州大学的 Perry Hu 博士以及香港科技大学的 Albert Park 教授也是项目的副负责人。南加州大学的 Eileen Crimmins 教授是 CHARLS 补充资助项目的副负责人, 主要负责血样分析数据(这部分的使用手册后期会陆续公布)。兰德公司的 James P. Smith 博士是国际顾问委员会的主席, 他在整个研究中发挥了重要作用。

CHARLS 的研究团队成员负责问卷不同模块的设计, 同时在项目的其他方面包括从培训材料准备、培训到实地调查, 他们都为之投入了极大的精力。除了项目的主要负责人和副负责人, 项目的研究团队还包括北京大学中国经济研究中心(CCER)的雷晓燕、沈燕和李力行教授, 清华大学的施新政教授, 中央财经大学的吴晓瑜教授以及世界银行的王德文博士。

CHARLS 实地调查工作由胡以松副研究员统筹管理。项目 IT 吴云在由兰德公司的 Albert(Bas) Weerman 培训后, 负责带领整个 IT 团队将问卷编到 CAPI 系统中。由北京大学中国社会科学调查中心丁华老师领导的督导团队以及实地调查助理刘嫚、袁畅负责访员的招募、培训和实地执行, 整个调查共有近 500 名调查员, 他们的辛勤工作换来了我们今天见到的高质量数据。王亚峰博士后负责实地的数据质量控制、后期的数据清理和样本权重构建。有十多位的学生参与了数据后期处理工作, 包括检查数据、清理数据、回访、对开放性问题编码和准备数据发布等。在此, 我们要谢谢他们的劳动和付出! 项目财务金海玉老师负责 CHARLS 的账户、调查员工资支付以及合作方的费用支付等工作。

CHARLS 项目得到了国内外机构的资助：美国健康研究院老龄化研究所（NIA）社会与行为研究部（项目编号：1-R21-AG031372-01、1-R21-AG033675-01-A1、1-R01-AG037031-01、1-R01-AG037031-03S1）、国家自然科学基金管理学部（项目编号：70773002、70910107022、71130002）、世界银行北京代表办公室（合同编号：7145915）以及世界银行 Knowledge for Change Program 项目（合同编号：7159234）。在此，我们要向资助方致以最诚挚的感谢！

此外，CHARLS 还得到了国内外许多机构和个人的重要帮助。感谢 NIA 社会与行为研究部的 Richard Suzman 博士、John Phillips 博士等，感谢国家自然科学基金管理学部李一军、高自有副主任，感谢以李海征为主任的自科基金数据项目学术委员会的大力支持。我们还要感谢北京大学副校长刘伟，感谢北京大学原副校长林建华教授，感谢校长助理、北京大学 985 项目办公室李晓明主任，感谢社科部和社会科学调查中心主任李强教授，感谢国发院名誉院长林毅夫教授、前任院长周其仁教授和现任院长姚洋教授，感谢国家统计局、卫生部、劳动和社会保障部、民政部、教育部和城市建设部的领导。

由兰德公司 James Smith 博士为主任的 CHARLS 国际、国内顾问委员会还给予 CHARLS 极大的帮助，他们是 James Banks、Lisa Berkman、David Bloom、Axel Borsch-Supan、Arie Kapteyn、Jinkook Lee、David Weir、Robert Willis、David Wise、周其仁、蔡昉、Scott Rozelle、邬沧萍、姚洋和左学金教授，在此一并表示感谢。

我们感谢北京大学中国社会科学调查中心的同事孙妍、张越、陈敏燕等为项目提供的支持和帮助，感谢中国疾病预防控制中心（CDC）的王临虹处长、施小明处长、尹香君副研究员和殷召雪助理研究员等在项目健康部分的设计和实地执行中提供的大力支持，感谢 28 个省、市疾病预防控制中心以及 150 个县、市疾控中心、卫生部门和 450 个村居协调人为 CHARLS 提供的重要帮助。深深地感谢您们！

当然，如果没有和受访者的理解和支持，整个调查是不可能这么顺利地进行。这套调查数据不仅能为中国老龄化问题研究提供基础，而且能为我国政府制定未来的社会保障制度提供参考。所以，我们还应感谢所有受访者的参与和支持！

## 目 录

前言 .....	2
致谢 .....	3
1. CHARLS 概况 .....	7
1.1 背景及重要性 .....	7
1.2 本书结构 .....	8
2. 家户调查总体介绍 .....	9
2.1 家户调查的内容 .....	9
2.2 抽样 .....	13
2.2.1 县级抽样 .....	13
2.2.2 村、居抽样 .....	14
2.2.3 家户抽样 .....	15
2.2.4 个人抽样 .....	16
2.3 实施情况 .....	16
2.3.1 实地执行情况 .....	16
2.3.2 问卷负担的说明 .....	17
3. 社区调查 .....	17
3.1 社区数据的基本形式 .....	17
3.2 社区调查问卷 .....	17
4. 个人模块间以及家庭和个人与社区间数据联系 .....	19
5. 职业和行业编码 .....	19
6. 样本权重 .....	20
附录 A 实地程序 .....	22
A.1. 问卷设计 .....	23
A.2. 建立计算机辅助个人调查 (CAPI) 系统 .....	23
A.3. 两次试调查 .....	23
A.4. 人员招募 .....	24
A.5. 培训和相关材料准备 .....	24
A.6. 体检及生物样本收集 .....	25
A.7. 实地调查 .....	25
A.8. 质量控制 .....	27
A.9. 数据清理 .....	28
附录 B CHARLS-GIS 介绍 .....	29
附录 C 代理受访者 .....	33
附录 D 体检详细描述 .....	34
表 1 家户问卷收集的数据 .....	38
表 2 测量指标和设备 .....	39
表 3 样本量和应答率 (%) .....	39
表 4 受访者的数量和年龄/性别构成 .....	40
表 5 分性别和年龄的代理率 .....	41
表 6 分性别/年龄体检率 .....	42
表 7 每个模块完成时间中位数 .....	43
表 8 是否参与体检的 Logit 回归 .....	44
表 A.1 研究组名单 .....	45

表 A.2 国际和国内顾问委员会成员 .....	46
--------------------------	----

## 1. CHARLS 概况

### 1.1 背景及重要性

我国是世界上老龄人口规模最大的国家，也是世界上老龄化速度最快的国家之一。据预测，到 2050 年，我国 60 岁以上的老年人占总人口的比重将由 2000 年的 10% 增加到 30%，而老年人口赡养比（25-64 岁的壮年人口规模/65 岁以上的人口规模）将由 2000 年的近 13:1 下降到 2.1:1（联合国, 2002）。

随着中国人口的快速增长，老龄人口的供养日益成为一个重要问题。经济快速发展的特征之一是年轻一代的终身收入显著地超过他们年老父母，老龄人口成为中国最大的弱势群体之一。与此同时，中国的生育控制政策意味着相比过去中国的老年人由更少的子女供养。如何解决为老年人提供良好生活保障的问题是近年来中国社会快速发展面临的巨大挑战之一。

应对这一挑战，中国政府采取强有力的措施。最近几年，一系列新的社会保障体系得以实施，特别是在健康服务领域。这类政策包括：最低生活保障体系，新型合作医疗制度，城市居民医疗保险制度，低收入居民医疗协助，等等。尽管此类政策并非专为老龄人口设计，但老龄人口无疑是最重要的受益群体之一。与众多其他政策相似，这类政策由中央政府发起，地方政府在实施过程中保留一定自主权。地方政府可就初步试验和全面推动的时间自主决策，也可采取不同的实施计划。CHARLS 在家庭和社区两个层面衡量这些社会保障体系的存在并对其进行分析，为政府修正和完善现行政策提供更加科学的基础。

目前，对中国老龄政策的科学研究还处在起步阶段，最大的障碍在于缺乏足够的微观、纵向的数据。现有的数据倾向于专业化，不具备社会科学分析必需的数据宽度。举例说，健康数据以健康测量为中心，缺乏社会经济的指标；另一方面，社会科学学者收集的数据则缺乏对健康的衡量。老年人的福利与他们的健康和社会经济状况密切相关，同时，健康和社会经济水平自身又相互关联，具有广度和精度的微观数据对于中国的老龄问题研究尤为必要。CHARLS 是填补这一空白的一次尝试。

## 1.2 本书结构

手册的第 2 部分介绍家庭调查，主要是问卷内容和抽样程序。这部分也包括在家庭和个人层面的实地调查成果的简要说明。

第 3 部分介绍如何将不同家户模块以及家户和社区模块的数据连接方法。第 4 部分描述讨论职业和行业编码。第 5 部分描述如何建立抽样权重。

附录 A 描述调查过程的细节，从问卷设计、试调查、调查员培训、实地流程、调查后期工作到数据公开发布的准备。附录 B 介绍 CHARLS 怎样使用 GIS 软件进行村、居内样本框的构建。附录 C 对怎样使用代理访问进行了描述。附录 D 对体检以及相关的设备进行了详细介绍。

对收入、财产、家庭支出，社区和县、区层面的社会、经济和政策环境数据后期会陆续发布，同时也会发布收入、资产和家户支出的插值变量。另外，血样的数据还在进行分析，这些测量结果后期会在另外一本用户使用手册中发布。

所有 CHARLS 收集的数据都将保存在北京大学的国家发展研究院的中国经济研究中心。同时这些数据将在 CHARLS 项目网站 [charls.ccer.edu.cn](http://charls.ccer.edu.cn) 上对全世界研究者开放。

2008 年 CHARLS 试调查数据从 2009 年 4 月份就公开发布，截止到目前数据使用者有 2600 人，其中 1/4 来自国外。如果您想从网上申请公布的数据，您可以在网上签署一份数据使用协议，并提供您的基本信息如姓名、地址、机构、联系电话和 email 等进行注册。3 个工作日内您的申请会被核查，如果被批准，您将收到一封 email 告诉您用户名和密码就可以进行数据下载。如果您对部分没有公开发布的数据感兴趣，请直接联系我们进行申请。



## 2. 家户调查总体介绍

### 2.1 家户调查的内容

中国健康与养老追踪调查(以下简称 CHARLS)旨在建立一个公开的、高质量的,具有全国代表性的微观数据库,提供一套关于老年家庭以及老年人及其配偶个人的全面信息。CHARLS 数据可以进行跨学科分析。CHARLS 问卷在设计上参考了与美国健康与养老调查(HRS)以及世界上其他老龄化调查项目(英国老年追踪调查(ELSA)以及欧洲的健康、老年与退休调查(SHARE)),具有国际可比性的同时,结合了中国的具体国情。

