

ВЫСШАЯ ШКОЛА ЭКОНОМИКИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

Р.И. Капелюшников

**РАНДОМИСТЫ:
НОВАЯ ЭКОНОМИКА РАЗВИТИЯ**

Препринт WP3/2022/07
Серия WP3
Проблемы рынка труда

Москва
2022

УДК 332.05

ББК 65.7

К20

Редактор серии WP3
«Проблемы рынка труда»
В.Е. Гимпельсон

Капелюшников, Р. И.

К20 Рандомисты: новая экономика развития [Текст] : препринт WP3/2022/07 / Р. И. Капелюшников ; Нац. исслед. ун-т «Высшая школа экономики». — М. : Изд. дом Высшей школы экономики, 2022. — 60 с. — (Серия WP3 «Проблемы рынка труда»). — 35 экз.

Работа посвящена достаточно необычному явлению в современной экономической науке — интеллектуальному противостоянию между сторонниками «новой» экономики развития («рандомистами») и ее противниками. Общий посыл нового подхода в том, что единственным действительно научным методом следует считать технику рандомизированных контролируемых экспериментов (randomized controlled trials — RCT) и что при изучении проблем развивающихся стран следует использовать только ее. RCT были признаны «золотым стандартом» при оценке эффективности программ борьбы с бедностью; сегодня это направление абсолютно доминирует в экономике развития. В работе рассматриваются главные отличительные черты экономики RCT: стремление подражать медико-биологическим исследованиям; атеоретический характер; идея иерархии методов, на вершине которой находится RCT; отказ от любых других методов помимо экспериментальных; установка на идентификацию каузальных эффектов без попыток выяснения, какие механизмы лежат в их основе; претензия на то, что при оказании помощи развивающимся странам только RCT позволяют отделять «то, что работает» от «того, что не работает»; переориентация от масштабных макроэкономических и институциональных реформ на точечные социальные и гуманитарные интервенции. Подробно обсуждаются проблемы внутренней и внешней валидности RCT; специальный раздел посвящен политико-экономическим аспектам рандомистской программы действий. Анализ позволяет сделать вывод, что с концептуальной точки зрения убедительнее выглядит позиция критиков: навязывание идеи иерархии методов ненаучно; никакого методологического «золотого стандарта» в экономическом анализе не существует и существовать не может; рандомизацию в реальных полевых условиях недопустимо смешивать с рандомизацией в гипотетических идеальных условиях; внутренняя валидность RCT далеко не безупречна и получаемые на их основе оценки нельзя считать несмещенными; RCT полностью бессильны перед проблемой внешней валидности; RCT чрезвычайно дороги и сталкиваются с большим числом этических проблем; из-за низкой внешней валидности политические рекомендации, вырабатываемые на основе RCT, имеют крайне ограниченную практическую ценность. Однако логика критиков оказалась слабее риторики рандомистов: авторитет экономики RCT в глазах научного сообщества, политиков, СМИ и широкой публики остается непоколебленным. Скорее всего, поток рандомистских исследований будет становиться все шире, а их интеллектуальное и политическое влияние все сильнее.

УДК 332.05
ББК 65.7

Ключевые слова: рандомисты, экономика развития, внутренняя валидность, внешняя валидность, рандомизированные контролируемые испытания

JEL: A12, B41, C93, I32, O20

Капелюшников Ростислав Исаакович (rostis@hse.ru) — член-корреспондент Российской академии наук (РАН), главный научный сотрудник ИМЭМО РАН, заместитель директора Центра трудовых исследований (ЦеТИ) Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ).

**Препринты Национального исследовательского университета
«Высшая школа экономики» размещаются по адресу: <http://www.hse.ru/org/hse/wp>**

© Капелюшников Р.И., 2022

Введение

«Рандомистами» Ангус Дитон окрестил сторонников нового направления в экономике развития (development economics), которые считают, что любые программы по борьбе с бедностью в развивающихся странах должны основываться на технике рандомизированных контролируемых испытаний (randomized controlled trials — RCT), причем не в лабораторных, а в полевых условиях [Deaton, 2009]. Взлет нового движения был стремительным: возникнув на волне «эмпирического поворота» в экономической науке на рубеже 1990–2000-х годов, оно быстро завоевало широчайшее признание не только среди академических исследователей, но также среди политиков, неправительственных организаций и международных агентств, занимающихся оказанием помощи бедным странам. Практически мгновенно оно стало не просто частью, но самым передним краем экономического мейнстрима. RCT были признаны «золотым стандартом» при изучении истоков бедности и оценке программ борьбы с нею. За считанные годы сформировалось глобальное сообщество рандомистов: существуют уже десятки международных центров, организующих и финансирующих проведение RCT. Сегодня этот подход полностью доминирует в экономике развития.

В последующем изложении мы будем обозначать эту исследовательскую программу терминами «новая экономика развития» и «экономика RCT», рассматривая их как синонимичные.

Успех нового направления был поистине ошеломляющим: по имеющимся оценкам, с начала 2000-х годов наблюдался 30-кратный рост числа академических публикаций с использованием рандомизированных контролируемых экспериментов. Сейчас ежегодно публикуется порядка 350 работ, где излагаются и анализируются результаты таких экспериментов [Ravallion, 2020]. Веб-сайт одного из главных рандомистских центров — Лаборатории борьбы с бедностью им. Абдул-Латифа Джамии (J-PAL), созданной в 2003 г. при Массачусетском технологическом институте, — содержит информацию о более чем 1000 RCT, проведенных в более чем 80 странах [de Cler, 2021]. В последнее время при оценке эффективности проектов, осуществляемых Всемирным банком, метод рандомизации использовался в каждом двух случаях из трех [Bédécarrats et al., 2019]. Требование проведения RCT все чаще выдвигается сегодня как обязательное усло-

вие при выделении средств на новые программы по борьбе с бедностью, разрабатываемые правительствами, международными институтами и частными благотворительными фондами. Оценивание проектов с использованием RCT в качестве идеальной модели получило настолько широкое распространение, что превратилось в настоящую индустрию, куда ежегодно вкладываются сотни миллионов долларов из самых разных источников [Bédécarrats et al., 2019].

Сторонники метода рандомизации расценивают его проникновение в экономические исследования как научную революцию: «Создание культуры, в которой поощряются, продвигаются и финансируются строгие рандомизированные оценки, может революционизировать социальную политику в XXI в., точно так же, как рандомизированные испытания произвели революцию в медицине в XX в.» [Duflo, Kremer, 2008, p. 117]. Иллюстрацией того, как далеко простираются притязания нового подхода, может служить название книги известного австралийского экономиста Эндрю Ли: «Рандомисты: как радикальные исследователи меняют мир» [Leigh, 2018]. Речь, как видим, идет не просто о новом понимании реальности, но также о ее радикальном преобразовании на практике. Прогресс в теории и политике развития рандомисты связывают с систематическим накоплением экспериментальных данных на уровне отдельных точечных интервенций. Конечной целью движения является создание универсальной базы данных о «хороших» и «плохих» формах политики развития с тем, чтобы правительства и частные спонсоры могли делать осознанный выбор, имея перед глазами полный список «того, что работает» и «того, что не работает» [Labrousse, 2020]. Когда благодаря RCT в распоряжении лиц, принимающих решения, появятся неопровержимые научные доказательства, им не останется ничего другого как проводить правильную политику: программы, которые «не работают», будут выводиться из употребления, а программы, которые «работают», повсеместно внедряться. По подсчетам самих рандомистов, благодаря их исследованиям удалось уже улучшить жизнь более чем 400 млн жителей развивающихся стран [Еникополов, 2020].

Публичным признанием научных и практических достижений новой экономики развития стало присуждение в 2019 г. Нобелевской премии по экономике трем ведущим рандомистам — Эстер Дюфло, Абхиджиту Банерджи и Майклу Кремеру. Премия была дана им «за их экспериментальный подход к сокращению глобальной бедности»,

а также за то, что они «превратили экономику развития — раздел, изучающий причины глобальной бедности и способы борьбы с ней, — в бурно растущую, преимущественно экспериментальную область исследований». По заключению Нобелевского комитета, «теперь у нас есть большое количество конкретных данных о конкретных механизмах, лежащих в основе бедности, и конкретных мерах по ее преодолению», что «значительно улучшило нашу способность бороться с бедностью на практике» [Royal Swedish Academy of Sciences, 2019].

Однако далеко не все ведущие экономисты согласны с такой панегирической оценкой. У новой экономики развития есть немало последовательных оппонентов, которые отвергают ее претензии на то, чтобы считаться единственно правильным и единственно научным подходом к решению экономических и социальных проблем развивающихся стран. Они подвергают ее жесткой критике по множеству оснований — от чисто технических до политических. Так, лауреат Нобелевской премии Джеймс Хекман характеризует энтузиазм по поводу RCT как «квазирелигиозный» [Heckman, 2020], а другой лауреат Нобелевской премии, Ангус Дитон, называет мышление рандомистов скорее «магическим», нежели научным, отмечая, что они ставят метод впереди содержания [Deaton, 2020]. Одному из ведущих авторитетов в области экономики развития, британскому экономисту Ланту Притчетту, всеобщее увлечение рандомизацией представляется «помешательством» (madness) [Pritchett, 2020]. Новозеландский экономист Роберт Пичиотто сравнивает рандомистов с адептами новой «истинной веры»: «Они испытывают моральную уверенность в своей правоте и не готовы принимать доказательства, противоречащие открывшейся им истине. Они исключают другие точки зрения, предпочитая общаться с другими истинно верующими и пытаясь преодолеть сопротивление неверующих путем их исключения» [Picciotto, 2020].

Интеллектуальное противостояние в экономике развития длится уже не одно десятилетие без видимых признаков к сближению позиций. Настоящая работа посвящена анализу этой достаточно необычной для современной экономической науки «патовой» ситуации. Как и почему такое могло произойти? Каковы перспективы выхода из создавшегося концептуального клинча? Что может ожидать экономику развития в обозримом будущем? Чтобы попытаться ответить

на эти вопросы, мы рассмотрим как методологические, так и содержательные аспекты экономики РСТ, а также представим основные контраргументы, выдвигаемые ее критиками. Анализ этого интеллектуального противостояния способен, на наш взгляд, пролить свет на то, какими сложными и неожиданными путями развивается подчас современная экономическая мысль.

Общая характеристика

Но прежде чем обращаться к обсуждению этих вопросов, сначала стоит пояснить на условном примере, что именно имеют в виду, когда говорят о рандомизированных контролируемых экспериментах, и как именно эти эксперименты обычно строятся¹.

Как хорошо известно, в развивающихся странах существует достаточно острая проблема, связанная с абсентеизмом школьных учителей, склонных манкировать своими обязанностями и нерегулярно появляющихся на рабочем месте. Допустим, мы решили провести РСТ в каком-то регионе какой-то развивающейся страны, где для одной случайно отобранной группы школ оставили все без изменения (контрольная группа — control group), а для другой (экспериментальная группа или группа воздействия — treatment group) ввели систему регистрации времени прихода и ухода учителей (скажем, с помощью камер наблюдения). Через какое-то время мы сравнили показатели абсентеизма учителей, а также показатели успеваемости учащихся в обеих группах. Если обнаружился значительный по величине и статистически значимый разрыв в пользу экспериментальной группы (в ней ниже абсентеизм и выше успеваемость), то значит, оказанное нами воздействие (контроль времени) стало *причиной* меньшего абсентеизма учителей и лучшей успеваемости учащихся. Нам удалось идентифицировать и количественно оценить *каузальный эффект* (treatment effect) вмешательства, иными словами — эффективность тестируемого проекта. Тогда результаты нашего эксперимента могут быть использованы в качестве весомого аргумента для внедрения системы контроля за временем прихода и ухода учителей в школах других регионов или даже всей страны.

¹ Описание наиболее известных рандомизированных контролируемых экспериментов в рамках экономики развития см.: [Банерджи, Дюфло, 2021; Еникопов, 2020].

Исходный импульс к проведению таких экспериментов в рамках новой экономики развития исходил от медико-биологических исследований, где новые лекарства и новые методы лечения перед тем, как быть рекомендованными к применению на практике, проходят обязательную проверку на эффективность в ходе двойных слепых рандомизированных испытаний: «Необходимо имитировать рандомизированные исследования, которые применяются в медицине для оценки эффективности новых лекарств» [Банерджи, Дюфло, 2021, с. 29]. Опыт медицинской науки исходно служил для рандомистов образцом и примером для подражания: «В XX веке клинические испытания, — отмечала Э. Дюфло, — произвели революцию в медицинской практике. К сожалению, это не относится к политическим мерам в области образования и здравоохранения. Часто такие меры не подвергаются тщательной оценке до того, как они начнут широко применяться... Можно, однако, при оценивании пилотных программ в области образования и здравоохранения черпать вдохновение из клинических испытаний» (цит. по: [Favereau, 2016])². Рандомисты убеждены, что подобно тому, как существует доказательная (evidence-based) медицина, точно так же должна существовать и доказательная (evidence-based) экономика, олицетворение которой они видят в RCT [Ibid.]³. Новая экономика развития выступает с претензией на то, что она способна положить конец затяжным теоретическим спорам по поводу стратегии помощи развивающимся странам, снабдив политиков этих стран надежным, научно обоснованным руководством к действию — конкретными рецептами, эффективность которых подтверждена экспериментально.

Интеллектуальный фон. Победному маршу экономики RCT способствовали несколько обстоятельств. Первое — это кризисное состояние традиционной («старой») экономики развития. Многочисленные макроэкономические исследования так и не смогли выявить

² Если экономика развития пойдет по этому пути, убеждена Э. Дюфло, она сможет повторить успех медицины: «Медицинская наука не предложила пока миру средства от всех болезней, но современные врачи спасают миллионы жизней каждый год. Мы, экономисты, должны добиваться того же» (<https://www.youtube.com/watch?v=0zvrgiPkVcs>).

³ По сути, речь идет ни много ни мало как о смене дисциплинарного идеала: на протяжении двух с лишним столетий экономическая наука мечтала стать как теоретическая физика, но в конце концов сделалась как медицинская статистика [Капелюшников, 2021].

какой-либо устойчивой связи между объемами помощи, типами проводимой экономической политики и темпами роста развивающихся стран. В ретроспективе экономика развития предстает как длинная череда интеллектуальных мод, когда одна предполагаемая волшебная палочка сменяла другую, что оборачивалось постоянным шарханьем из стороны в сторону: от планирования — к развитию инфраструктуры и человеческому капиталу, от них — к структурным реформам, улучшению здравоохранения и социальному капиталу, от них — к защите окружающей среды и вновь развитию инфраструктуры, причем не было похоже, чтобы этот процесс сопровождался прогрессирующим накоплением научного знания [Deaton, 2010]. Новая экономика развития отказалась участвовать в бесплодных теоретических дебатах, сосредоточившись на решении конкретных практических проблем, с которыми приходится сталкиваться развивающимся странам: «Слишком часто политика развития основывалась на очередной моде, тогда как рандомизированные оценки смогут помочь ей основываться на доказательствах» [Duflo, 2004].

