

# 减少全球贫困的实验性方法

——2019年诺贝尔经济学奖得主的贡献与评析

Understanding Development and Poverty Alleviation: A Review of the Contribution of the Winners of Nobel Prize for Economics in 2019

易红梅

YI Hong-mei

【中图分类号】F061.3 【文献标识码】A 【文章编号】1000-1549(2019)12-0134-07

DOI:10.19681/j.cnki.jcufe.2019.12.011

## 一、前言

贫困是人类长期面临的社会现象。在过去30多年,全球减贫事业取得了巨大的成就。到2015年,全球贫困人口<sup>①</sup>已经从1990年的36%下降到了10%<sup>[1]</sup>。但直到2018年,世界上仍有超过7亿人口生活在极端贫困当中,享受不到基本的医疗教育服务,没有干净的饮用水和卫生条件<sup>[2]</sup>。联合国可持续发展目标提出,到2030年,全球要消除所有的极端贫困。但是根据世界银行的分析,由于很多国家经济增速放缓,要如期实现这一目标面临着严峻的挑战<sup>[1]</sup>。

如何有效消除贫困一直是发展经济学的核心议题。2019年,麻省理工学院教授阿比吉特·巴纳吉(Abhijit Banerjee)和埃丝特·迪弗洛(Esther Duflo)以及哈佛大学教授迈克尔·克雷默(Michael Kremer)荣膺诺贝尔经济学奖,以表彰他们“以实验性方法致力于减轻全球贫困”。从20世纪90年代开始,三位学者突破传统的发展经济学的研究范式,将全球贫困从一个宏观的“大问题”分解成一个个更加具体的细微的问题,深入探究贫困的本质,并引入随机干预实验(randomized controlled trial, RCT)方法以评估相关政策或者干预对解决这些细微问题的有效性,推动了政府部门和实践者以事实为基础的决策的发展,为发展经济学转型与发展全球减贫事业做出了杰出的贡献。

但是,今年的诺贝尔经济学奖一发布就引发了国内学者对于诺奖获得者的贡献价值的热烈争议,早期国内外关于RCT方法的诸多争论也再次引起关注。其中,质疑者主要持有两个观点:一是认为三位诺奖得主的研究都集中在“细小”问题,不能从根本上解决全球大范围的贫困问题<sup>[3-5]</sup>。二是认为通过RCT得到的估计量最多只是一个无偏估计,相对于RCT的结论在外部有效性方面的局限而言,这一方法上的改进价值有限<sup>[6]</sup>。

本文的目标是通过介绍发展经济学近年的发展、RCT方法在影响评估中的应用,以及基于科学事实的政策决策和实践活动帮助读者从更广的背景了解三位诺奖获得者的学术贡献与局限,以及这场争论的来龙去脉,从而在今后的研究中能扬长避短。文章的第五部分还将简要回顾RCT在我国的实践,最后是总结和讨论。

## 二、发展经济学研究选题:从“大”到“小”

经济发展是所有中低收入国家所追求的目标。发展经济学作为一个独立的学科起始于20世纪50年代,研究对象就是一个国家从不发达向发达,特别是从低收入向高收入过渡的过程。发展经济学的研究内容包括中低收入国家所特有的各类问题。但在20世纪50、60年代,发展经济学主要围绕一些宏观问题开展研究,并产生了广泛的影响。例如,1979年,阿瑟·刘易斯(Arthur Lewis)和西奥多·舒尔茨

【收稿日期】 2019-11-10

【作者简介】 易红梅,北京大学现代农业学院副教授,博士生导师,北京大学中国农业政策研究中心研究员。

【基金项目】 国家自然科学基金面上项目“我国中等职业教育质量提升政策的实施及其影响研究”(项目编号:71573246);国家自然科学基金优秀青年科学基金项目“职业教育管理与政策”(项目编号:71922001)。

特别感谢厦门大学王亚南经济研究院的饶思航同学协助作者进行文献收集和编辑工作。

① 按照每人每天不低于1.90美元的贫困标准计算。

(Theodore Schultz) 因为在经济发展方面做出了开创性研究,特别是对中低收入国家所面临问题的研究而获诺贝尔经济学奖。但之后很长一段时间,发展经济学在宏观问题上的研究陷入了一个僵局。一方面,发展经济学家为世界各国的发展开出的“药方”并没有收到预期的效果;另一方面,当时新古典经济学的演绎逻辑推理方法无法满足发展经济学研究的现实问题的需要。

