

## Рандомизированные контролируемые исследования – золотой стандарт или иллюзия? Роль экспериментальных методов оценки социальной эффективности в некоммерческом секторе

Салли Капитт (Sally Cupitt), Глава Службы оценки для благотворительных организаций Национального совета добровольных объединений (The NCVO Charities Evaluation Services; Великобритания) Май 2015 г.

### 1. Введение

Демонстрация социального эффекта (impact) является сложной задачей для некоммерческих организаций (НКО) и добровольных объединений местного уровня. Как мы можем доказать, что желаемые изменения действительно произошли, и что они напрямую связаны с нашими собственными программами (вмешательствами)?

В медицинских учреждениях «рандомизированные контролируемые исследования» (РКИ/RCTs - randomised controlled trials) являются общепринятым способом определения эффективности (или неэффективности) лечения. РКИ также используются для оценки социального эффекта образовательных и социальных инициатив, включая самые разнообразные варианты – начиная от схем микрофинансирования, и заканчивая эффектом от введения школьного «дресс-кода». В данной статье рассматривается применение РКИ в учреждениях НКО-сектора.

Наш сектор прошел долгий путь в сфере оценки социальных результатов/изменений (outcomes) и устойчивых эффектов (impact). Двадцать пять лет назад консультанты-оценщики Службы оценки для НКО (CES - Charities Evaluation Services) часто встречали людей, которые просто не видели никакой пользы от оценки, не говоря уже о подтверждении социальных результатов. Сегодня оценочная практика (по большей части) в аргументах не нуждается, и акцент дискуссии переместился на вопросы повышения стандартов оценки/доказательной информации и соответствующих методологий, способных откликнуться на этот запрос.

Некоторые НКО делают первые практические шаги на пути к пониманию социальных изменений и эффектов, а другие организации сектора уже существенно продвинулись на этой ниве, разработав сложные подходы «широкого назначения». Эти институты уже более восприимчивы к таким техническим решениям, как «социальный возврат на инвестиции» (Social Return on Investment) и РКИ.

Question to ask before opting for an RCT...

What's your treatment?



We don't have one,  
we just heard RCTs were  
the gold standard.



freshspectrum.com

Главное преимущество РКИ заключается в том, что они дают возможность для изучения эффектов вмешательств (инициатив), позволяя избежать распространенных видов погрешностей. Их часто называют «золотым стандартом» научных исследований. Однако РКИ тоже не лишены некоторых проблем, которые чаще всего проявляются в ходе тестирования комплексных программ в сложных социальных ситуациях. Это означает, что на этапе разработки плана (дизайна) РКИ для организаций НКО-сектора возникает ряд трудностей, которые мы обсуждаем в этой статье.

## 2. Что такое рандомизированное контролируемое исследование?

### 2.1. Основные понятия

Чтобы объяснить, что такое РКИ, сначала необходимо разъяснить некоторые основные понятия:

- **Социальный эффект или воздействие на общество (Impact).** Существует много толкований данного понятия, но в контексте РКИ социальный эффект, как правило, определяется как состоявшееся устойчивое изменение за минусом того, что могло бы произойти при условии отсутствия интервенции (вмешательства). Изменение может включать как отрицательные, так и положительные эффекты.
- **Вмешательство (Intervention).** Под вмешательством (или интервенцией) мы понимаем то, что подвергается тестированию. В медицинских исследованиях это может быть лекарство или препарат для инъекций, либо что-то другое – например, обучение или терапия.
- **Погрешность (Bias).** Это ошибка, которая может возникнуть в том случае, если подход к проведению исследования влияет на достоверность выводов. В силу погрешности исследователи могут переоценивать или недооценивать эффекты вмешательства. Например, если сведения о социальных результатах собраны исключительно посредством онлайн опроса, то люди, не имеющие доступа в интернет, в аудиторию респондентов не входят. Пользователи интернета могут во многом отличаться от «не-пользователей», и эти отличия могут привести к погрешностям в выводах исследования.
- **Контрольная группа (Control group).** «Контролируемый» компонент понятия «РКИ» означает, что участников инициативы (испытуемых) сравнивают со второй группой, которая называется контрольной группой людей, не вовлеченных во вмешательство. Качественный дизайн РКИ означает, что различия между экспериментальной и контрольной группами сведены к минимуму. Рандомизация – основной способ решения этой задачи в рамках РКИ.
- **Рандомизация (Randomisation).** Произвольный отбор – когда люди распределяются по группам, получающим или не получающим вмешательство, «случайным» образом ('at random'). Представьте эксперимент, в котором люди из двух городов получают разные вмешательства. Любые различия между людьми из двух городов могут привести к результатам с погрешностью. Например, если люди из одного города будут старше среднего возраста, то могут возникнуть трудности с разделением возрастных эффектов и изучаемых эффектов интервенции. Однако, распределяя людей из двух городов для двух вмешательств методом случайного отбора, мы можем снизить возрастные (и любые другие) различия до среднего уровня. Это дает нам уверенность в том, что любые наблюдаемые нами эффекты произошли благодаря данному вмешательству.

