

社会调查质量研究： 访员臆答与干预效果*

严洁 邱泽奇 任莉颖 丁华 孙妍

提要：本文利用中国家庭动态跟踪调查并行数据和问卷数据，采用事后准实验设计方法，对面访调查中访员臆答及对臆答进行干预的效果进行了分析，指出访员臆答是面访调查中不可忽视的直接影响调查数据质量的因素；访员的职业特征与臆答的发生比有显著关联，计算机辅助面访调查可以有效监测并干预访员臆答；干预可使臆答显著减少，减少的幅度因访员年龄不同而有显著差异，但不因访员性别、教育和职业因素而变化。

关键词：臆答 问卷调查 质量监控 CAPI

面访调查作为一种有效、可靠的数据采集方法被广泛地应用在社会科学研究中。质量监控是面访调查中至关重要的环节之一。然而，面访调查的质量监控一直存在着难点和盲区，例如，访员在访问过程中是否诱导受访者，是否催促受访者，使受访者未加思考匆匆作答，是否对于问卷中的某些题目并不提问受访者而自填答案（本文称其为“臆答”）等。在纸笔方式的面访调查中，访员的这些做法很难被监测到。

本文通过对北京大学2010年“中国家庭动态跟踪调查”（CFPS）部分质量监控数据的分析发现，有8.4%的农村成人问卷至少发生了一道题目的访员臆答（curbstoning）。^①由此产生的一系列问题是：访员臆答在什么情况下容易发生？发生的比例有多高？怎样才能监测到访员臆答？在采访过程中如何干预和纠正访员的这种行为，以及干预的效果如何？

本文采用事后准实验设计（post quasi-experiment design）方

* 作者感谢匿名审稿人所提出的修改意见。本文系国家自然科学基金资助项目“并行数据和调查数据质量管理（71171004）”成果之一，特表感谢。

^① 感谢库伯（Mick Couper）教授提供臆答的英文类似术语。英文中的 curbstoning，直译为“取巧”，带有贬义，在调查中，不仅包括了这里的臆答，还包括了广义的访员自己作答的其他行为。尽管如此，在现有的文献中，对于 curbstoning 也鲜有论述。

法，运用 2010 年 CFPS 的并行数据^①和问卷数据，对以上问题进行分析，并进一步探讨面访调查质量监控方法的改进方向。

一、访员臆答释义

面访调查中，访员不规范的访问行为会对数据质量造成损害，故此一直被作为质量监控的重点。常见的访员不规范行为可以细分为：替换受访地址或受访户；替换受访者；访员不登门，自填整份问卷；访问过程中破坏中立原则，引导受访者选择答案；访问过程中用各种方式催促受访者，使受访者未加思考匆匆作答；通过跳转模式故意回避需要长时间作答的题组；有意地不提问某些题目而自行填答；系统地读错题目或选错答案中的代码等。

我们把访员有意不提问某些题目而自行填答的访问行为称为“访员臆答”。

访员臆答与访员不登门而自填问卷有所不同，后者是整份问卷都由访员自填。访员臆答是指访员询问了受访者一些个人特征的题目，和/或一些有可能被核查的事实性题目，而对一些难以核查的题目，或者不会被核查的题目，或者是受访者比较反感的题目，或者是自己能够不用提问就判断出答案的题目进行自填。

访员臆答在纸笔面访调查中很难被监测到。在纸笔访问模式中，现场督导拿到访员交回的问卷后，有办法监控替换受访者、不登门而自填问卷、采访用时短等不规范行为，也有办法核实受访者人口变量的准确性和真实性，但却没有办法判断问卷中哪些题目是访员自填的。即使采用电话回访或实地回访，也只能访问那些事实性的、受访者不易改变答案的题目，而对态度性的题目，很难通过前后两次访问的对比结果来判断访员是否自填。因此，在纸笔面访调查中，对于访员题目臆答的监控，始终没有一个切实有效的方法。

