

СТАНДАРТЫ ДОКАЗАТЕЛЬНОЙ ИНФОРМАЦИИ

Доказательство: то, что доказывает что-либо, или то, что генерирует подтверждающую информацию (Словарь Вебстера/Webster)

В отношении термина «программа с доказанной эффективностью» (Evidence-Based Programs) существует некоторое недопонимание. Оно вызвано тем, что разные реестры программ пользуются разными вариантами определения, а также тем, что у многих игроков сложилось неверное представление о минимальных требованиях, которые необходимо выполнить для надлежащего подтверждения эффективности программы. Чтобы внести ясность в отношении уровня доказательств, которому должны соответствовать инициативы, называемые программами с доказанной эффективностью, американская организация [Blueprints for Healthy Youth Development](#) («Программы воспитания жизнеспособного поколения молодежи») подготовила общий «портрет» Доказательного континуума ([Evidence Continuum](#)).



СЕРТИФИКАЦИЯ BLUEPRINTS

Соответствие самым высоким стандартам доказательной информации по итогам независимой экспертизы с участием выдающихся ученых национального уровня

СТАНДАРТЫ ДОКАЗАТЕЛЬНОЙ ИНФОРМАЦИИ

Доказанная эффективность	Доказательный континуум	Тип доказательства	Континуум доверия пользователей	Тип программы по классификации Blueprints
✓	Экспериментально подтвержденная эффективность (готовность к широкому внедрению)	Независимая репликация Многочисленные рандомизированные контролируемые исследования (РКИ)	Очень высокий уровень доверия	Программа «Модель плюс»
✓	Экспериментально подтвержденная эффективность (готовность к широкому внедрению)	Рандомизированные контролируемые исследования с репликацией	Высокий	Модельная программа
✓	Экспериментальные доказательства	Метод разрывной регрессии Анализ прерванных временных рядов Парное сравнение групп Корреляционное исследование Измерение результатов посредством начальных и заключительных тестов/опросов Оценка результатов посредством заключительного тестирования/опроса	Средний	Перспективная программа
	Доказательства, основанные на результатах исследований	Оценка удовлетворенности посредством опросов Индивидуальный опыт Отзывы / рекомендации специалистов	Низкий	
	Доказательства, основанные на изучении мнений		Очень низкий	

Перевод выполнен АНО «Эволюция и Филантропия» по заказу и в рамках программы «Семья и дети» Благотворительного фонда Елены и Геннадия Тимченко. Москва, 2016.

Доказательства эффективности программы, практики или политики представляют собой континуум, охватывающий пять уровней доверия - от самого низкого до самого высокого. Чем лучше (с научной точки зрения) проработаны дизайны оценочных исследований, чем выше число оценок с положительными выводами, и чем выше степень доверия пользователей, тем больше вероятность, что интервенция достигнет поставленной цели и окажет поддержку молодежи.

ДОКАЗАТЕЛЬСТВА, ОСНОВАННЫЕ НА ИЗУЧЕНИИ МНЕНИЙ

Доказательства, которые пользуются наименьшим доверием - это результаты оценок, основанных на изучении мнений ('opinion informed'). Они включают такую информацию, как факты, характеристики и/или индивидуальный опыт, приобретенный отдельными субъектами. Проведение опросов с целью оценки удовлетворенности лишь отчасти повышает планку доказательной информации, поскольку охватывает мнения только в пределах программы - даже если объем выборки достаточно велик. Данный тип подтверждений (несмотря на свою полезность на ранних этапах создания программы) не дает необходимых данных для системного изучения результатов целевой молодежной аудитории.

ДОКАЗАТЕЛЬСТВА, ОСНОВАННЫЕ НА РЕЗУЛЬТАТАХ ИССЛЕДОВАНИЙ

Доказательства, основанные на результатах исследований ('research-informed studies') - это сведения более высокого порядка, чем просто отзывы или мнения профессионалов. Для сбора информации о результатах/достижениях молодых людей проводятся опросы, изучаются материалы официальных органов и иные источники. Эти сведения в некоторой степени подтверждают эффективность программ, но уровень доверия к ним остается низким. Основная проблема заключается в том, что доказательства такого типа не «обособляют» социальный эффект (impact) инициатив от других возможных факторов, влияющих на результаты молодежной целевой аудитории.