CHARLS 问卷由八个模块组成:(a) 家户登记表,(b) 基本信息,(c) 家庭,(d) 健康状况与功能,(e) 医疗保健与保险,(f) 工作,退休与养老金,(g\h) 收入,支出与资产,(i/j) 住房特征和访员观察(表 1)。

以下是每个模块的主要内容:

#### 0. 过滤问卷

过滤问卷是整个问卷的基础,主要用来判断备选户是否有 45 岁及以上的受访者。如果住在样本住址里有多个家户,并且不止一个家户有适龄受访者,系统将随机抽取一户进行调查,然后列出此户所有的家户成员及其年龄,如果只有一个适龄受访者,他/她直接成为“主要受访者”,如果有多个适龄受访者,则随机抽取一个作为“主要受访者”。如果他/她有配偶,则不管配偶年龄多大也自动成为受访者。过滤问卷的回答者没有年龄限制,也不要求是家庭成员,只要熟悉家户情况即可。

过滤问卷部分有以下四个目标:

第一,确认家户是否是我们的样本户。通过核对绘图阶段采集到的姓名、地址和家户照片以及询问家户是否收到“致居民的一封信”实现这一目标。

第二,确认抽中住宅里的家户数量以及每个家户在同一村居范围内居住的住宅数量。这一信息用来做抽样以及调整抽样的权重。因为家户样本的抽样框是住宅抽样框,可能有多个家户住在同一个住宅里。这种情况下,我们抽取住宅里有适龄家户成员的一户做调查。

第三,列出随机抽取的家户里所有的家户成员并选择主要受访者。在家户成员里选择一个年龄在 45 岁及以上的人作为主要受访者,他/她的配偶(如果有的

话)自动成为受访者。主要受访者和配偶的基本信息都需要收集,包括姓名、性别、出生日期和婚姻状态等。

第四,确定模块的回答者。在家户调查中主要受访者及其配偶各自回答个人的模块,除了个人模块,问卷还有家庭层面的模块,需要熟悉家庭信息以及熟悉家庭收支状况的人各自回答。我们将这两部分问卷的回答者称为“家庭受访者”和“财务受访者”。家庭受访者需要回答“家户登记表”和“家庭”模块,通常是最为熟悉家庭情况和家庭经济往来信息的主要受访者和配偶之一。财务受访者需要回答家庭收入支出与资产模块,这个人必须最熟悉家庭收支情况的家庭成员。

### 主问卷模块

基线访问的 CAPI 系统将主问卷划分为有顺序的几个模块。通过过滤问卷确定了每个模块的受访对象后,CAPI 系统根据受访者的状态加载相应的模块和问题。访问中有一些模块依赖于其他特定模块,必须特定模块完成才能开始。有一些模块的问题需要引用家户登记表和家庭部分的信息,因此这两个模块必须先完成。完成它们之后可以按照方便的原则自由选择访问顺序。家户问卷更详细的信息请参见网站 (<http://charls.ccer.edu.cn/>)。

#### A. 家户登记表

这一模块由家庭受访者回答,收集除主要受访者及其配偶以外的其他家户成员的个人信息。主要内容包括家户成员的性别、出生日期、婚姻状态以及与受访者的关系,户口类型,户口所在地,教育程度以及有限的迁移史。

#### B. 基本信息

这一模块收集主要受访者及其配偶的个人信息,主要受访者及其配偶都需作答。问卷询问的内容主要是:出生日期与出生地、户口情况、受教育程度、相关的迁移史、居住情况和婚姻状态。对于离婚或者丧偶的受访者,问卷还将询问其前任配偶的基本信息:出生日期、教育背景、离婚/丧偶的时间、配偶去世的原因。

#### C. 家庭模块

这部分收集所有家庭成员(父母兄弟姐妹和子女)(不仅仅是家户成员)的个人信息,不管这些人是否住在这个家里。另外家庭成员之间时间和经济往来的

详细信息也需要收集。关于主要受访者及其配偶的父母和子女，问题包括是否健在、血缘关系（是否亲生父母/子女），出生地，成长环境，出生日期，属相，去世日期，教育背景，劳动力信息，居住地，户口状况以及是否拥有房产。另外收集受访者家庭与不住在一起的父母、子女、其他亲戚和非亲戚的朋友之间的实物以及金钱往来的信息，对孙子女以及父母提供的照料信息。

最后一部分问卷询问受访者的居住安排偏好，询问受访者：假定一个老年人有成年子女和配偶，另外一个老年人有成年子女但没有配偶，何种方式对他们的生活安排最好。

主要受访者和配偶均需要回答这部分问题。

#### D. 健康状况与功能

##### D1 健康状况：自评健康

这部分大量的题目是自评健康状况，包括自报一般健康状况、是否有医生诊断患有特定的慢性病，以及受访者是否遭遇过交通事故或跌倒。我们还询问与健康相关的行为，如吸烟、喝酒和身体活动（包括锻炼和日常身体活动），我们也关注日常生活活动（ADL）和工具性日常生活活动（IADL）以及身体功能的测量。ADL 和 IADL 部分收集帮助者信息，这部分也有精神健康（抑郁）以及认知能力问题。为了解释一般健康问题，我们随机抽取一半的家庭询问健康情景题，情景题用来得到受访者评价自己健康时参照的初始指标。最后，我们询问受访者活到某一特定年龄的主观期望。

##### D2 健康状况：体检

因为自报健康变量可能存在误差，我们对每一个受访者进行身体检查。包括：身高、体重、腰围、上臂长和膝高。我们还测量三次血压和脉搏、三次肺峰流速（使用肺风流仪）、每只手两次握力（使用握力器）、坐起站立时间、60 岁及以上受访者的走路速度以及平衡力（更多信息请参看表 2 和附录 D）。访员记录测试结果并反馈给受访者，如果有一些指标不正常，访员会建议受访者看医生进一步确认。

我们还会对签署知情同意书的受访者采集全血，全血储存在零下 80 度的低温冰箱中，会用来分析 C 反应蛋白，糖化血红蛋白，血糖，总胆固醇，高密度脂蛋白胆固醇，低密度脂蛋白胆固醇，甘油三脂，血红蛋白，胱抑素 C，血肌酐，尿素氮和

尿酸。血样分析完成后会对外发布并与 CHARLS 主问卷数据进行连接。这部分数据也将会发布另外的用户手册。

#### E. 医疗保健与保险

这部分收集医疗以及预防性的医疗信息，包括过去一年预防性医疗服务的利用、过去一个月的门诊治疗情况、过去一年的住院情况以及过去一个月的自我治疗。详细信息包括：治疗地点、距离、总花费、自付花费和医保报销的部分、药物花费等。每个受访者还会被问到现在和过去享受的医疗保险，以及是否没有医疗保险。通过这些信息研究者能够了解新的保险项目的覆盖率。

#### F. 工作退休和养老金

这部分记录受访者目前的工作状况（工作、现在不工作以前有工作以及一辈子从未工作过）并收集关于劳动力供给、工资和单位福利、从雇主那里得到的社会保险等。如果受访者有副业，关于这些工作的部分信息也会收集。对于现在不工作的受访者，收集上一份工作的信息。最后收集简短的工作史信息。我们会收集详细的退休信息，并区别正常退休（养老保险）和实际退休（退出劳动力市场），另外，还询问关于养老金的详细信息。

#### G. H. 收入支出与资产

##### GB, HA. 家户收入、支出与资产

这部分询问难以区分为某个人的家户层面的收入、支出与资产。家户总支出，包括自产产品的消费。家户总收入包括详细的农产品收入和支出，非农个体经营的净收入。资产分为：房产、土地、家庭设备和耐用消费品以及金融资产。这部分的资产仅指家户共同所有的资产，收集的信息包括除主要受访者及配偶之外的其他家户成员的财富。这部分的回答者是“财务受访者”，对家庭财务最为熟悉。

##### GA, HB. 个人收入支出

这部分询问主要受访者和配偶的个人收入和财富，询问自己单独拥有的资产，个人收入和个人债务。这部分还特别询问受访者是否在二十世纪九十年代按单位补贴价购买了房改房。

#### I. 住房特征

这部分询问房子的特征，包括房子的类型、有多少个台阶，是否有入户无障碍通道，是否有厕所、自来水和电、能否上网等以及房子的清洁度和温度。

## J. 访员观察

这部分主要记录访员对访问过程的个人观察，受访者回答问卷的配合程度以及是否在理解问卷上有困难。

## 2.2 抽样

CHARLS 基线调查在全国 28 个省 150 个县区的 450 个村、居开展。CHARLS 样本代表中国 45 岁及以上住户人群，机构中的老年人并没有进入抽样，但是基线的受访者在后续的调查中如果进入到机构中将会被随访。所有的样本通过 4 个阶段被抽出来。

### 2.2.1 县级抽样

第一阶段，除了西藏以外，全国其他所有的县区通过地区排序，在地区内按照城市和农村进行排序，然后再通过人均 GDP 进行排序。地区是基于国家统计局关于省级地区的分类。排序后，会列出每个县、区的人口以及累计人口。如果  $N$  是所有县级抽样单位的总人口，需要抽取的县区是 150 个，我们定义一个间隔为  $n=N/150$ 。第一个县区的抽取是先选择一个 0 到 1 之间的一个随机数，然后选择第一个累计人口大于  $r*n$  的县区作为第一个县区。然后将  $n$  加上  $r*n$  这个起点，第二县区就是累计人口大于  $r*n+n$  的第一个县区。第三个县区以此类推，在  $r*n+n$  的基础上再加上间隔  $n$ ，选择累计人口大于  $r*n+n+n$  的第一个县区。图 1 是 CHARLS 的县、区在全国的分布。