Второе — это глубокое разочарование в способности Всемирного банка, главного интеллектуального и финансового центра по оказанию помощи развивающимся странам, извлекать уроки из собственных проектов, оценивать их эффективность и предлагать убедительные доказательства того, что эти проекты действительно способны обеспечивать устойчивый экономический рост и сокращение бедности [Deaton, 2010]. Провал многих макрореформ в развивающихся странах привел к взрывному росту популярности микропроектов, оказавшихся в центре внимания новой экономики развития. В отличие от макрореформ такие узкоцелевые проекты поддаются предварительному тестированию с помощью RCT, что в глазах рандомистов служит важнейшим аргументом в их пользу.

Третье — это резко возросшее среди экономистов недоверие к традиционным (неэкспериментальным) методам эконометрического анализа, не оправдавшим возлагавшихся на них надежд из-за неспособности решить проблему идентификации, то есть выявления причинно-следственных связей [Bédécarrats et al., 2019]. Профессиональное сообщество было измучено десятилетиями непрекращающихся споров о потенциальных смещениях при использовании эконометрических методов [Ogden, 2017]. Возникла усталость от бесконечных межстрановых регрессий, которые никуда не вели и из кото-

рых было невозможно вывести никаких однозначных рекомендаций о том, каким образом развивающиеся страны могли бы переходить на траекторию устойчивого роста. Метод рандомизации открывал выход из этого тупика, поскольку обещал переход от фиксации корреляций к установлению реальных каузальных эффектов. И поскольку, как показал опыт, феномен экономического роста «сопротивлялся» получению однозначных каузальных объяснений, новая экономика развития сделала главный акцент на феномене бедности, который поддается им намного лучше.

В результате можно говорить о нескольких принципиальных отличиях «новой» экономики развития от «старой»: 1) переориентация от исследований преимущественно на макроуровне к исследованиям на микроуровне, от разработки и оценки крупномасштабных реформ к разработке и оценке точечных локальных интервенций («миниатюрных проектов», по выражению Дж. Хекмана [Heckman, 2020]); 2) переход от наблюдательных методов анализа, основывающихся на наблюдениях, к экспериментальным методам, основывающимся на полевых испытаниях, от «предположительных» (некаузальных) к «доказательным» (каузальным) выводам; 3) пересмотр общей стратегии развития, когда главным объектом интереса исследователей вместо проблемы экономического роста становится проблема бедности, что предполагает переадресацию основной части помощи — от *всего* населения развивающихся стран к их *беднейшим* слоям [Pritchett, 2020].

Есть еще одно важное отличие, прямо связанное с предыдущими. Дело в том, что «старая» и «новая» экономики развития строятся на разных — по природе и качеству — эмпирических данных. «Старая» экономика развития чаще всего оперировала агрегированными показателями, бравшимися в готовом виде из различных международных баз данных, причем из-за плохой организации статистики в развивающихся странах эти данные часто оказывались крайне сомнительного качества [Jerven, 2015]. Дефицит качественных данных был одной из причин, почему экономика развития не пользовалась среди экономистов большим авторитетом и по сравнению со многими другими разделами оставалась парией. В этих условиях молодые исследователи нечасто решались связывать с ней свою академическую карьеру. «Новая» экономика развития строится на использовании новых, не существовавших ранее индивидуальных данных, которые она получает сама из первых рук по ходу проводимых ею экспери-

ментов. Резкое улучшение качества эмпирических данных, с которыми имеют дело исследователи, способствовало перемещению этого раздела экономической науки с аутсайдерских позиций на лидерские, сделав его чрезвычайно привлекательным для молодых поколений экономистов. Сегодня экономика развития предстает как история успеха, как олицетворение несомненного и впечатляющего прогресса научного знания.

Экспериментальный характер. С методологической точки зрения экономику RCT можно рассматривать как наиболее полное и последовательное выражение тренда по превращению экономической науки в экспериментальную дисциплину — тренда, активно заявившего о себе в последние десятилетия [Капелюшников, 2021]. Современные экономисты полностью отказались от давнего представления о том, что эксперименты в экономике невозможны⁴. Многие из них готовы идти даже дальше, наделяя статусом научных фактов только и исключительно экспериментальные данные. В сегодняшней экономической науке развернулась настоящая «охота за причинами»: достоверное установление *причины* того или иного явления практически гарантирует исследованию широкое признание и высокий научный статус⁵.

Экономика RCT представляет собой едва ли не самое яркое проявление этого эксперименталистского крена, поскольку задает максимально высокие стандарты строгости и достоверности, аналогичные тем, что приняты в естественнонаучных дисциплинах. В отличие от стандартных (обсервационных) методов эмпирического анализа, для которых установление точных причинно-следственных связей

⁴ Активное проникновение в экономическую науку эксперименталистского подхода сопровождалось изменением ее рабочего языка. Из лексикона экономистов практически исчезло понятие «закона» и его место заняло понятие «причины». В самом деле, если раньше экономическая наука претендовала на открытие неких общих *законов*, то теперь ее сверхзадача свелась к выявлению конкретных *причин* тех или иных частных явлений, наблюдаемых здесь и сейчас. Значение этой переориентации трудно переоценить: установка на обнаружение универсальных закономерностей сменилась установкой на отыскание специфических причин, действие которых ограничено определенным локальным контекстом.

⁵ «Неотъемлемой частью профессиональной одержимости в экономической науке стало стремление к получению “каузальных эффектов”, даже если идентифицируемые эффекты не имеют социального значения и/или экономического смысла» [Neckman, 2020].

всегда оставалось нерешаемой задачей, рандомизированные испытания, как считается, способны достоверно выявлять и оценивать *каузальные* эффекты. Многие рандомисты воспринимают RCT не просто как самый научный, но как *единственно* научный метод из всех, имеющихся в распоряжении экономистов: «Когда мы говорим о строгих доказательствах, то имеем в виду... свидетельства рандомизированных экспериментов» [Banerjee, 2006, p. 8]. Например, в меню методов, одобряемых Лабораторией J-PAL, RCT не просто занимают верхнюю строчку: на самом деле, ничего кроме них в этом меню нет. Показательно в этом смысле, что когда Всемирный банк начал активно внедрять для оценки своих проектов RCT, в ведущем медицинском журнале *Lancet* появилась редакционная статья, извещавшая о том, что «Всемирный банк наконец-то обратился к науке» [Lancet, 2004, p. 731].

Рост популярности метода рандомизации был во многом спровоцирован активной методологической критикой, которой в последние десятилетия подверглись традиционные методы эконометрического анализа и которая резко подорвала доверие к их результатам. Согласно доминирующей точке зрения, RCT обладают по отношению к этой критике абсолютным иммунитетом, так как свободны от бесчисленных эконометрических проблем, с которыми сталкиваются иные подходы, требуют минимума предположений, не нуждаются в сборе обширной предварительной информации, не связаны ни с какой теорией и не зависят от идеологических предпочтений экспертов [Deaton, Cartwright, 2018].

Отсюда одна из их ключевых идей экономики RCT — идея *иерархии методов* [Imbens, 2010]. Речь идет о ранжировании методов эмпирического анализа в зависимости от того, насколько они способны без смещений идентифицировать каузальные эффекты. Исходя из этого критерия на вершине иерархии, естественно, оказываются RCT, тогда как все остальные методы, не прибегающие к рандомизации, рассматриваются как имеющие лишь небольшую научную ценность или не имеющие ее вообще. Согласно идее иерархии методов, доказательства без рандомизации недостойны считаться доказательствами или, по меньшей мере, «строгими» доказательствами [Deaton, 2020]. Девиз рандомистов — «один хороший эксперимент побивает тысячу плохих регрессий» [Pritchett, Sandefur, 2013]. Или, в иной формулировке: «Обсервационные исследования предполага-

ют, рандомизированные эксперименты располагают» [Deaton, Cartwright, 2018].

Как следствие, в глазах рандомистов RCT предстают как «золотой стандарт», которому все другие методы заведомо и безнадежно проигрывают, являясь недостаточно строгими и недостаточно научными [Favereau, Nagatsu, 2020]. Соответственно, в тех случаях, когда результаты применения иных методов расходятся с результатами применения RCT, предпочтение безусловно должно отдаваться последним. Неявно это подразумевает, что в идеале экономистам вообще следовало бы пользоваться только ими, полностью отказавшись от любых неэкспериментальных подходов, которые несостоятельны по определению (естественные эксперименты — единственный альтернативный метод, для которого рандомисты готовы делать исключение [Banerjee, 2006]). RCT — это настоящая наука, все остальное — не наука или, в лучшем случае, полунаука.

Установка на атеоретичность. В эпистемологической перспективе экономика RCT предстает как предельно атеоретический, ультраэмпирический подход [Favereau, 2016]. Свобода от теории рассматривается в ней не как недостаток, а как достоинство, поскольку позволяет предлагать эффективные средства по борьбе с глобальной бедностью, не погрязая в бесконечных концептуальных спорах [Deaton, Cartwright, 2018]. Теория, по их мнению, плохой советчик при проведении экспериментов: она только запутывает дело, мешая успешной борьбе с бедностью [Банерджи, Дюфло, 2021]. Идеал рандомистов — пусть данные говорят сами за себя [Deaton, 2020]⁶.

Превосходство RCT над другими методами чаще всего связывают именно с этим — с тем, что они позволяют идентифицировать каузальные эффекты, не прибегая ни к каким априорным предположениям и используя лишь закон больших чисел (подробнее об этом см. ниже) [Ravallion, 2020]. Утверждается, что для того, чтобы сделать из рандомизированного эксперимента достоверный каузальный вывод, не требуется никаких предварительных теоретических знаний. Это позволяет обходить одно из главных возражений против исполь-

⁶ «Необычно в сегодняшнем состоянии экономики развития не то, что в ней слишком мало теории, а то, что теория потеряла в ней позицию авангарда. Новые вопросы задаются эмпирическими исследователями и по большей части они приходят не из имеющегося корпуса хорошо разработанной теории: неясно даже, как начинать строить такую теорию» [Banerjee, 2005].

зования эконометрических методов анализа, состоящее в том, что они вынуждены опираться на неправдоподобные экономические модели, далекие от реальности [Deaton, 2010]. Сторонники метода рандомизации отказываются планировать и проводить эксперименты, отталкиваясь от каких-либо эксплицитно выраженных теоретических идей. Они не устают подчеркивать, что RCT имеют «беспредпосылочный» характер и не нуждаются в предположениях ни о структуре причинно-следственных связей, ни о включении или невключении в анализ различных переменных, ни о функциональной форме эмпирических зависимостей, ни о выборе той или иной теоретической модели [Deaton, Cartwright, 2018].

В менее жесткой формулировке можно сказать, что по вопросу о необходимости теории экономика RCT занимает агностическую позицию — во всяком случае, до получения необходимых экспериментальных данных [Favereau, Nagatsu, 2020]. По мнению рандомистов, если теория вообще нужна, то она должна следовать за результатами экспериментов, а не предшествовать им. Опыт RCT свидетельствует, что в экспериментальной среде многие интервенции демонстрируют значительные результаты, не ожидаемые и не объяснимые теоретически, или же, наоборот, демонстрируют отсутствие результатов, которые предполагаются общепринятой теорией [Ibid.]. Время для построения теорий, полагают сторонники RCT, наступит тогда, когда будет накоплен достаточный массив экспериментальных данных, что и даст возможность приступить к их обобщению [Banerjee, 2005]. Пока же усилия исследователей должны быть сосредоточены на получении таких данных: задача состоит в том, чтобы копить эмпирические факты, а уже потом пытаться теоретизировать на их основе [Favereau, 2016].

В тех достаточно редких случаях, когда рандомисты все же вспоминают о теории, они понимают под ней нечто иное по сравнению с тем, как она определяется обычно. Речь идет не о генерализациях высокого уровня, а скорее о психологических эффектах в духе поведенческой экономики. Наглядный пример дают знаменитые эксперименты Паскалин Дюпас с противомоскитными сетками, защищающими от заболевания малярией [Cohen, Dupas, 2010; Dupas, 2014]. Исследователи попытались оценить, как краткосрочные субсидии домохозяйствам могут отражаться на их долгосрочном спросе на такие сетки. Дело в том, что противомоскитные сетки нужно менять

каждые 6–12 месяцев. Поэтому важно знать, как поведут себя индивиды после того, как истечет срок действия программы помощи и им придется приобретать сетки уже за свой счет. Что эффективнее с этой точки зрения — раздача противомоскитных сеток полностью бесплатно или же их продажа по ничтожно низкой (иными словами — субсидируемой), но все же ненулевой цене? Если доминирует *эффект привыкания*, то приучившись получать сетки даром, индивиды станут в будущем отказываться покупать их за свой счет: бесплатная раздача будет действовать как отрицательный стимул. Если же доминирует *эффект обучения*, то убедившись на опыте в полезности противомоскитных сеток, индивиды станут охотнее приобретать их в будущем даже за полную цену: бесплатная раздача будет действовать как положительный стимул. Эксперименты П. Дюпас показали, что бесплатная раздача действительно активизирует долговременный спрос на противомоскитные сетки, откуда был сделан вывод о преобладании эффекта обучения над эффектом привыкания. В данном случае не так важно, что сами предполагаемые психологические эффекты привыкания и обучения прямо не тестировались и суждения о них выносились исходя из косвенных свидетельств. Важнее, что такого рода предположения едва ли могут называться «теорией» в традиционном смысле или, во всяком случае, «высокой теорией»: если это и теория, то, если можно так выразиться, «теория без теорем».

В итоге общая исследовательская установка рандомистов может быть описана как стремление «держаться от теории подальше» [Favereau, Nagatsu, 2020]⁷.

Практическая программа. Если говорить о практической программе новой экономики развития, то свою главную миссию она усматривает в том, чтобы находить эффективные средства для смягчения проблемы бедности в развивающихся странах, опираясь на эмпирические свидетельства рандомизированных экспериментов. «Старая» экономика развития действовала фактически вслепую, полагаясь на предложения, эффективность которых не была известна *ex ante* и, более того, чаще всего не оценивалась даже *ex post*. «Новая» экономика развития считает проверку любых потенциальных интервенций

⁷ Естественно, политикам и широкой публике идея о том, что практические рекомендации можно вырабатывать полностью «атеоретически» (не опираясь ни на какие теории или модели), должна казаться чрезвычайно привлекательной.

строго обязательной, причем производить ее следует не *ex post*, а *ex ante* — на материале пилотных проектов еще до того, как пытаться внедрять эти интервенции в полном масштабе. При этом, естественно, доверять следует только оценкам, следующим из рандомизированных экспериментов. Так, А. Банерджи обвинил в свое время Всемирный банк в «интеллектуальной лени» и «сопротивлении новым знаниям», указав на то, что деятельность Банка демонстрирует «поразительное отсутствие различения между стратегиями, основывающимися на твердых доказательствах, полученных с помощью RCT или естественных экспериментов, и всем тем, что таких доказательств лишено» [Banerjee, 2006].

Иными словами, новая экономика развития стремится к тому, чтобы получать надежные эмпирические свидетельства эффективности программ помощи развивающимся странам, а затем предлагать эти свидетельства в качестве руководства к действию правительствам, международным институтам и частным благотворительным фондам [Favereau, Nagatsu, 2020]. Рандомисты убеждены в том, что полевые эксперименты — это уникальный инструмент для оценки эффективности предлагаемых интервенций и получения научно обоснованных знаний в этой области [Picciotto, 2020]. Только с помощью RCT можно установить, какие программы не приведут к сокращению глобальной бедности, так что от них следует отказаться, а какие помогут ее сократить, так что их следует, наоборот, реплицировать и распространять.