Что касается корреляционных исследований (correlational studies), то они способны распознавать наличие или отсутствие связи между программой и желаемым результатом: это может быть положительная связь, отрицательная связь, или полное ее отсутствие. Тем не менее, идентификация связи еще не означает, что одно преобразование *послужило причиной* другого эффекта. Например, корреляционное исследование может показать, что участие молодежи в местной программе, нацеленной на предотвращение противозаконного поведения, имеет более значительное отношение к снижению показателя рецидивной преступности, чем коррекционная программа регионального уровня. Однако данный вывод еще не доказывает, что более низкий показатель рецидивизма образовался вследствие выполнения местной программы. Может оказаться, что судья направляет в коррекционную программу более серьезных нарушителей, а участники местной программы поступают из более благополучных домов и школ, либо входят в молодежные группы с более положительными моделями поведения. Так или иначе, но снижению рецидивизма может поспособствовать любая из этих программ. Корреляционные исследования не могут доказать, что именно эта (а не иная) программа стала причиной, повлиявшей на изменение наблюдаемого показателя.

Другие доказательства, основанные на результатах исследований, получают свидетельства эффективности посредством сбора данных в ходе опросов участников программы. Это может быть только заключительное тестирование (posttest), либо тестирование на начальном и заключительном этапах инициативы (pre-/post-tests). В силу того, что контрольная (или сравнительная) группа отсутствует, остается непонятным, в какой мере программе удалось (или не удалось) добиться тех изменений, которые были выявлены в ходе заключительного теста, либо произошли в период между начальным и итоговым тестированием. Иными словами, мы не можем связать имеющиеся результаты с данной программой, поскольку результаты могут возникнуть под воздействием других факторов.

Таким образом, доказательства, основанные на результатах исследований, не могут подтвердить наличие связи между интервенцией и полученным эффектом по причине отсутствия

Перевод выполнен АНО «Эволюция и Филантропия» по заказу и в рамках программы «Семья и дети» Благотворительного фонда Елены и Геннадия Тимченко. Москва, 2016.

соответствующей сравнительной группы. Исследовательская работа такого типа предлагает программам некоторую первичную поддержку, которая помогает обосновать потребность в более строгой экспериментальной оценке, но уровень доверия при этом остается низким.

ЭКСПЕРИМЕНТАЛЬНЫЕ ДОКАЗАТЕЛЬСТВА И ЭКСПЕРИМЕНТАЛЬНО ПОДТВЕРЖДЕННАЯ ЭФФЕКТИВНОСТЬ (ПРОГРАММЫ С ДОКАЗАННОЙ ЭФФЕКТИВНОСТЬЮ)

В верхнем сегменте континуума мы видим «экспериментальные доказательства» и «экспериментально подтвержденную эффективность» ('experimental' and 'experimentally proven' studies). Они являются атрибутами инициатив и практик, которые зачастую называются «программами с доказанной эффективностью» (EBPs). В процессе сертификации на получение статуса EBP практически все веб-реестры EBPs требуют от соискателей экспериментальных доказательств. Дизайн всех экспериментальных оценок включает использование сравнительных или контрольных групп. Если участники, получающие поддержку/услуги в рамках программы, показывают более высокие результаты по сравнению с членами контрольных групп, не вовлеченных в программу, то высока вероятность того, что программа действительно генерирует желаемый эффект (т.е., является причиной/движущей силой изменения). Однако уровни доверия и доказательств эффективности, характерные для экспериментальных оценок, могут варьироваться от среднего до самого высокого.

Средней степени доверия соответствует серия планов оценки, которую обычно называют квазиэкспериментальными дизайнами (QEDs - quasi-experimental designs). Три варианта, указанные в приведенной выше таблице, используются чаще всего. Этим дизайнам не свойственно применение метода случайной выборки, который является характерной чертой РКИ и дает уверенность в том, что группа, получающая вмешательство, и контрольная группа идентичны в начале исследования. На стартовом этапе QED сравнительные группы могут быть сопоставимыми по выявленным показателям, но при этом могут различаться по неуставленным характеристикам. Например, школы, осваивающие конкретную программу, могут быть сопоставимыми с другими школами по контрольным баллам (результатам проверки знаний) или по таким параметрам, как пол, социоэкономический и этнический состав участников. Однако может оказаться, что в школах, получающих вмешательство, число мотивированных сотрудников и учащихся выше, чем в других учреждениях. Из этого следует, что тест на эффективность может быть необъективным. Чем больше совпадений между группами, тем точнее может быть оценка программы. Однако уровень доверия все равно останется средним, поскольку сохраняется вероятность необъективных выводов.