这种僵局自20世纪90年代开始改变。三位诺奖得主的主要贡献之一就是发展经济学的研究视角从宏观层面转向了微观层面,将减贫这一“大问题”分解成个人或者群体层面上更微观的更容易处理的“小问题”进行研究,并通过大田实验(field experiment)来寻找解决方案。例如,三位诺奖得主认为人力资本匮乏是贫困的根本原因之一,并将如何提升中低收入国家的人力资本问题分解成具体的如何改善中低收入国家人口的教育、健康和营养等问题。其中,在研究教育问题时,他们又进一步通过RCT分别评估了一系列具体措施,例如增加投入、按照学生成绩对学生分班对改善学生学业表现的有效性<sup>[7-9]</sup>。瑞典皇家科学院常任秘书戈兰·汉松在诺奖新闻发布会上说,“在短短20年间,他们基于实验的新方法改变了发展经济学,如今这已是一个蓬勃发展的研究领域。”通过为这些具体的“小问题”寻找答案,三位诺奖得主积累了大量的关于贫困背后的具体作用机制和有针对性的减贫措施的有效性的证据,提出了当前条件下一些小的但是一定能取得进展的政策改革建议,为政策决策者和实践者在资源有限的情况下进行政策选择提供了科学的事实依据。

但是,很多学者也担心,仅关注“小问题”可能并不能从根本上解决贫困问题。世界银行经济学家 Ravallion<sup>[5]</sup>指出,回答“小问题”对于填补重大的知识空白和服务政府决策作用有限;并且如果存在贫困陷阱,依据对“小问题”研究得出的结论作决策可能只是让穷人陷入低水平均衡。要真正解决大多数人的贫困问题,还是需要从制度、经济发展等“大问题”着手。一个典型的例子是如何解释我国和其他国家与地区在减贫进程上的差异<sup>[4][5][10]</sup>。过去30多年,尽管世界贫困人口显著减少,但是减贫进程却存在巨大的区域和国别差异。我国是世界上减贫人口最

多并率先实现联合国千年发展目标的国家<sup>[11]</sup>。但是,在撒哈拉以南的非洲国家,贫困的发生率仍然很高,并且时有反弹<sup>[12]</sup>。为什么有的国家成功减贫?有的国家却进展缓慢?很多经济学家都认为,我国的政治经济制度、国家治理能力,以及在1970年代末开始的改革开放带来的经济发展无疑对减贫贡献很大,但这些因素都不是三位诺奖得主关注的“小问题”的范畴。

### 三、影响评估与RCT方法

#### (一) 影响评估与RCT方法的关系

三位诺奖得主的另一个主要贡献是将实验性方法引入发展经济学。RCT是影响评估方法的一种。政策影响评估是经济分析中最受关注的核心问题。在过去20年,影响评估已经逐渐成为很多国际组织、国家政府部门、发展机构设计和执行发展类项目或政策的重要部分。Sabet和Brown<sup>[13]</sup>对低收入和中等收入国家发展政策的影响评估研究进行了系统的回顾,他们发现在1981—2015年期间一共有4205篇公开发表的影响评估文章,其中2/3的文章是在2010—2015年期间发表的。而从国际影响评估组织(International Initiative for Impact Evaluation, 3ie)网站影响评估数据库<sup>①</sup>的最新信息来看,目前这一数量已经超过了4800篇,其中,采用RCT方法进行的影响评估约占42%。

影响评估的目标是在控制其他因素作用的情况下,分解出哪些影响(或变化)是由于某项(组)因素干预所引起的。但是社会经济是一个有机的整体,各类因素之间相互作用,错综复杂,很多经济事件或者政策往往都是内生的。因此,在评估一个项目或者政策的影响时进行归因常面临两个难题:项目安置和项目参与都不是随机决定的。影响项目安置和项目参与(自选择)的因素可能是可观测的因素,也可能是不可观测的因素,或者两者都有。此类不可观测的因素与项目安置或者是否参与项目相关时,普通最小二乘法得到的估计量不再是无偏估计量。

现代影响评估的概念是建立在反事实分析(Counterfactual Analysis)的框架上,核心问题是归因<sup>[14]</sup>。影响是指政策(或者干预)的作用对象与他在未接受政策(或干预)的情况下在结果上的差别。反

① 2003年,3ie创建了一个数据库(Impact Evaluation Repository)对在低收入和中等收入国家开展的发展项目的影响评估研究进行追踪和总结(Sabet和Brown,2018<sup>[13]</sup>)。

