

# 【思·享】大数据时代，我们为什么还要入户调查？

2017-04-25 王芷苓 黄埔大道西观点

## 大数据时代，我们为什么还要入户调查？

文/暨南大学经济与社会研究院助理教授王芷苓

在当下大数据风行的时代，我们依然在执行住户调查，而且一直在坚持。



图片来源于网络

社会在越来越走向数字化，互联网和智能手机的广泛普及交织出了一个大数据的新网络。我们个人实时的消费习惯、资产配置、出行模式、地理定位、社交关系、娱乐活动等，都被编码成不同形式的信息储存在各大商业公司的数据库中，为市场营销、金融咨询、甚至政治选举这样重大事件的预测提供了海量的信息基础。这股潮流近年来也渐渐地蔓延至了学术界，涌现出了很多颇为新颖的研究话题。在学术价值上，大数据宝贵的一点就

是极大程度将数据的收集化繁为简，省去了冗长的问卷调查过程，为研究者节省了不少精力和时间。

而此前的几十年时间，**住户抽样调查**一直都是社会研究的一个主要数据来源。如今在这样一个大数据风行的时代，传统的住户调查会逐渐地被淘汰吗？为什么我们现在依然在执行住户调查？笔者近期在参与中国乡城人口流动调查（RUMiC）督导培训会议中，有一些思考，在此抛砖引玉，与各位同行分享。

**首先，住户调查注重个体和家户行为的成因，而大数据更多的分析只是局限于相关性。**

研究者利用住户调查的问卷信息，探究某两个变量之间是否存在因果机制。比如，子女外出务工对留在家的老人精神状况的影响。而大数据的亮点之一，就是包含了某固定群体的全样本，**这有助于精准地分析某固定群体的总体特征**。比如，大众点评网的用户有五成以上是女性，年龄段集中在15至40岁之间，或者民谣歌手歌词中念叨“南方”比“北方”的频率高出六个百分点等等类似这样的描述性统计。这样的分析揭示的，**仅仅是某种趋势和规律**。

**其次，住户调查涉及的个人信息维度更广更深。**

问卷会收集个体及其家庭的生活条件、福利状况、教育历史、工作信息以及近年来逐渐增加的各种心理认知和健康方面的测度。同时，家庭中所有成员的特征、成员之间的关系和活动也会被收集。研究者不仅可以利用个人信息探索个人行为的成因，也可以探索代际的家户决策的成因，如教育投资、转移支付、老人照料等。而大数据，虽然目前社会上存在的数据集很多，但都只有各自侧重的维度。携程的出行数据、淘宝的交易数据、微信的社交数据，在目前的商业环境和数据保密协定的规制下，它们彼此很难完全融合，使得研究者可以全方位地了解个体。

**再次，住户调查能确保访问到当前大数据中遗漏的个体对象，同时进行逐年追踪。**

农村人口和城市外来务工人口，正是中国经济发展和城市化进程的学术研究中需要高度重视的两部分劳动力。这当中的许多人并不是互联网的频繁使用者，他们的家庭情况如何，需要我们抽取有代表性的样本进行问卷调查才能了解全貌。更重要的是，他们的流动性很大，必须通过访员孜孜不

倦的联系和沟通，才能保证我们可以追踪其家庭经济和社会状况的演变。



入户访问现场

大数据不是万能的，当然，住户调查也不是。**受访者拒访率的提高，越发增加了执行调查团队的工作难度。**在暨南大学经济与社会研究院正在进行的中国乡城人口流动调查（Rural-Urban Migration in China，简称RUMiC）培训会议期间，具有该项目多年调查经验的督导汇聚在一起。他们每一年都带领一队访员去实地采访新住户，并联系往年老住户进行追踪调查，也缘于此，他们与很多当地的外来务工人员相识。某地的督导讲述了一次自己冒着大雨给受访住户送去节日问候和调查致谢礼金的场景。

“调查不是冰冷的你问我答，不是填完问卷就走人。跟受访者的深度交流都是我们应该做到的。”

的确，传统的住户调查可以提升的空间还有很多。比如，今年RUMiC调查的执行方式由**纸版记录全面升级为平板电脑问卷**，社会调查中心的同事们每天加班加点编程修改程序模板，对问卷内容字斟句酌、调整逻辑跳转关系和导入老住户原有信息，确保平板操作的问卷方式能让访员和被访者都更好地接纳。因此，拥抱新的变化，对传统的住户调查着实重要。





#### RUMiC培训使用平板模拟访问

说到这里，我也去搜集了一些国内住户追踪调查的情况，发现优质的住户追踪调查依旧屈指可数，我也深感经济与社会研究院致力于建设高质量的微观数据库的大业任重而道远。目前我们承担的三大项目——中国家庭就业调查、中国乡城人口流动调查项目、儿童认知与非认知能力追踪调查都已全面启动，旨在提供优质的长期家户追踪数据，也请大家持续关注后期动态。

既需适应大数据的潮流，也执着于对厚数据的追求。经济与社会研究院在努力，而且会一直坚持下去。

---

制作：IESR宣传团队

作者：王芷苓

编辑：武茜