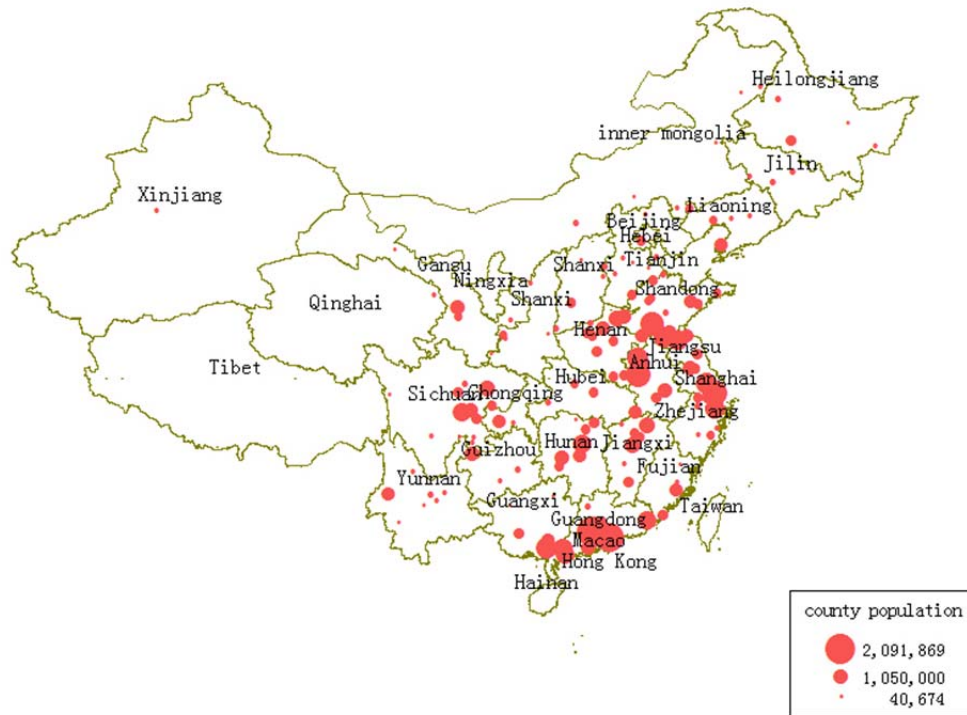


图 1 CHARLS 抽中县、区在全国的分布情况

### 2.2.2 村、居抽样

在农村地区样本选自行政村（村），城市地区以社区（社区）作为首选抽样单位（PSUs），社区包括一个或多个原居民委员会（居委会）。每个县级单位选取 3 个 PSU，使用 PPS（规模比例概率）抽样。需要注意的是，农村的县同时包括农村村庄和城市社区，城市地区同样可能包括农村村庄。对于每个县级单位，PSU 列表随机排序。随后列出每个 PSU 的人口与累积人口（该 PSU 人口加上前面所有 PSU 人口数）。如果  $N$  是这个县级单位的人口总数，抽样 PSU 数量是 3 个，则定义区间  $n=N/3$ 。第一个 PSU 在 0 到 1 之间随机选择数  $r$ ，选择第一个社区使累积人口大于  $r*n$ 。以  $r*n$  作为起点，加上区间  $n$ ，第二个 PSU 使累积人口大于  $r*n+n$ 。第三个 PSU 再加上区间  $n$ ，累积人口数大于  $r*n+n+n$ 。这一过程在 STATA 中用 `samplepps` 命令执行。

人口数量较多的社区（大于 2000 户），基于地图的抽样框架花费较高，允许负责人选择该社区的一个地理分区作为 PSU，如一个或多个原居委会。抽样需包含足够的子社区以保证足够数量的合格抽样受访者。子社区的选择应基于每个子社区的估计人口，我们有 30 个社区进行了拆分。

因为原先的样本框存在后来又有社区的合并和拆分情况，在抽中的 450 个村、居中，我们将 6 个村居进行了替换，2 个村是因为搬迁，1 个社区居委会升级成

为县级的区，2 个社区几乎全部为集体住所，1 个为大学的宿舍区，另外 1 个为监狱，这些村居并不是我们的样本。样本村居的替换也是和其他村居完全一样的抽样方法。有 6 个县区行政边界发生变化，所以抽中的村、居分配在两个县区中。我们没有替换这些村、居，结果最终的县区数为 156 个。

### 2.2.3 家户抽样

在没有 PSU 中，我们从样本框中选择住户样本，样本框是在当地协调人的帮助下由绘图员根据地图构建的。为了准确获得每个村、居中家户的样本框，我们 CHARLS-GIS 绘图/列表软件。对于每一个 PSU，一名绘图员首先会携带 GPS 到社区收集边界，然后 CHARLS 项目办公室会使用边界信息抓取 Google Earth 底图，以此作为绘图和列表的基础。然后，每个 PSU 中的所有建筑物都会进行拍照和 GPS 定位，并对每个建筑物中的住户进行列表。集体性住所如军队、学校、宿舍以及养老院被排除在样本框之外。

CHARLS 项目总部会对每个 PSU 中的样本框进行核查，保证村、居内的所有建筑物都包括在内。核对后，督导会使用 CHARLS-GIS 软件随机抽取 80 户，以小红旗的方式标记在地图上，并发送回实地的绘图员进行信息收集，包括住户中最长人的年龄、户主的姓名、电话以及是否空户。最后，实际抽中的数量在每个 PSU 中会超过目标样本 24 户，因为某些住户可能没有 45 岁以上的受访者、或者空户、无应答。以 80 户中的收集的信息为基础，包括最长人的年龄、户主的姓名、电话以及是否空户等，督导会使用 CHARLS-GIS 软件在每个社区、村居抽取一定数目的住户。开始的样本是一个 80 户的随机样本，从这些户中我们计算适龄率、空户率指标。然后分别估算每个村居的样本数量，从整个样本框中抽取样本。

每个 PSU 的最终抽样工作结束后，抽中的住户信息再次返给绘图员，绘图员将这些信息在 CHARLS-GIS 软件上进行加载。然后，绘图员会给抽中的住户送“至居民的一封信”。同时，CHARLS 项目的 IT 会把每个村、居抽中的住户列表和地址导入到访员的 CAPI 系统中。

我们对每个 PSU 中的所有能够找到并愿意参与调查的适龄户进行了访问。某些住户单元有多个家户住在一起。这种情况下，我们随机选择一个有适龄受访者的家户进行调查。因为 PSU 中的住户因为找到与否、是否有适龄受访者或者愿意参加与否的不同，所以每个 PSU 最后完成的受访户数量也不同。这些都会在样本

权重上进行调整。

#### 2.2.4 个人抽样

在每个抽中的家户中，会有一个短的过滤问卷来确认家户中是否有适龄要求的受访者。如果一个家户中有多个超过40以上的人，我们随机选择一个。如果抽中的人年龄在45岁以上，他/她作为主要受访者，并同时访谈他或她的配偶。如果抽中的人年龄在40-44岁之间，他/她作为预留样本以后进行调查。如果适龄成员无法回答问题，我们指定一个代理受访者帮助他或她回答问题，如果受访者有配偶或知情的成年子女，通常作为代理受访者。无45岁及以上成员的家庭没有对其进行访问。

A 部分的家户登记表、C 部分涉及家庭结构和财产转移的问题由“家庭模块受访者”回答，可以是主要受访者或其配偶，最能准确回答这部分问题的个人尽可能被选为受访者。

同样，“财产受访者”被选择回答有关家庭收入、支出和资产等问题。对于此类问题，任何年龄在 18 岁及以上的家庭成员（包括主要受访者和其配偶）都可被选为“财产受访者”，同样采用对此类事项最知情者为受访人的原则。

### 2.3 实施情况

#### 2.3.1 实地执行情况

中国健康与养老追踪调查（CHARLS）基线调查于 2011 年 5 月至 2012 年 3 月在全国 28 个省、市施行。调查覆盖 150 个县区的 450 个 PSU（村庄或社区），在 450 个基本抽样单位中，52.67%是农村地区，47.33%是城市地区<sup>1</sup>。

基线调查抽取的样本包括 23,422 户住户（见表 3）。其中，4341 户住房无人居住。在 19081 户家庭中，12740 的家庭有适龄成员<sup>2</sup>。和预期一样，农村家庭（73.80%）相比城市家庭（59.04%）有较高的适龄率。

在适龄的家庭估计总数中，我们设法联系并得到 10257 户家庭应答，约占 80.5%（表 3）。余下 19.5%同样为适龄家庭，但无法联系占 8.2%和拒绝回答有 8.8%，其他无应答有 2.5%。这一回答率相比 HRS 第一次调查较好，并且明显优于 HRS 和 SHARE 的最近一次调查。应答率农村明显高于城市，这与其他低收入国家调查

<sup>1</sup> 城市地区与农村地区的定义是基于国家统计局的定义。城市地区指的是位于城市或城市郊区、城镇或城镇郊区或其他非农产业占了 70% 以上的特殊区域比如特殊经济带，国有农业企业等。

<sup>2</sup> 我们有一些没联系上的家庭的适龄人数信息，适龄率的计算是依据和假定没联系上的家庭的适龄率与此相同。



的经验一致。

表 4 说明了 CHARLS 样本的年龄与性别构成。我们拥有 17708 份个人数据，其中 52.1%为女性。大部分样本为低龄老年人，40%的年龄在 60 岁及以上。样本中 91.3%为直接访谈，8.7%通过代理受访者访谈（表 5）。

表 6 描述了体检的完成率为 78.9%。部分受访者身体状况欠佳而不能完成生理指标部分，也有个人特别是稍年轻的男性由于工作、旅行等原因不在家，或者在访谈时繁忙不能完成体检。此外有拒绝访谈的情况。但与 HRS 相比，体检的完成率较好。从表 6 可以看出，较年轻男性和年纪较大的女性不愿意接受体检。

### 2.3.2 问卷负担的说明

中国健康与养老追踪调查（CHARLS）基线调查相当复杂，覆盖到个人生活的几乎所有方面。不仅对适龄受访者收集信息，而且收集其配偶和其他家庭信息。我们的问题是，调查收集信息较多，问卷的负担影响回答的质量。表 7 列出完成每个问卷模块的平均时间，由 CAPI 自动记录。时间数据由每个模块的结束时间减去开始时间得出，由 CAPI 自动记录。这一时间包括访问中的暂停和干扰。从表 7 可以看出，家庭模块（C）和家庭收入、支出、资产模块（G2&HA）一般比个人模块花费更多时间，模块 D 健康状况和功能在所有五个个人模块中花费最多时间。

## 3.社区调查

为合理分析生活标准和行为，在家庭和个人层面之外还需要在社会层面获得信息。有益的信息包括价格、当地基础设施和服务的便利性，在社区之间改变的政策环境。CHARLS 引入社区调查，以服务这些目标。以下部分将介绍社区调查的形式和主要内容。

### 3.1 社区数据的基本形式

我们让所有选择农村社区和城市社区完成社区调查。绘图员与居委会或村委会负责人共同执行社区调查。为完成调查问卷，要求受访者查找关于村庄或居委会的特定统计数据，诸如自然环境、就业情况、金融情况等。

### 3.2 社区调查问卷

社区问卷收集了整个社区的社会、经济和政策环境等情况，主要包括以下几个部分：

(a) 基本信息

这部份问卷收集了组织结构、实际面积、自然地域以及土地权（农村地区）等。

(b) 基础设施和公共设备

这部份问卷收集了关于社区基础设施、公共设施的信息，包括道路、学校、邮局、健康设施、娱乐设施和公共交通等详细情况。问卷还询问了有关水、燃料、电力等在社区的使用情况。