事实就是指干预的对象在没有接受干预的情况下的结果。因此,最理想化的方案是,对于同一个接受项目干预的主体,比较其在某一时点接受项目干预和不接受项目干预之间的结果差异。但这在现实中不可行,一个对象在某一时点要么接受了干预,要么没有接受干预,不可能两种状态同时存在。在这种情况下,我们通常考虑将一组接受干预的人群和一组与之(在统计意义上)类似但未接受干预的人群(反事实对照)进行对比,以此来评估项目或政策的真实影响。在没有干预的情况下,这两组人群的结果变量是相似的。

构建反事实对照通常有两个途径:第一个途径是通过统计方法构建一个“反事实对照”;第二个途径是在项目干预之前就通过改变项目的瞄准方式创建一个“反事实对照”。由此也相应地衍生出两类影响评估的方法:第一类是传统的计量经济学非实验性的或者拟实验的方法(quasi-experimental study),例如倍差法、匹配法、工具变量法、断点回归法等;第二类就是三位诺奖得主倡导的RCT的方法。RCT的方法在医学、农业领域已经有着很久的应用历史,但是作为一种影响评估的方法被引入经济学并得到推广却是开始于20世纪90年代。谈及RCT这一方法的应用和推广,我们不得不提及麻省理工学院阿卜杜勒-拉蒂夫-贾米尔(Abdul Latif Jameel)的贫困行动实验室(The Abdul Latif Jameel Poverty Action Lab, J-PAL)。J-PAL由阿比吉特·巴纳吉(Abhijit Banerjee)和埃斯特·迪弗洛(Esther Duflo),以及他们当时的同事塞德希尔·穆来纳森(Sendhil Mullainathan)<sup>①</sup>在2003年共同创建。J-PAL致力于通过设计和支持创新性的、大规模的RCT研究,进行政策倡导和开展培训项目为减贫提供基于科学依据的政策决策支持。截至2019年,J-PAL和它下属的研究人员已经在83个国家开展983个RCT,并且基于这些研究得出的结论制定的政策惠及了全球4亿多人<sup>[15]</sup>。

## (二) RCT方法简介

相对于第一类方法,RCT的一个主要突破是通过随机分配得到了因果效应(或处理效应)的无偏估计量。本文借鉴Duflo等人<sup>[16]</sup>列举的评估教科书对学生学业表现的影响案例来说明采用RCT方法进行因果效应估计的基本原理。 $Y_i$ 表示实际观察到的学校*i*的学生的平均成绩,用 $Y_i^T$ 表示如果得到免费教科书,学校*i*的学生的平均成绩;用 $Y_i^C$ 表示没有得

到免费教科书,学校*i*的学生的平均成绩。我们感兴趣的是获得免费教科书对学校*i*的影响,即 $Y_i^T - Y_i^C$ 。但是如前所述,我们并不能对同一所学校同时观察到两个结果。对总体而言,我们希望知道平均处理效应(Average Treatment Effect, ATE):  $E(Y_i^T - Y_i^C)$ 。在大样本的情况下,平均处理效应将收敛于对项目参与者的平均处理效应(Average Treatment Effect on the Treated, ATT),这里用 $D$ (式1)表示。

$$\begin{aligned} D &= E[Y_i^T | \text{得到免费教科书}] \\ &\quad - E[Y_i^C | \text{没有得到免费教科书}] \\ &= E[Y_i^T | T] - E[Y_i^C | C] = E[Y_i^T | T] \\ &\quad - E[Y_i^C | T] - E[Y_i^C | C] + E[Y_i^C | T] \\ &= E[Y_i^T - Y_i^C | T] + (E[Y_i^C | T] \\ &\quad - E[Y_i^C | C]) \end{aligned} \quad (1)$$

其中,第一项 $E[Y_i^T - Y_i^C | T]$ 表示项目参与者的平均处理效应,第二项 $(E[Y_i^C | T] - E[Y_i^C | C])$ 度量了选择偏误,即干预组学校和控制组学校在不接受干预时的结果上的差异。

