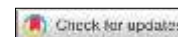


БИОИНФОРМАТИКА

BIOINFORMATICS



УДК 519.23
<https://doi.org/10.37661/1816-0301-2023-20-1-27-39>

Оригинальная статья
Original Paper

Оценка размера эффекта воздействия в квазиэкспериментальных исследованиях

О. В. Красько

*Объединенный институт проблем информатики
Национальной академии наук Беларуси,
ул. Сурганова, 6, Минск, 220012, Беларусь
E-mail: krasko@newman.bas-net.by*

Аннотация

Цели. Поставлена задача оценить размер эффекта в квазиэкспериментальных исследованиях.
Методы. Используются методы теории оценивания и методы математической статистики.
Результаты. Оценен размер эффекта на порядковой и бинарной шкалах в случае разнонаправленных эффектов в группах в квазиэкспериментальных исследованиях для аналитического метода «различие в различиях».
Заключение. В работе представлены подходы к оценке абсолютных и стандартизированных размеров эффектов в экспериментальных и квазиэкспериментальных исследованиях. Дан краткий обзор оценок абсолютных и стандартизированных размеров эффектов для количественных и бинарных переменных исследования. Рассмотрен практический подход к оценке размеров эффектов порядковой и бинарной переменных в случае разнонаправленных эффектов в группах в квазиэкспериментальных исследованиях для аналитического метода «различие в различиях». Приведен пример расчетов абсолютных и стандартизированных размеров эффектов количественной и бинарной переменных в квазиэкспериментальных исследованиях в клинической эпидемиологии.

Ключевые слова: экспериментальные исследования, квазиэкспериментальные исследования, абсолютный размер эффекта, стандартизированный размер эффекта, аналитический метод «различие в различиях»

Для цитирования. Красько, О. В. Оценка размера эффекта воздействия в квазиэкспериментальных исследованиях / О. В. Красько // Информатика. – 2023. – Т. 20, № 1. – С. 27–39.
<https://doi.org/10.37661/1816-0301-2023-20-1-27-39>

Конфликт интересов. Автор заявляет об отсутствии конфликта интересов.

Поступила в редакцию | Received 25.11.2022
Подписана в печать | Accepted 23.01.2023
Опубликована | Published 29.03.2023

Effect size assessment in quasi-experimental studies

Olga V. Krasko

*The United Institute of Informatics Problems
of the National Academy of Sciences of Belarus,
st. Surganova, 6, Minsk, 220012, Belarus
E-mail: krasko@newman.bas-net.by*

Abstract

Objectives. To estimate effect sizes in quasi-experimental studies.

Methods. Methods of the theory of estimation, methods of mathematical statistics.

Results. Estimation of the effect size on an ordinal scale, estimation of the effect size on a binary in the case of opposite direction effects in groups, in quasi-experimental studies for the analytical method "differences in differences".

Conclusion. The paper considers approaches to assessing absolute and standardized effect sizes in experimental and quasi-experimental studies. A brief review of the estimators of absolute and standardized effect sizes for quantitative and binary study variables is provided. The applied approach is proposed to assess the effect sizes of a binary variable in the case of opposite direction effects in groups within a quasi-experimental studies for the "differences in differences" analytical method. An example of assessment of absolute and standardized effect sizes of quantitative and binary variables in quasi-experimental studies in clinical epidemiology is considered.

Keywords: experimental study design, quasi-experimental study design, absolute effect size, standardized effect size, "difference-in-differences" analytical method

For citation. Krasko O. V. *Effect size assessment in quasi-experimental studies*. Informatika [Informatics], 2023, vol. 20, no. 1, pp. 27–39 (In Russ.). <https://doi.org/10.37661/1816-0301-2023-20-1-27-39>

Conflict of interest. The author declares of no conflict of interest.

Введение. Большие объемы медицинских данных предоставляют уникальные возможности для исторического анализа. Изучение баз медицинских данных может дать новые знания о диагностике и лечении различных заболеваний. На сегодняшний день изучение таких данных представляет собой отдельный аспект клинической эпидемиологии со своим дизайном и особенностями.

Поскольку данные историй болезней изначально не предназначались для исследования, а фиксировали историю лечения пациента, то использование их напрямую для научных исследований затруднительно из-за различных особенностей лечения, многочисленных конфаудеров и т. п. [1, 2]. Тем не менее существует возможность построения дизайна исследования и методов обработки данных для определения размера эффекта лечения или иного вмешательства для получения доказательств наличия или отсутствия эффектов.

В настоящей статье рассмотрены особенности анализа данных в квазиэкспериментальных исследованиях, дано понятие абсолютного и стандартизированного размера эффекта, на примере показаны простые расчеты размеров эффектов для непрерывных и бинарных откликов.

Особенности квазиэкспериментального дизайна исследований. Квазиэкспериментальный дизайн исследований имеет свою область применения, свои преимущества и недостатки [3–5]. Как и экспериментальный, квазиэкспериментальный дизайн исследования направлен на установление причинно-следственной связи между независимой и зависимой переменными. Однако в отличие от настоящего эксперимента при квазиэкспериментальном дизайне исследования не используется рандомизация. Вместо этого объекты исследования распределяются по группам на основе некоторых критериев, часто связанных с проявлением или течением заболевания. Квазиэкспериментальный дизайн исследования является полезным инструментом в ситуациях, когда настоящие эксперименты не могут быть использованы по этическим или практическим причинам.