### 2.1. Как работает РКИ

РКИ представляют собой разновидность экспериментальных исследований. Изначально они преимущественно использовались в сфере медицины, и в последнее время обрели популярность в качестве способа тестирования социальных программ и одной из форм оценки социального эффекта.

РКИ работают посредством сравнения изменений, произошедших в группе участников вмешательства, с результатами людей, которые воздействию не подвергались. При правильном

подходе РКИ способны генерировать убедительные доказательства того, что вследствие вмешательства произошли либо не произошли изменения. Однако для проведения РКИ необходимы некоторые «технологические» навыки. Мы не приводим здесь их полного описания, но, в первую очередь, предлагаем ознакомиться с ключевыми аспектами РКИ.

В самом начале необходимо составить описание изучаемого вмешательства, и, во многих случаях, разработать «теорию изменений» (theory of change). Затем, в процессе формирования выборок, следует распределить людей (или, в некоторых случаях - организации) в одну из двух групп (зачастую с привлечением третьей независимой стороны):

1. в «экспериментальную группу» (в которую входят участники вмешательства; иногда их называют группой испытуемых) и
2. в «контрольную группу» (в которую входят те, кто не получал воздействия).

РКИ помогает сконструировать то, что называется «контрфактуальным» ('counterfactual') – оценить то, что могло бы произойти с клиентами при условии отсутствия интервенции. Благодаря этому изменения в экспериментальной группе можно сравнивать с результатами контрольной группы и, затем, анализировать отличия.

Две группы должны быть настолько похожими, насколько это возможно – чтобы единственным отличием между ними было наличие интервенции. Тогда любое другое отличие можно будет расценивать как следствие вмешательства. Тот факт, что группы формируются произвольным образом, является ключевым моментом методологии, позволяющим снижать вероятность отличий между двумя группами и, соответственно, возможные погрешности.

Для снижения вероятности погрешностей рандомизация может осуществляться «вслепую». Это когда участники и практикующие специалисты не знают о том, в какую группу распределится тот или иной человек. Очевидно, что так намного проще действовать в медицинском учреждении (где вместо настоящих лекарственных препаратов можно выдавать плацебо), чем в рамках социальной инициативы. Если вы, например, тестируете обучающую программу, то вы не скроете от участников факт получения либо неполучения образовательной услуги. В подобных форматах слепой метод более применим к людям, которые занимаются анализом данных, т.к. они могут не знать о том, кто получал вмешательство.

После завершения интервенции социальные результаты измеряются в каждой группе; зачастую замеры также делаются до начала вмешательства с целью получения исходных данных (baseline). Несмотря на то что РКИ можно провести без исходных данных, для комплексных социальных программ такие сведения имеют большое значение – не только потому, что они облегчают анализ в подгруппах, но также потому, что с их помощью можно оценить изменения в сравнении с различными стартовыми параметрами разных подгрупп.

Зачастую отличия между экспериментальной и контрольной группами весьма незначительны и случайны. Следует обращать внимание на различия между группами, которые отмечены как «статистически значимые». Эта формулировка означает, что разница настолько велика, что изменение не могло произойти само по себе и, соответственно – с высокой степенью вероятности – является эффектом вмешательства.

### 2.3. Квазиэкспериментальные планы исследования

Если в рамках оценки произвольное распределение не используется, но при этом сравниваются группы, сформированные без применения принципа случайной выборки (например, школа может сопоставлять результаты двух разных классов), то такая оценка называется «квазиэкспериментальной», и обычно считается менее надежным способом выявления

истинных причин изменения. Потому что вероятность существенных различий между группами достаточно велика, и результаты людей, которые сделали выбор в пользу участия в инициативе, могут оказаться лучше – в силу того, что их мотивация к переменам была более значительной.