^① 并行数据是指在问卷主体调查内容之外，在采访过程中同时获得的数据(Couper et al., 1997)。最常见的有：访员在每个题目上的采访用时，每个题目上的键盘使用痕迹，联系受访者的时点、次数、结果，从样本发放、样本接触、样本调配到样本完成访问的历史，以及访问的录音文件等等。运用并行数据，在理论上，可以实现在任意一个调查时点上直接监测访员的行为，并在短时间内实现汇总、统计分析，并反馈给访员，以达到质量监控的目的。

二、研究方法与数据

(一) 研究内容的进一步界定

访员臆答分为多种情况,有的是访员故意为缩短访问时间而自行填答,有的则是访员根据观察到的或者已经采集到的有关受访者的信息而自行判断得出答案。前者通常会对调查质量产生影响,而后者则很难判断其影响。2010年的CFPS调查质量监测的对象是前者,^①所选择的监测题目均是访员无法通过自行判断能得出答案的题目,例如,一些只有受访者个人能回答的客观性题目和态度性量表题。

访员臆答在调查误差中属于测量误差,可能与不同的题目特征相关联,也可能对数据质量造成严重的影响,所以对研究社会调查数据的质量而言至关重要。但是,限于当前CFPS问卷数据的使用权限,本文将暂不讨论这两方面的内容,而是在现有资源的范围内,将研究焦点锁定在访员臆答的发生与访员特征的关系、访员臆答如何被监测、如何被干预、以及干预的效果方面。

访员臆答研究以访问用的问卷为载体。CFPS问卷结构复杂,包括住户过滤问卷、家庭成员问卷、家庭问卷、成人问卷、少儿代答问卷、少儿自答问卷、村居问卷共七大类。为了避免不同问卷之间的差异、城乡之间的差异可能造成的影响,本文的研究内容限定为农村成人问卷中发生的臆答,以及进行农村成人问卷访问的访员。

在CFPS的监测方式中,每周都汇总并向访员反馈质量监测的结果,因此访员因违规而被干预的次数是不等的,其中被干预多次的访员数量很少。假设首次干预最有效,因此,本文把研究焦点进一步锁定在首次干预。

通过对研究内容的进一步界定,本文关注的焦点可简述为:(1)对2010年CFPS农村成人问卷访员臆答的首次干预效果;(2)农村成人问卷访员臆答的发生与访员特征的关系。为分析便利,这里定义了两种不同的臆答卷比例作为分析对象。^②

臆答卷比例(I型):指访员在某个期间内出现的臆答卷数量占其同期完成卷数量的比例。本文关注首次干预,因此取第一次干预前的臆答卷比例,以及第一次干预后到第二次干预前的臆答卷比例作为研

^① 访员臆答本应监测所有的题目,但考虑到访问过程中的监测量、时效,可干预的内容,以及访员的积极性和情绪,在访问过程中,监测的题目选择了那些不可能通过自行判断而填答的题目。

^② 选用比例而非数量,是因为每个访员在监测期内的访问量不同,相对比较比绝对比较更符合客观状态。

究对象。

臆答卷比例（II型）：指访员在整个调查期间的全部臆答卷数量占全部完成卷数量的比例。这个比例用来分析臆答的发生与访员特征的关系。

（二）数据与研究方法

本文的数据来源于2010年CFPS数据仓库（包括并行数据和问卷数据）中的质量监测过程数据、题目采访用时数据，以及采访员特征数据，得到了北京大学中国社会科学调查中心授权使用。

在质量监测过程数据中，本文选取了各个监测时点的访员臆答卷数量、监测间隔内访员进行采访的天数、完成的农村成人卷总数，以及访员开始访问和结束访问的日期。在题目采访用时数据中，选取了每份成人问卷监测变量的采访用时，包括敏感问题、困难问题、态度量表问题、大幅跳转的节点题目，以及关键的个人特征题目等。^①在访员特征数据中则选取了访员性别、年龄、教育程度、职业、婚姻状况等变量。

