

УДК 167.7

DOI: 10.15372/PS20240308

EDN VLICJW

А.М. Слесарева

**«ЗОЛОТОЙ СТАНДАРТ» КЛИНИЧЕСКИХ ИССЛЕДОВАНИЙ:
РАНДОМИЗАЦИЯ И КАУЗАЛЬНОСТЬ**

Статья открывает серию работ, посвященных концептуальному анализу эпистемологических характеристик представлений о «золотом стандарте» методологии медицинских исследований. Приведены результаты генезиса представлений о связи концепций рандомизации и каузальности, анализируются обоснования рандомизации как средства установления причинных отношений.

Ключевые слова: рандомизация; контролируемые клинические исследования; причинность

A.M. Slesareva

**“GOLD STANDARD” OF CLINICAL TRIALS:
RANDOMIZATION AND CAUSALITY**

This article begins a series of works on the conceptual analysis of the epistemological characteristics of ideas about the “gold standard” of medical research methodology. The results of the genesis of ideas concerning the relationship between the concept of randomization and that of causality are presented, and the grounds for randomization as a means of establishing causal relationships are analyzed.

Keywords: randomization; controlled clinical trials; causality

Успехи в развитии фундаментальной медицины, превосходящие порою самые смелые ожидания, и стремительное воплощение этих достижений в клинической практике привлекают к себе значительное внимание широкой публики. Как это часто случается, философское осмысление возможности такого роста знания, обоснованности и надежности провозглашаемых достижений либо запаздывает, либо заглушается восторженными приветствиями достижений и ожиданиями невероятного прогресса в ближайшем будущем. В те нечастые моменты, когда «кухонные разговоры» ученых и филосо-

фов проникают в поле внимания образованной публики, они обычно имеют статус «скандала», весьма редко приводя к конструктивной дискуссии. Например, ворвавшийся в «интеллектуальный дискурс» пару лет назад кризис невоспроизводимости эксперимента очень скоро превратился из философско-методологической проблемы в проблему доверия к экспертизе, проблему обоснованности политики научных журналов, даже в ревизию карьер отдельных ученых. Эпистемологическая составляющая феномена очень быстро практически полностью вытеснилась более скандальной – социальной.

Но и подобного рода редкие вспышки внимания к теоретико-методологической стороне развития медицинского знания не особо ставят под сомнение широко разделяемое и не оспариваемое мнение о том, что наша эпоха – это время невиданного прежде прогресса, и не в последнюю очередь такое мнение сделала возможной прочная методологическая основа клинических исследований. Вместе с тем сомнения в последнем утверждении иллюстрируются широко известной ироничной «притчей». Два медика летают на воздушном шаре над сельской местностью. Они дрейфуют по ветру и наслаждаются пейзажем, но через пару часов понимают, что заблудились. Они видят кого-то внизу на земле и кричат: «Привет! Можете сказать нам, где мы?» Человек на земле отвечает: «Вы на высоте пятидесяти футов на воздушном шаре». Один врач поворачивается к другому и говорит: «Этот человек, должно быть, статистик». «Откуда вы знаете?» – послышалась удивленная реплика с земли. «Только статистик может дать ответ, который был бы абсолютно точным и в то же время совершенно бесполезным». На что статистик отвечает: «Очень хорошо. Но я также могу сказать, что вы оба – врачи. Только врач может так хорошо видеть местность и при этом не иметь ни малейшего представления о том, где он находится».

Ситуация, когда признание безусловных успехов клинических исследований, причиной которых является надежная методология, сопровождается ироничным замечанием об основаниях нашей уверенности в этом знании, требует объяснения. По меньшей мере она достаточно необычная для того, чтобы привлечь к себе внимание. Ниже мы попытаемся показать, что такая ситуация, являясь большим чем казус, имеет интересное эпистемологическое содержание, которое может быть значимым для оценки обоснованности общепринятого представления о надежности методологических стандартов клинических исследований.

Уже не раз упомянув пресловутое «общее мнение» (мы настаиваем на том, что это не просто фигура речи, призванная снять необходимость подтверждать заявленное), подчеркнем, что оно стало неотъемлемой частью сложившейся практики клинических исследований, известной как доказательная медицина. Каноническая история возводит ее появление к 1747 г., к проведенному Дж. Линдоном изучению цинги, которое оказалось первым контролируемым клиническим исследованием. До 1900 г. было проведено несколько контролируемых клинических исследований [3, р. 665], а с начала XX в. и на всем его протяжении наблюдается утверждение доказательной медицины как образца обоснованности и достоверности медицинского знания. Свидетельством признания доказательной медицины в качестве максимально соответствующей стандартам знания может служить, например, признание международного стандарта оценки эффективности лекарственных средств. [3, р. 667].