Отсюда еще одна ключевая идея рандомистов — идея о том, что именно RCT способны отделять «то, что работает» от «того, что не работает» [Pritchett, Sandefur, 2013]. Нет сомнений, что и политикам, и широкой публике подобная установка должна импонировать: фактически им внушают мысль, что арсенал новой экономики развития состоит исключительно из вмешательств с подтвержденной эффективностью, раз предварительно все они прошли успешную проверку в ходе рандомизированных испытаний. Сегодня, когда эксперты и их суждения вызывают у многих глубокое недоверие, RCT создают по контрасту впечатление полностью объективного знания, не зависящего от субъективных мнений экспертов [Deaton, Cartwright, 2018].

Но экономика RCT — это не просто определенная техника выработки, анализа и оценки рекомендаций: за ней стоит иная философия

развития, принципиально отличная от той, что доминировала на протяжении многих предшествующих десятилетий. Когда речь идет о реформах в масштабе всей экономики или целых секторов, проведение рандомизированных экспериментов становится технически, организационно и финансово невозможным. Соответственно, сфера применения РСТ ограничивается исключительно точечными локальными проектами, нацеленными на решение конкретных проблем конкретных групп населения (скажем, предоставление жителям тропических стран противомоскитных сеток, дегельминтизация детей, борьба с абсентеизмом школьных учителей и т.д.). Поэтому в новой экономике развития философия «больших дел», охватывающих все общество, оказывается вытеснена философией «малых дел», адресуемых исключительно беднейшим слоям населения⁸.

Стратегию постепенного накопления небольших изменений рандомисты характеризуют как «тихую революцию»: «Значительный прогресс в решении самой большой мировой проблемы [бедности. — *Р. К.*] может быть достигнут постепенно, небольшими шагами, каждый из которых хорошо продуман, тщательно протестирован и разумно реализован... Подобные постепенные изменения будут иметь постоянный и нарастающий характер. Они могут стать началом тихой революции» [Банерджи, Дюффо, 2021, с. 40, 395].

В итоге для стороннего наблюдателя экономика РСТ предстает как чрезвычайно привлекательное сочетание научной строгости, идеологической нейтральности и делового прагматизма.

Дизайн и логика РСТ

Дизайн и логика РСТ во многом отличны от дизайна и логики «стандартных» экспериментов.

Условно схему «стандартного» эксперимента можно описать так. Обозначим состояние объекта до оказания на него некоего воздействия T как A_0 , а состояние объекта после того, как оно было на него оказано, как A_1 . Сравнивая состояния объекта в моменты времени t_0

⁸ «Миниатюрные исследования получили признание как идеал строгой эмпирической экономики. Стремление задавать большие и важные вопросы и пытаться отвечать на них было дискредитировано в погоне за четкими ответами на мелкие вопросы, не имеющими большого значения с точки зрения политики» [Нескман, 2020].

и t_1 , мы получаем ответ на вопрос, произвело ли воздействие T какое-то изменение в объекте A или нет. Разница между его состояниями в моменты времени t_0 и t_1 будет мерой каузального эффекта, вызванного T . Конечно, при проведении эксперимента мы должны будем изолировать объект от любых иных потенциальных воздействий помимо воздействия T : только в таком случае будет корректно сказать, что оказанное нами воздействие есть причина, а измененное состояние объекта — следствие. Без изоляции от внешней среды выявить каузальный эффект T будет невозможно. Успех любого эксперимента критически зависит от того, в какой мере при его проведении нам удалось обеспечить подобную изоляцию.

RCT строятся иначе. Вместо того, чтобы сравнивать состояние объекта A в момент времени t_1 с его состоянием в момент времени t_0 , мы сравниваем состояние объекта A в момент времени t_1 с состоянием аналогичного объекта B («объекта-близнеца», не имеющего систематических отличий от A) также в момент времени t_1 . При этом по условиям эксперимента в момент времени t_0 мы оказываем воздействие только на A , а с его «объектом-близнецом» B никаких манипуляций не производим, предоставляя его «самому себе». Сравнивая состояния A и B в момент времени t_1 , мы получаем ответ на вопрос, было ли воздействие T причиной измененного состояния A . Каузальный эффект (E), измеряющий силу воздействия T на A , будет определяться как разность между значениями A и B в момент времени t_1 :

$$E = A_1 - B_1.$$

Огромное преимущество такого «нестандартного» дизайна заключается в том, что он не требует изоляции объекта от влияния сторонних факторов, которые могут заявлять о себе параллельно с воздействием T (во многих случаях это бывает физически невозможно). Хотя они могут действовать одновременно с T , это не будет порождать смещений, поскольку вызванные ими изменения в «объектах-близнецах» A и B будут идентичными. Соответственно, вычитая из значения A в момент времени t_1 значение B в момент времени t_1 , мы сможем полностью элиминировать возможные влияния любых других привходящих факторов. Поскольку изначально между сравниваемыми «объектами-близнецами» не было никаких систематических различий, остается единственная причина, которая была спо-

собна породить возникшую между A и B разницу, — воздействие T , оказанное на объект A , но не оказанное на объект B .

Ключевая проблема, возникающая в подобном контексте, заключается в том, каким образом мы можем сформировать «объекты-близнецы», между которыми в момент времени t_0 не будет существовать никаких систематических различий (если какие-то различия и будут наблюдаться, они будут чисто случайными). В полевых экспериментах формирование «объектов-близнецов» как раз и обеспечивается методом рандомизации. Во-первых, из рассматриваемой популяции случайным образом формируется выборка. Во-вторых, эта выборка также случайным образом разбивается на две группы — экспериментальную, на которую затем оказывается воздействие, и контрольную, на которую воздействия не оказывается⁹. При условии, что выборка достаточно велика, закон больших чисел гарантирует, что до оказания воздействия сравниваемые группы окажутся идентичными по средним значениям их наблюдаемых и ненаблюдаемых характеристик, если не считать стохастических ошибок. По сути, рандомизация действует как процедура устранения систематических различий между экспериментальной и контрольной группами.

Фактически рандомизация создает группы, характеристики которых являются статистически выровненными. Соответственно любое различие, возникшее после оказания воздействия, может быть отнесено только на его счет. Таким простым способом RCT позволяют избавляться от статистических искажений, неизбежно возникающих при использовании других эконометрических методов. Вычитая из результата для экспериментальной группы результат для контрольной группы, мы получаем несмещенную (то есть очищенную от влияния всех прочих факторов) оценку каузального эффекта, порожденного нашим вмешательством¹⁰. Ведь между сравниваемыми

⁹ Когда мы говорим о случайном характере формирования выборки и случайном характере ее разбивки на экспериментальную и контрольную группы, то имеем в виду использование процедур, основанных на случайности, — наподобие бросания игральных костей, вращения колеса рулетки, вытаскивания из урны черных и белых шаров или генератора случайных чисел.

¹⁰ Еще одно очевидное преимущество RCT заключается в том, что они легко обходят проблему обратной каузальности, выступающую камнем преткновения для многих других эконометрических методов, когда неясно, то ли A есть причина B , то ли B есть причина A . Экспериментальный дизайн позволяет однозначно идентифицировать направление причинно-следственной связи, не полагаясь на громоздкие

группами нет никаких различий, кроме одного, — подвергались они воздействию или нет. Это открывает возможность оценивать как величину среднего эффекта воздействия (average treatment effect — ATE), так и его статистическую значимость, не прибегая ни к каким сложным техническим процедурам. ATE рассчитывается просто как разность в средних значениях двух распределений (для экспериментальной и контрольной групп). Таким образом, чтобы сделать из RCT каузальный вывод, достаточно простого сравнения двух средних значений.

Мы можем представить внутреннюю логику RCT в более формализованном виде [Deaton, 2009]. Пусть случайным образом сформирована выборка, которая затем также случайным образом разбита на экспериментальную (*e*) и контрольную (*c*) группы. Любой член выборки *i* может принимать два возможных значения в зависимости от того, подвергся он воздействию или нет, — либо Y_{1i} , либо Y_{0i} . При этом каждое *i* мы наблюдаем только в одном из двух возможных состояний — 1 или 0, но никогда одновременно в обоих. Для членов экспериментальной группы мы видим только Y_{1i} и не видим Y_{0i} , а для членов контрольной группы видим только Y_{0i} и не видим Y_{1i} . Если нас интересует средний эффект воздействия, то мы должны были бы сравнить средние значения, которые наблюдались бы у членов экспериментальной группы при наличии и при отсутствии воздействия:

$$ATE = \tilde{Y}_1^e - \tilde{Y}_0^e. \quad (1)$$

Но мы этого сделать не можем, так как не видим членов экспериментальной группы в состоянии 0, когда бы на них не оказывалось воздействия: Y_{0i}^e ненаблюдаемы. Все, что мы можем наблюдать в данных, это разность в средних значениях между экспериментальной группой при наличии воздействия и контрольной группой при отсутствии воздействия. Как ни странно, этого оказывается достаточно. Вычтем из обоих членов уравнения (1) среднее значение для контрольной группы при отсутствии воздействия \tilde{Y}_0^c :

$$ATE = \tilde{Y}_1^e - \tilde{Y}_0^e = [\tilde{Y}_1^e - \tilde{Y}_0^c] - [\tilde{Y}_0^e - \tilde{Y}_0^c]. \quad (2)$$

и не вполне надежные процедуры. В самом деле, было бы абсурдом предполагать, что изменение, произошедшее с экспериментальной группой, могло бы каким-то образом породить наше вмешательство *T*.

Выражение во вторых квадратных скобках равно нулю в силу рандомизации, гарантирующей, что при отсутствии воздействия средние значения всех характеристик экспериментальной и контрольной групп должны быть идентичными. Таким образом мы получаем возможность оценить средний эффект воздействия несмотря на то, что наблюдаем членов экспериментальной группы только в одном состоянии:

$$ATE = \tilde{Y}_1^e - \tilde{Y}_0^e = \tilde{Y}_1^e - \tilde{Y}_0^e.$$

Здесь, однако, следует уточнить, что утверждение об идентичности экспериментальной и контрольной групп не предполагает полного совпадения между ними на уровне индивидуальных единиц, входящих в их состав: речь идет лишь о равенстве на уровне их средних характеристик (математического ожидания). В этом смысле метафора про «объекты-близнецы» является, конечно, не вполне точной. RCT позволяют оценивать *средний эффект* воздействия, но оставляют за скобками вопрос о его *индивидуальных эффектах*. Одно и то же значение АТЕ может быть получено при самых разных вариантах распределения индивидуальных эффектов: и если интервенция оказала положительное воздействие на всех участников экспериментальной группы, и если она оказала положительное воздействие (только более значительное) лишь на одну их половину, но никак не отразилась на другой, и если она оказала положительное воздействие (еще более сильное) на одну половину и отрицательное (слабое) на другую, и т.д.

По этой причине в интерпретации результатов RCT имеет место интересная асимметрия [Deaton, 2010]. Если оценка АТЕ значимо отличается от нуля, то это свидетельствует о том, что нам удалось идентифицировать каузальный эффект интервенции T . Но обратное неверно. Если АТЕ равен нулю, то отсюда не следует, что воздействие T никак не отразилось на поведении индивидов из экспериментальной группы: возможно, на одну их половину оно повлияло со знаком «плюс», а на другую со знаком «минус» и соответствующие эффекты нейтрализовали друг друга.

Нетрудно догадаться, почему оценки среднего эффекта воздействия обладают огромной привлекательностью для всех, кто вовлечен в разработку и реализацию программ для развивающихся стран, — политиков, государственных чиновников, международных агентств,

частных благотворительных фондов. Во-первых, такие оценки оставляют впечатление точного экспериментального знания причинно-следственных связей, позволяющего надежно отделять работающие программы от неработающих, причем это знание оказывается доступно *ex ante*, а не *ex post* — в рамках пилотных проектов еще до их внедрения в полном масштабе (см. выше). Во-вторых, смысл показателя среднего эффекта воздействия чрезвычайно прост и легкодоступен пониманию практически каждого — политиков, СМИ, широкой публики. В конце концов, что такое среднее арифметическое, известно всем, кто когда-то учился в школе. Нельзя не согласиться с высказыванием Э. Дюфло о том, что по сравнению с любыми другими методами RCT являются «более прозрачными и простыми для объяснения» [Duflo, 2017, p. 17]. Эта простота на фоне переусложненности альтернативных подходов — одно из несомненных преимуществ RCT. Конечно же, неэкономистам намного легче понять суть метода рандомизации, чем смысл каких-то других эконометрических методов, становившихся с течением времени все более громоздкими, технически изощренными и интуитивно неочевидными.

Итак, рандомизация создает ситуацию, при которой все прочие факторы, способные повлиять на исход, в равной мере воздействуют на обе сравниваемые группы, так что любое появившееся между ними различие можно с уверенностью приписать нашему вмешательству. Привлекательность рандомизированных испытаний заключается именно в этом — в том, что исследователям не нужно понимать и контролировать все факторы, способные порождать смещения: рандомизация автоматически обеспечивает их полную нейтрализацию. Благодаря RCT можно делать выводы о причинно-следственных связях, ничего не зная об этих искажающих факторах.

Проблема внутренней валидности

При оценивании качества RCT различают их внутреннюю и внешнюю валидность: внутренняя оценивает корректность выводов *в* эксперименте (были ли при их получении устранены все возможные смещения), тогда как внешняя — корректность выводов *из* эксперимента (будут ли они воспроизводиться в иных условиях). В первом случае речь идет о статистической чистоте *получения* результатов, во

втором — о допустимой широте их последующего *использования*. Внутренняя валидность отвечает на вопрос о выявляемости каузальных эффектов, внешняя — на вопрос об их генерализуемости.

Общепризнанно, что с точки зрения внутренней валидности рандомизированные контролируемые испытания превосходят любые другие методы. Высокая внутренняя валидность RCT связана с их способностью давать несмещенные оценки причинно-следственных связей. Отсюда их квалификация как «экспериментального идеала» [Angrist, Pischke, 2010]. Дизайн RCT обеспечивает идентичность экспериментальной и контрольной групп по всем наблюдаемым и ненаблюдаемым характеристикам, кроме одной, в измерении которой мы заинтересованы, и позволяет на этой основе выявлять и оценивать связь между произведенным воздействием и его последствиями: «Невозможно спорить с тем, что тщательно спланированный эксперимент дает нам несмещенную оценку воздействия конкретной интервенции, которая нас интересует. Используя общепринятый жаргон, можно сказать, что экспериментальные результаты являются внутренне валидными» [Banerjee, 2005]. Именно это дает право называть RCT «золотым стандартом» эмпирического анализа (см. выше).

Однако с вердиктом об абсолютном превосходстве RCT над любыми другими методами с точки зрения их внутренней валидности согласны далеко не все. Критики считают его явно преувеличенным. По их мнению, это не более чем риторический прием, используемый рандомистами для продвижения своих идей, поскольку сама по себе рандомизация не дает гарантии безошибочной верификации каузальных эффектов [Deaton, 2010].