使用上述数据，采用事后准实验设计方法进行分析，即：（1）将发生了臆答的访员和未发生臆答的访员分为两个组别，前者为实验组，后者为对照组，将访员特征作为控制变量，分析访员特征与臆答发生比之间的关系；（2）针对干预效果，把进行过干预和没有进行过干预的访员分别列为实验组和对照组，将访员特征作为控制变量，分析干预效果与访员特征之间的关系。

三、访员臆答的监测方法与监测结果

（一）监测方法

计算机辅助面访（Computer Assistant Personal Interviewing, CAPI）技术的应用，给质量监控方法带来了飞跃（Baker et al., 1995; Couper et al., 1997）。运用计算机辅助技术可以及时、有效地获得可用于监测访问过程的并行数据，及时对访员臆答进行干预。

欧美国家使用计算机辅助技术的知名调查项目很多，如在美国，几乎所有大型调查中都采用了计算机辅助技术甚至混合调查技术，如

^① 不包括那些可以通过自行判断得出答案的题目。

家庭收入动态调查 (PSID, 1968-), 健康与养老调查 (HRS, 1992-), 青少年健康调查 (Add Health, 1994-); 在其他国家, 如英国家庭动态跟踪调查 (BHPS, 1991-), 理解社会调查 (Understanding Society, 2009-), 德国的社会经济跟踪调查 (SOEP, 1984-) 等等。中国首次将 CAPI 调查技术和计算机辅助调查管理技术应用于全国性调查的则是北京大学的 CFPS。^①

针对访员臆答, CFPS 调查通过电话回访、录音回放, 访问过程键盘回放, 以及对题目采访用时的统计分析来发现问题。

首先, 计算各类问卷^②监测题目的采访用时。在这个环节, 第一步是确定监测题目列表, 即按照问卷类型、模块、区段、间隔方式挑选监测问题, 主要针对容易出现臆答的题目, 例如态度量表题目、敏感题目、访问困难题目等; 第二步是计算监测题目的最短读题时间作为合理用时。^③最后用每个访员在每个样本的各个监测题目上的用时与合理用时进行比较, 低于合理用时的定义为疑似臆答样本。

其次, 对于疑似臆答样本进行录音回放, 确认是否没有提问。

第三, 对于没有录音的样本, 进行电话回访, 询问受访者是否提问了某个题目。

最后, 对于既没有录音也没有联系电话的样本, 进行访问过程键盘回放, 以确认是否臆答。

在时效性策略方面, 收到调查数据的第 2 天, 便开始题目采访用时的统计分析。录音回放、电话回访也在第一可能时间点上开始。

(二) 访员臆答的监测结果

CFPS 的目标样本家庭数为 16,000 户, 户内的全部成员都要接受访问。截至 2010 年 9 月 15 日, 参加访问的访员共有 438 名, 完成住户过滤问卷、家庭成员问卷、家庭问卷、成人问卷、少儿问卷共计 84,385 份。^④

依据 CFPS 调查质量核查方案, 50% 的问卷被随机核查到。在被抽查的问卷中, 根据对每份问卷监测题目的采访用时与合理采访用时的

^① CFPS 调查于 2009 年首次将 CAPI 调查技术应用于实地访问。2010 年 4 月 5 日正式启动全国范围的入户面访调查。截至 2010 年 9 月 15 日, 完成了除上海部分村居以外的全部访问。

^② 中国家庭动态跟踪调查分为住户过滤问卷、家庭成员问卷、家庭问卷、成人问卷、少儿问卷、村居问卷等共七类问卷。

^③ 5 个经过培训的督导员用最快的方式读题, 取最短时间为合理用时。

^④ 其中, 住户过滤问卷 14,965 份, 家庭成员问卷 14,343 份, 家庭问卷 14,193 份, 成人问卷 32,108 份, 少儿问卷 8,776 份。