(c) 人口和劳动力

这部份问卷询问了社区人口结构以及劳动力的详细信息，包括社区居民的教育水平分布和劳动力迁移情况。

(d) 企业和工资

这部分收集社区内的企业数量，询问分男女和教育程度的工资水平。

(e) 迁移

这部分收集迁移人群的地域分布，已经不同性别迁移人群的平均工资水平。

(f) 健康和保险

这部份问卷包括了社区所有的健康设施以及可及性，记录了社区居民享有医疗保险的登记手续以及社区内医疗保险的覆盖率。

(g) 社会政策

这部分收集的政策包括失业补助、最低工资、其他补助、独生子女补助、农业补助、林业补助、65 岁以上老人补助、80 岁以上高龄补助、新农保和老农保等。

(h) 社区历史

这部份会召集一些老年人进行座谈收集社区历史信息。问卷试图通过这社区经历的政策改革来描述这个社区政策变化的历史。问卷记录了社区不同政策改革的起始时间，如家庭联产承包责任制改革和计划生育政策。另外，也收集了知青下乡和赤脚医生的信息。

(i) 流行病和自然灾害

这部分收集过去 5 年和从 1945 年来发生的自然灾害和流行病，记录每个事件开始的年份和受影响的人数。

(j) 产品、收入和物价水平

这部份问卷收集了调查当时的市场价格，包括食物、能源和住房等与当地居民生活密切相关的物品价格。

(k) 访问员观察

这部份问卷是由绘图员主观作答。问题包括这个社区的社会经济状况（从贫穷到富裕），街道的整洁、拥挤程度，残疾人设施和普通话的流利程度等，每一问题设置了七个等级。

#### 4. 个人模块间以及家庭和个人与社区间数据联系

不同个人和家庭模块间数据可以通过家庭或个人在每个数据文件中使用的标示连接。“循环”文件包括了家庭层面的数据，而“主调查”文件包括了个人层面的数据。

家庭标识符是一个 9 位数号码。这是虚构的数字，前两位代表省份，接下来两位数字代表省内的市级代码，第 5 和 6 位数字代表市内的县区。用户使用县的固定或随机效应模型，可以从前 6 位数字区分不同县的家庭。第 7 位数字代表县内的村级编码。

xx	Xx	xx	x	xx
省	市	县	村	家庭

家庭 ID

同样，这些都是 CHARLS 中虚构的数字。前 7 位数字是村级的唯一代码，同时作为我们的社区 ID。社区 ID 用以将家庭和个人数据连接到社区模块。最后，第 8 和第 9 位数字是社区内家庭编号。个人识别码以家庭编号加上两个额外的数字来区分在家庭内受访者或配偶。

xx	Xx	xx	x	xx	Xx
省	市	县	村	家庭	个人

个人 ID

#### 5. 职业和行业编码

在劳动情况模块，用受访者的原话描述其所在的工作行业。调查员没有必要

立即对职业进行分类，因为由调查员做好这件工作也是非常困难的。

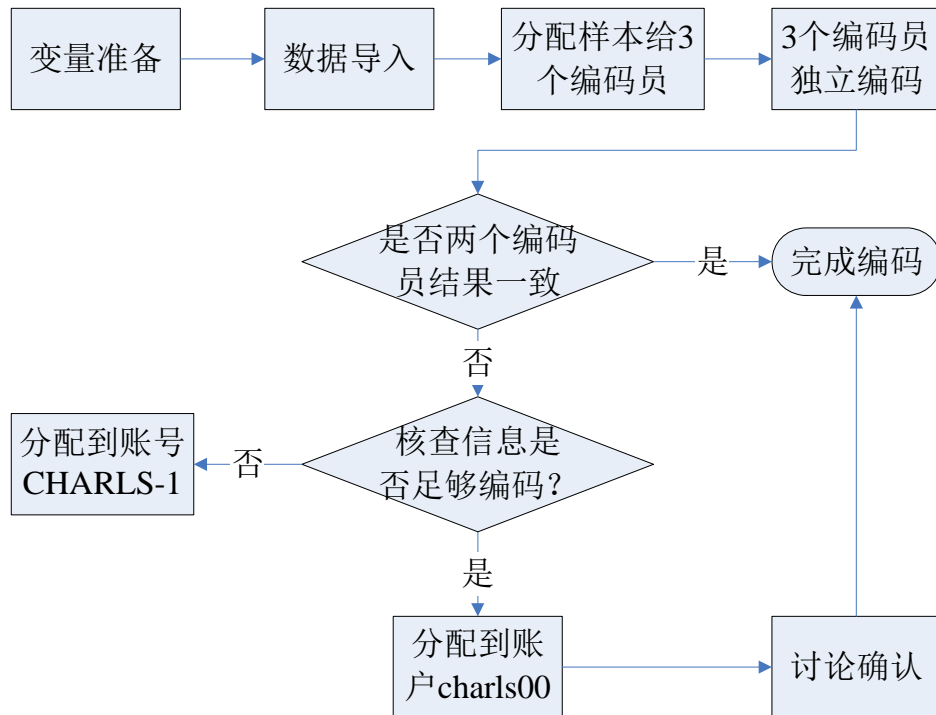


图 2 职业和行业编码流程图

调查结束后，我们开发了网上的编码系统进行编码工作。首先，将工作模块的数据导入到编码系统，然后在编码审核员的管理下，编码员进行编码。通常情况下，3个编码员作为一组同时、独立的对一组数据进行编码。当该组完成工作后，编码审核员在系统中进行比对。如果对于一个记录三个编码员中有至少两个编码员编码一致，我们就接受这个编码。如果三个人的编码都不一致，这条记录就会进行核查，经3个有经验的编码员讨论后进行最终的编码。编码时采用标准的国标6位的编码分类表，我们对不能很好编码的职业加入了一些新的职业编码。同时，行业编码为2位数编码。我们在职业的中国编码和国际编码间做了一个对应。我们在数据发布了这两套编码。

## 6. 样本权重

我们构建了家庭和个人层面的样本权重。家庭层面的权重分为两类：一个校正了无应答，一个没有校正无应答。基于是否校正家庭以及个人层面的无应答，我们提供了三类个人层面的权重。同时针对体检数据提供了一个校正了体检无应答的个人权重。所有的样本权重都是直接通过抽样概率构建的。

鉴于 150 个县区是通过标准的 PPS 方法抽中的，某县区被抽中的概率就是该县区的人口除以中国的总人口，然后乘以 150。然后，我们使用相同的公式计算特定村级单位在选定县域内被抽中的条件概率，即该村占全县人口的比例乘以 3，3 是每个县内抽取的村的数量。如果首选抽样单元是子社区，则根据子社区人口在村内所占比例做进一步的调整。下一步，我们计算一个家庭在特定村内被选中的条件概率。这需要几个步骤。首先，我们用抽中的住房数除以村内住房总数，然后乘以拥有适龄成员的家庭的抽样比例，最后乘以住房内的家庭抽样率，这是由于同一住宅内可能包含多个拥有适龄成员的家庭。我们最后将这些条件概率（县的选择，村的选择和家庭的选择）相乘，并取其倒数来获得家庭的样本权重。

这一样本权重没有考虑访问中可能的无应答，但我们仍然提供这一权重以供用户选择，因为某些研究者可能倾向于不使用无应答校正，因为这一校正的渐近一致性依赖于一些比较强的假设（比如非混淆性）。我们利用逆概率加权因子乘以上述的家庭样本权重校正家庭层面的无应答。<sup>3</sup>

逆概率加权因子构建的方法如下，首先估计一个因变量为家户是否应答自变量为关于 PSU 的虚拟变量的 logit 回归（样本为所有抽中的适龄家户）<sup>4</sup>。我们原本可以在自变量中加入家庭和户主的特征，然而对于那些我们无法找到或者拒绝访问以致没有开始过滤问卷的家庭，这些特征都无法观察。我们进而根据这一 logit 回归预测每个家庭的应答概率，并取其倒数。为剔除过度加重的情况，我们在该应答概率倒数的第 99 百分位进行截取，从而得到逆概率加权因子。将其与上述未校正无应答的权重相乘就可以得到我们的无应答校正权重。这一校正是有用的，但其一致性需要一个选择是基于可观测变量的强假定（即没有基于不可观测变量的样本选择，Wooldridge, 2002 年）。

个人权重使用家庭权重作为其基础，在其所属家庭被抽中的情况下，除以个人被抽中的概率。可以使用两类家庭权重（校正了无应答的和没有校正无应答的）中的任何一个作为基础，获得两类个人权重。

个人在家庭内被抽中的条件概率取决于其年龄、婚姻状况及其所属家庭中适龄的成员数量。以仅有一个适龄成员的家庭为例，该成员被选中的概率是 1，因

---

<sup>3</sup> 参见 Jeffrey Wooldridge, 2002, *Econometric Analysis for Cross Section and Panel Data*, Cambridge: MIT Press

<sup>4</sup> 因为可能某些村庄内所有的家庭都参加或者都没有参加调查，所以我们需要加总一些村庄虚拟变量。在这种情况下，我们把一些附近的村庄或者至少在一个县的村庄加总进去了。

此将该家庭的基础权重除以 1 以得到个人权重。现在假设家庭内有两个 45 岁以上的未婚成员，每位成员被选中的概率是 1/2，所以将该家庭权重除以 1/2 以得到个人的适当权重。现在假设一个家庭拥有 2 名成员，都适龄并且已婚，每个成员都有 1/2 的概率被选为主要受访者，同时每人有 1/2 的概率被选为配偶。所以每人被选中的总概率为 1，我们用家庭权重除以 1 得到每个人的权重。最后，考虑拥有 3 个成员的家庭：A 是 65 岁的未婚女性，而 B 是一名 70 岁男性，C 是一个 60 岁的女性并且与 B 结婚。A、B、C 三人都三分之一的机会被选为主要受访者，但是 B 和 C 有另外 1/3 的机会被选为配偶。所以 A、B、C 三人被抽中的总概率分别是 A: 1/3，B、C 都是 2/3。我们将基础（倒数）家庭权重除以这些概率以分别得到 A、B、C 的个人权重。

我们也提供了个人水平上的无应答校正，我们使用同样的方法在个体水平上构建了逆概率加权因子。首先进行 logit 回归估计，以个人是否应答（使用应答家户中的所有适龄个体作为观察）作为因变量，以性别虚拟变量、年龄信息的虚拟变量、年龄（如果知道）、婚姻状态的虚拟变量以及关于 PSU 的虚拟变量作为自变量。然后根据该回归预测个人的应答概率，取其倒数并在第 99 百分位截取以得到个人水平的逆概率加权因子。