Перечислим основные особенности квазиэкспериментального дизайна исследования в клинической эпидемиологии:

1. Назначение пациентов в группы лечения (или лечения и контроля) происходит неслучайным образом, в отличие от настоящего экспериментального исследования в нем отсутствует рандомизация.

2. Часто используются исторические данные, лечение не проектируется, а назначается исходя из экспертного мнения специалиста. Таким образом, исследователь может не разрабатывать лечение, а только анализировать исторические результаты.

3. Исследуются группы пациентов, которые не могли попасть в рандомизированные клинические испытания (РКИ) по причине ограничений в нормативных актах, например запрета проведения РКИ на беременных, детях и т. п.

Квазиэкспериментальные исследования могут иметь более низкую внутреннюю и более высокую внешнюю валидность, чем настоящие эксперименты, поскольку есть возможность использовать ситуации реального мира вместо искусственно созданных условий эксперимента [3]. Отсутствие рандомизации, а также неспособность контролировать с помощью дизайнера уровень воздействия изучаемого фактора и иных связанных с исследованием факторов делают данный вид исследования менее желательным для установления причинно-следственной связи между фактором риска и исходом. Однако он является более распространенным в клинической эпидемиологии, когда врач, занимающийся клинической врачебной практикой, переходит к научным исследованиям. Типичный пример – это исследования, которые планируются не заранее, а исходя из опыта врача-клинициста. Например, врач-психолог несколько лет ведет прием пациентов с повышенной тревожностью. Лечение назначается по стандартной методике. Исходом является состояние пациента спустя год после начала лечения. Накопив практический опыт, врач понимает, что методику можно улучшить. Следующие несколько лет новые пациенты проходят лечение по измененной методике, также фиксируется их состояние спустя год после начала лечения. Фактический материал, накопленный по двум группам (стандартного и экспериментального лечения), должен послужить доказательством, что измененная методика дает лучшие результаты лечения.

Описанный тип исследования хотелось бы назвать клиническими испытаниями (нерандомизированными), однако клинические испытания являются разновидностью когортных исследований, поэтому наиболее точное определение таких исследований – когортные исследования. Лечение по разным методикам будет рассматриваться как фактор, влияющий на исход заболевания. Исследование будет квазиэкспериментальным, и после учета всех влияющих на него факторов можно будет оценить эффект от изучаемого фактора (измененной методики лечения).

Таким образом, методы и приемы обработки и анализа данных в квазиэкспериментальном исследовании актуальны и востребованы в прикладных исследованиях.

Одним из распространенных аналитических методов квазиэкспериментального исследования является метод «различие в различиях» (difference-in-differences, DID) – аналитический подход, который облегчает вывод о причинно-следственной связи, даже когда рандомизация невозможна. В самом простом случае применения DID сравниваются изменения результатов с течением времени между двумя группами: популяцией, участвующей в эксперименте (экспериментальная группа), и популяцией, не участвующей в нем (группа сравнения).

РКИ и исследования по методу DID имеют некоторые общие характеристики: они включают четко определенную исследуемую популяцию и набор условий лечения, где легко отличить группу лечения от контрольной группы, а также периоды времени до лечения и после. Наиболее важным их различием является то, что условия воздействия случайным образом распределяются между группами в РКИ, но не в исследованиях по методу DID. При случайном назначении воздействие лечения статистически не зависит от любого (измеренного или неизмеренного) фактора, который также может повлиять на результаты. В исследованиях по методу DID невозможно полагаться на случайное распределение, чтобы избежать систематической ошибки из-за неизмеренных вмешивающихся факторов. Вместо этого вводятся предположения, которые ограничивают множество возможных вмешивающихся факторов. В частности, метод DID предполагает, что вмешивающиеся факторы, имеющие неоднородность в различных группах, не изменяются во времени, а изменяющиеся во времени вмешивающиеся факторы инвариантны относительно групп. Эти утверждения называются общим предположением о трендах [1, 3–5].

Рассмотрим метод DID для простейшего исследования, где анализируются две группы, экспериментальная и контрольная, до начала исследования и после его окончания.

Сделаем некоторые предположения. Пусть $g = 1, 2$ – две группы, которые наблюдаются в два периода времени $t = 1, 2$, а $Y(1)_{gt}$ представляет собой интересующий результат для группы g в период t при гипотетическом сценарии, в котором воздействие было активным в группе g в момент времени t . Величина $Y(0)_{gt}$ является результатом при той же единице времени при альтернативном сценарии отсутствия воздействия. Эффект воздействия для этой конкретной группы и периода времени определяется выражением $\Delta_{gt} = Y(1)_{gt} - Y(0)_{gt}$, которое представляет собой разницу в значениях переменных интересующего результата для одной и той же группы при двух гипотетических сценариях.