В различных областях медицины рандомизированные контролируемые исследования распространены в разной степени. Так, в кардиологии и онкологии их количество намного больше, чем, например, в нефрологии, неврологии или эндокринологии [4]. Несмотря на это, такой вид исследований признается эталонным. Например, философ науки Дж. Уоррал заявляет: «Результаты двойных слепых рандомизированных контролируемых испытаний почти повсеместно рассматриваются как “золотой стандарт” доказательства в медицине» (цит. по: [5, р. 680]). Или известный специалист в области медицинской статистики Дж.В. Тьюки утверждает, что «почти единственным источником надежных доказательств [в медицине]... является то, что получено путем тщательно проведенных рандомизированных исследований» (цит. по: [2, р. 692]). В. Герберт, известный специалист в области исследований вирусов, передающихся через кровь, убеждает, что «единственным источником надежных доказательств, достигающих уровня доказательства полезности любой новой терапии, является только то, что получено в результате хорошо спланированных и тщательно проведенных рандомизированных и, где это возможно, двойных слепых клинических исследований. [Другие] исследования могут указывать направление, но не могут быть доказательствами в том смысле, в котором юристы используют термин “доказательства” для обозначения чего-то доказательного... [то есть] стремящегося доказать или фактически доказывающего» [2, р. 690].

Эти мнения специалистов из трех разных областей легко можно дополнить аналогичными утверждениями других ученых, в поле интереса которых попадает методология клинических исследований. Институциональным утверждением «золотого стандарта» клинических исследований стало движение в поддержку доказательной медицины, учрежденное в 1960 г. университетом Макмастера. Иными словами, доказательная медицина, основанная на контролируемых исследованиях, эталоном которых являются двойные слепые рандомизированные исследования, стала общепризнанным каноническим обоснованием и надежностью медицинского знания.

Поэтому редкие возражения, высказываемые не столько против самой методологии двойного слепого рандомизированного исследования, сколько против интерпретации такового как единственно надежного метода клинических исследований или против необоснованного приписывания ему абсолютной надежности и общезначимости, воспринимаются как скандальные. И если философам, таким как Н. Картрайт [1], сомнения в обоснованности восприятия эпистемологических достоинств рандомизированных клинических исследований простительны по причине специализации, то критические заявления клиницистов (например, заявления бывшего директора Центра по контролю и профилактике заболеваний в США Т. Фридена) воспринимаются как «скандал в благородном семействе».

Дж. Линдон понятия не имел о достоинствах рандомизации при проведении клинических исследований, когда совершил свое ставшее знаменитым достижение, понуждаемый не столько эпистемологическими или методологическими размышлениями, сколько стесненными условиями врачевания на корабле. Нынешние приверженцы рандомизации из числа клиницистов осознанно предпочитают полагать соответствующее понятие примитивным, интуитивно ясным. Смеем предположить, что познакомившись с историей возникновения и становления концепции рандомизации в связи с психофизиологическими исследованиями спиритизма, они были бы удивлены странной траекторией развития концепции, возникшей в столь сомнительной области и резвившейся до неотъемлемого элемента стандарта надежного вывода в клинических исследованиях.

В клинической практике рандомизация применяется как средство обнаружения причинных связей. Зачастую именно такой способ обнаружения причинности считается абсолютно надежным. Вместе с тем критика необоснованного, как полагает критикующая

сторона, абсолютизации рандомных контролируемых исследований как «золотого стандарта» методологии клинических исследований претендует на то, чтобы показать необоснованность возлагаемых на рандомизацию надежд именно как на средство выявления причинности. Одна из первых широко известных дискуссий, в ходе которой была поставлена под сомнение возможность выявления причинности в клинических исследованиях средствами рандомных контролируемых исследований, была инициирована П. Урбахом [6; 7], заявившим о неспособности рандомизированных исследований обеспечить проверку каузальных гипотез. Обсуждение выдвинутых им возражений было важной вехой в формулировании господствующей в медицинском сообществе точки зрения о рандомизации как средстве установления и проверки каузальных гипотез. В основании утвердившейся точки зрения лежат обнаруженное в ходе дискуссии различие между экспериментальной рандомизацией и случайной выборкой и та роль, которую играет рандомизация в выведении причинных связей из объективных вероятностей.