1. Рандомисты без колебаний приписывают любую разницу в средних результатах между экспериментальной и контрольной группами производимым интервенциям. По категорическому утверждению А. Банерджи и Э. Дюфло, «поскольку до начала эксперимента индивиды, назначенные для различных видов «лечения», были совершенно сопоставимы (потому что были отобраны случайным образом), любое возникающее между ними различие является результатом внешнего воздействия» [Банерджи, Дюфло, 2021, с. 39]. Однако на практике достигать посредством рандомизации абсолютной сбалансированности между экспериментальной и контрольной группами никогда не удастся. Поэтому претензия RCT на то, что они гарантированно обеспечивают полное отсутствие каких-либо систематических разли-

чий между экспериментальной и контрольной группами, лишена оснований. Красноречивой иллюстрацией может служить одно исследование, проведенное в Дании [Vass, 2010]. Была сформирована случайная выборка из примерно 1000 пожилых индивидов, которые были затем случайным образом разбиты на две группы. При этом *никакому реальному воздействию ни одна из них не подвергалась*. Тем не менее через полтора года между ними обнаружилась значительная по величине разница в показателях смертности, статистически значимая на очень высоком уровне существенности ($p = 0,003$).

2. Рандомисты редко вспоминают, что RCT гарантируют точную верификацию каузальных эффектов и, соответственно, могут считаться внутренне валидными только в идеальных условиях — при очень значительном размере выборки (в пределе — стремящемся к бесконечности) и при ее многократных повторных разбивках на экспериментальную и контрольную группы (в пределе — тоже стремящихся к бесконечности) [Deaton, Cartwright, 2018]. Только в таком идеальном случае оценки каузального эффекта, получаемые с помощью RCT, можно считать свободными от смещений¹¹. Идеальная схема RCT основывается на законе больших чисел, из которого следует, что при рандомизации индивидуальные характеристики будут равномерно распределяться между двумя группами, но такое распределение будет достигаться, во-первых, только в среднем, во-вторых, только при очень большом числе наблюдений и, в-третьих, только при многократно повторяющейся разбивке выборки на разные по составу экспериментальные и контрольные группы [Ibid.]. Поэтому исследователям, описывающим достоинства метода рандомизации, постоянно приходится делать оговорки о том, что он обеспечивает идентичность сравниваемых групп только «теоретически», только «в среднем», только «чисто статистически», только «для бесконечно больших выборок» [de Cler, 2021].

Естественно, реальная схема RCT никогда не совпадает с идеальной. На практике рандомистам чаще всего приходится работать с не очень большим или даже очень небольшим числом наблюдений. Ограниченность бюджета при высоких финансовых, временных и организационных издержках вынуждает их довольствоваться скромными

¹¹ Причина, по которой RCT возводят на пьедестал «золотого стандарта», как раз и заключается в предполагаемой несмещенности результатов в *идеальных условиях* [Ravallion, 2020].

по размеру выборками [Ravallion, 2020]. Если наблюдательные исследования чаще всего строятся на уже имеющихся базах данных (выборочных обследований или административных регистров), то RCT, как уже отмечалось, всегда требуют новых уникальных данных, что делает их чрезвычайно дорогостоящими. Ситуация усугубляется тем, что эффективный размер выборок, с которыми приходится работать исследователям, нередко оказывается в десять или даже сто раз меньше их номинального размера (из-за истощения выборок, неполноты данных и т.д.) [Ioannidis, 2018]¹².

По тем же причинам почти никогда не практикуется и многократная рандомизация на основе одной и той же выборки с формированием в каждом раунде своих особых, отличающихся по составу экспериментальной и контрольной групп. Это серьезное отклонение от идеальной схемы: в RCT все различия между группами элиминируются при условии усреднения результатов по бесконечному числу (чисто гипотетических) повторений, так что среднее значение АТЕ для всего множества единичных рандомизаций может рассматриваться как «истинное». Хотя любое отдельное испытание не в состоянии обеспечить полного баланса факторов, способных порождать систематические различия между экспериментальной и контрольной группами, отклонения от него будут варьироваться то в пользу первой, то в пользу второй. Поэтому в длинной серии воображаемых экспериментов эти отклонения будут компенсироваться [Deaton, Cartwright, 2018].

Однако при проведении «разовой» рандомизации — то есть на практике — такой компенсации не происходит. Баланс между экспериментальной и контрольной группами всегда остается неполным, так что разность в средних результатах между ними оказывается равна сумме: средний эффект воздействия + ошибка, порождаемая несбалансированностью прочих факторов. Совокупное влияние этих факторов может быть сильнее для экспериментальной, а может для контрольной группы, но в любом случае об идентификации «истинной» причинно-следственной связи говорить уже не приходится: эффект, приписываемый воздействию, в действительности может быть

¹² Но даже резкое увеличение размера выборки ничего не гарантирует: если имеется очень большое число факторов, способных влиять на исход, то обеспечение баланса между экспериментальной и контрольной группой по всем этим факторам может оставаться все равно недостижимым [Deaton, Cartwright, 2018].

следствием неполной сбалансированности сравниваемых групп по каким-то прочим факторам [Deaton, Cartwright, 2018].

Никакая «розовая» рандомизация не способна автоматически выравнивать все характеристики экспериментальной и контрольной групп [Deaton, 2009]. Она не может гарантировать, что все параметры будут сбалансированы хотя бы в среднем, даже если эти группы отбирались случайным образом. В исследовательской практике для борьбы с такими дефектами в ход идут перевзвешивания, кластеризация, оценивание стандартных регрессионных моделей, которые наряду с какими-то контрольными переменными дополнительно включают фиктивную переменную, отражающую принадлежность индивидов к экспериментальной или контрольной группам, и т.д. [de Cler, 2021]. Очевидно, однако, что такие компромиссные решения лишают RCT декларируемых преимуществ над другими методами, так что идея «золотого стандарта» повисает в воздухе.

3. Как и многие другие методы, RCT не застрахованы от проблемы аутлайерства, то есть появления выбросов в экспериментальной или контрольной группах [Deaton, 2020]. Допустим, проводится экспериментальная проверка эффективности каких-то программ — по микрокредитованию или улучшению качества школьного образования. Достаточно, чтобы в экспериментальную группу попало несколько аутлайеров — сверхталантливых предпринимателей или сверходаренных детей, чтобы прийти к ложному выводу о том, что интервенции успешны и приводят к желаемым результатам. И, наоборот, достаточно, чтобы несколько таких аутлайеров попало в контрольную группу, чтобы был сделан ошибочный вывод о неэффективности этих интервенций.

4. Достаточно искусственно выглядит излюбленная рандомистами аналогия с клиническими испытаниями в медицине. Как отмечает А. Дитон, их постоянные отсылки к опыту доказательной медицины носят по большей части риторический характер [Deaton, 2010].

В отличие от рандомизированных экспериментов в медицине рандомизированные эксперименты в экономике не являются двойными слепыми [Ravallion, 2020]. При тестировании новых лекарств испытуемые не знают, получили ли они действующее вещество или плацебо, и точно так же эксперты не знают, кто из испытуемых попал в группу воздействия, а кто в группу сравнения. В экспериментах, проводимых экономистами, эти условия не соблюдаются: любой из участ-

ников знает, подвергся он воздействию или нет (получил или не получил он противомоскитную сетку, был или не был дан ему микрокредит, прошел или не прошел он курс санитарного просвещения и т.д.), а исследователям точно так же известно, кто из участников к какой группе принадлежит. Отсюда высокий риск возникновения эффекта самоотбора: пользуясь доступной им информацией, участники эксперимента могут переходить из одной группы в другую, если надеются, что их положение от этого улучшится. Сходным образом исследователи, заинтересованные в успехе эксперимента, могут по-разному взаимодействовать с участниками из разных групп, сознательно или бессознательно подталкивая их к желательным для себя результатам [Deaton, 2020]. Как следствие, разбивка выборки на две группы — экспериментальную и контрольную — утрачивает случайный характер.

Допустим, проводится эксперимент с тем, чтобы выяснить, как на успеваемость школьников влияет размер класса. В одной половине отобранных школ мы делим все классы пополам, уменьшая их численность вдвое, а в другой оставляем все, как было. Естественно ожидать, что, реагируя на это вмешательство, родители из наиболее образованных и состоятельных семей начнут делать все возможное и невозможное (вплоть до смены места жительства), чтобы перевести своих детей в школы с уменьшенными классами. Из-за этого между экспериментальной и контрольной группами возникнут систематические различия и полученные оценки окажутся смещенными. Эффект, который мы будем склонны приписывать различиям в размерах классов, на самом деле будет объясняться различиями в социально-экономическом положении семей школьников.

Рандомизированные эксперименты с неодушевленными предметами — это далеко не то же самое, что рандомизированные эксперименты с людьми. Земельные участки не способны реагировать на предстоящее внесение удобрений и не станут пытаться от него уклониться [Deaton, 2010]. Достаточно странно, но о том факте, что в отличие от экспериментов в медицине эксперименты в экономике не являются слепыми, в литературе по RCT почти никогда не упоминается.

5. Рандомисты далеко не во всех случаях прибегают к процедуре истинной рандомизации. Нередко разбивка на экспериментальную и контрольную группу производится не по индивидам, а по целым

кластерам — деревням, школам, больницам и т.д. Однако разбивка на две случайные группы *кластеров* не гарантирует разбивки на две случайные группы *индивидов* [Ravallion, 2020]. Понятно также, что при использовании подобной процедуры у участников появляется возможность переходить из кластера в кластер — из экспериментальной группы в контрольную или наоборот, если это отвечает их интересам. (Скажем, менять школы для своих детей, как в приведенном выше примере.)

Более того: многие известные исследования с использованием РСТ ограничиваются псевдорандомизацией, когда, скажем, участники делятся на группы воздействия и сравнения в алфавитном порядке — по первым буквам их фамилий [Deaton, 2010]. Наблюдаются и случаи открытого пренебрежения принципом случайного отбора, когда участниками эксперимента становятся только индивиды, изъявившие желание присоединиться к нему, или только индивиды, находящиеся в зоне доступности для экспериментаторов [Deaton, Cartwright, 2018]. Очевидно, что все такие практики создают угрозы для точной верификации каузальных эффектов и поэтому достаточно далеки от провозглашаемых рандомистами высших стандартов строгости и научности.

6. Даже в идеальных условиях РСТ подходят для оценки только одного показателя — среднего эффекта воздействия — и ничего не говорят о других характеристиках распределения, которые могут представлять интерес для политиков или спонсоров программ помощи. Но любое вмешательство чаще всего порождает как выигрывающих, так и проигрывающих — в зависимости от характеристик индивидов и социальных групп, к которым они принадлежат [Ravallion, 2020]. Политикам важно знать, какая доля населения получит от данной интервенции выгоды, а какая понесет потери, какие социально-демографические группы окажутся в выигрыше, а какие в проигрыше и т.д. Говоря иначе, им необходимо иметь представление об общей кривой распределения исходов тестируемой интервенции. Совершенно не исключена ситуация, когда интервенция продемонстрирует положительный средний эффект воздействия, хотя практически все население будет от нее проигрывать и лишь несколько человек станут получать от нее очень крупный выигрыш. РСТ не позволяют выявлять такие ситуации с катастрофическими последствиями для большей части общества [Deaton, 2010]. Лекарство, которое помога-

ет выжить в 60% случаев, но в 40% случаев приводит к смерти пациентов, едва ли будет рекомендовано к применению. Но социальный проект с аналогичным распределением выигравших и проигравших вполне может быть признан успешным исходя из результатов RCT и начнет внедряться повсеместно.

Хотя в принципе RCT могут давать несмещенные оценки среднего эффекта воздействия, другие характеристики распределения — например, медианный эффект или эффекты по квантилям — не идентифицируются из них напрямую. Для оценки таких показателей приходится использовать отличные от RCT методы, требующие большего числа дополнительных специальных предположений [Deaton, 2020]. С этой точки зрения никаких преимуществ перед другими методами у рандомизированных экспериментов нет.

7. Как правило, RCT ограничиваются оценкой краткосрочных эффектов вмешательства, игнорируя долгосрочные. Объясняется это двумя факторами: во-первых, тем, что эксперименты, растягиваемые на длительное время, требуют резкого увеличения бюджета, и, во-вторых, тем, что в таких случаях начинается постепенное истощение выборки. Это приводит к тому, что рандомисты достаточно часто фиксируют внимание не на конечных, а на промежуточных результатах вмешательства. Как следствие, их оценки эффективности тестируемых проектов могут оказываться неполными или вообще ошибочными [Eble et al., 2014]. Кроме того, действие далеко не всех причинно-следственных механизмов идет по линейной траектории: положительные эффекты на одних этапах могут сменяться отрицательными на других [Woolcock, 2013]. Ограниченность временного горизонта большинства RCT не позволяет учитывать эту потенциальную нелинейность. Метод рандомизации годится для оценки только краткосрочных эффектов в коротких причинно-следственных цепочках.

В качестве примера того, что внутренняя валидность даже эталонных исследований с использованием RCT нередко оставляет желать лучшего, можно сослаться на знаменитую работу Э. Мигеля и М. Кремера, посвященную оценке эффективности программ по дегельминтизации школьников [Miguel, Kremer, 2004]. Опубликованная в престижнейшем журнале *Econometrica*, она была признана абсолютной классикой новой экономики развития (так, в числе других работ М. Кремера она упоминается в решении о присуждении ему Нобелевской премии по экономике).

Известно, что в развивающихся странах миллионы детей страдают от заражения гельминтами. Э. Мигель и М. Крамер предположили, что из-за этого многие из них вынуждены пропускать уроки в школе, что должно негативно отражаться на их успехах в учебе и будущих перспективах на рынке труда. Для проверки этого предположения в конце 1990-х годов они провели эксперимент в Кении, в котором участвовали учащиеся 75 местных школ. Детям из экспериментальной группы на протяжении двух лет давали в школах препарат от глистов, которого не получали дети из контрольной группы. Эксперимент показал, что дегельминтизация улучшает здоровье учащихся и снижает частоту пропусков ими школьных занятий (на 7,5%), хотя при этом никак не отражается на их академических результатах. Хотя авторы обнаружили значительный прямой эффект от программы, еще более сильным оказался ее внешний эффект, связанный с тем, что избавление от гельминтов детей из экспериментальной группы снижало риск заражения при общении с ними у детей из контрольной группы. Отсюда Э. Мигель и М. Кремер сделали вывод, что дегельминтизация — это самый эффективный способ повышения посещаемости школ в развивающихся странах. Их исследование получило широкую известность и стало одним из важных импульсов к развертыванию мировой кампании по дегельминтизации детей в бедных странах.

Однако когда группа авторитетных эпидемиологов провела повторный анализ микроданных Э. Мигеля и М. Кремера, попытавшись реплицировать их результаты, они выявили множество неточностей и нестыковок [Aiken et al., 2015; Davey et al., 2015]. После устранения обнаруженных ошибок оценка общего влияния дегельминтизации на частоту пропусков школ снизилась вдвое по сравнению с заявленной изначально. Прямой эффект (для вылеченных детей) остался значительным, но косвенный (для детей, не получавших лечения) исчез. Хотя некоторое влияние дегельминтизации на посещаемость по-прежнему прослеживалось, оно представляло уже как очень слабое и, скорее всего, смещенное. Более того, его не удавалось однозначно приписать программе дегельминтизации, потому что в рамках проекта параллельно с ней осуществлялась еще одна программа — по санитарному просвещению детей (как важно мыть руки и т.д.), которая и могла стать главным фактором, ответственным за произошедшие изменения. Общий вывод, к которому приш-

ли критики, был негативным: каких-либо убедительных свидетельств в пользу предположения о положительном влиянии дегельминтизации на посещаемость школ нет, а анализ Э. Мигеля и М. Кремера не соответствует научным стандартам, принятым для РСТ в сфере здравоохранения.