Существует целый ряд разнообразных квазиэкспериментальных планов (дизайнов) исследования. Оценщики, в частности, могут отдать предпочтение сравнительной группе, которая имеет значительные сходства с экспериментальной группой – например, другой группе молодых людей-участников программы того же возраста с аналогичным гендерным составом. Чтобы скорректировать сравнительную группу, устранить некоторые различия и приблизить ее к группе испытуемых, иногда используются статистические методы. Одним из наиболее знакомых квазиэкспериментальных планов является изучение входных и выходных данных (pre-post study), когда сбор сведений от клиентов происходит на начальном и завершающем этапах вмешательства, и, затем, результаты сопоставляются. Несмотря на то что такой способ сбора данных весьма полезен, он считается наименее строгим среди других квазиэкспериментальных методологий. Его можно усилить за счет сбора информации по ряду позиций на протяжении всего вмешательства.

### 3. Почему сегодня РКИ воспринимаются как новшество?

Первый план оценки, разработанный в Соединенном Королевстве, изначально был весьма близок к научной модели, ориентированной на точность измерений и объективность выводов. Однако в 1980-1990-х годах многие оценщики отдали предпочтение «качественным» подходам (qualitative approaches), акцентирующим внимание на описании и интерпретации многочисленных перспектив. Они в большей мере отвечали духу НКО-сектора и больше подходили для развития культуры самооценки, требующей меньших технологических знаний.

Тем не менее, интерес к научно-техническим моделям оценки (включая РКИ) недавно возродился вновь. Возникли опасения, что методы, применяемые организациями НКО-сектора, не отвечают критерию научной достоверности, необходимой для установления причинно-следственных связей между социальными результатами и вмешательствами и, соответственно, для подтверждения того, что изменения являются следствиями деятельности НКО или оказания услуг.

Наши коллеги в сфере международного развития тоже осваивали экспериментальные методологии в целях решения проблемы распознавания причин и следствий. В частности, на веб-сайте Лаборатории по изучению проблем бедности им. Абдула Латифа Джамилы (Poverty Action Lab; США) публикуется значительное число примеров РКИ, проведенных в рамках программ развития, зачастую масштабных и значимых, но оценивающих предельно простые интервенции – такие как распространение антимоскитных сеток в целях снижения заболеваемости малярией. В целом мы обнаружили очень мало сведений, указывающих на то, что британские международные НКО используют РКИ достаточно активно.

Правительство воспринимает сведения, полученные в ходе РКИ, как образцовую практику оценки социальной эффективности, например – эффекта «облигаций, содействующих преобразованию общества» (Social Impact Bonds). Правительство также объединяет результаты РКИ по различным направлениям социально-экономической политики. В Третьем секторе Фонд Большой Лотереи (The Big Lottery Fund) финансирует программу, демонстрирующую применение экспериментальных методов для оценки инициатив в сфере профилактики молодежной преступности. А в Лондоне Проект Oracle (Project Oracle), формирующий пул достоверных данных о детских и молодежных проектах, придает большое значение экспериментальным методам, расценивая их как важный компонент высоких доказательных стандартов.

Следует признать, что для британского НКО-сектора и даже для сферы международного развития РКИ пока являются, скорее, исключением, чем правилом. Мы не нашли убедительных подтверждений тому, что НКО часто используют РКИ, а те примеры, которые мы обнаружили, оказались не особо успешными. Эти изыскания созвучны с последним Обобщающим анализом, проведенным Проектом Oracle ([Project Oracle Synthesis Study](#)). Оно показало, что в Лондоне было зафиксировано всего лишь 11 РКИ с участием молодежи; и эти исследования, преимущественно, проводились на базе школ. А в НКО-секторе было поддержано всего лишь несколько подобных инициатив.

## 4. Трудности для НКО-сектора

РКИ являются признанной действенной методологией, предлагающей один из лучших способов получения доказательств или максимального приближения к подтверждению эффективности или неэффективности программ. Однако существует ряд трудностей, связанных с использованием РКИ в НКО-секторе:

1. Масштабы и сроки
2. Технологический навык
3. Этические аспекты
4. Контроль различий между группами
5. Природа вмешательств
6. Обобщение
7. Потребность в других подходах к оценке, помимо РКИ.

### 4.1. Масштабы и сроки

#### **Затраты**

РКИ, организованные по всем правилам, являются дорогостоящим занятием. И не всегда есть уверенность в том, что полученная польза будет сопоставима с затратами - особенно если вспомнить о том, что результаты оценки – всего лишь один из факторов, влияющих на решения о финансировании (или не финансировании) программ (интервенций). Квазиэкспериментальные методы нередко обходятся дешевле и могут послужить источником существенной информации для принятия решений. Помимо этого, РКИ вряд ли заменят «внутреннюю самооценку» (internal self-evaluation), которую уже проводят компетентные НКО – скорее всего, они могут стать дополнением к тому, что есть.