比较,并辅助以录音回放、电话回访和键盘回放加以确认,共有 7,914 份问卷被检测出至少有 1 个以上题目是臆答的(以下简称“臆答卷”^①),占全部问卷总量的 18.8%。

在农村成人问卷^②中,共有 946 份问卷被检测出至少存在 1 个以上题目的臆答,占农村成人问卷总量(22,522 份)的 8.4%。这 946 份问卷涉及到 165 个访员,占访问过农村成人问卷的访员总数(318 名)的 52%。表 1 显示,在发生臆答的 165 名访员中,50%的访员有 3 份及以下的问卷属于臆答卷,10%的访员其臆答卷在 12 份以上,臆答卷最多的一个访员达到 46 份,有 30%的访员其臆答卷占其完成问卷总量的 9%以上。

表 1 农村成人问卷中的臆答卷数量和比例

臆答卷	臆答卷数量(份)	臆答卷比例(II 型)
访员数量	165	165
平均值	5.73	.08
中位值	3	.05
最大值	46	.75
总和	946	
20%分位值	1	.02
40%分位值	2	.04
60%分位值	4	.07
70%分位值	7	.09
80%分位值	9	.13
90%分位值	12.4	.21

(三) 访员特征与臆答发生比

此项研究中,以 II 型臆答卷比例为因变量,对被干预过的访员,将其在调查期间完成的全部问卷作为分母,将全部的臆答卷作为分子。对没有被干预过的访员,采用同类的计算方法,对没有被检测出臆答的访员,设置其臆答卷比例为 0。

我们首先通过单因素方差分析方法,检验出不同访员的 II 型臆

^① 原则上来说,臆答题目较多的问卷会被视作废卷,不过 CFPS 调查中没有发生这类问卷。对于臆答题目较少的问卷则安排责任访员重新补访这些题目,对于无法补访的,则在数据库中标识变量,提示题目的疑点,提示数据使用者可将其视为缺失数据(missing data)。

^② 总量中有 50%的问卷被核查到。

答卷比例的均值差异显著 ($F=75.661, p<0.001$)。表 2 归纳了三类人群臆答卷比例的基本情况。从中可以看出, II 型臆答卷比例的均值在干预之后已经减低为 6.8%, 而未被干预的访员,^①其臆答卷比例的均值为 11.5%。

表 2 干预前后的臆答卷与无臆答卷的基本情况

访员类型	臆答卷比例的均值	访员数	臆答卷总数	农村成人卷总数
有臆答未干预	.115	57	316	3541
有臆答已干预	.068	108	630	10418
没有臆答	.000	153	0	8563
合计		318	946	22522

为验证访员特征与臆答发生比是否有显著关联, 我们把访员分为未发生臆答和发生过臆答两类人群, 用二分变量的逻辑斯蒂回归模型 (Binary Logistic Regression) 检验访员性别、年龄、职业、受教育程度对臆答发生比的影响。表 3 显示, 访员当前的职业^②能够显著地区分出访员是否会发生臆答。

在 95% 的置信水平下, 计生干部、村居委会干部臆答的发生比是学生 3.740 倍, 无业/待业人员臆答的发生比是学生 2.770 倍, 政府企事业单位职员臆答的发生比是学生 2.495 倍, 而中小学教师等专业技术人员臆答发生比与学生无显著性的差异。此外, 在控制了职业的变化之后, 即对于同一类职业的访员而言, 访问量的增加会导致臆答发生比的增加。

^① 在 165 个发生臆答的访员中, 有 35% 的访员 (57 个) 共计 316 份问卷由于已经结束访问而没有来得及进行干预。没有实现干预的原因有四类: 有 13% 的访员 (21 个) 历经了每周的质量监测, 但是直到最后一次监测才发现臆答; 有 6% 的访员完成的样本量很少, 且执行时间集中在 1-2 周内, 当质量监测人员发现其臆答时, 访问已经结束; 有 1% 的访员开始访问的时间临近调查季结束, 发现其臆答时, 已经结束访问; 有 15% 的访员 (24 个) 因录音核查、访问过程键盘回放的延时和滞后, 导致发现问题晚, 从而未能及时实施干预。前三类存在不可避免性, 第四类则是在调查中需要进一步解决的问题, 本文在此暂不作讨论。