П. Урбах является сторонником байесовского подхода [7] и полагает, что случайность выборки не играет роли при выводе объективных вероятностей из выборочных данных. Поэтому его удивляет распространённый в медицинском сообществе энтузиазм в отношении рандомизации как средства выявления причинных связей. Одним из основных соображений, по которым Урбах отрицает способность рандомизации придавать большую объективность причинным выводам, является то, что решая, какие факторы следует «рандомизировать», экспериментатор неизбежно будет опираться на личные предположения о том, какие переменные могут влиять на выздоровление пациента.

Представим себе, что для лечения ранее неизлечимой болезни вводится новая терапия (Т) и оказывается, что вероятность выздоровления (R) в группе в целом выше у тех, кто получает Т, чем у тех, кто его не получает:

$$\text{Prob}(R/T) > \text{Prob}(R/\neg T).$$

Такая корреляция является *prima-facie* причиной полагать, что Т определяет R. Но, возможно, эта корреляция ложная: возможно, те, кто получает лечение, как правило, например, моложе (Y) и поэтому более склонны к выздоровлению в любом случае, а само ле-

чение не имеет отношения к выздоровлению. Тем не менее мы можем легко это проверить: мы можем рассматривать молодых и старых людей отдельно и наблюдать, коррелирует ли выздоровление с лечением в каждой группе:

$$\begin{aligned} \text{Prob}(R/T \text{ and } Y) &> \text{Prob}(R/-T \text{ and } Y), \\ \text{Prob}(R/T \text{ and } -Y) &> \text{Prob}(R/-T \text{ and } -Y). \end{aligned}$$

Если ни одно из этих неравенств не выполняется, т.е. если обнаружится, что T не оказывает вероятностного влияния на R ни среди молодых людей, ни среди пожилых, то мы можем заключить, что T не вызывает R и что исходная корреляция была связана с тем, что лечение было более успешным у молодых людей, которые в любом случае выздоравливают, чем у пожилых.

С другой стороны, если неравенства (2) выполняются, мы не можем заключить, что T действительно вызывает R . Возможно, какая-то другая сопутствующая причина ответственна за начальную корреляцию T - R (1). Возможно, люди, которые получают лечение, как правило, имеют более высокий уровень общего фонового здоровья (H) независимо от того, молоды они или нет, и по этой причине чаще выздоравливают. Мы можем это проверить: учитывая некоторый индекс общего фонового здоровья, мы можем рассматривать здоровых и нездоровых людей отдельно и наблюдать, имеет ли лечение значение в каждой группе. Если нет, то начальная корреляция T - R раскрывается как ложная и мы можем заключить, что T не вызывает R . С другой стороны, если лечение все еще имеет значение в каждой группе, то и т.д. Проверка всех возможных сопутствующих факторов, которые могут быть ответственны за начальную корреляцию T - R , оказывается нескончаемой, ведь мы можем постоянно вводить все новые переменные, обозначающие возможные факторы, определяющие выздоровление в группах.

Консенсус по этому поводу состоит в том, что проблема решается рандомизированным экспериментом. Берем выборку людей с интересующим нас заболеванием. Делим исходную группу на две случайным образом. Затем даем лечение одной группе, отказывая в нем другой, и на основе результатов определяем, выше ли вероятность выздоровления в одной из групп. Если это так, то мы имеем основания утверждать, что T является причиной R , поскольку рандомизация устранила возможность любых смешивающих факторов,

которые могут быть ответственны за ложную корреляцию. Экспериментальная рандомизация реализует функцию установления причинности после того, как мы сформировали суждения об объективных вероятностях и готовы определить, какие из них свидетельствуют о каузальных связях. Иными словами, рандомизированный эксперимент предполагает объективные вероятности предварительно установленными. Здесь мы имеем дело с предположением, что событие, такое как лечение T , является причиной события, такого как выздоровление R , если и только если существуют контексты (возможно, включающие другие неизвестные факторы), в которых T фиксирует выше среднего единичную объективную вероятность для R . По сути, это означает, что лечение вызывает выздоровление только в том случае, если есть некоторые типы пациентов, у которых лечение увеличивает шансы на выздоровление.