RCT的方法就是通过随机分配将研究对象随机分配到干预组或者控制组来减少或者消除可能的选择性偏误。常用的随机分配的方法包括过度招募后随机分配(oversubscription)、随机逐步推进项目(randomized phase-in)、组内随机(within-group randomization)和随机鼓励参与(encouragement design)等,关于这些方法的详细介绍可以参考Duflo等人的介绍<sup>[16]</sup>。<sup>②</sup>随机分配生成的干预组和控制组在没有干预的情况下的结果变量理论上是没有显著差异的,即式(1)中的第二项等于0。分配完成后,干预组的学校接受干预(得到免费教材),控制组的学校则不接受干预(没有免费教材)。通过比较两组在干预后均值的差异可以得到政策或者项目的因果效应的无偏估计量(式(2))。

$$\hat{D} = \hat{E}[Y_i | T] - \hat{E}[Y_i | C] \quad (2)$$

在大样本情况下, $D$ 收敛于式(3)中的差值。如果样本为独立同分布,即个体*i*是否参与项目不影响其他任何个体的情况下,ATT=ATE。

$$D = E[Y_i^T | T] - E[Y_i^C | C] \quad (3)$$

通过公式(4)中的计量经济学回归模型也能计

<sup>①</sup> 穆来纳森现在就职于芝加哥大学布斯商学院。

<sup>②</sup> 这些方法的选择很大程度上取决于项目的预算和实地实施项目的可行性。

算出  $D$  的无偏估计量。其中,  $T$  是一个反映随机分配结果的二元变量, 通常干预组为 1, 控制组为 0。采用普通最小二乘回归的方法就能得到估计量  $\hat{\beta}_{OLS} = \hat{E}[Y_i^T | T] - \hat{E}[Y_i^C | C]$ 。

$$Y_i = \alpha + \beta T + \varepsilon_i \quad (4)$$

相对于第一类影响评估方法而言, 通过 RCT 得到的因果效应的估计结果对于一般公众而言更直接也更容易理解。这也是 RCT 在社会经济研究中迅速崛起的原因之一。但是, 要运用 RCT 方法开展研究并不是一件简单的事情。在技术层面, 开展 RCT 研究还需要至少考虑两个问题。一个是样本量、实验设计和统计效力问题。通常要求, 如果项目有影响, 样本量至少要在给定的显著性水平下, 能够以相当的概率 (统计效力) 拒绝干预没有影响的假说。另一个是要对干预的依从率进行监测。依从是指研究对象服从随机分配, 即干预组对象接受干预, 控制组对象不接受干预。如果存在不依从, 需要结合工具变量的方法估计项目的局部处理效应。

### (三) RCT 方法的优势与局限性

与第一类方法相比, RCT 方法具有三个方面的优势。第一个优势是估计量的内部有效性。内部有效性通常指一个统计量是否是无偏估计量和有效估计量<sup>①</sup>。第一类方法只有在满足条件独立假定 (Conditional Independence Assumption) 的情况下, 才有可能得到无偏估计<sup>[17]</sup>。而 RCT 方法只要满足弱假定——项目安置的条件外生 (Conditional Exogeneity of Program Placement) 就能够得到无偏估计量<sup>[18]</sup>。因此, 相对于第一类方法, RCT 方法得到因果关系的无偏估计要求的假定条件更少, 虽不必然但通常能得到无偏的估计结果。有学者指出, 即便是完全的随机分配 (或者纯随机分配), 在单次实验中, 仍然可能出现随机分配导致两个组在某些重要的因素上不平衡, 从而最后的影响可能是干预和该因素共同作用的结果<sup>[6]</sup>。如果随机分配过程采用的是部分随机分配, 即基于一些可观测的特征进行的随机分配, 这种情况下得到的估计量可能只是一致性估计量<sup>[18]</sup>。并且, 一个估计量是无偏估计量并不一定意味着它是有效估计量。也就是说, RCT 在进行统计推断时面临的问题并不一定比第一类方法少。但总体而言, 目前大部分学者都认为通过 RCT 方法得到的估计量的内部有效性更加可

靠, 因此被认为是政策影响评估的“黄金标准”。

此外, RCT 还具有以下两个优势: (1) 对于一般公众而言, 采用 RCT 方法估计的因果关系比采用其他方法得到的结果更容易理解。正如 Ravallion<sup>[5]</sup> 课堂上的学生所言“我采用 RCT 是因为这是我父母能理解的方法”。影响评估在近些年得到国际组织、政府部门、非政府机构和公众的认可和接受很大程度上也得益于 RCT 方法的普及。(2) 政策影响的对象是有限理性的决策人或者具有各种决策行为偏差的决策人, RCT 让研究者有机会在政策实施前通过小范围的实验检验不同政策设计的效果, 帮助完善政策设计, 提高决策质量。