Впрочем, этот критический разбор несколько не поколебал звездного статуса работы Э. Мигеля и М. Кремера: в восприятии экономистов она так и осталась классикой РСТ.

Проблема внешней валидности

Внешняя валидность отвечает на вопрос о степени универсальности результатов РСТ, возможности их переноса на иной, более широкий контекст. Успешный рандомизированный эксперимент говорит лишь о том, что для данной группы населения данного региона данной страны программа, подвергнутая экспериментальной проверке, оказалась эффективной. Но он ничего не говорит о том, есть ли смысл распространять эту программу на другие группы населения других регионов других стран: приведет ли она там к тем же результатам, если попытаться ее внедрить? Результат, верный в одном месте в один момент времени при одних обстоятельствах, не обязательно будет верным в другом месте в другое время при других обстоятельствах.

Это проблема внешней валидности: окажется ли интервенция, хорошо зарекомендовавшая себя в экспериментальной среде, столь же эффективной в какой-то иной, неэкспериментальной среде? Где еще, помимо самого проведенного испытания, будут выполняться его результаты — повсеместно или в большинстве случаев, или только в небольшом числе случаев при соблюдении каких-то особых условий, или вообще больше нигде [Deaton, Cartwright, 2018]? Очевидно, что для политиков и спонсоров международной помощи результаты РСТ будут представлять практический интерес, только если они обладают высокой внешней валидностью и не зависят жестко от специфического контекста, в котором были получены¹³.

¹³ Когда мы говорим о «контексте», то имеем в виду множество самых разных вещей: идентичность агента, осуществлявшего вмешательство (академические исследователи, НПО, правительственные агентства); квалификацию нанятого пер-

Представим, что в ходе пилотного проекта в какой-то развивающейся стране мы экспериментально установили, что использование в школах флипчартов повышает успеваемость учащихся. Следует ли отсюда, что международные агентства должны покупать и устанавливать флипчарты во всех школах бедных стран? Понятно, что эффект от флипчартов зависит от множества факторов — уже имеющегося у учащихся уровня образования, качества преподавания, содержания учебных программ, методов обучения и т.д. Поэтому неудивительно, что контекст может иметь значение: в индийских и мексиканских школах «прочие условия» нельзя считать априорно «равными», так что интервенция, сработавшая в первых, может ничего не дать во вторых.

У внешней валидности можно выделить два измерения — горизонтальное и вертикальное. В первом случае речь идет о переносе программы, апробированной в ходе эксперимента в какой-то одной специфической местности или каком-то одном специфическом регионе, в какую-то другую местность или какой-то другой регион — вправе ли мы ожидать, что получим там сходные результаты? Разные контексты (местности, регионы, страны) всегда отличаются друг от друга по множеству параметров, так что результаты, полученные в одних условиях, могут не воспроизводиться в других. Иногда достаточно лишь немного изменить дизайн программы или перенести ее на какую-то иную группу населения, чтобы она перестала работать. Так, в индийском штате Раджастан было проведено успешное RCT, в рамках которого проверялось, какое влияние на абсентеизм учителей начальных школ могла бы оказать установка камер наблюдения, фиксирующих время их прихода и ухода [Duflo et al., 2012]. Были получены положительные результаты: абсентеизм учителей резко снизился. Но когда в клиниках того же индийского штата был проведен аналогичный эксперимент для медицинских сестер, были получены совершенно иные результаты [Banerjee et al., 2008]. Наблюдалось лишь кратковременное снижение прогулов, которое затем быстро сошло на нет: медсестры нашли пути обхода системы контроля, недоступные школьным учителям.

социала; особенности местной, региональной и национальной культуры, обычаев и институтов; технику опроса и даже погоду и время суток [Ogden, 2020].

Во втором случае речь идет о масштабировании проекта: вправе ли мы ожидать, что при переносе программы с местного или регионального уровня на национальный получим сходные результаты? Когда масштаб программы меняется, существует риск возникновения дополнительных («общеравновесных») эффектов, которые могут сводить полученные в ходе эксперимента результаты на нет [Bédécarrats et al., 2020; Deaton, Cartwright, 2018]. Масштабирование высвобождает причинные силы, отсутствовавшие в эксперименте. Как следствие, эффекты, наблюдаемые для небольших выборок, часто исчезают при переходе к более крупным агрегатам.

Допустим, мы провели в каком-то регионе рандомизированный эксперимент, проверяя, как на уровень образования детей из бедных домохозяйств повлияет предоставление обусловленных денежных трансфертов их родителям, выплата которых производится в зависимости от того, отправляют они своих детей в школу или нет. Допустим далее, что, как показал эксперимент, эта программа увеличила продолжительность пребывания детей в школе на два года, способствуя повышению их заработков после выхода на рынок труда и вытягивая их таким образом из бедности. Однако если попытаться распространить эту программу на всю страну, то тогда предложение работников с более высоким образованием резко возрастет, что неизбежно потянет вниз уровень их оплаты, и из-за этого они, возможно, так и не смогут выбраться из состояния бедности. Одна и та же интервенция, внедренная локально и внедренная глобально, будет иметь совершенно разные последствия.

Принято считать, что внутренняя валидность — самая сильная сторона рандомизированных контролируемых испытаний (по этому параметру они превосходят любые другие методы), тогда как внешняя — их самая слабая сторона (по этому параметру они уступают большинству других методов). Действительно, в исследованиях рандомистов поражает контраст между тщательной проработкой вопросов внутренней валидности и почти полным пренебрежением к вопросам внешней валидности [Deaton, 2020]. В подавляющем большинстве работ с использованием RCT о проблеме внешней валидности даже не упоминается, как если бы ее не существовало.

В тех редких случаях, когда о ней все же вспоминают, это выглядит скорее как формальность. Так, А. Банерджи предложил удивительный по простоте выход: считать результаты любого рандомизи-

рованного эксперимента по умолчанию общезначимыми и независимыми от контекста, если только нет свидетельств, явно говорящих об обратном [Banerjee, 2005]. Очевидно, что такая механическая экстраполяция очень далека от стандартов строгости и научности, приращенность которым декларируют рандомисты¹⁴.

Другой вариант не намного конструктивнее: многократно повторять один и тот же эксперимент в разных контекстах (местностях, регионах, странах) до тех пор, пока мы либо не убедимся, что эффект воздействия достаточно стабилен, либо обнаружим, что это не так [Ibid.]. Когда выяснится, что повторные испытания, несмотря на различия в условиях, приводят к более или менее идентичным результатам, это будет свидетельствовать об их внешней валидности. Коллекционирование таких метаоценок рекомендуется как наилучший способ, позволяющий преодолевать проблему внешней валидности и разрабатывать на их основе «доказательные» рекомендации для политиков: «Единственный эксперимент не дает окончательного ответа на вопрос, будет ли программа универсально “работать”. Но мы можем провести серию экспериментов, отличающихся либо местом проведения, либо проверяемым видом вмешательства (или и тем, и другим). Подобная серия экспериментов позволит нам проверить надежность наших выводов» [Банерджи, Дюфло, 2021, с. 39].

Но и это лишь видимость решения: из того, что какая-то интервенция оказалась успешной в двух разных странах, никак не следует, что она будет успешной и в третьей. Накопление кейсов само по себе ничего не гарантирует. Сторонники RCT утверждают, что для того, чтобы удостовериться, что что-то действительно «работает», следует просто много раз повторять один и тот же эксперимент в разных условиях. Но сколько таких повторений необходимо, чтобы окончательно убедиться, что это что-то «работает»? Ответа на этот вопрос у них нет. Если в десяти случаях в десяти разных местах интервенция *A* оказалась лучше интервенции *B*, это еще не значит, что она окажется лучше и в одиннадцатом. Здесь возникает классическая проблема индукции, которая, как хорошо известно из философии науки, не имеет удовлетворительного решения¹⁵. Без теории, которая

¹⁴ В другой работе он заявил, что оценка внешней валидности результатов RCT всегда будет оставаться чисто «субъективной» [Banerjee et al., 2016].

¹⁵ В этом контексте некоторые критики справедливо вспоминают знаменитый пример Бертрана Рассела про курицу, которая после сотни повторений вывела за-

могла бы стать основой для прогнозирования результатов (а рандомисты ее избегают), все сводится к простому перечислению: первый лебедь — белый, второй — белый, третий — белый, но каким окажется четвертый, никому не известно [Deaton, Cartwright, 2018].

К этим концептуальным соображениям добавляются чисто практические. Для исследователей стимулы к тому, что заниматься репликациями экспериментов, уже проведенных кем-то другим, гораздо слабее стимулов к тому, чтобы браться за новые. Только оригинальный эксперимент способен обеспечить публикацию в топовом научном журнале и ее последующее активное цитирование. Идея подвергать все RCT многократным репликациям выглядит утопически, поскольку репликации ценятся научным сообществом не слишком высоко и мало помогают академической карьере. Когда репликации все же проводятся, они практически никогда не бывают точными, отличаясь от исходного исследования по множеству деталей формирования выборки, дизайна эксперимента, техники оценивания и т.д. Еще большие затруднения возникают при попытках масштабирования. Показательно, что из всего массива завершенных рандомизированных экспериментов, информация о которых содержится на веб-сайте Лаборатории J-PAL, масштабированию подвергались только 2% [Bédécarrats et al., 2019].

Метаанализ рандомизированных оценок среднего эффекта воздействия, проведенный Э. Вивальт, свидетельствует об их существенном разбросе в зависимости, во-первых, от размера выборок и, во-вторых, от идентичности экспериментаторов [Vivalt, 2020]. Оценки, полученные на малых выборках, оказываются намного выше, чем полученные на больших, а полученные академическими исследователями и НПО — намного выше, чем полученные государственными агентствами. Денежные и политические стимулы, с которыми сталкиваются государственные служащие из развивающихся стран, мало напоминают стимулы, которые направляют деятельность дипломированных экспертов из развитых стран [Deaton, 2020].

В качестве примера можно сослаться на школьный эксперимент, проведенный в Кении одной неправительственной организацией. Случайно отобраным школам были предоставлены средства для

кономерность, что стоит только ей слышать шаги фермера, как у нее тут же появляется корм, пока в день накануне Рождества после послышавшихся шагов фермера ей не свернули шею [Deaton, 2020].

найма по краткосрочному контракту дополнительного учителя [Duflo et al., 2015]. Дети, занимавшиеся с дополнительными учителями, продемонстрировали на тестовых испытаниях значительно лучшие результаты, чем дети, занимавшиеся без них. Однако когда другая группа исследователей [Bold et al., 2013], используя ту же схему, попыталась воспроизвести результаты этого эксперимента, только на этот раз с дополнительными учителями, нанятыми не НПО, а правительством, эффект исчез. Стало ясно, что успеваемость школьников повышалась не по причине найма дополнительных учителей, а по причине того, что организацией и мониторингом всего процесса занимались высокомотивированные представители неправительственной организации [Ravallion, 2020].

Чтобы удостовериться в инвариантности полученных результатов, они должны быть привязаны к параметрам внешней среды [Deaton, Cartwright, 2018]. Многие каузальные процессы работают только при наличии строго определенных экономических, социальных и культурных структур. Для разных типов причинно-следственных отношений требуются разные наборы поддерживающих факторов, так что программа, успешная при одном сочетании этих факторов, может потерпеть неудачу при другом. Иными словами, чтобы иметь возможность использовать результаты RCT, нужно знать условия, при которых они будут сохраняться вне исходного контекста. Это требует информации об окружающем мире, выходящей далеко за пределы непосредственно самих RCT, причем основная часть такой информации добывается с помощью других методов. Без нее мы лишаемся не только возможности переносить результаты эксперимента куда-то еще, но и возможности делать нормативные выводы о том, как эти результаты могут повлиять на благосостояние людей [Ibid.].

Неспособность новой экономики развития обеспечить внешнюю валидность RCT напрямую связана с ее общей атеоретической ориентацией — с тем, что ее интересует только вопрос «что» и не интересует вопрос «почему». Дизайн RCT строится по принципу «черного ящика»: известно, что находится на входе и что получается на выходе, но то, что происходит внутри, остается за кадром [Deaton, 2010]. Метод рандомизации способен (в идеальных условиях) идентифицировать каузальный эффект, вызванный определенной интервенцией, но он ничего не может сказать о том, какие конкретные *при-*

чинно-следственные механизмы за этим стоят [Bédécarrats et al., 2020; Deaton, 2010]. Чтобы ответить, почему работает та или иная программа, нужна теория: нужно понимание *характера связи* между оказанным воздействием и полученным от него эффектом. Только теория способна подтвердить инвариантность полученных результатов, открывая возможность для их переноса в иные контексты. В экономике RCT интерполяции и экстраполяции заменяют прогнозирование на основе моделей [Heckman, 2020].

Отсутствие у рандомистов интереса к раскрытию устройства причинно-следственных механизмов выглядит парадоксально для подхода, который видит своей главной задачей установление причин явлений [Bédécarrats et al., 2020]. Это резко ограничивает практическую полезность RCT, идет ли речь о воспроизведении их результатов в других локальных контекстах или об их масштабировании. По мнению М. Равайона, в утверждениях рандомистов о том, что RCT абсолютно надежны, что их результаты не зависят от контекста и что поэтому в качестве «золотого стандарта» они занимают вершину иерархии исследовательских методов, больше веры, чем науки [Ravallion, 2020].

Чтобы быть уверенными в том, что программа сработает в других контекстах, нам необходимо знать, *почему* она сработала в *данном* контексте, то есть иметь представление о механизме, посредством которого воздействие T вызвало эффект E . Именно наличие теории, раскрывающей характер причинно-следственной связи, способно обеспечить внешнюю валидность, давая основания ожидать, что результаты, полученные в одном контексте, с высокой вероятностью будут воспроизводиться в других: «Для того, чтобы рандомизированное исследование давало “полезное знание” за пределами локального контекста, оно должно отражать некую общую тенденцию, некий общий эффект, являющийся продуктом действия какого-то механизма, который, насколько можно судить, имеет более широкую сферу приложения» [Deaton, 2010].

Идентификация того или иного локального каузального эффекта сама по себе недостаточна, чтобы считать его внешне валидным [Deaton, Cartwright, 2018]. В лучшем случае RCT могут давать несмещенные оценки интересующих нас эффектов, но из-за слабой внешней валидности практическая полезность таких оценок ограничена. Внешнюю валидность невозможно обеспечить изнутри самого экс-

перимента: для этого, как уже отмечалось, нужны данные, лежащие за его пределами. Как интервенция взаимодействует с другими факторами? Насколько сильно данный контекст отличается от других? Какие общеравновесные эффекты могут возникать при увеличении масштабов программы? В рамках RCT все эти вопросы остаются без ответа.