Пусть во втором периоде воздействие происходит в группе $g = 2$, а не группе $g = 1$ и фиктивная переменная $T_g = 1[g = 2]$ идентифицирует наблюдения в группе 2, $T_g = 0[g = 1]$. Переменная T_g не имеет индекса времени, поскольку членство в группах постоянно и не меняется во времени. Пусть фиктивная переменная $P_t = 1[t = 2]$ идентифицирует наблюдения из периода времени 2, а переменная $P_t = 0[t = 1]$ – из периода времени 1. Тогда переменная P_t не имеет группового индекса, поскольку периоды времени не различаются между группами. В таком исследовании переменная воздействия является произведением двух фиктивных переменных: $S_{gt} = T_g \times P_t$. Для гипотетического результата в каждой группе в каждый момент времени можем записать $Y_{gt} = Y(0)_{gt} + (Y(1)_{gt} - Y(0)_{gt}) \times S_{gt}$. Данное формальное представление описывает проблему контрфактического вывода, которая возникает в каждом исследовании причинно-следственного вывода. В исследовании имеется доступ к данным Y_{gt} и S_{gt} , и одна из его задач – установить среднее значение $D = Y(1)_{gt} - Y(0)_{gt}$ в рамках изучаемой когорты. Однако для группы без воздействия (контрольной) невозможно оценить размер эффекта воздействия, но возможен расчет, связанный с изменением результата во времени (D_k). Общее предположение о трендах в этом случае позволяет трактовать оценку изменения D_k как связанную с периодом наблюдения. Для группы с воздействием (экспериментальной) невозможно рассчитать чистый размер эффекта воздействия, но возможен расчет, связанный с изменением результата во времени, который ассоциирован с воздействием (D_s). Если предположить аддитивность эффектов, связанных с воздействием и временем, то $D = D_s - D_k$. Конкретная формула расчета для D_k и D_s в данном случае зависит от шкалы, в которой измеряется соответствующая переменная. Это может быть разница в пропорциях для бинарной переменной, разница в порядках для порядковой шкалы или разница в измерениях до и после исследования на некоторой физической шкале. Расстояние между D_k и D_s может служить оценкой размера эффекта при соблюдении предположения о параллельных трендах. Далее приводятся основные формулы для расчетов различий в группах в разные моменты времени (повторные измерения) и для различных групп в некоторый момент времени.

Метод DID отражает логику построения оценки размера эффекта, связанного с воздействием. Далее будет показан практический пример использования данного метода в квазиэкспериментальном исследовании. Полный теоретический подход, связанный с несколькими группами и периодами наблюдения, а также возможными конфаундерами, не рассматривается в настоящей статье, поскольку ее цель – показать, что исследования, не связанные с РКИ, могут представлять интерес для специалистов в предметной области при оценке размера эффекта, и привести пример простого исследования для освоения основ данного метода.

Понятие размера эффекта. Размер эффекта – это численная оценка изменений, связанных с воздействием (лечением) (d -семейство оценок), или численная оценка связи переменных в исследовании (r -семейство оценок). Соответственно этому определению оценки размера эф-

фекта делятся на два больших семейства: d -семейство оценок, рассчитывающее размер различий в форме абсолютной разницы, стандартизированной разницы или отношения, и r -семейство оценок, оценивающее тесноту связи между переменными.

Согласно работе [6] размер эффекта определяется как количественное отражение величины некоторого явления, которое используется для решения интересующей задачи. На сегодняшний день существует более сотни различных мер оценки размеров эффектов для разных типов переменных и для ответов на научные вопросы.

Далее рассматриваются основные меры d -семейства оценок размера эффекта как наиболее распространенные в рандомизированных контролируемых испытаниях и квазиэкспериментальных исследованиях.

Абсолютные размеры эффектов. В экспериментальных и квазиэкспериментальных исследованиях традиционно используется d -семейство оценок для демонстрации лечебного эффекта (воздействия). Рассматривается исход в экспериментальной группе и группе контроля. Различия оцениваются в виде точечной оценки 95 % доверительного интервала (ДИ). Однако нельзя не отметить, что наряду с оценкой абсолютных различий на основе ДИ оценка стандартизированного размера эффекта в последнее время приобретает популярность в научных публикациях.

Рассмотрим сначала оценки абсолютных различий для количественной и бинарной переменных, которые чаще всего используются для оценки исхода воздействия в экспериментальных и квазиэкспериментальных исследованиях (табл. 1). В квазиэкспериментальных исследованиях необходимо в обязательном порядке контролировать исходные состояния наблюдаемых в группах, поскольку они могут значительно различаться.