但是, RCT 也具有很大的局限性。第一个局限是采用 RCT 方法得出的研究结论的外部有效性, 即研究结论是否或者在多大程度上适用于其他人群。研究结论的外部代表性主要取决于研究样本对总体的代表性。RCT 结论的外部有效性在现实中常面临很大的挑战。造成这一问题的主要原因可以归纳为四个方面: 一是相对于第一类方法可采用的样本量, RCT 的样本量通常比较小, 并且项目安置也并不是完全随机的, 本身能代表的总体有限。二是项目从小范围实验到大面积推广可能会打破原来的市场均衡, 最后在大面积推广时产生不同于 RCT 阶段的影响。三是由于“霍桑效应” (Hawthorne Effect), 参与项目这个事实可能让干预组参与者的行为发生改变, 导致 RCT 方法估计的影响可能要大于项目实地实施时的影响。四是 RCT 的“干预”通常由非政府组织或者研究者自己来实施, 其组织方式和执行能力与很多中低收入国家的政府部门差异较大, 从而导致一些 RCT 研究发现有影响的政策或项目到大范围推广时并不能产生预期的效果。例如, Mo<sup>[19]</sup> 的研究就发现, 同样的项目干预, 政府部门负责执行时产生的影响要显著降低。

RCT 方法的第二个局限在于并不是所有的问题都适合采用 RCT 的研究方法。一些宏观性的“大问题” (例如政治问题、产业政策) 是没有办法开展 RCT 的。因此, 目前的 RCT 基本上是以研究和解决“小问题”为特点, 从而导致其研究对知识积累或者重大创新的贡献可能有限。同时, 由于研究伦理的要求, 一些可能会让研究对象受到损害的 RCT 研究是不被允许的。

RCT 方法还因为未能对项目或政策产生影响的

① 无偏性是指估计量抽样分布的数学期望等于总体参数的真值; 精度或者有效性是指估计量与总体参数的离散程度。

作用机制进行解释曾备受争议。早期的 RCT 主要致力于回答项目或者政策是否有影响的问题,因此,一些学者认为 RCT 得出的结论缺乏理论根基,是一个“黑匣子”<sup>[6][20][21]</sup>。但是,现今很多 RCT 研究,包括三位诺奖得主的研究都是基于变化的理论(theory of change)绘制了干预与结果之间清晰的因果链,深入理解了干预是如何通过改变信息、激励或环境限制等因素影响到结果变量的作用机制<sup>[9][22]</sup>。以理论为基础的因果链的建立帮助 RCT 在回答是否有影响的同时还解释了“为什么”,这也将进一步促进 RCT 的发展<sup>[6]</sup>。

#### 四、基于事实的政策决策

三位诺奖得主的第三个贡献是以他们为代表的一批发展经济学家采用 RCT 方法研究贫困问题,极大地促进了中低收入国家以事实为基础的政策决策和实践活动。一个旨在减少非洲儿童辍学率的项目启动前,政府部门或者投资者常常面临着很多困惑:目前的这个方案能够减少辍学吗?还有更有效的方案吗?遗憾的是,影响评估兴起之前,很少有事实能够帮助回答这些问题。但现在,随着影响评估的发展,以及以影响评估为基础的成本有效性分析和荟萃分析(meta-analysis)等研究工具的结合,研究者们已经积累了大量的事实为回答这些问题提供参考。

RCT 应用的领域也从早期仅关注教育和健康拓展到更多的发展问题。从 20 世纪 90 年代开始,三位诺奖得主采用 RCT 的方法从如何改善中低收入国家人口的教育、健康等人力资本的研究入手,逐步扩展到对行为偏误、性别与政治、信贷等领域的研究,为相关领域决策积累了丰富的经验。在这股力量的带动下,全球影响评估实践也在 2000 年之后迎来了多领域百花齐放的时期。根据 3ie 影响评估数据库的信息,1990—2015 年,健康、教育和社会保障领域的研究约占影响评估研究的 65%,农业与农村发展、私营部门发展、信息技术等领域的影响评估自 2000 年以来得到了较快的发展<sup>[13]</sup>。

以这些研究为基础提供的事实依据具有两个显著特点:一是相关政策都不是平地而起,基本是对既有政策或制度有效性的检验,并在此基础上加以微小的改进;二是结论清晰,政策建议明确,易于操作。以三位诺奖得主在教育领域的研究为例,Kremer 和其合作者早期在肯尼亚的 RCT 发现,仅增加投入(提供免费教科书、活动挂图)可能可以提高学生的出

勤率,但是难以保证提高学生的学业成绩<sup>[7][8][23]</sup>。这些研究不仅对以增加投入为主要政策工具的发展政策发出了预警,还极大地促进了研究者们将目光从关注教育对中低收入国家人口的可达性转向关注中低收入国家的教育质量。之后,三位诺奖得主通过一系列的 RCT 评估了分快慢班、改变班级规模、课外补习、计算机辅助学习、增强教师激励、改革教师聘用制度等具体政策措施对提高学生学业表现的影响<sup>[9][24-27]</sup>。