我们需要一组不同的权重来分析个人的生理指标，因为超过 20% 的个人没有参加体检。我们利用与家庭层面类似的逆概率加权进行体检的无应答校正。在这种情况下，logit 回归的样本是所有主要的受访者及其配偶，如果他们体检了则因变量等于 1。因为我们有这些个人的信息，从而可以控制更加丰富的自变量。我们仍然使用村一级的虚拟变量，但现在还包括受访者年龄的虚拟变量、性别及他们的交互项，教育水平的虚拟变量。回归结果列于表 8。老年男子和获得高中及以上教育的人参加体检的概率较小。将这一回归的预测概率取倒数并且在第 99 百分位对其截取，获得逆概率加权因子，最后用校正了家庭层面无应答的个人权重除以这一加权因子以得到体检权重。

## 附录 A 实地程序

本附录讨论了中国健康与养老跟踪调查（CHARLS）基线调查的实地程序和实地调查前的步骤。在基线调查实地执行之前，经过了一年半的准备工作。CHARLS

研究小组于 2009 年底在 2008 年的试调查基础上开始进行调查问卷的修订工作，在此期间，组织了一些小规模的小规模试验调查来测试和改进调查问卷和调查流程。2010 年 8 月和 12 月，我们使用 CAPI 系统进行了两次正规的试调查，然后对基线调查问卷进行了最终确定。有了这些试调查的经验，研究组在 2011 年 4 月完成了问卷和实地调查流程更新。此后，项目小组开始招募绘图员和访员，并编写培训教材。从 2011 年 5 月到 7 月，我们在北京大学进行了 10 期培训班。在 5 期的绘图培训班中，平均每个培训班有 3 个老师培训 30 名绘图员。另外 5 期的访员培训班中，平均有 6 个老师培训 60 名访员。当选定的调查员完成了培训，将他们派往全国各地完成实地调查。约 90% 的实地调查工作在 2011 年 9 月底完成，但是整个调查工作比预计的时间长，最终于 2012 年 3 月底全部完成。

### A. 1. 问卷设计

研究组非常认真的完成最终的问卷。在 2008 年甘肃、浙江试调查之后，研究组进行了数据分析，并在数据使用的基础上更新了问卷。在 2010 年北京和廊坊的两次正式的预调查后，问卷又做了修改。在最后编写 CAPI 程序和培训过程中，都对问卷做出了进一步修改完善。

### A. 2. 建立计算机辅助个人调查（CAPI）系统

CHARLS 项目使用的 CAPI 系统项目由兰德公司专家 Bas Weerman 和其训练的程序员吴云在 MMIC(LINUX 系统)下完成编写。我们的程序员不断重新调整程序，以对应于试调查后调查问卷的修改。在廊坊正式试调查后，问卷进行更新后，CAPI 程序也相应进行了更新和重复测试。全国基线调查的 CAPI 程序在 2011 年 6 月最后确定。

### A. 3. 两次试调查

我们在最终确定基线调查问卷之前，利用计算机进行了两次正规的试调查。2010 年 8 月 5-21 日，我们在北京是海淀区的两个社区（一个城市、一个农村）进行了一次正规的试调查。研究组的关键人员包括赵耀辉、John Strauss、施小明以及其他项目重要的成员都参加了这些试调查。我们招募了 15 名学生作为访员。整个培训持续了 12 天，现场调查工作持续了 9 天，共完成 29 户访问。试调查后，我们计算了问卷的时长，在此基础上对问卷进行了删减。2010 年 11 月 5

日到 12 月中旬，我们在河北廊坊市进行了另外一次的正式试调查。这次试调查的设计是整个基线调查工作的排练。我们选择了 3 个县区，每个县区一个村进行测试。我们从县区 CDC 招募访员，和我们实际的实地工作基本一样，每个县区包括一名绘图员，两名访员，一名血液分析人员以及一名协调人员。唯一不同的是，我们多加了一名血样收集人员。在培训访员之前，我们先培训绘图员。访员的培训持续了 9 天，访问工作几乎用了一个月的时间。我们完成了 110 户的访问，应答率为 86.6%。这些样本将用于我们第二轮调查的试调查追踪。廊坊试调查后，问卷最终确认。更新后问卷也得到了北京大学生物医学伦理委员会的批准。

#### A. 4. 人员招募

两次试调查后，我们决定单独招募调查员。我们每个县区需要 1 名绘图员，2 名访员，所以我们一共需要招募 150 名绘图员，300 名访员。为了使招募流程自动化，我们设计了网上招募系统，申请人可以在网上进行信息填写，我们进行初步筛选。北京大学中国社会科学调查中心的工作人员帮助我们在求职网站上（包括覆盖全国的网站、省级网站以及校园的 BBS 上）发布广告。很多情况下，我们地方高校的同事招募学生作为访员。大部分的绘图员和访员是年纪在 20 多岁的高校学生。选择绘图员和访员的标准是他们有努力工作的意愿，而且曾有过调查的经验。同时，也会强调沟通的能力，特别是要熟悉实地可能碰到的当地方言。

#### A. 5. 培训和相关材料准备

绘图员和访员分别分成 5 期培训班在北京进行培训。绘图员的培训是在 2011 年 5 月进行，每期 5 天。每天上午是理论课，下午是在附近的社区进行实地绘图练习。我们特意选择了一个建筑物复杂的社区。每天上午会进行测试，检验对前一天的课程的掌握情况。绘图员培训后会立刻下实地开展绘图工作。

访员的培训是在 2011 年 6 月和 7 月进行，每期 9 天，利用 DVD 教学。前 7 天是理论课和教室内练习。每天在课程快结束时，会有方便样本供学生进行现场练习。和绘图员培训一样，每天上午会进行测试，检验对前一天课程的掌握情况。培训第 8 天，和实地调查一样，访员会到附近的社区进行现场访问。培训的最后一天，他们会回到教室进行总结。质控组会根据从访员电脑传到我们服务器上的真实数据简单总结他们的访问效果。然后，访员们会交流前一天的调查经验。最



后，会对访员颁发培训证书和聚餐。当有访员退出或者被解雇后，我们会招募额外的访员进行额外的培训。

因为我们在实地工作开始前需要进行很多次培训，我们需要一套标准化的培训教材，保证所有访员无论何时何地接受同样的培训。研究组和执行组从 2010 年 5 月就开始准备培训手册。他们将每个培训部分的课程写了脚本，并经过多次修改完善。在 2010 年 8 月北京的试调查我们使用了这些培训材料。研究组培训过程进行了录像，所说的语言进行了记录并和书面的脚本进行比对修订。在对脚本进行编辑后，在北京大学电视台的帮助下，我们制作了一套 DVD。2010 年 11 月在廊坊的试调查，我们主要使用培训光盘进行培训以检验 DVD 培训的效果。廊坊试调查后，我们将每个 DVD 的时间进行了调整，改变了培训内容的顺序，优化整个流程。所有的培训脚本进行了更新，DVD 也由北京大学电视台重新制作。

#### A. 6. 体检及生物样本收集

甘肃和浙江 2008 年的试调查中，我们收集了详细的人体测量，包括身高、体重、腰围、膝高、上臂长（从肩到腕），以及几项反映力量和运动能力的体能测量，包括最大呼吸峰流速，握力，5 次重复站立的时间以及血压。利用其他资助，我们利用可携带的设备检测了血脂（总胆固醇和高密度脂蛋白胆固醇），并收集了干血斑进行了 C 反应蛋白和血红蛋白的检测。全国基线调查中，我们加入了平衡能力测试和步行速度测试，同时我们利用国内的资助收集了全血而不是干血斑。在北京的试调查中，我们包括了人体测量，体能测量和血压，没有收集血液样本。在 2010 年 11 月的廊坊试调查中，我们加入了血样的收集、运输和分析。我们在 2011 年 3 月通过竞争获得 R01 的血样分析资助。

基线血液采集的人员通过县区 CDC 进行招募、遴选，由国家 CDC 单独进行培训，因为这些人员需要学习血样收集的整个流程。采血人员的培训安排在 2011 年 6-7 月份，分别按照地区远近在 4 个省进行。当每个县区的访问工作结束后，我们将同意抽血的人员名单交给中国 CDC，县区 CDC 然后组织血样收集工作。

#### A. 7. 实地调查

##### 1) 获得伦理委员会批准和知情同意

项目组 2011 年 1 月向北京大学生物医学伦理委员会申请了伦理审查。在对知情同意部分进行修订后，CHARLS 项目获得了批准。

在实地期间，每个同意参与调查的受访户都必须签署两份知情同意，纸质的知情同意保留在 CHARLS 办公室，并进行扫描后存成 PDF 文件供以后必要的审查。我们对 4 个部分进行了知情同意，包括：问卷调查、体检、血样收集和血样储存分析。

## 2) 实地调查工作

2011 年 3 月 23 日 28 个省级 CDC 的启动会在北京举行，会上我们将 CHARLS 的工作计划进行了介绍。在 28 个省级 CDC 的帮助下，北京的 CHARLS 办公室收集了县区 CDC 联系人的名单和联系方式。这些县区联系人通过沟通和培训对 CHARLS 的目的和工作计划进行了了解。在绘图员到达抽中的 3 个村居进行绘图前，县区联系人和抽中的 3 个村居的负责人进行沟通联系，帮助绘图员协调当地村居干部，并协助绘图和列表过程中的给予支持和帮助。如果条件允许，县区联系人会协助收集村居建筑物的分布信息图。

在每个县区，一个绘图员会花一个月的时间进行 3 个村居的绘图/列表工作。在社区干部或村居联系人的帮助下，绘图员的主要工作有两个：1) 绘图/列表，2) 社区调查。本来的计划是在绘图员去现场前，根据村居的名称确定 Google Earth 的底图，然后将底图加载到绘图电脑上。然而，Google Earth 底图并没有包括村居边界信息，很难知道村居的底图应该包括多大面积。所以，我们决定首先让绘图员去实地利用 GPS 获得村居边界，然后在此基础上抓取底图。这个工作需要将图片文件来回在现场绘图员和项目总部之间发送很多次，同时进行质量核查。考虑到某些村居没有用于绘图的 Google Earth 底图，我们为这些地区单独制作了培训材料。最后，450 个村居，379 个 (84.2%) 有清楚可用的 Google Earth 底图，66 个 (14.7%) 村居底图不清楚，5 个 (1%) 没有底图。在这些底图不清晰或不存在的地方，绘图员利用 CHARLS-GIS 软件绘制底图，每个绘图员都进行了培训。绘图员还需要对村居的负责人进行访谈完成社区问卷调查。

