卫生发展瞭望 Health Development Outlook

第 77 期　2021 年 4 月　北京大学中国卫生发展研究中心 简报

随机干预实验的外部有效性问题：国际经验和中国实践

随机干预实验(Randomized Controlled Trials, RCTs)因其在克服内生性上的优势，不仅在医学领域而且在社会科学领域尤其是发展经济学研究中进行因果推断和有效性分析的“黄金准则”。2019 年诺贝尔经济学奖授予三位利用该方法研究全球减贫问题的发展经济学家（Abhijit Banerjee、Michael Kremer、Esther Duflo），更是让随机干预实验方法走入大众的视野。

然而，该方法自进入社会政策研究的视野之初便饱受争议。**争议的焦点在于：随机干预实验相当于一个“黑箱”，使用这种方法的研究只能回答 “什么有效”，无法回答“为什么有效”。这就意味着人们既无法从中进一步了解人类行为的一般规律，也不确定这种政策能否推广到其他样本中(即“外部有效性”问题)，因而极大地削弱了随机干预实验研究的价值。**

本文梳理国际随机干预实验研究文献中探讨外部有效性问题的经典案例，并结合课题组所在团队在中国开展随机实验研究的经验，探讨随机干预实验的外部有效性问题。

国际经验

自 2006 年小额信贷之父 Muhammad Yunus 因其在孟加拉国向贫困人口提供信用贷款及服务的贡献获得了诺贝尔和平奖，**小额信贷（Microcredit）作为一种常见的扶贫减贫干预手段得到了很大的应用**，到 2012 年全球已经有超过 3000 家在中低收入国家提供小额信贷的机构。

然而，在孟加拉国被证明有效的小额信贷近几年来在中低收入国家开展的随机干预实验中并没有复制其早期的成功。2015 发表于《美国经济杂志：应用经济学》的一项基于 6 个中低收入国家的随机干预实验表明，为穷人提供贷款虽然在某种程度上增加了他们进行商业活动的积极性，但是这些投资和商业活动并没有显著提高家庭的收入和消费水平；进一步数据分析显示，小额贷款在女性赋能和对儿童的人力资本投资上也没有显著积极影响。很多贷款被用来消费而非投资。

中国实践

作者所在的课题组从发展经济学的视角近十年来一直在利用 RCTs 开展缩小城乡间人力资本不平等（特别是健康不平等）的研究。早期的工作主要集中于农村地区中小学生的近视问题，利用随机干预实验（RCT），探索提高农村近视学生眼镜配戴率的干预手段，进而提高近视学生学习成绩，缩小城乡间代际人力资本差距，前期成果发表于英国医学杂志（The BMJ）等国际期刊上。

自 2012 以来，课题组已成功开展了 5 项农村地区视力健康干预的 RCT 研究。近年来，课题组也在思考 RCT 研究的外部有效性问题，即：**同样的干预手段，为什么在某些地区有效，在另一些地区却无效？造成这种干预效果差异的愿意究竟是什么？**

为了回答以上问题，课题组对比了两个视力干预的随机干预实验，**两个实验采用相同的随机分配方案**（学校层面）、**相同的干预手段**（免费眼镜加健康宣教）、**相同的干预时间**（一学年）、**相同的结局变量**（标准化考试成绩），**并由相同的课题团队实施**。唯一的不同是两个实验开展的**情境（context）的差异**，一个是在甘肃和陕西的贫困农村的小学实施（下文简称**“西部项目”**），另一个是在上海和苏州的民办打工子弟小学开展（项目简称**“东部项目”**）。

对比两个实验结果，在控制了基线成绩和学生及家庭变量之后，**西部项目中的干预组相比于对照组成绩提高了 0.14 个标准差，而东部项目中干预组相比对照组成绩没有统计上显著的变化。**

	Western China Program		Eastern China Program	
	Unadjusted	Adjusted	Unadjusted	Adjusted
Dep. Var: Endline Standardized Math Score	(1)	(2)	(3)	(4)
Treatment (1 = yes)	0.157*	0.137*	0.043	0.046
	(0.072)	(0.074)	(0.708)	(0.673)
Baseline standardized math score controlled	Yes	Yes	Yes	Yes
Additional controls	No	Yes	No	Yes
Mean of dep. var. in control group	0.149		-0.618	
Observations	653	617	706	603
R^2	0.420	0.438	0.087	0.155