^② 在 CFPS 调查中, 绝大多数访员都保有原来的职业, 业余时间兼职做我们的采访员。438 名访员中, 有 13% 来自于计生干部, 11% 来自于中小学教师, 34% 来自于政府企事业单位的职员, 25% 为在读的大学生, 17% 为无业或待业人员。

表3 访员特征与是否发生臆答的分析结果 (Binary Logistic Regression)

解释因素	模型 I		模型 II	
	回归系数	回归系数	标准误	发生比率
职业 (参照类: 学生)				
计生、村居委干部	1.385**	1.319**	.428	3.740
中小学教师等专业人员	.597	.489	.432	1.630
政府企事业单位职员	.899*	.914**	.324	2.495
无业/待业	.937*	1.019**	.388	2.770
年龄 (参照类: 18-24 岁)				
25-34 岁	-.274			
35 岁及以上	-.398			
性别 (0=女; 1=男)	.034			
教育 (0=大学及以上, 1=大专及以下)	.329			
访问量	.012***	.012***	.003	1.012
常数	-1.495***	-1.412***	.304	.244
模型整体:				
-2LL	388.375	391.228		
伪确定系数 Nagelkerke R ²	.168	.157		
模型正确预测百分比	64.0	63.0		

注: * p<0.05, ** p<0.01, *** p<0.001。

四、对访员臆答的干预效果

(一) 对访员臆答的干预

在 CFPS 的质量监测中,对于发生了臆答的访员,第一步会由调查执行督导通知访员改正行为,包括命令其不得臆答,明确重申臆答严重会进行经济处罚和取消访员资格,要求访员电话回访受访者、询问臆答题目的真实答案等。

第二步则对这些访员的后续所有样本都进行重点监测和督导。

如果首次干预之后臆答仍有发生,第三步则除了由督导通知其纠正访问行为之外,还会通知访员要对多少份问卷进行多大比例的经济

处罚（即扣除一定比例的问卷劳务费）。经济处罚是在调查结束后的结账阶段进行的，没有在访问过程中执行经济处罚的主要原因在于质控进程与财务管理进程不一致。^①

第四步，对臆答最严重的情况则取消访员资格。不过，在 CFPS 调查中，并没有发生臆答非常严重的情况，因此没有访员因臆答而被取消资格，也没有发生由于臆答题目比例大而进行重访的情况。^②

本文关注首次干预的效果，对于所有臆答访员所进行的首次干预是一致的，即由执行督导通知访员改正行为。接下来，我们来检验这种首次干预对臆答卷比例所产生的影响。

（二）首次干预效果

检验首次干预的效果要以 I 型臆答卷比例作为研究对象，这类臆答卷比例在每个访员那里因干预的次数不同而存在多个状态。这里，我们选第一次干预前的臆答卷比例，以及第一次干预后到第二次干预前的臆答卷比例作为因变量，采用配对样本的 t 检验和一般线性模型（GLM）下的重复测量多因素方差分析（repeated measures analysis of variance）（刘红云、张雷，2005:18-38）方法进行干预前和干预后的比较分析。^③

分析结果显示，首次干预能够使得访员臆答卷的比例显著降低。干预前和干预后的臆答卷比例均值差异显著（参见表 4， $t=6.134$ ， $p<.0001$ ）。干预之前臆答卷比例的均值为.255（标准误为.020），干预之后则下降到.098（标准误为.018）。

表 4 干预前后 I 型臆答卷比例均值差异的配对样本 t 检验结果

	配对统计量					t 值	自由度	显著度 (双侧检验)
	均值	标准差	标准误	95% 置信区间				
				下限	上限			
干预前 - 干预后	.157	.266	.026	.106	.207	6.134	107	.000