一般情况下，2 个访员完全负责一个县区内 3 个村居的家户访问。访员利用绘图员提供的样本列表进行调查。在实地工作中，我们遇到一些合作困难的情况。有 23 个村居不愿意参与访问，我们利用个人关系和村居领导进行协调。最终，我们获得了 150 个抽中县区的合作。

当家户拒绝访问时，会有很多的访员派去解决这些困难，这种情况涉及 52

个村居。最困难的村居，我们最多派出了 6 批访员。在访问攻关的最后阶段，CHARLS 办公室的员工组队去了 11 个村居进行访问攻关。如图 3 显示，大部分的访问在 2011 年 8 月份前完成，但是基线调查工作一直延续到 2012 年 3 月底。

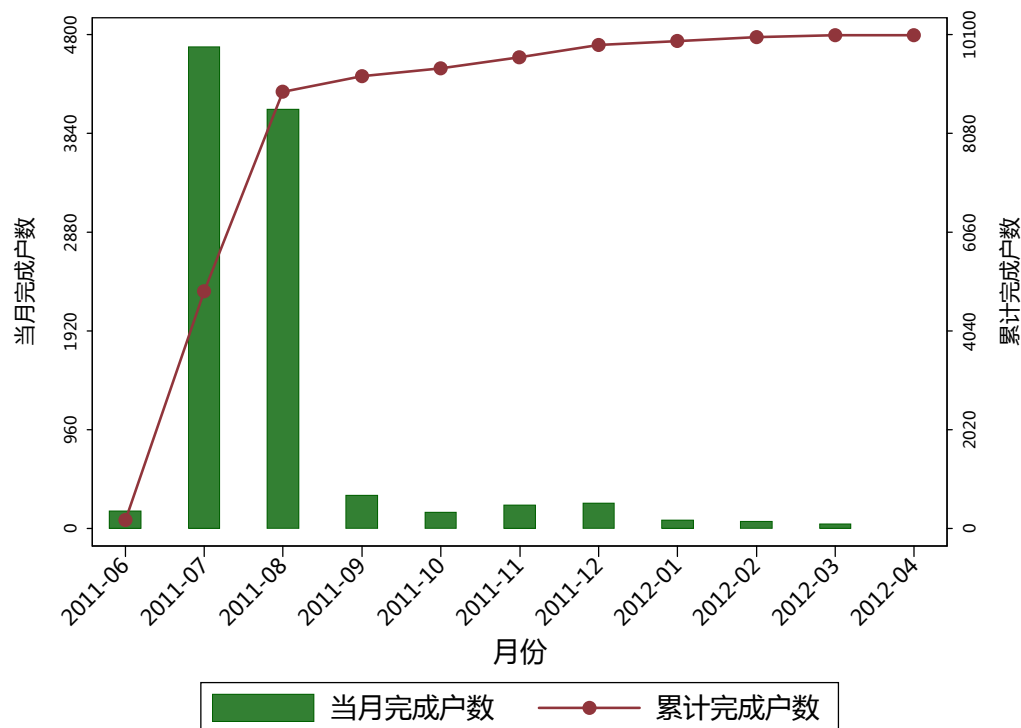


图 3 不同时间的 CHARLS 访问进展情况

当每个县区的访问工作结束后，我们将同意抽血的人员名单交给县区 CDC，县区 CDC 然后组织血样收集工作。

我们在调查中使用了计算机辅助个人面访系统（CAPI），每个访员有一个小电脑在访问的时候可以录入数据。我们不使用纸质问卷。CAPI 系统的能够显著的提高现场错误的发现。当访员录入一个逻辑错误或不正常的值时，系统会有弹出框提醒访员。CAPI 也会大量减少问卷的不正确跳转。访员在每天的实地工作结束后，会将数据上传到北京办公室的数据服务器上。

### A. 8. 质量控制

我们充分利用 CAPI 系统在实地期间进行质量核查，主要包括 4 个方法：GPS 比对、数据核查、录音核查以及电话核查。

GPS 比对是比较绘图和访员阶段收集的 GPS 信息。我们也使用绘图和访员阶段的照片以保证我们找对了抽中的家户。

在访问期间，CHARLS 总部的程序员会核查所有访员的数据，查看数据缺失过多和访员时间过短的情况。另外，所有访员的最初两户访问都会进行录音核查，从而确实进行了访问以及某些部分的题目询问是正确的。如果因为技术问题或其他原因不能进行录音核查，项目总部的人员会打电话到这些家户进行电话核查。例如，数据可能显示某些受访者一生都没有工作过，但这可能是因为访员错误。另外一个例子是某个受访者的整个模块都缺失，例如财产部分。

家户因为技术问题或其他原因不能进行录音核查的，可能的话会进行电话核查，确认实际上有没有访员去访问和进行数据收集。如果在数据核查、录音核查或电话核查时发现任何作弊行为，我们会对这个访员扩大样本核查。

### **A. 9. 数据清理**

除了上述讨论的职业和行业编码外，我们在实地调查工作结束后，做了一些基础的数据清理和数据核查。

我们核查了数据的跳转，某些问题基于之前的答案不应该再问。举例来说，一旦受访者告诉我们医生诊断了他或她的高血压，受访者是否知道他或她有高血压的问题就应该跳过。我们检查了数据中的所有跳转并纠正了不正确跳转的数据（如果存在的话，一般是因为 CAPI 系统问题）。

在此之后，我们检查数据了不一致。一个很好的例子是一个受访者在工作模块报告工资收入，但在个人收入模块没有报告收入。当我们获得用户的反馈后，我们会做进一步的核查。

## 附录 B CHARLS-GIS 介绍

在住户抽样阶段，国内没有高质量的村居水平抽样框。我们将绘图/列表电子化，并开发了 CHARLS-GIS 系统。

### B. 1 CHARLS-GIS 一般介绍

通过使用 Google Earth 和其他来源的底图作为底图，可以利用 CHARLS-GIS 软件完成建筑物绘制、住户列表、随机抽样，记录每个建筑物和每个住户单元的 GPS 信息，并对建筑物和住户单元的门拍照。软件优于纸笔绘图/列表的地方是更加精确，方便定位抽中住户，容易质量控制。除了 CHARLS-GIS 软件，在绘图/列表过程中使用的设备包括 HP mini 5102 型上网本（有触屏和摄像头，电池时间长），一个 USB 接口的 GPS 仪。

### B. 2 CHARLS-GIS 的工作流程

从一个 PSU 中进行最终抽样需要经过 4 步，包括：边界绘制、绘图/列表、初次抽样和最终抽样，见图 4。

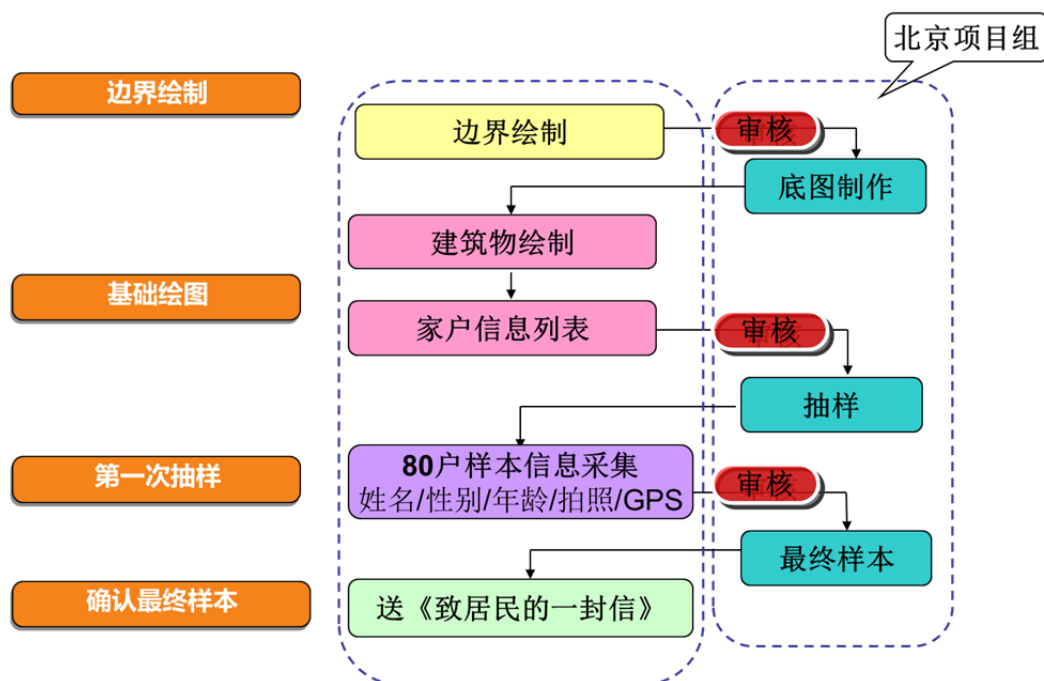


图 4 使用 CHARLS-GIS 的工作流程图

对于每个村和社区（PSU），绘图员携带 EPC 和 GPS 仪先围绕村、社区走一圈确定每个 PSU 的边界。在 CHARLS 项目总部办公室进行质量核查后，边界图和 GPS 信息（图 5）会发送到我们的合作方（中科院地理所）抓取每个 PSU 的底图（图 6）。底图来源于 Google 地图或其他途径。

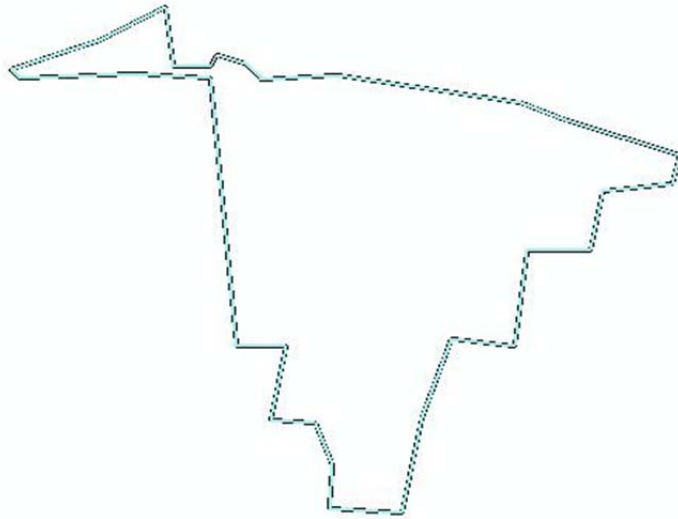


图 5 一个 PSU 带有 GPS 信息的边界图



图 6 带有边界的底图



图 7 绘图和列表后的地图

然后，PSU 中的所有建筑物都进行了标记，而且建筑物中所有住户单元都利用标准的方法进行了列表和编码。图 7 显示的是经过绘图/列表工作后标有建筑物的样图。图中不同的颜色表示不同的建筑物，例如绿色是住户，黄色是学校。对于每个建筑物，我们收集建筑物的类型、名称、地址、照片以及 GPS 信息（图 8）。