Доказательства более высокого уровня появляются там, где проводятся рандомизированные контролируемые исследования, для которых случайное распределение участников в экспериментальную (получающую вмешательство) и контрольную группы является обязательным условием. Рандомизация гарантирует, что обе группы (контрольная группа и группа, получающая терапию/услуги/ поддержку) в начале исследования практически идентичны как по выявленным, так и по неуставленным параметрам. Выравнивание изначально существующих характеристик этих групп посредством рандомизации становится возможным только благодаря случайному выбору участников и предлагаемых испытаний. Значимость различий в результатах и возможностях можно оценить/исключить с помощью статистического теста.

Доверие самого высокого уровня проистекает из многократных РКИ, демонстрирующих полезность программы, наблюдаемой в разных выборках случайно назначенных объектов. Непротиворечивые выводы многих РКИ, полученные на разных площадках, существенно снижают вероятность того, что наблюдаемые различия между группой, подвергнутой вмешательству, и контрольной группой можно объяснить случайным стечением обстоятельств. Согласованные заключения многих РКИ, также повышают «обобщаемость» (generalizability) результатов исследований, равно как и вероятность того, что программа будет эффективной в случае проведения вмешательства на разных видах площадок и с участием разных типов объектов. Только программы с многократными РКИ могут рассчитывать на уровень доверия, необходимый для применения практики в широких масштабах.

Дополнительным фактором, способным повысить уровень доверия, могут быть РКИ, проведенные финансово незаинтересованными исследователями независимо от разработчика программы.

НАПРАВЛЕНИЕ ЭФФЕКТОВ

Направление эффектов тоже требует отдельного внимания. Исследования могут демонстрировать положительные эффекты, отсутствие эффектов и вредоносные эффекты. Заявление о том, что программа неэффективна или опасна, требует доказательств и доверия того же уровня, который предъявляется к подтверждениям эффективности программы.

Континуум доверия позволяет утверждать, что экспериментальные доказательства результатами исследований «высокого качества». Строгий процесс экспертной оценки программ, используемый командой Blueprints for Healthy Youth Development, отличает данный ресурс от других похожих реестров. В ходе внутреннего рассмотрения программ осуществляется предварительная идентификация практик, соответствующих установленным критериям. Затем, высококвалифицированный и авторитетный совет экспертов рассматривает документы и формулирует итоговые заключения, рекомендуя для сертификации только те программы, которые прошли экспериментальную проверку и продемонстрировали эффективность в ходе оценочных исследований «высокого качества». Модельные программы располагают результатами многих РКИ или одного РКИ в сочетании с одним квазиэкспериментальным дизайном. Как правило, такие программы обеспечены результатами РКИ. Перспективные программы предлагают, как минимум, результаты одного РКИ и двух квазиэкспериментов. Многие экспериментальные оценки не соответствуют стандартам качества Blueprints. Список самых распространенных недоработок, препятствующих сертификации программы на основе стандартов Blueprints, включают следующие проблемы: (а) отсутствие данных о том, что рандомизация обеспечила исходную эквивалентность экспериментальной и контрольной групп, (b) неспособность показать, что отток или убыль объектов не ставит рандомизацию под сомнение, (с) отсутствие сведений об использовании надлежащих статистических показателей, которые, например, позволяют привести объекты рандомизации в соответствие с необходимым уровнем (уровень школы и др.), и (d) неспособность обеспечить сопровождение и анализ всех объектов согласно исходному назначению испытаний (intent to treat). Не только тип дизайна имеет значение для Blueprints, но также общее качество исследования и его способность распознавать причинно-следственные связи.

Статья подготовлена организацией Blueprints for Healthy Youth Development в сотрудничестве с рабочей группой Blueprints Policy Group 11 ноября 2015 г:

<http://www.blueprintsprograms.com/standards-of-evidence>