Чтобы ответить на них, нужно понять, почему и при каких условиях работает та или иная программа, а чтобы понять это, необходима теория¹⁶. Заявления рандомистов о внешней валидности эффектов, обнаруживаемых с помощью RCT, лишены научного основания и, по существу, выдают желаемое за действительное. Как выразился по этому поводу Дени Родрик, «обещание рандомистов превратить RCT в машину познания без теории является несбыточным» (цит. по: [Deaton, Cartwright, 2018]). Тот факт, что для новой экономики развития вопросы внутренней валидности обладают абсолютным приоритетом над вопросами внешней валидности, имеет негативные практические последствия: отсутствие у RCT внешней валидности неизбежно обесценивает политические рекомендации, вырабатываемые на их основе.

Во многом это связано с тем, что сторонники экономики RCT исходят из узкого понимания каузальности, пытаясь идентифицировать причины напрямую из эмпирических данных без привлечения каких-либо теоретических идей или дополнительных предположений. Современная философия науки считает такой подход несостоятельным: «Каузальные выводы не следуют дедуктивно из данных без строго определенного набора вспомогательных допущений, причем эти допущения сами по себе также не являются дедуктивными следствиями из данных». Поэтому, строго говоря, никакую оценку, полученную в каком-то одном исследовании, даже если это RCT, некорректно называть «каузальной»: «Только целая совокупность свидетельств приводит к вердикту о каузальности. ... Каузальность — это научный вывод, *теоретическое* утверждение, и как таковое оно выходит за границы любого отдельно взятого исследования» [Broadbent et al., 2017].

¹⁶ По поводу атеоретической установки экономики RCT Д. Родрик заметил как-то, что прежде чем выписывать рецепт, сначала хорошо было бы ставить диагноз [Favereau, 2016].

В этом критическом пункте позиции «старой» и «новой» экономики развития радикально расходятся. Как полагают критики, мечта рандомистов о получении чисто эмпирических доказательств без всякой теории представляет собой утопию [Easterly, 2009]: секреты экономического развития невозможно раскрыть, если не руководствоваться какой-то общей теорией, открытой для пересмотра и обновления [Deaton, 2010]¹⁷.

Политико-экономические аспекты

Отправным пунктом для рандомистской политической программы стала крайне скептическая оценка результатов, достигнутых «старой» экономикой развития, которая видела свою главную задачу в выведении развивающихся стран на орбиту устойчивого экономического роста. «Неприятная правда заключается в том, что нет никакой формулы роста, — отмечают А. Банерджи и Э. Дюфло в статье, написанной как своеобразный политический манифест движения, — ученые не понимают, почему одни экономики растут, а другие топчутся на месте. Не существует общепринятого рецепта, как можно подтолкнуть бедные страны к достижению стабильно высоких темпов роста, не существует общих принципов роста и не существует двух полностью одинаковых примеров экономического роста» [Banerjee, Duflo, 2020].

¹⁷ Многие наблюдатели обращают также внимание на этическую уязвимость проводимых рандомистами экспериментов [Abramowicz, Szafarz, 2020]. Так, в них не соблюдается принцип информированного согласия, строго обязательный для клинических испытаний в медицине, когда испытуемые в явной форме подтверждают свое согласие на участие в эксперименте. Нарушается и принцип равного отношения ко всем участникам эксперимента из-за фактической дискриминации, которой во многих случаях подвергаются индивиды из контрольной группы, когда им отказывают в получении помощи (например, таблеток от гельминтов), в благотворности которой уверены экспериментаторы: «Рандомизаторы готовы пожертвовать благополучием участников исследования ради получения новых знаний» [Ziliak, Teather-Posadas, 2016]. В медицинских клинических испытаниях это также считается недопустимым. Достаточно часто в РСТ люди, нуждающиеся в помощи, ее не получают, в то время как люди, не нуждающиеся в ней, ее получают [Ravallion, 2020]. Наконец, этический дискомфорт вызывает сама ситуация, когда успешные академические исследователи из развитых стран используют бедняков из развивающихся стран для улучшения своего профессионального CV [Deaton, 2020].

Но это не значит, что развивающимся странам ничем нельзя помочь: несмотря на то, что проблема экономического роста остается загадкой, бремя невзгод, падающее на жителей бедных стран, можно значительно уменьшить. Для этого, по мнению сторонников новой экономики развития, есть все необходимое. Во-первых, в отличие от проблемы роста эксперты знают, как добиваться прогресса по различным конкретным аспектам жизни людей в развивающихся странах (бедность, образование, санитария и т.д.). Во-вторых, интервенции по таким локальным проблемам имеют измеримые цели и их эффективность может быть оценена количественно. В-третьих, сегодня эксперты располагают надежным научным инструментарием для получения подобных оценок — техникой рандомизированных контролируемых испытаний, благодаря которым они могут экспериментировать, отказываясь от программ, которые не работают, и продвигая программы, которые работают [Banerjee, Duflo, 2020].

Экономисты старой школы полагали, что основной упор следует делать на выработку правильной макроэкономической политики и создании «хорошей» институциональной среды, не утруждая себя поиском микродоказательств [Banerjee, 2009]. Экономисты новой школы исходят из противоположной установки: главное — это микродоказательства, поскольку без них невозможно знать, принесут самоочевидные, на первый взгляд, рекомендации пользу или вред. Безусловное преимущество RCT в том, что мы можем начинать с конкретного политического вопроса, а затем начинать искать доказательства [Ibid.].

Как подчеркивают рандомисты, наиболее крупные успехи последних десятилетий были связаны не с ускорением роста, а с улучшением конкретных сторон жизни даже тех стран, которые были и остаются очень бедными: «Никто не знает, когда в той или иной стране заработает локомотив роста; но когда это произойдет, бедняки с большей вероятностью смогут запрыгнуть в этот поезд, если они будут иметь приличное здоровье, уметь читать и писать и заботиться не только о своих неотложных нуждах. Не нужно напрасно пытаться ускорить рост: следует сосредоточиться на том, к чему он должен приводить, — к улучшению благосостояния бедного населения. Устранение таких проблем, может быть, и не подтолкнет страны к устойчивому ускорению роста, но оно сможет значительно повысить благосостояние их граждан. В отсутствие волшебного зелья для запуска

процесса развития лучший способ коренным образом изменить миллионы жизней не в том, чтобы безуспешно пытаться ускорить рост. Он заключается в том, чтобы сосредоточиться непосредственно на том, что рост должен улучшать: на благосостоянии бедных» [Banerjee, 2009]¹⁸.

Рандомисты уверены, что реальный прогресс в развивающихся странах возможен только с помощью получаемых ими «строгих» доказательств. В перспективе должен быть создан специальный орган, который мог бы создать исчерпывающий список «хороших» и «плохих» политик развития и затем действовать как координационный центр, предоставляя всем заинтересованным сторонам ответы на вопросы «что работает», а «что не работает» на практике [Labrousse, 2020].

Такой подход можно назвать технократическим, поскольку он стремится к деполитизации процесса развития — выведению его из сферы политических и идеологических разногласий. Как неявно предполагают рандомисты, главный источник подобных разногласий — отсутствие строгих научных доказательств. Когда такие доказательства будут предъявлены, политические страсти улягутся сами собой. Иными словами, РСТ отводится роль беспристрастного арбитра при выработке и оценке государственной политики. Де-факто это отдает деятельность государства в руки экспертов, рекомендациям которых должны строго следовать политики и чиновники: «Нам требуется превратить государственную политику из искусства в науку» [Leigh, 2018].

Критикуя рандомистскую программу, Л. Притчетт выдвигает несколько контраргументов [Pritchett, 2014]. Во-первых, множеству стран удалось достичь высокого уровня жизни и искоренения крайней бедности в те времена, когда о рандомизированных экспериментах еще ничего не было слышно. Никому никогда не приходило в голову утверждать, что эти страны стали развитыми и процветающими, потому что при формулировании своей политики использовали РСТ. Во-вторых, две страны, которые в последние десятилетия внесли наибольший вклад в сокращение глобальной бедности, — Китай

¹⁸ Ср.: «Не очевидно, что лучший способ добиться роста — это проводить политику роста. Возможно, обеспечение роста находится по большому счету вне нашего контроля. ... Возможно, мы никогда не узнаем, с чего он начинается и что заставляет его продолжаться» [Banerjee, 2009].

и Индия, добились этого за счет масштабных экономических реформ, а не точечных интервенций в области здравоохранения или образования¹⁹. И если им удалось радикально снизить бедность без помощи РСТ, то на чем основаны утверждения, что их использование — самый эффективный способ борьбы с ней? Имеются ли какие-либо свидетельства того, что страны, активно прибегавшие к РСТ, сокращали бедность быстрее, чем страны, этого не делавшие? По-видимому, нет: «Никогда не существовало никакой теории и никаких свидетельств о том, что ключевым или хотя бы важным ограничением для политики развития является отсутствие строгих доказательств причинно-следственных связей или что получение таких доказательств изменило бы практику. Нам предлагают принимать это просто на веру» [Pritchett, 2014]. В-третьих, рандомисты почему-то игнорируют первую волну увлечения РСТ в США, которая пришла на 1970–1980-е годы, когда для оценки эффективности социальных программ, связанных с развернутой президентом Л. Джонсоном программой «Борьба с бедностью», начали активно использоваться рандомизированные эксперименты. Нынешняя волна, охватившая экономику развития, является, таким образом, второй. Однако рандомисты ничего не пишут о первой волне РСТ и причинах разочарования в них.

Л. Притчетт выделяет две конкурирующие стратегии, существующие в современной экономике развития, — стратегию «национального развития», которая имеет своим объектом все общество и нацелена на решение глубинных экономических и социальных задач, и стратегию «урезанного развития» (kinked), которая имеет своим объектом отдельные социальные группы и нацелена на решение ограниченных, не требующих фундаментальных изменений задач

¹⁹ Ср. с аналогичными рассуждениями А. Дитона: «Великие эпизоды сокращения материальной бедности в мире, особенно в Китае и Индии, были вызваны экономическим ростом и глобализацией. Общий рост шел вместе с ростом в отдельных сферах — вместе с увеличением количества рабочих мест, расширением возможностей, разрастанием дорожной сети, появлением лучших по качеству и более многочисленных школ и клиник, но в экономиках с хорошими институтами, где шел быстрый рост, все такого рода вещи возникали более или менее спонтанно. Насколько мне известно, никто не объявлял РСТ ключом к экономическому росту и трудно вообразить, чтобы они имели хоть какое-то отношение к сокращению бедности в Китае. Испытания на микроуровне часто бывают успешными, но их решающая роль в снижении уровня бедности в мире остается не более чем символом веры» [Deaton, 2020].

[Pritchett, 2020]. Стратегия «урезанного», или «кинки-развития», была выдвинута рандомистами в качестве альтернативы более традиционной стратегии «национального развития». Согласно этой точке зрения, смысл процесса развития состоит в достижении минимально приемлемых показателей в определенных сферах жизни общества («искоренение крайней бедности», «всеобщее начальное образование», «обеспечение доступа к чистой воде» и т.д.). Такие цели можно назвать «урезанными» в том смысле, что они устанавливают какую-то произвольную планку по определенному аспекту благосостояния людей, а затем направляют весь процесс развития на ее достижение [Kenny, Pritchett, 2013].

Так, внимание рандомистов обращено почти исключительно на беднейших из бедных — живущих в крайней нищете (меньше чем на 1,9 долл. по ППС в день), полностью неграмотных, лишенных доступа к чистой питьевой воде и т.д. Свои эксперименты они проводят практически только на них и свои рецепты адресуют практически только им. В рамках такой кинки-стратегии повышению благосостояния людей, находящихся выше этого порога, де-факто придается нулевая ценность. Вопрос о том, как можно было бы повысить уровень жизни *всего* населения развивающихся стран, сторонников экономики RCT, строго говоря, не интересует. Они озабочены не тем, чтобы сдвинуть вверх всю кривую распределения доходов в развивающихся странах, но лишь тем, чтобы приподнять ее левый хвост, где концентрируются беднейшие из бедных [Pritchett, Sandefur, 2013]. Но даже здесь кинки-стратегия довольствуется малым: хотя меры, рекомендуемые рандомистами, могут, конечно, улучшать положение наиболее уязвимых групп населения развивающихся стран, в большинстве случаев они все равно оставляют эти группы *под* чертой бедности [Ogden, 2020]. Для сокращения бедности, убеждены критики, важен экономический рост, а не «строгая» оценка тех или иных локальных проектов.

RCT подходят лишь для узкого круга потенциальных интервенций и даже тогда они не отвечают на многие вопросы, которые могут быть важны политикам. Они неприложимы к решению важнейших проблем развития, таких, например, как роль институтов, разработка правильной макроэкономической политики, улучшение корпоративного управления, рационализация системы налогообложения, участие в международной торговле, и это лишь некоторые из них

[Bédécarrats et al., 2019]. Ключевых проблемы, связанные, в терминах Притчетта, с целями «национального развития», остаются вне зоны досягаемости RCT.

Рандомизации лучше всего поддаются сравнительно простые микропроекты с четким разделением на группы участников и неучастников, с коротким временным горизонтом, с однонаправленными линейными причинно-следственными связями (*A* вызывает *B*) и с отсутствием внешних эффектов, затрагивающих третьих лиц. Это узкоцелевые программы, получившие в исследовательской литературе название «программ туннельного типа» [Bernard et al., 2012]. Если мы хотим, чтобы люди спали под противомалерийными сетками, лучше их раздавать бесплатно или продавать по символической цене? Могут ли условные денежные трансферты подтолкнуть матерей из бедных семей к тому, чтобы они начали посылать своих детей в школу? Усилят ли гендерные квоты при выборах в местные органы власти политическое влияние женщин? Как отмечает Д. Родрик, «неопрровержимые доказательства, которые генерируют многие рандомизированные эксперименты, относятся, как правило, к вопросам, которые настолько узко ограничены по охвату и значению, что сами по себе могут представляться неинтересными» [Rodrik, 2009]. Кроме того, даже если «что-то» работает, совсем не очевидно, что желательно именно это «что-то», а не какое-то другое, которое работает еще лучше.

Отсюда — практически нулевая полезность RCT для лиц, ответственных за принятие решений. Когда известного индийского экономиста Арвинда Субраманиана, в течение трех лет занимавшего пост главного экономического советника правительства Индии, спросили в интервью, как дорогостоящие рандомизированные эксперименты повлияли на государственную политику страны, он не смог вспомнить ни одного случая, когда бы RCT хоть чем-то помогли в решении десятков неотложных вопросов, с которыми ему пришлось сталкиваться [Ravallion, 2020].

В то же время рандомизация неприменима для оценки комплексных долговременных проектов в нестабильной среде [Picciotto, 2020]. Говоря иначе, наиболее подходящим для RCT объектом выступают частные блага, получаемые отдельными домохозяйствами, но не общественные блага, достаемые всем или большей части общества [Hammer, 2017]. Так, она крайне затруднена в случае средних и круп-

ных инфраструктурных проектов, а в случае масштабных реформ на уровне целых секторов или всей экономики она, по-видимому, вообще невозможна. Это по определению исключают огромное количество самых различных направлений политики развития. По некоторым оценкам, RCT поддаются менее 5% всех потенциальных мер политики в этой области [Bédécarrats et al., 2019]. В результате сфера применения «строго доказательного подхода» к проблемам развития оказывается «исчезающе мала» [Pritchett, Sandefur, 2013].