Таблица 1
Основные оценки абсолютных размеров эффектов в экспериментальных и квазиэкспериментальных исследованиях

Table 1

Main absolute effect sizes in experimental and quasi-experimental studies

Количественная переменная <i>Quantitative variable</i>		
Расстояние между двумя независимыми группами	Нормальное распределение переменной в каждой группе	Оценка среднего разности между двумя случайными величинами $d = \bar{x}_2 - \bar{x}_1$ с расчетом ДИ
	Распределение, отличное от нормального	Оценка Ходжеса – Лемана (Hodges-Lehmann Estimator) с расчетом ДИ [7]
Расстояние между двумя зависимыми группами (парные сравнения), рассчитывается переменная различий между измерениями в каждой паре	Для переменной различий в случае ее нормального распределения	Оценка среднего \bar{x} с расчетом ДИ среднего
	Для переменной различий в случае ее распределения, отличного от нормального	Расчет псевдомедианы (оценка Ходжеса – Лемана для одной выборки) с расчетом ДИ
Бинарная переменная <i>Binary variable</i>		
Различия в пропорциях между двумя независимыми группами	Биномиальное распределение	Оценка разницы в пропорциях с расчетом ДИ
Различия в пропорциях между двумя зависимыми группами (парные сравнения)		

Оценка Ходжеса – Лемана [7, 8] определяет разницу между двумя популяциями. Для двух наборов данных с m - и n -наблюдениями набор составленных из них двухэлементных наборов представляет собой декартово произведение, содержащее $m \times n$ пар точек (по одной из каждого набора). Каждая такая пара определяет одну разность значений. Оценка Ходжеса – Лемана является медианой разностей набора $m \times n$. Псевдомедиана (оценка Ходжеса – Лемана для единственной популяции) определяется следующим образом. Пусть данные содержат n измерений. Определим $n(n+1)/2$ возможных пар. Для каждой пары рассчитаем среднее

$(x_i + x_j)/2, i \leq j, i=1, \dots, n, j=1, \dots, n$. Медиана всех возможных средних является псевдомедианой. Данная величина называется одновыборочной оценкой Ходжеса – Лемана и не оценивает ни среднее значение, ни медиану.

Стандартизированные размеры эффектов. Преимущество стандартизированного размера эффекта заключается в возможности сравнения различных исследований, а также унификации интерпретации.

Самая распространенная оценка стандартизированного размера эффекта количественной переменной в двух группах – это индекс d Коэна [9]. Он рассчитывается как $d = \frac{\bar{x}_1 - \bar{x}_2}{s}$, где \bar{x}_1 – оценка среднего в первой группе; \bar{x}_2 – оценка среднего во второй группе; s – общее стандартное отклонение двух групп, которое рассчитывается при равном размере групп как $s = \sqrt{\frac{s_1^2 + s_2^2}{2}}$, где s_1 и s_2 – стандартные отклонения в соответствующих группах.

Помимо индекса d Коэна существуют еще две похожие оценки: индекс Δ Гласса (Glass' Δ) и индекс g Хеджеса (Hedges' g). Они отличаются от индекса d Коэна способом расчета знаменателя. В случае индекса Δ Гласса это стандартное отклонение контрольной группы, а в случае индекса g Хеджеса – взвешенное стандартное отклонение двух групп.

Гласс предложил индекс Δ для случаев, когда две выборки представляют экспериментальную и контрольную группы, и рекомендовал использовать для стандартизации стандартное отклонение контрольной группы [10].

Индекс g Хеджеса [11] обеспечивает лучшую оценку размеров эффекта, если размеры выборки различаются, поскольку при этом используются размеры выборки в качестве весов при определении общего стандартного отклонения: $s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$. Также эта оценка рекомендуется для малых выборок.

Одной из интересных оценок размера эффекта, которая построена на совершенно другом подходе (вероятностном), является индекс δ Клиффа (Cliff's δ) [12]. Он не зависит от вида распределения и предназначен для измерений в шкале не ниже порядковой. Пусть существуют выборки $X = \{x_1, \dots, x_m\}$ и $Y = \{y_1, \dots, y_n\}$. Тогда дельта-функция определяется как

$$\delta(i, j) = \begin{cases} +1, & x_i > y_j; \\ -1, & x_i < y_j; \\ 0, & x_i = y_j, \end{cases}$$

а размер эффекта – как $\delta = \frac{1}{m \times n} \sum_{i=1}^m \sum_{j=1}^n \delta(i, j)$.

Согласно определению $-1 \leq \delta \leq 1$. Значения, близкие к ± 1 , означают минимальное перекрытие двух распределений, в то время как значения, близкие к 0, означают значительное перекрытие наблюдений в двух выборках.

Еще одна уникальная оценка размеров эффектов – ранговая бисериальная корреляция [13–15], которая принадлежит к d -семейству эффектов (эффектов различий), и r -семейству эффектов (эффектов связи): $r_{rb} = \frac{2}{n_1} \times \left(\bar{Y}_2 - \frac{n+1}{2} \right) = \frac{2}{n_2} \times \left(\frac{n+1}{2} - \bar{Y}_1 \right)$, где \bar{Y}_1 и \bar{Y}_2 – среднее рангов в первой и второй группах соответственно; n_1 и n_2 – размер первой и второй группы соответственно, $n = n_1 + n_2$. Эта оценка используется для определения размера эффекта между двумя группами для показателей, имеющих шкалу измерений: порядковую, шкалу интервалов или шкалу

отношений. В случае двух групп ранговая бисериальная корреляция совпадает с индексом δ Клиффа [16].

Для оценки стандартизированных размеров эффектов бинарных переменных при рассмотрении различий в пропорциях в двух независимых группах можно использовать индекс h Коэна (Cohen's h) [9], который рассчитывается как $h = 2 \times (\arcsin \sqrt{p_1} - \arcsin \sqrt{p_2})$, где p_1, p_2 – пропорции интересующего исхода (бинарной переменной) в двух независимых группах.