^① 2010 年 CFPS 调查的财务由北京大学中国社会调查中心和北京大学社会科学部，北京大学财务部共同管理。

^② CFPS 调查中有取消访员资格的情况，是因为发现有访员委托没有经过培训的人代替采访，执行机构的处理措施是更换访员重访，并且废掉了原来的问卷。本文使用的数据中排除了这种情况。

^③ 选择这种方法是因为干预前后的访员是一样的，且分析目标在于干预前后的臆答卷比例的均值是否有显著差异。

这种幅度的比例变化对调查质量来说是非常可喜的。虽然 CFPS 调查并没有在访问过程中实行更严厉的质量干预措施，但可以假设，持续的质量关注和更为有效的干预手段会大幅度减少臆答卷的比例。

(三) 访员特征与首次干预效果

为分析干预前后臆答卷比例在不同特征访员中是否有显著的差异，我们把访员特征变量的主效应和交互效应纳入分析中，采用重复测量的多因素方差分析方法进行了检验。在通过球形假设检验^①保证因变量满足方差齐性的前提条件之后，发现干预的主效应显著（参见表 5）。

表 5 干预前后 I 型臆答卷比例均值差异的方差分析结果

主体内效应的检验 (Tests of Within-Subjects Effects)							
源		III 型平方和	自由度	均方差	F 值	显著度	偏 Eta 方
臆答卷比例	球形假设	1.326	1	1.326	37.620	.000	.260
	格林豪斯-吉泽尔参数	1.326	1	1.326	37.620	.000	.260
	辛-费德特参数	1.326	1	1.326	37.620	.000	.260
	下限	1.326	1	1.326	37.620	.000	.260
误差 (臆答卷比例)	球形假设	3.771	107	.035			
	格林豪斯-吉泽尔参数矫正	3.771	107	.035			
	辛-费德特参数	3.771	107	.035			
	下限	3.771	107	.035			

注：格林豪斯-吉泽尔参数 (Greenhouse-Geisser)、辛-费德特参数 (Huynh-Feldt)、下限 (Lower-bound) 均为球形假设 (Sphericity Assumed) 偏离条件下的矫正估计的方法。本文检验的臆答卷比例由于通过了球形假设的检验，因此格林豪斯-吉泽尔参数、辛-费德特参数、下限的矫正值和球形假设下的估计值是相等的。

接下来还要检验的是，在不同访员之间，干预效果是否有显著差异。在分析模型中纳入访员基本特征变量后的分析结果发现：访员的性别、受教育程度、职业的主效应及它们之间的二维、三维交互效应

^① 重复测量的多因素方差分析的前提假设要求个体观测之间独立，多次测量服从正态分布，并且方差齐性。球形检验的目标是检查是否满足方差齐性的条件。

对干预前后的 I 型臆答卷比例的影响均不显著（略微显著但未能通过方差齐性检验和组间协方差矩阵相等检验）。

但不同年龄组的访员之间，干预的效果有显著的差异。鉴于访员的年龄分布状况，把访员的年龄分为 18-24 岁，25-34 岁，35 岁以上三个年龄组。各年龄组访员的 I 型臆答卷比例的均值情况如表 6。

表 6 干预前后不同年龄组访员的 I 型臆答卷比例均值的基本情况

	年龄组	臆答卷比例的均值	臆答卷比例的标准差	访员数量
干预前	18-24 岁	.308	.222	43
	25-34 岁	.220	.205	52
	35 岁以上	.214	.182	13
	合计	.255	.212	108
干预后	18-24 岁	.062	.168	43
	25-34 岁	.133	.212	52
	35 岁以上	.077	.124	13
	合计	.098	.188	108

三个年龄组样本协方差矩阵相等检验通过（Box's M =12.428, F=1.982, p>.05），球形假设检验也通过，即符合纳入模型进行方差分析的前提假设条件。