Как только мы сформулировали предположение о причинно-следственной связи, становится возможным понять, почему экспериментальная рандомизация имеет значение. Поскольку это предположение подразумевает, что если $\text{Prob}(R/T)$ больше, чем $\text{Prob}(R/-T)$, т.е. если объективная вероятность выздоровления для пациентов, получавших лечение, в целом выше, чем для тех, кто его не получал, то это либо потому, что T само по себе вызывает R , либо потому, что T коррелирует с одним или несколькими другими факторами, которые вызывают R . В первом случае корреляция T - R будет обусловлена, по крайней мере частично, тем фактом, что T само по себе фиксирует выше среднего единичную вероятность для R . Во втором случае корреляция T - R будет обусловлена тем фактом, что T коррелирует с другими причинами, даже если само T не оказывает непосредственного причинного воздействия. Проблема заключалась в том, что разница между $\text{Prob}(R/T)$ и $\text{Prob}(R/-T)$ в целом не позволяет провести различий между этими двумя возможностями, в частности она не позволяет нам исключить вторую возможность в пользу гипотезы о том, что T действительно вызывает R .

Именно рандомизированный эксперимент позволяет провести такое различие. Предположим, что $\text{Prob}(R/T)$ и $\text{Prob}(R/-T)$ представляют собой вероятностную разницу между двумя группами пациентов в рандомизированном эксперименте. Поскольку лечение было назначено случайным образом (в том смысле, что все пациенты независимо от их других характеристик имеют абсолютно одинаковую объективную вероятность получения лечения T), мы можем быть

уверены, что T само по себе объективно не коррелирует с какой-либо другой характеристикой, которая влияет на R . Таким образом, мы можем исключить возможность ложной корреляции и быть уверенными, что T действительно вызывает R .

Одной из главных причин, по которой П. Урбах отрицает способность рандомизации быть основанием для каузальных выводов, является то, что решение о том, какие факторы следует «рандомизировать», неизбежно будет зависеть от личных предположений экспериментатора относительно того, какие «лишние переменные» могут влиять на выздоровление пациента. В процессе обсуждения этого возражения были достигнут консенсус, согласно которому, рандомизированные эксперименты важны не потому, что они помогают справляться с «лишними переменными», о которых мы знаем, а потому, что они позволяют справляться со всеми теми, о которых мы не знаем. Если мы можем определить некоторую конкретную переменную N , которая может влиять на выздоровление, мы не нуждаемся в проведении рандомизированного эксперимента, здесь нам достаточно статистических вероятностей. Именно тогда, когда у нас нет никаких дополнительных идей о том, какие N обуславливают выздоровление, рандомизированные эксперименты оказываются востребованными.

В вопросе о том, является ли лечение T причиной R , мы можем выделить два аспекта. Во-первых, это вопрос о том, являются ли условные вероятности $\text{Prob}(R/T)$ и $\text{Prob}(R/-T)$ в базовом вероятностном пространстве неравными. Если это так, то возникает второй вопрос: является ли это следствием того, что T вызывает R ? Первый вопрос – это вопрос статистического вывода. Мы хотим из частот, с которыми R , T и $-T$ встречаются вместе в нашей выборке, сделать вывод о базовых вероятностях R при условии T , а не $-T$. Но второй вопрос остается открытым. Предположим, что мы приходим к выводу, что базовая вероятность разная при T и не- T . Каким бы способом мы ни пришли к этому выводу, мы все равно столкнемся с дополнительным вопросом: является ли эта разница следствием того, что T вызывает R ? Так как остается возможность того, что T объективно коррелирует с R , но не потому, что оно само вызывает R , а потому, что оно объективно коррелирует с чем-то другим, что вызывает выздоровление. Именно в этом случае становится востребованным рандомизированный эксперимент, поскольку он «переключит» базовое вероятностное пространство на такое, в котором

T не будет объективно коррелировать ни с какими другими причинами.

Таким образом, есть два совершенно разных способа, которыми случайность может присутствовать в рассматриваемом типе исследования. Во-первых, случайная выборка: общая выборка может быть составлена случайным образом, т.е. каждый человек может иметь одинаковую объективную вероятность попасть в общую выборку. Во-вторых, рандомизированное экспериментирование: лечение может быть назначено случайным образом, т.е. как бы ни была сформирована выборка, все выбранные люди могут иметь одинаковую объективную вероятность получить лечение. Даже если мы настаиваем, как сторонники классического подхода, на случайных выборках при обследовании неэкспериментальной группы для оценки основных вероятностей, это не означает, что лечение было назначено случайным образом (и, стало быть, не дает никаких оснований для заключения о том, что корреляция T-R указывает на причинно-следственную связь). И наоборот, даже сторонник байесовского подхода, который готов делать статистические выводы из неслучайных выборок, все еще имеет все основания требовать, чтобы лечение было назначено случайным образом (поскольку в рамках этого подхода нужна дополнительная информация, чтобы исходя из корреляций, сделать вывод о причинных связях).