При оценке размера эффекта как для зависимых групп, так и для независимых возможно применение стандартизированного размера эффекта, который рассчитывается аналогично индексу d Коэна или индексу Δ Гласса с учетом различий в расчете стандартного отклонения в зависимых и независимых группах для бинарного отклика.

В качестве оценки стандартизированного размера эффекта для зависимых групп Коэн предложил использовать индекс g (Cohen's g) [9]. Изначально Коэн разработал эту оценку для сравнения одновыборочной пропорции с вероятностью 0,5. Для зависимых групп в этом случае определяются несогласованные пары (пары, которые изменили значение бинарного признака). Среди них выделяется число однонаправленных пар и определяются пропорции $p_{0 \rightarrow 1}$ и $p_{1 \rightarrow 0}$, $p_{0 \rightarrow 1} + p_{1 \rightarrow 0} = 1$. Одна из пропорций сравнивается с 0,5. Коэн называет эту оценку индексом размера эффекта, который может быть направленным (g определяется как $g = p - 0,5$) или ненаправленным (g определяется как $g = |p - 0,5|$), где p – одна из двух пропорций $p_{0 \rightarrow 1}$ или $p_{1 \rightarrow 0}$. Ненаправленный индекс может принимать значения от 0 (полная симметрия, равные пропорции) до 0,5 (полная асимметрия).

Интерпретация оценок размера эффекта. Помимо количественной оценки различий часто необходимо понять, насколько эти различия интересны с точки зрения предметной области, а не только с точки зрения статистической значимости. Является размер эффекта большим или малым, – это вопрос интерпретации специалиста предметной области.

Основные пороги интерпретации стандартизированного размера эффекта количественных показателей приведены в табл. 2 и 3, а бинарных показателей – в табл. 4.

Таблица 2

Интерпретация оценок стандартизированных размеров эффектов количественных показателей (индекса d Коэна, индекса Δ Гласса и индекса g Хеджеса) в различных научных источниках

Table 2

Effect size interpretation for Cohen's d , Glass' Δ , and Hedges' g in various scientific sources

[9]	[17]	[18]	[19]
$d(\Delta, g) < 0,2$ Очень малый	$d(\Delta, g) < 0,1$ Крошечный	$d(\Delta, g) < 0,15$ Очень малый	$d(\Delta, g) < 0,2$ Очень малый
$0,2 \leq d(\Delta, g) < 0,5$ Малый	$0,1 \leq d(\Delta, g) < 0,2$ Очень малый	$0,15 \leq d(\Delta, g) < 0,36$ Малый	$0,2 \leq d(\Delta, g) < 0,41$ Малый
$0,5 \leq d(\Delta, g) < 0,8$ Средний	$0,2 \leq d(\Delta, g) < 0,5$ Малый	$0,36 \leq d(\Delta, g) < 0,65$ Средний	$0,41 \leq d(\Delta, g) < 0,63$ Умеренный
$d(\Delta, g) \geq 0,8$ Большой	$0,5 \leq d(\Delta, g) < 0,8$ Средний	$d(\Delta, g) \geq 0,65$ Большой	$d(\Delta, g) \geq 0,63$ Большой
–	$0,8 \leq d(\Delta, g) < 1,2$ Большой	–	–
–	$1,2 \leq d(\Delta, g) < 2$ Очень большой	–	–
–	$d(\Delta, g) \geq 2$ Огромный	–	–

Таблица 3

Интерпретация оценок стандартизированных размеров эффектов ранговой бисериальной корреляции в различных научных источниках

Table 3

Standardized effect size interpretation for rank biserial correlation in various scientific sources

[20]	[19]	[9]	[18]	[21]
$r_{pb} < 0,05$ Крошечный	$r_{pb} < 0,1$ Очень малый	$r_{pb} < 0,1$ Очень малый	$r_{pb} < 0,12$ Очень малый	$r_{pb} < 0,2$ Очень слабый
$0,05 \leq r_{pb} < 0,1$ Очень малый	$0,1 \leq r_{pb} < 0,2$ Малый	$0,1 \leq r_{pb} < 0,3$ Малый	$0,12 \leq r_{pb} < 0,24$ Малый	$0,2 \leq r_{pb} < 0,4$ Малый
$0,1 \leq r_{pb} < 0,2$ Малый	$0,2 \leq r_{pb} < 0,3$ Умеренный	$0,3 \leq r_{pb} < 0,5$ Умеренный	$0,24 \leq r_{pb} < 0,41$ Умеренный	$0,4 \leq r_{pb} < 0,6$ Умеренный
$0,2 \leq r_{pb} < 0,3$ Средний	$r_{pb} \geq 0,3$ Большой	$r_{pb} \geq 0,5$ Большой	$r_{pb} \geq 0,41$ Большой	$0,6 \leq r_{pb} < 0,8$ Сильный
$0,3 \leq r_{pb} < 0,4$ Большой	–	–	–	$r_{pb} \geq 0,8$ Очень сильный
$r_{pb} \geq 0,4$ Очень большой	–	–	–	–

Таблица 4

Интерпретация оценок стандартизированных размеров эффектов бинарных переменных

Table 4

Standardized effect size interpretation for binary variables

Индекс Коэна [9] Cohen's [9]	
h	g
$h = 0,2$ Небольшой	$g < 0,05$ Очень малый
$h = 0,5$ Средний	$0,05 \leq g < 0,15$ Малый
$h = 0,8$ Большой	$0,15 \leq d < 0,25$ Средний
–	$d \geq 0,25$ Большой

Индекс δ Клиффа не имеет авторской интерпретации, однако если предположить нормальное распределение исследуемого показателя в двух группах, то при индексе d Коэна, равном $d = 0,2$, $\delta = 0,147$; при $d = 0,5$ $\delta = 0,33$; при $d = 0,8$ $\delta = 0,474$.