#### **Большой количественный состав**

Крупные выборки дают более точные результаты и позволяют демонстрировать даже неочевидные (гибкие) эффекты. Для многих НКО масштаб вмешательства – возможно в силу бюджетных ограничений – может означать, что для проведения полноценного РКИ число клиентов может оказаться недостаточным.

#### **Фактор времени и сроки**

Как и многие другие методы оценки, РКИ тоже сталкиваются с проблемой фактора времени. Чтобы желаемые изменения проявились, должно пройти достаточно времени; и не все НКО могут получить необходимые результаты в период финансирования инициативы. Определение наиболее подходящего момента для сбора данных (не слишком поздно после завершения вмешательства, и не слишком рано - чтобы изменения успели произойти) является проблемным моментом для всех видов «оценки социальных результатов» (outcome evaluation).

Производство результатов в рамках РКИ может занять определенное время, что идет вразрез с заинтересованностью НКО в проведении оценки в режиме реального времени, когда результаты оценки поступают к поставщикам услуг на постоянной основе в виде обратной связи, позволяя вносить в работу оперативные изменения.

### 4.2. Технологический навык

РКИ – это сложное дело, поэтому маловероятно, что какая-либо НКО сможет организовать РКИ без внешней поддержки. А это, в свою очередь, тоже требует затрат. Например, некоммерческим институтам, желающим провести РКИ, нужен член команды, являющийся профессионалом в сфере статистики.

В силу очень высокой технологичности методологии, участие стейкхолдеров в РКИ, а также степень их ответственности за результаты оценки, могут оказаться менее значительными. Это то, что мы постоянно разъясняем с момента создания Службы CES в 1990 году.

### 4.3. Этические аспекты

Самая обсуждаемая проблема РКИ – особенно для НКО-сектора, ориентированного на создание социальных ценностей – лежит в этической плоскости. РКИ может потребовать отказа от проведения интервенции тем людям, которые в ней нуждаются. Более того, эти люди должны нуждаться во вмешательстве в той же степени, что и члены экспериментальной группы. При ином раскладе сходство между группами будет недостаточным. В связи с этим, специалисты, работающие в непосредственном контакте с клиентами, испытывают трудности.

Тем не менее, существует несколько вариантов урегулирования подобных вопросов. Контрольной группе, например, можно предложить интервенцию, но в более поздний срок. Еще один аргумент заключается в том, что польза от вмешательства пока неизвестна и, значит, отказа в предполагаемой пользе, по сути, нет. Также следует отметить, что трудности этического характера могут быть весьма ощутимыми – особенно в тех ситуациях, когда люди находятся в кризисе или подвергаются риску (например, в случаях жестокого обращения с детьми или домашнего насилия) - несмотря на то что обе группы, вовлекаемые в эксперимент, в той или иной мере получают основной комплекс услуг.

### 4.4. Контроль различий между группами

Основная трудность РКИ связана с контролем различий между экспериментальной и контрольной группами. Люди живут не в лабораторных условиях, и поэтому изучаемые интервенции, как правило, затрагивают целый комплекс переменных величин, влияющих на жизни людей.

Следует отметить, что работа с контрольной группой важна не меньше, чем взаимодействие с группой испытуемых. Например, получив отказ в получении интервенции, некоторые члены контрольной группы могут попытаться найти другой вид вмешательства в другом месте. И если они поступят подобным образом, то это может сильно отразиться на результатах исследования.

Представим себе РКИ, которое стремится выяснить, способствуют ли решению проблемы детского ожирения кулинарные курсы для родителей. Если некоторые родители обучаются в помещении школы, то после занятий они могут обмениваться рецептами со своими знакомыми. Значит, часть информации, озвученной на курсах, может оказаться достоянием контрольной группы. Вполне возможно, что такой обмен хорошо скажется на режиме детского питания, но не на результатах эксперимента.

#### **Отсев**

Особенностью большинства социальных программ является добровольное участие, что также означает, что люди могут выбывать из программы по своему усмотрению. И может случиться так, что те, кто выбывает (например, по причине менее значительной мотивации к переменам), отличаются от тех, кто остается. И это может нарушить баланс «случайности» в выборке исследования.

Некоторые РКИ рассматривают результаты только тех участников, которые полностью прошли интервенцию, но для большей надежности в обычной практике также учитываются результаты

выбывших людей. Такой подход можно считать оправданным, если с помощью результатов вы намерены обосновать целесообразность распространения интервенции в широких масштабах.