С этим же связана отраслевая асимметрия в использовании RCT. По данным Всемирного банка, при оценке эффективности его программ в области образования метод рандомизированных контролируемых испытаний использовался в 58% случаев, программ в области здравоохранения, питания и контроля рождаемости — в 83% случаев, программ в области ИКТ — в 67% случаев, программ в области в области санитарии и водоснабжения — в 72% случаев. В то же время в программах, относившихся к сельскому хозяйству, экономической политике, энергетике, охране окружающей среды, поддержке бизнеса, транспорту и развитию городов, он использовался менее чем в трети случаев [Ravallion, 2020].

В тех сферах, где RCT невозможны, единственным доступным инструментом анализа и оценки оказываются обсервационные исследования, но рандомисты отвергают их как не соответствующие «золотому стандарту» строгости и научности. В результате множество потенциально эффективных интервенций выпадают из поля зрения государственных структур развивающихся стран и спонсоров международной помощи. Пропаганда RCT как наилучшего или даже единственно научного метода при оценке программ помощи ведет к серьезным деформациям в борьбе с глобальной бедностью. В частности, она чревата отказом международного сообщества от оказания помощи развивающимся странам в разработке крупномасштабных национальных и региональных программ, прежде всего — по реформированию их социально-экономических систем [Bédécarrats et al., 2020].

Ни один из типичных объектов RCT, будь то обусловленные денежные трансферты, микрофинансирование, снабжение домохозяйств более качественными кухонными плитами, дегельминтизация и т.д., невозможно отнести к числу ключевых драйверов экономического роста [Pritchett, 2020]. Успешное национальное развитие зави-

сит в первую очередь от качества институтов и качества государственной политики, а по этим вопросам РСТ мало что способны сказать. В то же время экономический рост неизбежно сопровождается ускоренным прогрессом во всех сферах, привлекающих внимание рандомистов, — здравоохранении, образовании, обеспечении доступа к воде, улучшении качества жилья и т.д. По оценке Л. Притчетта, эффекты узкоцелевых программ не идут ни в какое сравнение с эффектами инклюзивного роста. Практически по любому аспекту благосостояния возможный выигрыш от прогресса в масштабе всей страны или целых секторов на порядок превосходит возможный выигрыш от узкоцелевых программ [Kenny, Pritchett, 2013].

Технократический подход сторонников РСТ также вызывает серьезные возражения. Он удивляет, с одной стороны, своей политической наивностью, когда политическим конфликтам и влиянию групп давления не придается никакого значения, а с другой — своей слепой верой в безошибочность и могущество экспертного знания. А. Дитон видит в такой установке «первородный грех экономики развития»: «Согласно этой доктрине... глобальная бедность отступит при применении правильных технических средств, одно из которых — принятие РСТ в качестве фундамента при разработке доказательной (evidence-based) политики. Игнорирование политической сферы рассматривается как добродетель, а не как порок. ... Технократы считают, что могут способствовать развитию других стран извне, потому что знают, как находить то, что работает. По меньшей мере, в данном отношении они не видят особой разницы между проектированием какого-нибудь гаджета и проектированием социальной политики. И то и другое — упражнения для инженеров. Но последовательное сокращение бедности невозможно без вовлечения в политику» [Deaton, 2020].

По большому счету рандомистский подход отражает запросы международных неправительственных организаций и частных благотворительных фондов, которые ориентированы на оказание помощи беднейшим из бедных и стремятся к отбору наиболее эффективных форм гуманитарного вмешательства [Pritchett, Sandefur, 2013]. Экономика РСТ предоставляет им необходимые для этого инструменты. Но интересы НПО не обязательно совпадают с интересами правительств и уж тем более с интересами рядовых граждан развивающихся стран: «Благотворительные фонды и альтруисты часто “знают”,

что хорошо для бедных людей, и имеют самые лучшие намерения, но предоставляют мало доказательств того, что бедные люди согласны с их оценками или одобряют предлагаемые ими рецепты, так что их интересы могут легко приходиться в противоречие с интересами тех, кому они пытаются помогать» [Deaton, 2020].

Сторонники новой экономики развития утверждают, что RCT способны произвести революцию в социальной политике, сделав ее значительно более эффективной. Но, как иронически замечает Уильям Истерли, само это утверждение находится в противоречии с их же собственными стандартами строгости и научности, поскольку не поддается проверке с помощью RCT [Easterly, 2009].

Заключение

Какие выводы можно сделать из многолетнего интеллектуально-противостояния рандомистов и их критиков?

С концептуальной точки зрения более убедительной выглядит аргументация последних. Трудно не согласиться с их утверждениями о том, что навязывание идеи иерархии методов опасно и ненаучно, потому что она освобождает исследователей от согласования результатов, получаемых разными методами; что никакого методологического «золотого стандарта» не существует и существовать не может; что возможности рандомизации не следует переоценивать, заявляя, что по критериям строгости и научности она намного превосходит любые другие методы; что рандомизацию в реальных полевых условиях нельзя смешивать с рандомизацией в гипотетических идеальных условиях; что на практике формирование экспериментальной и контрольной групп далеко не всегда строится по принципу случайного отбора; что внутренняя валидность RCT совсем не безупречна и что получаемые на их основе оценки нельзя считать полностью свободными от смещений; что никакое отдельное рандомизированное испытание не может претендовать на точную идентификацию каузального эффекта; что RCT совершенно бессильны перед проблемой внешней валидности, так что попытки переноса результатов, полученных в том или ином специфическом контексте, на какие-то другие контексты всегда остаются субъективными и произвольными; что атеоретическая установка сторонников RCT представляет собой

концептуальный тупик; что рандомизированные исследования ничего не могут сказать о внутренних механизмах, стоящих за результатами экспериментов; что напрямую из рандомизированных контролируемых испытаний идентифицируется только один показатель — средний эффект воздействия, тогда как все остальные, не менее важные, остаются за кадром²⁰.

Некоторые из ограничений, с которыми сталкиваются RCT, являются их собственными, но многие оказываются общими с другими эконометрическими методами. Им нельзя приписывать какой-то привилегированный статус, поскольку они подвержены тем же проблемам, с которыми традиционно пыталась бороться эконометрика, и нет ничего такого, что бы было под силу только им [Deaton, 2020]. В определенных условиях при решении определенных проблем observationalные исследования могут давать более надежные результаты, так что широко разрекламированное превосходство RCT над всеми другими методами является «вопросом скорее веры, чем науки» [Ravallion, 2020]. Претензии рандомистов на создание «доказательной экономики» по аналогии с «доказательной медициной» выдают желаемое за действительное²¹.

Рандомизированные эксперименты чрезвычайно дороги и сталкиваются с большим числом этических проблем. Из-за крайне низкой внешней валидности политические рекомендации, вырабатываемые на основе RCT, имеют ограниченную практическую ценность и всегда сопряжены с риском, что интервенция, сработавшая в каких-то специфических условиях, не сработает при попытке ее более широкого внедрения. Плата за формальный успех RCT — зауженный фокус: они применимы только к небольшим «туннельным» проектам

²⁰ Ср. с вердиктом А. Дитона и Н. Картрайт: «Мы категорически не согласны с часто высказываемой идеей, будто АТЕ, рассчитываемый на основе RCT, автоматически достоверен, будто рандомизация обеспечивает автоматический контроль ненаблюдаемых факторов и, что хуже всего, будто получаемое таким образом значение АТЕ соответствует его истинному значению» [Deaton, Cartwright, 2018].

²¹ Большинство исследований, выполненных рандомистами, не соответствуют их собственным критериям строгости и научности, в чем они сами неявно признаются. «Экономисты, — пишут Банерджи и Дюфло, — больше похожи на водопроводчиков: мы решаем проблемы, сочетая интуиции, опирающиеся на науку, какие-то догадки, подкрепленные опытом, и множество проб и ошибок в чистом виде» [Banerjee, Duflo, 2019, p. 7]. Так все-таки — интуиции и догадки или «золотой стандарт» строгости и научности?

и ничем не могут помочь при проведении крупных социальных реформ [Deaton, 2010]. С быстрым ростом популярности РСТ связано постепенное вытеснение философии «больших дел» философией «малых дел» в рамках экономики развития: экономисты начали отказываться от изучения фундаментальных проблем, жизненно важных для развивающихся стран, переключаясь на изучение «миниатюрных» проблем, интересных прежде всего им самим. Результатом такой переориентации стала деформация усилий международного сообщества по оказанию экспертной, политической и финансовой поддержки развивающимся странам, отказ от помощи им при проведении крупномасштабных социальных и экономических реформ.

Вместе с тем нет сомнений, что формирование новой экономики развития стало одним из наиболее ярких событий в эволюции экономической мысли начала XXI в. Она выступила важнейшим каналом, по которому в экономическую науку начали проникать экспериментальные методы, аналогичные методам естественнонаучных дисциплин. В обозримом будущем этот эксперименталистский крен, скорее всего, будет только усиливаться и, значит, новая экономика развития будет по-прежнему находиться на переднем крае научных поисков. Экономика РСТ подняла совершенно новый пласт проблем, остававшихся раньше вне поля зрения исследователей, и предложила новый подход к их решению, опирающийся на экспериментальные данные. Во многом благодаря ей изменился ландшафт современной экономической науки и даже само мышление современных экономистов. Она выдвинула на первый план вопросы оценки эффективности программ помощи развивающимся странам, а также необходимости подотчетности при их реализации. Результатом этого стало критическое переосмысление опыта, связанного с попытками сокращения глобальной бедности: экспериментальная проверка выявила несостоятельность многих идей, не вызывавших никаких сомнений и казавшихся заведомо эффективными. Экономика РСТ дала мощный импульс сбору новых качественных данных непосредственно самими исследователями, что резко расширило пространство эмпирического изучения проблем развития. Благодаря РСТ экономика развития смогла спуститься с академических высот и приблизиться к реальной жизни рядовых людей из развивающихся стран: главным объектом ее внимания стали их насущные заботы и нужды. Она сделала сам исследовательский процесс более структурирован-

ным, потребовав придерживаться при проведении экспериментов строгого протокола, и более транспарентным, потребовав выкладывать в открытый доступ исходные данные и рабочие файлы. Это способствовало утверждению более высоких академических стандартов, которым должны удовлетворять эмпирические исследования, чтобы их результаты могли признаваться достоверными. Сегодня и теория и практика развития выглядят совершенно иначе, чем два или два с половиной десятилетия тому назад.

Однако критики не отрицают высокого аналитического потенциала метода рандомизации. Как и другие методы, RCT часто бывают полезны, но и, как и другие методы, имеют свои недостатки и подводные камни [Deaton, 2020]. С чем не согласны критики, так это с утверждениями о том, что RCT находятся на вершине иерархии методов и что рандомизация гарантирует получение несмещенных оценок каузальных эффектов. Они полагают, что методологический плюрализм обладает преимуществом перед жесткой приверженностью единственной модели оценивания социальных программ [Picciotto, 2020]. В конечном счете выбор метода всегда определяется природой изучаемых проблем. Какой метод с наибольшей вероятностью приведет к правильному каузальному выводу, зависит от того, что является целью исследования, а также от того, какой информацией о предмете исследования мы уже располагаем. Лучшим методом всегда следует считать тот, который способен дать наиболее убедительные и релевантные ответы в данном конкретном контексте [Deaton, 2020].

В итоге в интеллектуальном противостоянии между рандомистами и их критиками сложилась патовая ситуация, когда каждая сторона, несмотря ни на что, остается при своем мнении. Во многом это связано с тем, что коммуникация между ними была односторонней, поскольку сторонники RCT фактически отказывались вступать в концептуальный диалог со своими оппонентами, обходя их аргументы молчанием [Ogden, 2020]. Ретроспективно такая дискурсивная стратегия выглядит вполне успешной: критика, какой бы справедливой она ни являлась, была бессильна подорвать позиции экономики RCT и остановить ее победное шествие. Сегодня она абсолютно доминирует как в экономике, так и в политике развития, а голоса несогласных едва слышны. Контраст между узостью RCT и их научным, медийным и политическим успехом свидетельствует об

успешности рандомистской риторики. Можно сказать, что логика критиков оказалась слабее риторики рандомистов.

У этого есть как психологические, так и социологические причины. По-видимому, у человеческого сознания есть встроенный механизм, когда в любом эмпирическом результате (тем более — экспериментальном) оно склонно видеть не единичный изолированный случай, а проявление некой общей закономерности. К каждому отдельно взятому факту подсознательно примысливается его универсальная значимость. Именно так чаще всего происходит с результатами рандомизированных экспериментов. Когда экономисты вставляют в названия своих работ слово «каузальный», они явно имеют в виду нечто большее, чем какой-то единичный случай, наблюдаемый при каком-то уникальном стечении обстоятельств. Всегда есть соблазн взять исследование, давшее впечатляющие результаты, и распространить их далеко за границы исходного контекста, в котором они были получены. Как показывает опыт, противостоять этому искушению не способны ни сами исследователи, ни те, кому они дают советы [Deaton, 2020]. Против такой встроенной психологической установки любая концептуальная критика бессильна.

Другой психологический фактор — иллюзия «беспредпосылочности» RCT. Рандомисты утверждают, что рандомизированные эксперименты позволяют получать несмещенные оценки «истинного» каузального эффекта, поскольку они основываются только на законе больших чисел и полностью свободны от каких бы то ни было гипотез, теорий и предположений²². Не удивительно, что люди склонны доверять таким оценкам больше, чем оценкам, которые сопровождаются длинным списком условий, оговорок и допущений. Человеческое сознание готово автоматически принимать их на веру. Опять-таки противостоять этому искушению психологически чрезвычайно трудно.

Большой вклад в популяризацию и распространении RCT внесли неправительственные организации и частные благотворительные фонды, активность которых в последние десятилетия резко возросла. Это излюбленные партнеры для экономистов-экспериментато-

²² В действительности свобода RCT от каких-либо априорных предположений является иллюзией. Один исследователь насчитал 26 предпосылок, которые должны выполняться, чтобы отдельный эксперимент мог давать несмещенную оценку каузального эффекта [Cook, 2018].

ров. С одной стороны, объектом финансирования со стороны НПО чаще всего становятся сравнительно небольшие гуманитарные интервенции, для оценки эффективности которых идеально подходят RCT. С другой стороны, сами рандомисты предпочитают иметь дело с НПО, потому что по сравнению с правительственными агентствами они являются более гибкими, менее бюрократизированными и более открытыми для инноваций [Cohen, Easterly, 2010]. Можно даже говорить о своеобразном разделении труда, сложившемся между представителями старой и новой школ экономики развития: экспертные услуги первых оказываются чаще востребованы государственными структурами развивающихся стран, тогда как вторые — неправительственными организациями и частными благотворительными фондами развитых стран.

Что касается привлекательности RCT для академических исследователей, то она объясняется двойной мотивацией, стоящей за их использованием. С одной стороны, чисто проведенный рандомизированный эксперимент практически гарантирует автору публикацию в топовом журнале и, значит, вхождение в лигу суперпрофессионалов экономического сообщества. С другой стороны, это позволяет ему осознавать себя в качестве супергуманиста, проявляющего заботу о наиболее страдающей и обездоленной части человечества, живущей в нищете. Естественно, для новых поколений экономистов такая двойная мотивация обладает неотразимой силой.