Связь между индексом δ Клиффа и индексом d Коэна определяется выражением $\delta = 2 \times P(X \leq d/2) - 1 / P(X \leq d/2)$, где $P(X \leq x)$ – кумулятивная функция распределения Гаусса, $N(0,1)$.

Связь между индексом δ Клиффа и индексом d Коэна при распределении исследуемого показателя, отличного от нормального, рассмотрена в работах [22–24].

Пример квазиэкспериментального исследования. Рассмотрим условное квазиэкспериментальное исследование, в котором лечение назначалось с учетом некоторого бинарного признака и согласия пациента на его участие в исследовании и лечении. Контрольная группа формировалась из пациентов, которые не давали согласие на лечение, а соглашались только на участие в исследовании. Также мониторился уровень качества жизни пациента по 10-балльной визуально-аналоговой шкале (ВАШ). Контроль осуществлялся через три месяца после начала лечения. В каждой группе было 30 пациентов, группы не различались по возрасту, полу и степени тяжести заболевания.

Примером квазиэкспериментального исследования может служить исследование эффективности некоторого лечения, связанного с нарушением деятельности желудочно-кишечного тракта. Уровень боли оценивается по ВАШ (используется в медицинских исследованиях с 1921 г. [25]). Наличие рефлюкса желудка у пациента является бинарной переменной, которая определяет нерандомизированность назначения лечения.

Код примера использования воспроизводим в статистическом пакете R, для которого приведены коды генерации исходных данных, а также коды расчетов в виде отдельных чанков (см. https://drive.google.com/file/d/16bGni_smkDZchxCR4vN_p62tKJeP7WRF/view). Все интервальные оценки даны с уровнем доверия 95 %.

Анализ результатов исследования количественной переменной. Рассмотрим изменение показателя ВАШ в основной и контрольной группах. Учитывая, что этот показатель измеряется в балльной (порядковой) шкале, выбираем для его анализа непараметрические методы сравнения. Проверка распределений интересующих нас переменных показала, что распределения отличны от нормального.

Рассчитаем абсолютные и стандартизированные размеры эффектов в группах и различия в эффектах между группами («различия в различиях») (табл. 5 и 6). На основании расчетов различия в баллах между началом и концом исследования по ВАШ боли в контрольной группе (D_k) составили 3 (95 % ДИ 1,5; 3,5), а в экспериментальной группе (D_3) – -1 (95 % ДИ -2; 0). Искомый эффект «различия в различиях» $\Delta D = D_3 - D_k$ по ВАШ боли составил -2 (95 % ДИ -3; -1).

Таблица 5

Результаты оценки абсолютных и стандартизированных размеров эффектов количественной переменной в группах

Table 5

The assessment of the absolute and standardized effect sizes of the quantitative variable within groups

Группа Group	Статистическая значимость по критерию Вилкоксона при парных измерениях Statistical significance, paired Wilcoxon test	Абсолютный размер эффекта между началом и концом исследования, псевдомедиана (95 % ДИ) Absolute effect size within the study, pseudomedian (95 % CI)	Стандартизированный размер эффекта между началом и концом исследования, ранговая бисериальная корреляция (95 % ДИ) Standardized effect size within the study, the rank-biserial correlation coefficient (95 % CI)
Экспериментальная	0,083	-1 (-2; 0)	-0,45 (-0,76; 0,02)
Контрольная	<0,001	3 (1,5; 3,5)	0,90 (0,77; 0,96)

Таблица 6

Результаты оценки различий в эффектах между группами

Table 6

The assessment of the absolute and standardized effect sizes of the quantitative variable between groups

Экспериментальная группа, размер эффекта, псевдомедиана (95 % ДИ) Experimental group effect size, pseudomedian (95 % CI)	Контрольная группа, размер эффекта, псевдомедиана (95 % ДИ) Control group effect size, pseudomedian (95 % CI)	Абсолютный размер «различия в различиях» ΔD , оценка Ходжеса – Лемана (95 % ДИ) Absolute "difference-in-difference" distance ΔD , Hodges-Lehmann Estimator (95 % CI)	Стандартизированный размер «различия в различиях», ранговая бисериальная корреляция (95 % ДИ) Standardized "difference-in-difference" distance, the rank-biserial correlation coefficient (95 % CI)
-1 (-2; 0)	3 (1,5; 3,5)	-2 (-3; -1), $p < 0,001$	-0,62 (-0,77; -0,40)

Статистически значимого эффекта по снижению боли в экспериментальной группе не наблюдается, но вывод о его отсутствии некорректен. Учитывая усиление боли в контрольной группе, можно заключить, что воздействие (лечение) является эффективным для контроля уровня боли при данном заболевании с абсолютным размером эффекта по ВАШ до двух баллов. Интерпретация стандартизированных размеров эффекта может быть взята из табл. 3, но клинически более понятно, что при лечении уровень боли снижается в целом на один-три балла.