## 4.5. Суть вмешательства

Природа вмешательства, которое вы намереваетесь оценить, имеет значение при выборе плана оценки, включая РКИ.

### Природа вмешательства

Некоторые виды деятельности подходят для применения подхода, основанного на проведении РКИ. Простые, линейные интервенции, в которых изменения не подвергаются воздействию со стороны многих других переменных, нередко считаются особенно привлекательными. В отличие от них, для комплексных многогранных инициатив с участием многих субъектов более предпочтительной может оказаться оценка, которая нацелена на определение вклада интервенции в конкретное изменение, а не на поиск принадлежности результата конкретной инициативе.

### Контроль за проведением вмешательства



Представьте, что мы проводим РКИ для интервенции, которая предлагается по всей стране в разных регионах. Все эти услуги, в случае сравнения, должны быть одинаковыми во всех местах. Каждый, кто был когда-либо вовлечен в социальную программу, выполняемую на многих площадках, знает, насколько трудно гарантировать точное соответствие услуг согласованной модели. Например, практикующие специалисты нередко (и по понятным причинам) стремятся подстроить свою работу под местные потребности – и это может создавать проблемы при проведении РКИ.

### Стадия развития вмешательства

В ходе РКИ сама интервенция тоже должна быть подвергнута пристальному мониторингу – чтобы убедиться в том, что она осуществляется в соответствии с утвержденным стандартом. Это значит, что стадия развития проекта, подлежащего оценке, является важным моментом. Если проект пребывает в зачаточном состоянии, а изучаемое вмешательство не успело показать надежность своих методов, то проведение РКИ может оказаться преждевременным.

## 4.6. Обобщение

Существенная проблема РКИ заключается в том, что на основании результатов нескольких или отдельных исследований достаточно сложно делать обобщения. РКИ могут генерировать образцовые доказательства, подтверждающие эффективность вмешательства для конкретной группы людей в конкретном контексте и, благодаря этому, становятся убедительными аргументами в пользу продолжения выполненной работы. Их ограничение заключается в том, что они не могут сказать, будет ли та же самая интервенция эффективной в другом контексте, и применительно к другой группе людей. Если вас интересует эффективность вмешательства в самых разных контекстах с участием людей, которые в значительной мере отличаются друг от друга, то вам, возможно, потребуется провести более одного исследования.

#### 4.7. Потребность в других подходах к оценке, помимо РКИ

Для преодоления трудностей, связанных со спецификой РКИ, исследователи обычно сочетают РКИ с другими формами оценки – такими, например, как «оценка процессов» (process evaluation) – особенно в рамках комплексных программ.

В весьма полезном *Пособии по организации РКИ (How to guide on RCTs)*, разработанном Национальным исследовательским фондом сферы образования Англии и Уэльса ([The National Foundation for Educational Research in England and Wales](#)), говорится о том, что для проверки точности и качества оценки очень важно сначала оценить процессы, и только после этого переходить к формулировке допущений и толкованию результатов. Если способы проведения интервенции не идентичны, то группы не подлежат сравнению.

Для усиления РКИ и ограничения их проблемных моментов также можно использовать теорию программы и опираться на тщательно проработанную «теорию изменений» (theory of change).

## 5. Заключение

РКИ могут быть действенной методологией, но не всегда подходят для оценки устойчивых эффектов социальных программ. Их практическое применение в британском НКО-секторе в настоящий момент, как минимум, ограничено. Ресурсы и навыки, необходимые для проведения качественного РКИ, в совокупности с этическими проблемами и потребностью в дополнительных сведениях о контекстах и процессах (которые требуются для понимания результатов), могут ограничить их значимость для организаций Третьего сектора. Другие методы, которые не являются золотым стандартом, тоже могут генерировать полезные результаты. Поэтому очень важно помнить о том, что выбор дизайна оценки – это принятие решения о методах, подходящих для получения ответов на вопросы оценки, а также для формирования отчетливого представления о вмешательстве, которое предстоит изучить.

Дальнейшую дискуссию следует посвятить роли квазиэкспериментальных методов в практике оценки социальной эффективности (impact evaluation). Помимо этого, в более глубоком рассмотрении также нуждаются вопросы применения качественных методов оценки социального эффекта с целью проведения контрфактуального анализа. Более подробно об этом можно будет прочитать в статье Службы CES, которая выйдет в ближайшее время и будет посвящена использованию качественных подходов для изучения воздействия на общество.