故将干预主效应、干预与年龄组交互效应放进分析模型，结果如表 7。从表中可知，干预的主效应（F=28.111, p<.0001），以及干预与年龄组的交互效应（F=4.531, p<.05）对干预前后的臆答卷比例的变化有显著的影响。

表 7 不同年龄组访员干预前后 I 型臆答卷比例均值差异的方差分析结果

主体内效应的检验 Tests of Within-Subjects Effects							
源		III 型平方和	自由度	均方差	F 值	显著度	偏 Eta 方
臆答卷比例	球形假设	.929	1	.929	28.111	.000	.211
臆答卷比例 * 年龄组	球形假设	.300	2	.150	4.531	.013	.079
误差(臆答卷比例)	球形假设	3.472	105	.033			

注：本文检验的臆答卷比例由于通过了球形假设的检验，因此格林豪斯-吉泽尔参数、辛-费德特参数、下限的矫正值和球形假设下的估计值是相等的。此表中略去。

在三个年龄组中，18-24 岁的访员一开始的臆答卷比例偏高，干预后迅速降到较低的位置。如图所示，25 岁以上的访员在干预前的臆答卷比例大致相当，干预后，35 岁以上访员的臆答卷比例比 25-34 岁的访员下降的幅度更大些。总体上看，干预后臆答卷比例仍然相对偏高的是 25-34 岁的访员。^①

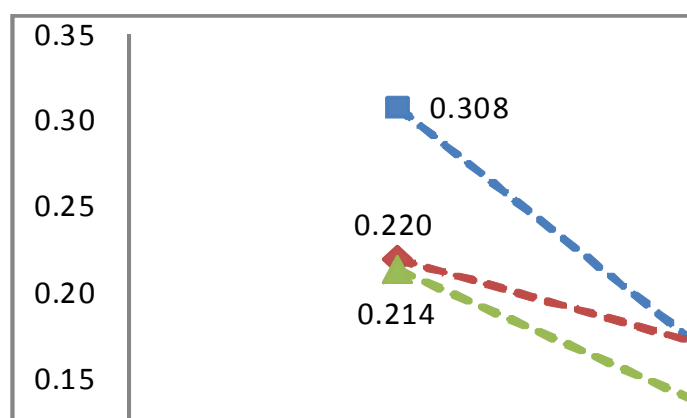


图 1 不同年龄组访员干预前后臆答卷比例均值的差异

从上述讨论可以看出，首次干预能获得显著效果，但干预效果不因访员的性别、职业、受教育程度而不同，却因年龄组不同而有显著差异，对 25-34 岁访员进行首次质量干预后的效果相对比其他年龄组要弱，臆答卷比例降低的幅度不大。降低幅度最大是 18-24 岁之间的访员。需要注意的是，对这个年龄组访员的监测和干预尤为重要，因为这类访员的臆答卷比例在干预之前是最高的。

五、结论与讨论：质量监控的挑战与改进策略

在 CFPS 质量监控环节发现的访员臆答是早期其他调查质量控制

^① 如前所述，在发生过臆答的访员中，访员职业与干预前后的臆答卷比例没有显著相关性，需要关注的是，在这个年龄组，却发现了比较集中的职业组，这些访员中，有 46.2% 的人是政府、企事业单位的从业人员兼职做访员，17.7% 为计生干部或村居委会的成员，13.5% 为无业或待业人员。这将是需要进一步探讨的问题。

中未曾也无法发现的现象，也是纸笔方式调查质量监控的盲点，需要引起对调查质量监控的足够重视。

通过 CFPS 的质量监控实践，看到了对臆答进行质量监控的显著效果。在未发生臆答、发生臆答未被干预和被干预的三类人群中，臆答卷比例均值的差异显著。在发生臆答并且被实施干预的人群中，首次干预可以显著地降低臆答卷比例，降低的幅度在职业、性别、教育等方面没有显著差异，但是在年龄组上有差异，25-34 岁访员臆答卷比例降低幅度最小，纠正的效果不如其他年龄组访员。在干预前，18-24 岁访员的臆答卷比例最高，但是在干预后下降得最快，最低。因此，质量监测与及时干预对于这些访员来说至关重要。更加值得关注的是访员臆答的普遍性，有臆答的各类问卷占总问卷量的比例平均为 18.8%，虽然农村成人问卷的臆答卷比例为 8.4%，但却涉及到 52% 的访员。对有臆答的访员特征分析显示，访员职业特征会影响着臆答的发生比，和学生相比，计生干部、村居委会干部、无业待业人员、政府企事业单位职员更容易发生臆答。