图 8 建筑物绘制收集的建筑物信息

房屋编号	是否走访	是否抽中
43-45号03单元04层02号	是	否
43-45号03单元05层01号	是	是
43-45号03单元05层02号	是	否
43-45号03单元06层01号	是	否

图 9 列表时建筑中的住户单元信息

每个村、居的绘图/列表工作完成后，所有相关的数据都会送到北京 CHARLS 项目总部。PSU 的抽样框会进行核查，督导在 CHARLS-GIS 软件中随机抽取 80 户。这些样本然后会发给实地的绘图员收集抽中家户的信息，包括最长人的年龄，户主的姓名、电话号码、是否空户等，以及每个建筑物的 GPS 和照片信息（图 9）。图 10 显示的是标有小红旗的 80 户地图。



图 10 初次抽样标有小红旗的 80 户地图

实际抽中的数量在每个 PSU 中会超过目标样本 24 户，因为某些住户可能没有 45 岁以上的受访者、或者空户、无应答。以绘图员收集的信息为基础，包括适龄率以及空户率，督导会使用 CHARLS-GIS 软件在每个社区、村居抽取一定数目的住户。在这个 PSU 的最终抽样工作完成后，具体的家户信息会返给绘图员送“至居民的一封信”。同样，CHARLS 总部办公室的 IT 人员会将特定 PSU 的抽中住户列表和地址导入访员的 CAPI 系统中。



## 附录 C 代理受访者

在本次基线调查中，在非常特殊的情况下受访者不能完成调查时，允许知情的代理受访者代为回答。当只有部分模块由其他人回答时，我们记录每一个模块回答的方式，包括代理受访者的信息。当选定受访者不在场或完全不能回答问题时，访员需要电话北京总部的指定工作人员申请代理码，如果申请被批准，访员会收到一个代理码，则可进入完全代理模式，然后 CAPI 系统在进入第一个模块之前自动选择代理模式。与正常模式相比，代理受访者只会被询问部分问题。如表 5，总的代理率为 8.7%。男性代理率（11%）高于女性（6%）。在男性中，代理率最高的是小于 50 岁的稍年轻的男性，可能是因为在工作不能接受访问的原因。在女性中，代理率最高的是超过 75 岁的人群，可能因为这些人回答问题时需要帮助。

## 附录 D 体检详细描述

负责一个县区访问的访问员同时会携带体检仪器在受访户家中进行体检，体检的项目包括身高、体重、腰围、膝高、上臂长、肺功能、握力、重复从椅子上站起的时间、血压、步行速度和平衡能力测试。表 2 显示测量的项目和仪器。每两个访员共用一箱体检设备完成所有项目的体检。CAPI 系统随机选择一些体检过程拍照进行质量控制。大部分的体检利用标准方法进行测量。在每个测量之前，会询问受访者他们是否理解测量说明以及他们在接受测量时是否感觉安全。如果受访者觉得没有这两个问题才会进行测量。同样，如果受访者觉得不安全，访员则不进行测量。告知受访者在测量的时候，不能吃饭、饮水、抽烟、嚼口香糖或刷牙。体检测量、方法和 HRS 几乎一样。

### D. 1 血压和脉搏

样本：除非受访者报告手臂上有皮疹、肿胀、或在血压计袖带接触的地方有开放性溃疡或伤口、或严重瘀伤，所有其他满足上述标准的受访者均包括在内。

测量：重复测量 3 次，间隔 45 秒，测左臂血压。每次测量记录的数据包括收缩压、舒张压、脉搏和测量时间。

方法：让受访者坐下，双脚着地，左臂舒服的放置（如桌子），手心朝上。受访者需要卷起袖子，除非是短袖衬衫或薄衬衫。袖带和受访者的上臂相适应，保证和皮肤直接接触，袖带的下缘大约在肘部上半英寸，空气管对着受访者的手臂中间。访员按下开始按钮。袖带自动充气，在显示收缩压、舒张压和脉搏后放气。访员记录收缩压、舒张压和脉搏以及时间。访员使用秒表等待 45-60 秒后开始下一次测量。

### D. 2 肺功能

样本：所有满足上述标准的受访者均包括在内。

测量：重复测量 3 次，间隔 30 秒。

方法：访员给受访者一个肺峰流速仪和一次性吹嘴，让受访者将吹嘴固定在肺峰流速仪上。让受访者站立，深呼吸，用口包住吹嘴，用最大的力气和最快的速度吹气。访员记录肺峰流速仪的数值后将指针归位。访员使用秒表等待 30 秒后开始下一次测量。

### D. 3 握力

样本：除非受访者在过去的 6 个月里有双手有过手术或者发生任何肿胀、发炎、严重疼痛或受伤，所有其他满足上述标准的受访者均包括在内。如果有一只手存在这些症状，则测量另外一只手。

测量：每只手测量两次，轮流测量。

方法：调节握力计和手部的大小相适应。受访者用惯用手先练习一次，站立，肘部在体侧呈 90 度角。让受访者使用最大力气握握力器，保持几秒钟，然后松开。等练习结束后，让受访者换到另外一只手。每只手测量 2 次，轮流测量。每次测量后，访员记录结果并把握力计递给受访者。

#### **D. 4 平衡测试**

样本：除非受访者在没有帮助的情况不能站立一分钟，所有其他满足上述标准的受访者均包括在内。所有受访者均试着进行双脚半前后站立测试。如果能站立 10 秒钟，受访者则进行双脚前后呈一条线站立测试。年龄超过 70 岁的受访者进行 30 秒的平衡能力测试，而 70 岁以下的受访者则进行 60 秒的平衡能力测试。如果受访者双脚半前后站立不能保持 10 秒钟，他们则进行双脚平行站立。如果受访者报告最近手术、受伤或其他可能导致他们不能从椅子上站起保持平衡的问题，访员需要和受访者讨论进行每项平衡测试的能力。

测量：三项测量双脚前后呈一条线站立、双脚半前后站立和双脚并拢站立要测量两项。

双脚半前后站立方法：让受访者站立，用一只脚的大脚趾贴着另外一只脚的中间，保持 10 秒钟。受访者可以把任何一只脚放到前面，也可以利用手臂、屈膝或晃动身体保持平衡，但是脚不能移动。如果必要，访员轻轻的扶着受访者的胳膊帮他们先用双脚半前后站立的姿势站好。访员应该站在受访者边上，如果受访者失去平衡，则进行帮助。让受访者尽可能保持这个姿势直到被叫停。访员 10 秒钟后、受访者挪动位置或者抓住访员胳膊的时候停止秒表。

双脚前后呈一条线站立方法：除了让受访者站立时一只脚的足跟挨着另外一只脚的脚趾保持 30/60 秒外，这种站姿和双脚半前后站立方法一样，访员 30/60 秒钟后、受访者挪位或者抓住访员胳膊的时候停止秒表。

双脚并拢站立方法：除了让受访者站立时双脚平行并拢保持 10 秒钟外，这种站姿和双脚半前后站立方法一样，访员 10 秒钟后、受访者挪位或者抓住访员胳膊的时候停止秒表。

#### **D. 5 步行速度**

**样本：**所有其他满足上述标准的 60 岁以上受访者均进行步速测试，但除外因为近期手术、受伤或健康问题可能不能行走的受访者。另外，这个测试需要足够的空间。所以需要有一个干净、没有地毯、大约 2.5 米长区域作为走道。

**测量：**受访者在 2.5 米长的距离上走两次（一来一回），并每次计时。

**方法：**访员将卷尺放在地上测量总长度，设一个走道，将约一米长的胶带分别粘在走道的两端。然后访员收起卷尺让受访者站在走道的起点。访员说“预备，开始”提醒受访者开始行走。一旦受访者的脚迈过了开始线完全着地的时候，访员开始计时。让受访者以正常的步速通过走道的终点。当受访者的脚完全迈过标示终点的胶带且完全着地时，访员停止计时。然后访员将秒表归零，让受访者往回走。同时访员记录第二次行走的时间和信息。

#### **D. 6 上臂长**

**样本：**所有满足上述标准的受访者均包括在内。

**测量：**使用马丁尺测量受访者的上臂长度。

**方法：**受访者侧面朝向访员，站直，双脚均衡受力，右肘呈 90 度夹角。将马丁尺的固定端放在尺骨鹰嘴处，滑动滑块至肩胛骨的肩峰处，读数。

#### **D. 7 膝高**

**样本：**所有满足上述标准且能坐起来的受访者均包括在内。

**测量：**使用马丁尺测量受访者的膝高。

**方法：**受访者坐在椅子上，右膝弯曲成 90 度。访员需要蹲在右侧，将马丁尺的固定端放在脚踝下方的脚底，然后滑动滑块至髌骨的近端，读数。

#### **D. 8 重复从椅子上站起的时间**

**样本：**所有满足上述标准且能坐起来的受访者均包括在内。

**测量：**重复 5 次站起坐下。

**方法：**让受访者两臂交差与胸前坐在项目配备的标准凳子上。当访员说“预备，站起！”，受访者开始用最快的速度站直坐下 5 次，中间不停而且不能用手撑。如果受访者完成了测量，访员记录使用的时间，否则记录完成次数。

#### **D. 9 身高**

**样本：**所有满足上述标准且能站起来的受访者均包括在内。

**测量：**利用身高仪测量受访者身高。

方法：受访者脱掉鞋子在身高仪的踏板上站直，背部靠在身高仪的垂直柱。受访者两脚均匀负重。两脚足跟并拢贴着垂直板，足尖分开呈 60 度。受访者的头保持法兰克福水平面位置（外耳门上缘点和眼眶下缘在同一水平）。访员向下滑动头板轻触受访者的头部，记录读数。

#### **D. 10 体重**

样本：除非受访者自报目前体重超过 150 公斤或者不能站立，所有其他满足上述标准的受访者均包括在内。

测量：让受访者站在体重秤上测量体重。

方法：选择一个适当的地点放置体重秤，最好没有地毯。让受访者脱去鞋子、厚重的外套以及去除口袋里重物。访员踏一下体重秤等显示屏上显示“0.0”。然后受访者站上体重秤，直到显示屏上显示体重数字。访员记录受访者体重。