Анализ исследования результатов бинарной переменной. Код генерации данных приведен в приложении, распределение в группах исследования бинарной переменной (наличие рефлюкса) представлено в табл. 7.

Таблица 7

Бинарная переменная, распределение в группах в начале и конце исследования, размер эффектов в группах

Table 7

Binary variable distribution in groups at baseline and end of study, within groups effect sizes

Экспериментальная группа <i>Experimental group</i>		Конец исследования, n (%) <i>End of the study, n (%)</i>		Эффекты <i>Effects</i>	
		Да 14 (47 %)	Нет 16 (53 %)	Абсолютный размер эффекта между началом и концом исследования, D_3 (95 % ДИ)	Стандартизированный размер эффекта между началом и концом исследования, индекс g_3 Коэна (95 % ДИ)
Начало исследования, n (%) <i>Experimental group, start of the study, n (%)</i>	Да 28 (93 %)	14	14	-0,46 (-0,65; -0,29) $p < 0,001$	0,50 (0,28; 0,50)
	Нет 2 (7 %)	0	2		
Контрольная группа <i>Control group</i>		Конец исследования, n (%) <i>End of the study, n (%)</i>		Эффекты <i>Effects</i>	
		Да 20 (67 %)	Нет 10 (33 %)	Абсолютный размер эффекта между началом и концом исследования, D_k (95 % ДИ)	Стандартизированный размер эффекта между началом и концом исследования, индекс g_k Коэна (95 % ДИ)
Начало исследования, n (%) <i>Control group, start of the study, n (%)</i>	Да 14 (47 %)	13	1	0,20 (0,03; 0,37) $p = 0,033$	0,38 (0,03; 0,48)
	Нет 16 (53 %)	7	9		

Видно, что лечение преимущественно назначалось пациентам с наличием бинарного признака (рефлюкса) при отношении шансов 15, $p < 0,001$. В конце исследования группы в процентном соотношении по бинарной переменной сравнивались ($p = 0,192$). Однако данное исследование является квазиэкспериментальным, поэтому необходимо учитывать состояние пациента не только в конце исследования, но и в его начале.

Для расчета размера эффекта с учетом начальных условий исследования следует оценить разницу в эффектах. Сначала рассчитаем размеры эффектов в группах как различия в пропорциях до начала и после окончания исследования с учетом повторных измерений.

Для определения вариации оценки абсолютного эффекта как пропорции в группе используем формулу Мак-Немара для парных пропорций [26]: $Var = (p_{0 \rightarrow 1} + p_{1 \rightarrow 0}) - (p_{0 \rightarrow 1} - p_{1 \rightarrow 0})^2$.

Стандартная ошибка составит $se = \sqrt{\frac{Var}{n}}$, где $p_{0 \rightarrow 1}, p_{1 \rightarrow 0}$ – доли несогласованных пар; n – размер выборки.

Абсолютное значение различия в эффектах («различия в различиях») оценивается как разность размеров эффектов в группах. Однако с учетом того, что в одной из групп эффект отрицательный, будем использовать для расчета формулы, приведенные ниже.

Для нахождения абсолютного размера эффекта «различия в различиях» воспользуемся формулой $\Delta D = D_3 - D_k$, стандартная ошибка по правилу распространения ошибки составит $se = \sqrt{se_3^2 + se_k^2}$. Тогда ДИ может быть рассчитан как $\Delta D \pm z_{1-\alpha/2} \times se$, где $z_{1-\alpha/2}$ – значение квантиля стандартного нормального распределения, α – уровень ошибки первого рода.

Для определения стандартизированного размера эффекта «различия в различиях» применим формулу для индекса h Коэна с учетом знака полученного эффекта в группе:

$$h = 2 \times \left(\text{sgn}(D_s) \times \arcsin \sqrt{|D_s|} - \text{sgn}(D_k) \times \arcsin \sqrt{|D_k|} \right).$$

Для нахождения стандартной ошибки результата углового преобразования воспользуемся формулой $se = \sqrt{\frac{1}{4n}}$ [27] и по правилу распространения ошибки рассчитаем ошибку стандартизированного размера эффекта «различия в различиях».

Результаты расчета абсолютных и стандартизированных размеров эффектов в группах приведены в табл. 7, результаты расчета «различия в различиях» – в табл. 8.