Отсюда достаточно парадоксальный финал интеллектуального противостояния, обсуждению которого посвящена наша работа: хотя с чисто научной точки зрения аргументы критиков выглядят, по видимому, предпочтительнее, с практической абсолютными победителями из этого спора вышли рандомисты. Никакая критика не смогла подорвать их авторитет в глазах научного сообщества, политиков, СМИ и широкой публики. Скорее всего, поток исследований с использованием RCT будет становиться все шире, их репутация в академической среде все выше, а их интеллектуальное и политическое влияние все сильнее.

Литература

Банерджи А., Дюфло Э. (2021). Экономика бедных. Радикальное переосмысление способов преодоления мировой бедности. М.: Издательство Института Гайдара.

Ениколопов Р.С. (2020). Доказательная экономика развития: Нобелевская премия по экономике 2019 года // Вопросы экономики. № 1. С. 5–17.

Капелюшников Р. И. (2021). О современном состоянии экономической науки: полу-социологические наблюдения // Капелюшников Р. И. Экономические очерки. История идей. Методология. Неравенство. Рост. М.: Изд. дом Высшей школы экономики.

Abramowicz M., Szafarz A. (2020). Ethics of RCTs: Should Economists Care about Equipoise? // RCT in Development: A Critical Perspective / F. Bédécarrats, I. Guérin, F. Roubaud (eds). Oxford: Oxford University Press. P. 280–292.

Aiken A. M., Davey C., Hargreaves J. R., Hayes R. J. (2015). Re-analysis of Health and Educational Impacts of a School-Based Deworming Programme in Western Kenya: A Pure Replication // International Journal of Epidemiology. Vol. 44. No. 5. P. 1572–1580.

Angrist J. D., Pischke J.-S. (2010). The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics // Journal of Economic Perspectives. Vol. 24. No. 2. P. 3–30.

Banerjee A. V. (2005). ‘New Development Economics’ and the Challenge to Theory // Economic and Political Weekly. October 1. P. 4340–4344.

Banerjee A. V. (2006). Making Aid Work // Boston Review. Vol. 31. No. 4. P. 7–9.

Banerjee A. V. (2009). Big Answers for Big Questions: The Presumption of Growth Policy // What Works in Development?: Thinking Big and Thinking Small / J. Cohen, W. Easterly (eds). Washington: Brookings Institution Press.

Banerjee A., Duflo E., Glennerster R. (2008). Putting a Band-Aid on a Corpse: Incentives for Nurses in the Indian Public Health Care System // Journal of the European Economic Association. Vol. 6. No. 2–3. P. 487–500.

Banerjee A., Chassang S., Snowberg E. (2016). Decision Theoretic Approaches to Experiment Design and External Validity. NBER Working Paper No. 22167. Cambridge, MA: NBER.

Banerjee A., Duflo E. (2019). *Good Economics for Hard Times Better Answers for Biggest Problems*. N.Y.: Public Affairs.

Banerjee A., Duflo E. (2020). *How Poverty Ends. The Many Paths to Progress — and Why They Might Not Continue* // *Foreign Affairs*. Vol. 99. No. 1. P. 22–29.

Bédécarrats F., Guérin I., Roubaud F. (2019). *All that Glitters is not Gold. The Political Economy of Randomized Evaluations in Development* // *Development and Change*. Vol. 50. No. 3. P. 735–762.

Bédécarrats F., Guérin I., Roubaud F. (2020). *Editors Introduction. Controversies around RCTs* // *RCT in Development: A Critical Perspective* / F. Bédécarrats, I. Guérin, F. Roubaud (eds). Oxford: Oxford University Press. P. 1–28.

Bernard T., Delarue J., Naudet J.-D. (2012). *Impact Evaluations: A Tool for Accountability? Lessons from Experience at Agence Française de Développement* // *Journal of Development Effectiveness*. Vol. 4. No. 2. P. 314–327.

Bold T., Kimenyi M., Mwabu G., Ng’ang’a A., Sandefur J. (2013). *Scaling Up What Works: Experimental Evidence on External Validity in Kenyan Education*. CGD Working Paper No. 321. Washington: Center for Global Development.

Broadbent A., Vandenbroucke J. P., Pearce N. (2017). *Formalism or Pluralism? A Reply to Commentaries on Causality and Causal Inference in Epidemiology* // *International Journal of Epidemiology*. P. 1–12.

Cohen J., Dupas P. (2010). *Free Distribution or Cost-Sharing? Evidence from a Randomized Malaria Prevention Experiment* // *The Quarterly Journal of Economics*. Vol. 125. No. 1. P. 1–45.

Cohen J., Easterly W. (2010). *What Works in Development?: Thinking Big and Thinking Small*. Washington: Brookings Institution Press.

Cook T. (2018). *Twenty-six Assumptions That Have to Be Met if Single Random Assignment Experiments Are to Warrant ‘Gold Standard’ Status: A Commentary on Deaton and Cartwright* // *Social Science and Medicine*. Vol. 210. P. 37–40.

Davey C., Aiken A. M., Hayes R. J., Hargreaves J. R. (2015). *Re-analysis of Health and Educational Impacts of a School-Based Deworming Programme*

in Western Kenya: A Statistical Replication of a Cluster Quasi-Randomized Stepped-Wedge Trial // *International Journal of Epidemiology*. Vol. 44. No. 5. P. 1581–1592.

de Cler A.-P. (2021). *The Rise and Fall of a Gold Standard The Case of Randomized Controlled Trials within the Experimentation for Youth Fund*. Master's Thesis. Paris: SciencesPo. School of Public Affairs.

Deaton A. (2006). Evidence-Based Aid Must not Become the Latest in a Long String of Development Fads // *Boston Review*. Vol. 31. No. 4. P. 13–14.

Deaton A. (2010). Instruments, Randomization, and Learning about Development // *Journal of economic literature*. Vol. 48. No. 2. P. 424–455.

Deaton A. (2020). *Randomization in the Tropics Revisited: A Theme and Eleven Variations / RCT in Development: A Critical Perspective*. Oxford: Oxford University Press. P. 29–46.

Deaton A., Cartwright N. (2018). Understanding and Misunderstanding Randomized Controlled Trials // *Social Science and Medicine / F. Bédécarrats, I. Guérin, F. Roubaud (eds)*. Vol. 210. P. 2–21.

Duflo E. (2004). *Evaluating the Impact of Development Aid Program: The Role of Randomized Evaluations*. Paper presented at the 2-nd the AFD-EUND Conference. AFD-EUND: Paris.

Duflo E. (2017). The Economist as Plumber // *American Economic Review*. Vol. 107. No. 5. P. 1–26.

Duflo E., Kramer M. (2008). *Use of Randomization in the Evaluation Development Effectiveness // Reinventing Foreign Aid / W. R. Easterly (ed.)*. Cambridge, MA: MIT Press.

Duflo E., Hanna R., Ryan S. P. (2012). Incentives Work: Getting Teachers to Come to School // *American Economic Review*. Vol. 102. No. 4. P. 1241–1278.

Duflo E., Dupas P., Kremer M. (2015). School Governance, Teacher Incentives, and Pupil–Teacher Ratios: Experimental Evidence from Kenyan Primary Schools // *Journal of Public Economics*. Vol. 123. P. 92–110.

Dupas P. (2014). Short-Run Subsidies and Long-Run Adoption of New Health Products: Evidence from a Field Experiment // *Econometrica*. Vol. 82. No. 1. P. 197–228.

Easterly W. (2009). Development Experiments: Ethical? Feasible? Useful? (<https://www.nyudri.org/aidwatcharchive/2009/07/development-experiments-ethical-feasible-useful>)

Eble A., Boone P., Elbourne D. (2014). Risk and Evidence of Bias in Randomized Controlled Trials in Economics. Unpublished Working Paper. Providence, RI: Brown University.

Favereau J. (2016). On the Analogy between Field Experiments in Economics and Clinical Trials in Medicine // *Journal of Economic Methodology*. Vol. 23. No. 2. P. 203–222.

Favereau J., Nagatsu M. (2020). Holding Back from Theory: Limits and Methodological Alternatives of Randomized Field Experiments in Development Economics // *Journal of Economic Methodology*. Vol. 27. No. 3. P. 191–211.

Hammer J. (2017). Randomized Control Trials for Development? Three Problems // *Brookings Institution Blog Post*, May 11.

Heckman J. J. (2020). Epilogue: Randomization and Social Policy Evaluation Revisited // *RCT in Development: A Critical Perspective* / F. Bédécarrats, I. Guérin, F. Roubaud (eds). Oxford: Oxford University Press. P. 304–330.

Imbens G. (2010). Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009) // *Journal of Economic Literature*. Vol. 48. No. 2. P. 399–423.

Ioannidis J. (2018). Randomized Controlled Trials: Often Flawed, Mostly Useless, Clearly Indispensable: A Commentary on Deaton and Cartwright // *Social Science Medicine*. Vol. 210. P. 53–56.

Jerven M. (2015). *Africa: Why Economists Get It Wrong (African Arguments)*. L.: ZedBooks.

Kenny C. P., Pritchett L. (2013). *Promoting Millennium Development Ideals: The Risks of Defining Development Down*. Center for Global Development, Working Paper.

Labrousse A. (2020). The Rhetorical Superiority of Poor Economics // *RCT in Development: A Critical Perspective* / F. Bédécarrats, I. Guérin, F. Roubaud (eds). Oxford: Oxford University Press. P. 227–255.

Lancet (2004). The World Bank is Finally Embracing Science // *The Lancet*. Vol. 364. P. 731–732.

Leigh A. (2018). *Randomistas: How Radical Researchers Are Changing Our World*. New Haven and London: Yale University Press.

Miguel E., Kremer M. (2004). Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities // *Econometrica*. Vol. 72. No. 1. P. 159–217.

Ogden T. (2017). *Experimental Conversations: Perspectives on Randomized Trials in Development Economics*. Cambridge, MA: MIT Press.

Ogden T. N. (2020). RCTs in Development Economics, Their Critics, and Their Evolution // *RCT in Development: A Critical Perspective* / F. Bédécarrats, I. Guérin, F. Roubaud (eds). Oxford: Oxford University Press. P. 126–151.

Picciotto R. (2020). Are the Randomistas' Evaluators? // *RCT in Development: A Critical Perspective* / F. Bédécarrats, I. Guérin, F. Roubaud (eds). Oxford: Oxford University Press. P. 256–279.

Pritchett L. (2014). An Homage to the Randomistas on the Occasion of the J-PAL 10th Anniversary: Development as a Faith-Based Activity. Center for Global Development. (<https://www.cgdev.org/blog/homage-randomistas-occasion-j-pal-10th-anniversary-development-faith-based-activity>)

Pritchett L. (2020). Randomizing Development: Method or Madness? // *RCT in Development: A Critical Perspective* / F. Bédécarrats, I. Guérin, F. Roubaud (eds). Oxford: Oxford University Press. P. 79–107.

Pritchett L., Sandefur J. (2013). Context Matters for Size: Why External Validity Claims and Development Practice Don't Mix. Center for Global Development Working Paper No. 336.

Ravallion M. (2020). Should the Randomistas (Continue to) Rule // *RCT in Development: A Critical Perspective* / F. Bédécarrats, I. Guérin, F. Roubaud (eds). Oxford: Oxford University Press. P. 47–78.

Rodrik D. (2009). The New Development Economics: We Shall Experiment, but How Shall We Learn? // *What Works in Development?: Thinking Big and Thinking Small* / J. Cohen, W. Easterly (eds). Washington: The Brookings Institution. P. 24–47.

The Royal Swedish Academy of Sciences. (2019). Press release: The Prize in Economic Sciences 2019. (<https://www.nobelprize.org/prizes/economic-sciences/2019/advanced-information/>)

Vass M. (2010). Prevention of Functional Decline in Older People: The Danish Randomised Intervention Trial on Preventative Home Visits. Doctoral Dissertation. Copenhagen: University of Copenhagen. Faculty of Health Science.

Vivalt E. (2020). How Much Can We Generalize From Impact Evaluations? // Journal of the European Economic Association. Vol. 18. No. 4. P. 3045–3089.

Woolcock M. (2013). Using Case Studies to Explore the External Validity of ‘Complex Development Interventions’ // Evaluation. Vol. 19. No. 3. P. 229–248.

Ziliak S. T., Teather-Posadas E. R. (2016). The Unprincipled Randomization Principle in Economics and Medicine // Oxford Handbook on Professional Economic Ethics / G. DeMartino, D. McCloskey (eds). Oxford: Oxford University Press.

Kapeliushnikov, R. I.

Randomistas: A new development economics [Text] : Working paper WP3/2022/07 / R. Kapeliushnikov ; National Research University Higher School of Economics. – Moscow : HSE Publishing House, 2022. – 60 p. – (Series WP3 “Labour Markets in Transition”). – 35 copies. (In Russian)

The paper explores a curious phenomenon in the modern economic science — the intellectual confrontation between proponents of the “new” development economics (“randomistas”) and its opponents. The general message of the new approach is that the technique of randomized controlled experiments (RCT) should be considered the only truly scientific method and only this technique should be used in studying the problems of developing countries. RCTs have been recognized as the “gold standard” in evaluating the effectiveness of anti-poverty programs; today this approach absolutely dominates in development economics. The paper discusses the main features of the RCT economics: an inclination to imitate biomedical research; atheoretical character; the idea of a hierarchy of methods, at the top of which is the RCT; refusal of any other methods besides experimental ones; a focus on identification of causal effects without trying to discover what mechanisms underlie them; the claim that, in aid to developing countries, only RCTs allow to separate “what works” from “what doesn’t work”; reorientation from large-scale macroeconomic and institutional reforms to targeted social and humanitarian interventions. The problems of internal and external validity of RCT are highlighted; a special section is devoted to politico-economic aspects of the randomist practical agenda. A general conclusion is that conceptually counter-arguments of critics look more convincing: the idea of a hierarchy of methods is unscientific; no methodological “gold standard” exists in economic analysis; randomization under real field conditions should not be confused with randomization under hypothetical ideal conditions; the internal validity of the RCT is far from perfect, so that their estimates cannot be considered unbiased; RCTs are completely powerless over the issue of external validity; RCTs are extremely expensive and face a lot of ethical issues; due to low external validity, policy recommendations derived from RCTs are of very limited practical value. However, logic of critics has been trumped by rhetoric of randomistas: the reputation of the RCT economics for the scientific community, politicians, media and general public remains unshakable. Most likely, randomized experiments will escalate, and their intellectual and political influence grow.

Keywords: randomistas, development economics, internal validity, external validity, RCTs

JEL: A12, B41, C93, I32, O20

Препринт WP3/2022/07
Серия WP3
Проблемы рынка труда

Капелюшников Ростислав Исаакович

Рандомисты: новая экономика развития

Публикуется в авторской редакции

Отпечатано в типографии
Национального исследовательского университета
«Высшая школа экономики» с представленного оригинал-макета
Формат 60×84 1/16. Тираж 35 экз. Уч.-изд. л. 3,3.
Усл. печ. л. 3,5. Изд. № 2690. Заказ №

Национальный исследовательский университет
«Высшая школа экономики»
125319, Москва, Измайловское шоссе, 44, стр. 2
Типография Национального исследовательского университета
«Высшая школа экономики»