我们已经看到,三位诺奖得主的很多研究成果已经转化为政策,对中低收入国家的普通民众产生了广泛的影响。例如,根据 Kremer 及其合作者关于驱虫对学生健康和学业的影响及如何有效驱虫的系列研究<sup>[23][28-30]</sup>,世界卫生组织在 2017 年建议利用大规模驱虫方式改善儿童健康和营养<sup>[31]</sup>。而 Banerjee 和 Duflo 等人对教师和学生水平匹配对学生学业表现影响的研究结论也推动教育部门改变了传统的按学生年龄分班的做法,实施将学生按照学习水平分班,目前这项政策已经覆盖了印度和非洲 6 000 万学生<sup>[25][32]</sup>。根据 J-PAL 网站的信息,截至 2019 年,全球有超过 4 亿人从 J-PAL 的研究中获益<sup>[15]</sup>。

#### 五、RCT 方法在我国的实践

RCT 方法在我国社会经济科学中的应用最早开始于健康领域。20 世纪 90 年代初,Wang 等<sup>[33]</sup>就在上海 21 家工厂和 6 所中学开展了一项 RCT 评估丈夫参与健康教育对降低女性怀孕和流产率的影响。但直到 2000 年代中期,我国采用随机干预试验方法开展的社会经济研究仍然都集中在健康领域。

过去 10 余年,RCT 方法的应用在我国教育领域的研究中得到了迅速发展,成为仅次于健康领域的第二大领域。截至 2019 年 11 月,3ie 网站影响评估数据库一共收录了 2 030 项采用 RCT 方法开展的影响评估研究,其中 129 项来自我国,占总数的 6%,其中涉及教育主题的研究是 51 项<sup>[34]</sup>。51 项研究中,超过 70% 的研究都是农村教育行动计划(REAP)研究团队完成的。他们第一次将 RCT 引入对我国教育问题的研究,并在 2007—2019 年期间围绕防止学生辍学、提高初高中阶段教育质量、提高教师教学水平、儿童早期发展、营养与健康、计算机辅助学习、儿童视力保护与学生学业表现等议题开展了 70 多个 RCT,为我国政府改进和完善相关教育政策、促进教育公平和提高教育质量做出了贡献。

但总体来说,目前我国 RCT 方法的应用领域还相对比较窄。3ie 网站影响评估数据库中收录的在我国开展的 129 项 RCT 研究中,124 项研究关注的主题都是健康和教育,仅个别研究来自于农业与农村发展、保险等领域<sup>[34]</sup>。除了对方法的认识存在局限外,RCT 在我国应用不广的一个重要原因在于开展 RCT 的实地工作量远远大于其他方法的工作量,尤其是 RCT 的“干预”除了需要大量的资金投入,更需要执行机构的支持。就像 J-PAL,如果没有像 Pratham 这样的非政府组织的支持,他们也无法开展如此大量的 RCT 研究。

## 六、总结与讨论

综上所述,三位诺奖得主将贫困的“大问题”分解成一个个微观的“小问题”进行研究,并将实验经济学方法引入发展经济学,不仅促进了发展经济学的转型与发展,还为世界反贫困行动提供了大量基于事实的科学依据。虽然关注“小问题”可能并不能从根本上解决贫困问题,RCT 的方法也并不是完美的,但是不可否认,三位诺奖得主打破了发展经济学早期研究范式的局限,引领了发展经济学研究的转型,推动了全球基于事实的政策决策发展,让数以亿计的中低收入国家的贫困人口因他们的研究受益,在学科发展、研究方法和反贫困事业上都做出了杰出的贡献。

没有一种研究或者方法能够解决所有的发展问题。正如 Imbens<sup>[35]</sup>所言,我们不能因为 RCT 的不足就忽略它在实证研究方法上向前迈出的的一大步。这种方法与其他研究方法相互补充,能够帮助我们回答更

多的问题<sup>[6]</sup>。我们在研究中需要问“正确的问题”,需要根据我们的研究问题选择适合的研究方法,并正确认识每种方法的优缺点,不断改进,才能真正促进学术研究的繁荣和发展。