#### **D. 11 腰围**

样本：受访者满足上述标准，能够站立而且可以抬起手臂将卷尺放在腰部。

测量：在肚脐位置测量受访者腰围。

方法：让受访者站立，去除厚重的外衣。让受访者指出肚脐的位置，访员将卷尺在肚脐的位置围在腰上。访员需要核查确认卷尺在腰四周是水平的，舒服不要太紧。让受访者吸气，慢慢呼出，在呼气结束时屏住呼吸。必要时可以调整卷尺，屏住呼吸时进行腰围测量。

**表 1 家户问卷收集的数据**

---

<b>人口学信息（主要受访者和配偶）</b>	<b>工作、退休和养老金</b>
出生日期和出生地	目前工作状态
居住和迁移	工作史
户口信息	目前主要工作的详细信息
教育程度	不工作和寻找工作
婚姻状态和历史	最近的工作（如果现在不工作）
	退休
	养老金
<b>家户登记表（不包括受访者及配偶的家户成员）</b>	<b>收入、消费和资产</b>
性别、出生日期和婚姻状态	家庭收入和消费
和主要受访者的关系	家庭资产
户口信息	个人资产
教育程度	
<b>家庭</b>	<b>住房特征</b>
所有父母和孩子：	建筑材料
人口学	屋内设施
教育程度	整洁度和温度
职业	
去世者：死亡时间	
兄弟姐妹：简单的综合信息	
每个家庭成员的联系：	
照顾父母的时间	
孩子前来探望	
父母和孩子的互相的经济往来	
<b>健康状况与功能</b>	<b>访员观察</b>
自报一般健康	访问期间的干扰
医生诊断有慢病或传染性疾病的情况	受访者的态度和理解力
眼睛、听力、口腔以及疼痛	
事故、跌落、骨折	
生活方式和行为，包括睡觉、体力活动、社会交往、饮食、吸烟和饮酒	
功能性障碍和辅助	
认知和抑郁	
健康情景	
<b>医疗保健与保险</b>	
目前和过去的医疗保险	
医疗保健利用：住院和门诊	
医疗花费和支付方式	

---

**表 2 测量指标和设备**

测量指标	测量次数	设备	厂商/来源
血压	3 次, 间隔 45 秒	Omron™ HEM-7112 血压计	中国大连欧姆龙有限公司
脉搏	3 次, 间隔 45 秒	Omron™ HEM-7112 血压计	中国大连欧姆龙有限公司
肺峰流速	3 次, 间隔 30 秒	爱普肺峰流速仪, 一次性吹嘴	上海爱普医械塑胶有限公司
握力	每只手 2 次	悦健 WL-1000 握力器	南通悦健体测器材有限公司
平衡能力	1 次	3 张标明脚放置位置的图和秒表	
步行速度	2.5米来回走两次	卷尺, 胶带和秒表	
上臂长	1 次	东方牌 XTCL-I 马丁尺	石家庄市东方金属制品有限公司
膝高	1 次	东方牌 XTCL-I 马丁尺	石家庄市东方金属制品有限公司
重复站立	重复 5 次	47cm 高凳子和秒表	
身高	1 次	Seca™213 身高仪	中国 seca (杭州) 有限公司
体重	1 次	Omron™ HN-286 体重秤	可瑞尔科技(扬州)有限公司
腰围	1 次	软尺	
静脉血样	8ml	标准采血耗材	
抑郁	1 次	10 项 CES-D 量表	
认知	1 次	从电话访问认知状态表的几项测量: 自评记忆、今天的日期、星期几、目前的季节; 回忆和延迟回忆 10 个词, 100-7 计算系列; 绘制两个重复的 5 边图。	HRS; 参考 McArdle, J.J., Fisher, G.G. and Kadlec, K.M. (2007)

**表 3 样本量和应答率 (%)**

	总计	农村	城市
抽中总家户	23,422	10,597	12,825
空户	4,341	1,914	2,401
目标家户	19,081	8,683	10,424
适龄率	66.77%	73.80%	59.04%
适龄家户	12,740	6,408	6,154
应答家户	10,257	6,033	4,224
应答率	80.51%	94.15%	68.63%

\*应答率计算基于适龄家户

**表 4 受访者的数量和年龄/性别构成**

年龄组	合计	性别		户口		居住地	
		男性	女性	城市	农村	城市	农村
-50	25.77	23.42	27.91	23.79	26.56	27.35	24.18
51-55	15.49	16.00	15.02	14.06	16.07	15.11	15.87
56-60	19.00	19.32	18.69	18.68	19.12	18.65	19.34
61-65	13.88	14.78	13.07	14.13	13.78	13.19	14.58
66-70	9.62	10.20	9.08	9.82	9.53	9.02	10.21
71-75	7.17	7.84	6.56	9.51	6.23	7.64	6.70
76-80	4.67	4.73	4.61	5.32	4.40	4.60	4.73
80+	4.41	3.71	5.05	4.69	4.30	4.44	4.38
合计	17,587	8,436	9,151	3,872	13,715	7,106	10,481

此表删除了年龄、性别和户口信息缺失的个体。



表 5 分性别和年龄的代理率

年龄组	合计		自答		代理	
	频数	比例(%)	频数	比例(%)	频数	比例(%)
合计	17,708	-	16,169	91.31	1539	8.69
- 50	3,927	22.18	3,541	90.17	386	9.83
50-59	6,098	34.44	5,551	91.03	547	8.97
60-69	4,727	26.69	4,434	93.80	293	6.20
70+	2,923	16.51	2,621	89.67	302	10.33
合计	8,476	-	7,545	89.02	931	10.98
-50	1,647	19.43	1,395	84.70	252	15.30
50-59	2,991	35.29	2,619	87.56	372	12.44
60-69	2,379	28.07	2,213	93.02	166	6.98
70+	1,448	17.08	1,311	90.54	137	9.46
合计	9,232	-	8,624	93.41	608	6.59
-50	2,280	24.7	2,146	94.12	134	5.88
50-59	3,107	33.65	2,932	94.37	175	5.63
60-69	2,348	25.43	2,221	94.59	127	5.41
70+	1,475	15.98	1,310	88.81	165	11.19

表 6 分性别/年龄体检率

年龄组	合计		男性		女性	
	频数	比例 (%)	频数	比例 (%)	频数	比例 (%)
合计	13974	78.91	6532	77.06	7442	80.61
-50	2945	74.99	1144	69.46	1801	78.99
50-59	4765	78.14	2274	76.03	2491	80.17
60-69	3923	82.99	1953	82.09	1970	83.90
70+	2326	79.58	1158	79.97	1168	79.19

表 7 每个模块完成时间中位数

	年龄组	中位数（分钟）		
		男性	女性	合计
过滤问卷	合计			9.43
	-50	4.43	4.55	4.5
基本信息	50-59	4.62	4.57	4.6
	60-69	4.97	4.83	4.92
	70+	5.63	5.75	5.68
	合计	4.82	4.78	4.8
	-50	23.32	24.8	23.96
家庭	50-59	26.47	26.58	26.52
	60-69	29.23	29.38	29.33
	70+	32.81	32.46	32.62
	合计	27.53	27.68	27.65
	-50	3.47	3.77	3.62
医疗保健	50-59	3.75	4.04	3.88
	60-69	4.44	4.33	4.38
	70+	4.75	4.45	4.55
	合计	4.03	4.1	4.07
	-50	19.13	20.27	19.82
健康	50-59	21.13	21.48	21.3
	60-69	23.41	22.57	22.95
	70+	23.97	23.07	23.65
	合计	21.88	21.67	21.75
	Under 50	17.13	17	17.02
家庭收入与资产	50-59	16.84	16.66	16.77
	60-69	16.94	15.89	16.43
	70+	14.53	12.96	13.82
	Total	16.52	16.02	16.25
	-50	3.95	3.82	3.88
个人收入与资产	50-59	4.07	3.89	3.98
	60-69	4.25	3.83	4.05
	70+	4.25	3.82	4.03
	合计	4.12	3.85	3.98
	-50	8.07	6.35	7.15
工作与退休	50-59	7.47	5.83	6.53
	60-69	7.13	5.38	6.08
	70+	6.48	5.33	5.83
	合计	7.37	5.73	6.4

表 8 是否参与体检的 Logit 回归

自变量	回归系数	标准误
女性	0.650***	0.088
年龄组		
50- (参照组)		
50-59	0.331***	0.08
60-69	0.802***	0.09
70+	0.681***	0.104
女性分年龄组		
50-59, 女性	-0.319***	0.113
60-69, 女性	-0.470***	0.126
70+, 女性	-0.764***	0.139
教育程度		
文盲 (参照组)		
能读、写	0.014	0.075
小学	-0.051	0.073
初中及以上	-0.239***	0.073
截距	1.900***	0.387
社区哑变量		有
样本量		17708

表 A.1 研究组名单

职位	Name	Institute
项目负责人	赵耀辉	北京大学国家发展研究院
项目负责人	John Strauss	美国南加州大学 经济学院
项目负责人	杨功焕	中国医学科学院、北京协和医学院
项目副负责人	Peifeng Hu	美国加州大学洛杉矶分校医学院
项目副负责人(血样分析)	Eileen Crimmins	美国南加州大学/加州大学洛杉矶分校
项目副负责人	Albert Park	香港科技大学
项目副负责人	John Giles	世界银行
主要研究成员	王德文	世界银行驻北京办公室代表
主要研究成员	曾毅	北京大学 国家发展研究院
主要研究成员	沈艳	北京大学 国家发展研究院
主要研究成员	雷晓燕	北京大学 国家发展研究院
主要研究成员	李力行	北京大学 国家发展研究院
主要研究成员	施新政	清华大学
主要研究成员	吴晓瑜	中央财经大学
主要研究成员	尹香君	中国疾病预防控制中心

表 A.2 国际和国内顾问委员会成员

类别	姓名	机构
<i>国际顾问委员会</i>	James P. Smith	兰德公司中国老年研究中心主任
	James Banks	伦敦大学学院
	Lisa Berkman	哈佛大学公共卫生学院
	David Bloom	哈佛大学公共卫生学院
	Axel Borsch-Supan	曼海姆大学
	Arie Kapteyn	兰德公司劳动与人口研究室主任
	Jinkook Lee	俄亥俄州立大学
	David Weir	密歇根大学
	Robert Willis	密歇根大学
	David Wise	哈佛大学肯尼迪政府学院
<i>Domestic Advisory Board</i>	周其仁	北京大学 国家发展研究院
	蔡昉	中国社会科学院 人口与劳动经济研究所
	Scott Rozelle	斯坦福大学
	邬沧萍	中国人民大学
	姚洋	北京大学 国家发展研究院
	左学金	上海社会科学院