На первый взгляд представление о случайном назначении лечения в пределах неслучайной выборки может показаться проблематичным. Но это типичная ситуация в медицинских исследованиях. Выборка пациентов, страдающих от заболевания, собирается относительно бессистемно. Но как только эта выборка сформирована, принимаются все меры для того, чтобы гарантировать, что все ее члены имеют одинаковую объективную вероятность получить лечение. Изначальная неслучайность выборки не является препятствием для ее использования с целью оценки объективных условных вероятностей. Но именно последующее рандомизированное назначение лечения имеет решающее значение для наших выводов о наличии причинно-следственной связи из этих условных вероятностей.

Обсуждение поставленных П. Урбахом проблем привело к формированию представлений о рандомизации как о надежном способе выявления причинно-следственных связей в клинических исследованиях, которые стали, по сути, концептуальным ядром интерпретации рандомизированного контролируемого исследования

в качестве «золотого стандарта» клинических исследований. Вместе с тем генерализация и абсолютизация эпистемологических преимуществ рандомизации привели к искаженному представлению о ее реальных эпистемологических характеристиках, которое, в частности, ведет к недооценке некоторых из этих характеристик и переоценке других. Именно эта «эпистемологическая девиация» станет предметом дальнейшего исследования.

Литература

1. Cowen N., Baljinder V., Mascarenhas-Keyes S., Cartwright N. Randomized controlled trials: How can we know “what works”? // *Critical Review*. 2017. Vol. 29. № 3, P. 265–292. DOI: 10.1080/08913811.2017.1395223.
2. Herbert V. Acquiring new information while retaining old ethics // *Science*. 1977. Vol. 198. P. 690–693.
3. Opinel A., Trohler U., Gluud C., Gachelin G., Smith G.D., Podolsky S.H., et al. Commentary: The evolution of methods to assess the effects of treatments, illustrated by the development of treatments for diphtheria, 1825–1918 // *International Journal of Epidemiology*. 2013. Vol. 42 (3). P. 662–676.
4. The Cochrane Collaboration. The Cochrane Library. 2024. URL: <http://www.thecochranelibrary.com/view/0/indexhtml> (дата обращения:)
5. Tukey J.W. Some thoughts on clinical trials, especially problems of multiplicity // *Science*. 1977. Vol. 198. P. 679–684.
6. Urbach P. Random sampling and methods of estimation // *Proceedings of the Aristotelian Society*. 1989. Vol. 89. P. 143–164.
7. Urbach P. Randomization and the design of experiments // *Philosophy of Science*. 1985. Vol. 52. P. 256–273.

References

1. Cowen, N., V. Baljinder, S. Mascarenhas-Keyes & N. Cartwright. (2017). Randomized controlled trials: How can we know “what works”? *Critical Review*, Vol. 29, No. 3, 265–292. DOI: 10.1080/08913811.2017.1395223.
2. Herbert, V. (1977). Acquiring new information while retaining old ethics. *Science*, 198, 690–693.
3. Opinel, A., U. Trohler, C. Gluud, G. Gachelin, G.D. Smith, S.H. Podolsky et al. (2013). Commentary: The evolution of methods to assess the effects of treatments, illustrated by the development of treatments for diphtheria, 1825–1918. *International Journal of Epidemiology*, 42 (3), 662–676.
4. The Cochrane Collaboration. (2024). The Cochrane Library. Available at: <http://www.thecochranelibrary.com/view/0/indexhtml> (date of access: 15.05.2024).
5. Tukey, J.W. (1977). Some thoughts on clinical trials, especially problems of multiplicity. *Science*, 198, 679–684.
6. Urbach, P. (1989). Random sampling and methods of estimation. *Proceedings of the Aristotelian Society*, 89, 143–164.

7. Urbach, P. (1985). Randomization and the design of experiments. *Philosophy of Science*, 52, 256–273.

Информация об авторе

Слесарева Алина Михайловна. Новосибирский государственный медицинский университет. (630091, Новосибирск, Красный проспект, 52).
alina.slesareva2018@yandex.ru

Information about the author

Slesareva, Alina Mikhailovna. Novosibirsk State Medical University. (630091, Novosibirsk, Krasny Prospekt, 52).

Дата поступления 05.06.2024