表 1. 西部项目和东部项目干预效果比较

是否是两个项目的学生和家长特征不同？

首先值得思考的是两个项目样本中的学生和家长的不同特征造成了西部项目显著提高了学生成绩，而东部项目没有显著结果。比如，家庭对于教育的重视和投入程度的不同造成了最终的学习成绩的差异。虽然无法对该方面进行直接测量，课题组收集影响家庭对于教育的重视和投入程度的重要因素（如父母受教育水平、家庭资产）。**经对比发现，东部项目样本中的家长受教育水平和家庭资产反而要显著高于西部项目的样本，理论上说，其对家庭教育的重视和投入程度应该高于西部项目，其最终结局学习成绩的提高也应该高于西部项目，这和两个项目的干预结果不符。**

是否是两个项目的依从性不同？

项目的结局指标是标准化考试成绩，在干预因果链的中最重要的中介变量即学生是否配戴眼镜。如果学生和家长之间的差异不足以解释项目的干预项目在两个不同的情境下产生的干预效果的差异，那么下一个值得分析的就是配戴眼镜依从性的差异。经对比发现，在基线时西部和东部样本中的干预和对照组的近视学生眼镜配戴率皆为 18%，在一年的干预后，西部地区的配戴率上升为（干预组 49%、对照组 26%），而东部地区的配戴率上升为（干预组 68%、对照组 24%）。也就是说，**就干预的依从性而言，东部地区的眼镜配戴率提升的更高，其最终结局学习成绩的提高也应该高于西部项目，但这和两个项目的干预结果还是不符。**

是否是溢出效应（spillover effect）造成的？

如果直接对比样本中的受干预的近视学生得到的效应是反直觉的（东部地区有更好的家庭条件和更高的戴镜依从性，但其干预效果反而更低），那么是否存在干预对样本外群体的溢出效应？在这里，溢出效应是指对于视力正常学生的学习成绩的间接影响：比如由于近视学生戴上了眼镜可以看清黑板，老师可能较以往花更少的时间辅导其学习，从而转移更多的时间到视力正常学生的身上；此外，由于近视学生戴上了眼镜可以看清黑板，其可能较有更少的可能性打扰到视力正常的学生学习等等。**以上任何一种情况都会使得对于近视学生的干预间接产生对于不近视的视力正常的学生的学习成绩的溢出效果。然而，对比发现，在控制了基线成绩和学生及家庭变量之后，干预在西部和东部项目都没有产生溢出效应。**

学校教育质量不同所照成的干预效果的异质性

基于以上的分析，相同的干预设计在不同的情境下产生不同的干预效果，其背后的原因不是不同样本中的个体和家庭层面的差异、也不是有依从性的差异、亦不存在溢出效应，那么两个样本中干预效果的异质性（heterogenous treatment effect）有可能是各自情境中（公办农村小学 VS. 民办打工子弟小学）的教育质量差异造成的。**综合数据，即使民办打工子弟小学在发达的上海和苏州地区，其在教师的从教年数、教师职称、学校有编制教师比例、学校的历史等四个变量上都落后于经济不发达的甘肃和陕西的公办农村小学。我们生成了一个基于四个变量的学校教育质量指数，定量结果发现相同的干预设计在西部和东部产生的干预效果的差异有一半可以归因于校教育质量差异。**

对未来研究的启示

课题组近十年来在中国不同地区开展RCTs的经验表明，从成功的干预手段到可复制可推广的政策有很长的因果链条，其中任何一个环节的偏离都会对干预效果产生影响。近十年来发展经济学的研究者们不仅寻求识别政策是否有效、效果大小，也越来越重视对政策作用机制的探索，了解政策有效或无效的原因。众多文献的共识是,**基于理论的影响评估（Theory-based Impact Evaluation,TBIE)有助于回答“为什么有效”这一问题。通过建立从投入到产出再到影响的因果链（Casual Chain），使用实证数据检验在因果链中可能起作用的理论和潜在假设是否成立,来达到厘清干预项目作用机理的目的。**（马晓晨）

注：本文摘译自课题组最新发表的研究：Ma X, Wang H, Shi Y*, Sylvia S, Wang L, Qian Y, Rozelle S. Improving Learning by Improving Vision: Evidence from Two Randomized Controlled Trials of Providing Vision Care in China. Journal of Development Effectiveness, 2021 13:1, 1-26, doi: 10.1080/19439342.2021.1876139