我们还发现尽管 CAPI 调查的质量监测可以监测到臆答，但监测的滞后会导致发现问题却来不及实现干预。在 2010 年 CFPS 的初访中，有 165 名访员有臆答，其中有 24 名访员就因监测滞后导致没有来得及干预。因此，掌握时效性对质量监控而言，尤显重要。对臆答监测的滞后主要源于录音回放的时效性。在所有核查方式中，录音回放是监测臆答的有效工具，改进录音回放的方式，提高录音回放的时效性，是提高调查数据质量的有效途径。

采访过程中的质量干预方式虽然可以有很多种，但是在采访过程中，访问进程管理、访员绩效管理、质控管理、数据管理和财务管理是交互错综的复杂系统，有些干预措施（例如，经济处罚）往往会滞后于质控的要求。

除此以外，在 CAPI 模式中，最迫切的也是最复杂的问题还是在于，如何将监测题目合理化，以及在保证监测题目具有高信度和效度的前提下，如何将质量监测结果直接和数据质量建立联系。

格鲁夫斯曾指出反映数据质量的四个误差来源 (Groves, 1989): 覆盖范围 (coverage)、无回答 (non-response)、抽样误差 (sampling error)、测量误差或者回答误差 (measurement or response error)。访员的臆答是测量误差中的一种重要来源，目前，我们还不能简单地从一份问卷的臆答数量来判断问卷的整体质量，也不能从某个题目是臆答就判断此题无效，除非是逐条题目，逐个受访者地一一对应加以

判断。那么如何将臆答与每一个变量的数据质量联系起来？哪些题目更容易产生臆答？哪些题目臆答会严重影响对总体的推断？等等。这些问题不是质量干预可以解决的，有待进一步探讨。

综上所述，从纸笔面访调查过渡到 CAPI 调查是为了提高调查质量做出的一种选择。在 CAPI 模式的质量监控环节，监测工具的精细化、监测设计的科学性、监测时效性的提高、监测结果与数据质量的联系，以及对质量监控人员职业素质的迫切需求，对访员招聘和管理上的分人群、分方式的要求，都是未来在质量监控方面所面临的新挑战和努力方向。

参考文献：

- 刘红云、张雷, 2005, 《追踪数据分析方法及其应用》, 北京: 教育科学出版社。
- Baker, Reginald P., Norman M. Bradburn & Robert A. Johnson 1995, "Computer-assisted Personal Interviewing: An Experimental Evaluation of Data Quality and Costs." *Journal of Official Statistics* 11(4).
- Couper, Mick P., Sue Ellen Hansen & Sally A. Sadosky 1997, "Evaluating Interviewer Performance in A CAPI Survey." In Lars Lyberg, Paul Biemer, Martin Collins, Edith De Leeuw, Cathryn Dippo, Norbert Schwarz & Dennis Trewin (eds.), *Survey Measurement and Process Quality*. New York: John Wiley & Sons.
- Groves, Robert M. 1989, *Survey Errors and Survey Costs*. New York: John Wiley & Sons.

作者单位：北京大学政府管理学院（严洁）
北京大学社会学系、
中国社会与发展研究中心（邱泽奇）
北京大学中国社会科学调查中心（任莉颖、丁华、孙妍）
责任编辑：杨典

文章来源：《社会学研究》2012年第2期

中国社会学网 www.sociology.cass.cn