目前,国内对于 RCT 方法的系统性介绍不够,导致学术界、项目资助方和社会公众对 RCT 方法都存在一些误解。一个极端是认为 RCT 方法完美无缺,是解决所有问题的“万能药”,对 RCT 方法盲目推崇。尤其是这种方法很容易通过“生动的故事”让资助方观看到项目的成效,更容易受到过度的追捧。另一个极端是认为 RCT 研究的都是“显而易见的常识问题”,没有价值。然而,Banerjee 和 Duflo 在《贫穷的本质》<sup>[36]</sup>中的论述正是围绕此类“常识问题”展开,他们通过回答为什么贫困人口并没有如我们预期的那样遵从“常识”生活,深入剖析了贫困人口的行为机制,通过政策设计的微小改变改善了贫困人口的教育与健康。

最后,RCT 方法在医学等领域应用的经验与教训也为该方法在发展经济学领域的未来发展提供了一些有价值的借鉴。一是考虑到 RCT 的资助方和研究者对“积极的干预结果”的过度追求<sup>[37]</sup>,建议采用 RCT 方法的研究最好能提前发布主要的分析计划,避免过度进行“数据挖掘”。二是要加强对社会经济学中开展的 RCT 的伦理审查。从影响评估的趋势来看,未来会有更多的研究者采用 RCT 的方法对贫困等发展问题开展研究,但是目前国内很多高校等研究机构都没有相应的专业伦理审查委员会,对此的监控还有待加强。

## 参考文献

- [1] The World Bank. Poverty Overview [EB/OL]. (2019-10-02) [2019-11-10]. <https://www.worldbank.org/en/topic/poverty/overview>.
- [2] United Nations. Goal 1: End Poverty in All Its Forms Everywhere [EB/OL]. [2019-11-10]. <https://www.un.org/sustainabledevelopment/poverty/>.
- [3] Sen A. Health in Development, Kenote to 52<sup>nd</sup> World Health Assembly [C/OL]. World Health Organization, (1999-05-18) [2019-11-10]. <https://apps.who.int/iris/handle/10665/79397>.
- [4] 林毅夫. 新结构经济学 [J]. 经济学(季刊), 2010, 10(1): 1-32.
- [5] Ravallion M. Fighting Poverty One Experiment at a Time: A Review of Abhijit Banerjee and Esther Duflo's Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty [J]. Journal of Economic literature, 2012, 50(1): 103-114.
- [6] Deaton A, Cartwright N. Understanding and Misunderstanding Randomized Controlled Trials [J]. Social Science & Medicine, 2018, 210: 2-21.
- [7] Glewwe P, Kremer M, Moulin S. Many Children Left Behind? Textbooks and Test Scores in Kenya [J]. American Economic Journal: Applied Economics, 2009, 1(1): 112-135.
- [8] Glewwe P, Kremer M, Moulin S, Zitzewitz E. Retrospective vs. Prospective Analyses of School Inputs: The Case of Flip Charts in Kenya [J]. Journal of Development Economics, 2004, 74(1): 251-268.
- [9] Duflo E, Dupas P, Kremer M. Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya

- [J]. *American Economic Review*, 2011, 101 (5): 1739-1774.
- [10] 张晓波. 诺贝尔经济学实验方法的实验方法可借鉴, 但没必要迷信 [N/OL]. 新京报, (2019-10-14) [2019-11-10]. <http://www.bjnews.com.cn/finance/2019/10/14/636373.html>.
- [11] United Nations. The Millennium Development Goals Report 2015 [R/OL]. (2015-07-01) [2019-11-10]. <https://www.un.org/millennium-goals/reports.shtml>.
- [12] World Bank. Global Monitoring Report 2015/2016: Development Goals in an Era of Demographic Change [M]. Washington, DC: World Bank, 2016.
- [13] Sabet S M, Brown A N. Is Impact Evaluation Still on the Rise? The New Trends in 2010-2015: For Submission to the Journal of Development Effectiveness [J]. *Journal of Development Effectiveness*, 2018, 10 (3): 291-304.
- [14] Rubin D B. Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies [J]. *Journal of Educational Psychology*, 1974, 66 (5): 688.
- [15] J-PAL. Release: J-PAL Co-Founders Abhijit Banerjee and Esther Duflo Awarded Nobel Memorial Prize in Economics [EB/OL]. (2019-10-15) [2019-11-10]. <https://www.povertyactionlab.org/news/release-j-pal-co-founders-abhijit-banerjee-and-esther-duflo-awarded-nobel-memorial-prize>.
- [16] Duflo E, Glennerster R, Kremer M. Using Randomization in Development Economics Research: A Toolkit [J]//Schultz T P, Strauss J A. *Handbook of Development Economics*. Amsterdam: Elsevier B. V., 2007, 4: 3895-3962.
- [17] Rosenbaum P R, Rubin D B. Reducing Bias in Observational Studies Using Subclassification on the Propensity Score [J]. *Journal of the American Statistical Association*, 1984, 79 (387): 516-524.
- [18] Khandker S, Koolwal G B, Samad H A. *Handbook on Impact Evaluation: Quantitative Methods and Practices* [M]. Washington, DC: World Bank, 2009.
- [19] Mo D. The Impact of Government Upscaling Efforts on the Effectiveness of Computer Assisted Learning Programs in Rural China [J]. REAP working paper, 2015.
- [20] Imai K, Keele L, Tingley D, Yamamoto T. Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies [J]. *American Political Science Review*, 2011, 105 (4): 765-789.
- [21] Cartwright N. A Philosopher's View of the Long Road from RCTs to Effectiveness [J]. *The Lancet*, 2011, 377 (9775): 1400-1401.
- [22] Bonell C, Fletcher A, Morton M, Lorenc T, Moore L. Realist Randomised Controlled trials: A New Approach to Evaluating Complex Public Health Interventions [J]. *Social Science & Medicine*, 2012, 75 (12): 2299-2306.
- [23] Miguel E, Kremer M. Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities [J]. *Econometrica*, 2004, 72 (1): 159-217.
- [24] Duflo E, Dupas P, Kremer M. School Governance, Teacher Incentives, and Pupil-teacher ratios: Experimental Evidence from Kenyan Primary Schools [J]. *Journal of Public Economics*, 2015, 123: 92-110.
- [25] Banerjee A V, Cole S, Duflo E, Linden L. Remedying Education: Evidence from Two Randomized Experiments in India [J]. *The Quarterly Journal of Economics*, 2007, 122 (3): 1235-1264.
- [26] Duflo E, Hanna R, Ryan S P. Incentives Work: Getting Teachers to Come to School [J]. *American Economic Review*, 2012, 102 (4): 1241-1278.
- [27] Glewwe P, Ilias N, Kremer M. Teacher Incentives [J]. *American Economic Journal: Applied Economics*, 2010, 2 (3): 205-227.
- [28] Kremer M, Miguel E. The Illusion of Sustainability [J]. *The Quarterly Journal of Economics*, 2007, 122 (3): 1007-1065.
- [29] Bundy D A, Kremer M, Bleakley H, Jukes M C, Miguel E. Deworming and Development: Asking the Right Questions, Asking the Questions Right [J]. *PLoS neglected Tropical Diseases*, 2009, 3 (1): e362.
- [30] Baird S, Hicks J H, Kremer M, Miguel E. Worms at Work: Long-run Impacts of a Child Health Investment [J]. *The Quarterly Journal of Economics*, 2016, 131 (4): 1637-1680.
- [31] 世界卫生组织. 世卫组织建议利用大规模驱虫方式改善儿童健康和营养 [EB/OL]. (2017-09-29) [2019-11-10]. <https://www.who.int/zh/news-room/detail/29-09-2017-who-recommends-large-scale-deworming-to-improve-children-s-health-and-nutrition>.
- [32] Banerjee A, Banerji R, Berry J, Duflo E, Kannan H, Mukherji S, Shotland M, Walton M. Mainstreaming an Effective Intervention: Evidence from Randomized Evaluations of "Teaching at the Right Level" in India [J]. *National Bureau of Economic Research*, 2016.
- [33] Wang C C, Vittinghoff E, Hua L S, Yun W H, Rong Z M. Reducing Pregnancy and Induced Abortion Rates in China: Family Planning with Husband Participation [J]. *American Journal of Public Health*, 1998, 88 (4): 646-648.
- [34] International Initiative for Impact Evaluation (3ie). *Impact Evaluation Repository* [EB/OL]. (2019). <https://www.3ieimpact.org/evidence-hub/impact-evaluation-repository>.
- [35] Imbens G W. Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009) [J]. *Journal of Economic Literature*, 2010, 48 (2): 399-423.
- [36] Banerjee A V, Duflo E. *Poor economics: A Radical Rethinking of the Way to Fight Global Poverty* [M]. New York: PublicAffairs, 2011.
- [37] Bothwell L E, Greene J A, Podolsky S H, Jones D S. Assessing the Gold Standard-Lessons from the History of RCTs [J]. *The New England Journal of Medicine*, 2016, 374 (22): 2175-2181.

(责任编辑: 李 晟 张安平)