Таблица 8
Оценка размера эффекта «различия в различиях» для бинарной переменной

Table 8
Estimate of the “differences in differences” effect size for binary variable

Экспериментальная группа, абсолютный размер эффекта, D_s (95 % ДИ) <i>Experimental group, absolute effect size, D_s (95 % CI)</i>	Контрольная группа, абсолютный размер эффекта, D_k (95 % ДИ) <i>Control group, absolute effect size, D_k (95 % CI)</i>	Абсолютный размер эффекта «различия в различиях», ΔD (95 % ДИ) <i>Absolute “difference-in-difference” distance, ΔD (95 % CI)</i>	Стандартизированный размер эффекта «различия в различиях», индекс h Коэна (95 % ДИ) <i>Standardized “difference-in-difference” distance, Cohen’s h (95 % CI)</i>
-0,46 (-0,65; -0,29)	0,20 (0,03; 0,37)	-0,67 (-0,91; -0,42)	-2,43 (-2,68; -2,18)

Интерпретацию результатов анализа можно выбрать согласно табл. 3 и 4.

Закключение. Аналитический метод DID является достаточно популярным на практике для оценки результатов с использованием данных наблюдений. Фактически этот метод пытается имитировать РКИ, изучая дифференциальное влияние воздействия (лечения, иного вмешательства) в естественных экспериментах. В своей канонической форме метод DID определяет средний эффект воздействия на испытуемых, участвующих в эксперименте, путем сравнения разницы в исходах до и после лечения двух групп: получающей и не получающей воздействие (экспериментальная и группа сравнения соответственно). Чтобы придать причинно-следственную интерпретацию оценкам метода DID, используется безусловное предположение о параллельных трендах: в отсутствие воздействия средние результаты для групп эксперимента и сравнения следовали бы с течением времени параллельными путями. Хотя данное предположение принципиально не поддается проверке, его правдоподобие обычно подвергается сомнению, если наблюдаемые характеристики, которые, как считается, связаны с эволюцией исхода, не сбалансированы между экспериментальной группой и группой сравнения. В таких случаях исследователи обычно включают ковариаты до лечения в анализ DID и считают, что предположения о трендах могут быть удовлетворены только после включения этих ковариат в более сложные модели оценки результатов [28, 29], которые опираются на линейную регрессию с взаимодействием и контролем ковариат и чаще применяются в эконометрике и социологии при анализе панельных данных. Переменная отклика (исхода) в таких моделях имеет количественную природу.

Медицинские данные не часто отвечают предположениям, лежащим в основе линейной регрессии, поэтому для оценки результатов могут выполняться более понятные расчеты с учетом баланса групп по ковариатам, принятым в клинических исследованиях. Таким образом, алгоритмы подбора групп, например propensity score matching (PSM) [30], могут быть использованы до начала анализа по методу DID.

Более того, чаще исход фиксируется как бинарная переменная, и эффект на основе, например, логистической регрессии (отношения шансов) не дает такой же клинической картины результатов лечения, как обычные пропорции.

В клинической практике размер эффекта играет важную роль. На основе результатов исследований принимаются решения, которые должны вести к более эффективным и качественным тактикам лечения, к выбору оптимальных подходов и получению клинически благоприятных исходов заболевания.

Статистические методы оценки результатов применения определенных методов и способов лечения пациента в конечном счете должны давать практикующему врачу понятные данные об эффектах лечения (воздействия). Безусловно, абсолютный размер эффекта воспринимается врачом легче, чем стандартизированный, однако для понимания того, насколько велик размер эффекта, необходима его стандартизация. Это помогает проводить мультицентровые исследования, сравнения различных тактик и т. п. Насколько важен результат различий, может решить только специалист в предметной области.

References

1. Lederer D. J., Bell S. C., Branson R. D., Chalmers J. D., Marshall R., ..., Vincent J.-L. Control of confounding and reporting of results in causal inference studies. Guidance for authors from editors of respiratory, sleep, and critical care journals. *Annals of the American Thoracic Society*, 2019, vol. 16, no. 1, pp. 22–28.
2. Fletcher R. H., Fletcher S. W., Wagner E. H. *Clinical Epidemiology: The Essentials*, 3rd ed. Baltimore, Williams & Wilkins, 1996, 276 p.
3. Campbell D. T., Riecken H. W. Quasi-experimental design. *International Encyclopedia of the Social Sciences*, 1968, vol. 5, no. 3, pp. 259–263.
4. Pearl J. Causal inference in statistics: An overview. *Statistics Surveys*, 2009, vol. 3, pp. 96–146.
5. Shadish W. R., Luellen J. K. Quasi-experimental design. *Handbook of Complementary Methods in Education Research*. Mahwah, Routledge, 2012, pp. 539–550.
6. Kelley K., Preacher K. J. On effect size. *Psychological Methods*, 2012, vol. 17, no. 2, pp. 137–152.
7. Hodges J. L., Lehmann E. L. Estimates of location based on rank tests. *The Annals of Mathematical Statistics*, 1963, vol. 34, pp. 598–611.
8. Hollander M., Wolfe D. A., Chicken E. *Nonparametric Statistical Methods*, 3rd ed. John Wiley & Sons, 2013, 848 p.
9. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, Lawrence Earlbaum Associates, 1988, 590 p.
10. Glass G. V. Primary, secondary, and meta-analysis of research. *Educational Researcher*, 1976, vol. 5, no. 10, pp. 3–8.
11. Hedges L. V. Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 1981, vol. 6, no. 2, pp. 107–128.
12. Cliff N. *Ordinal Methods for Behavioral Data Analysis*, 1st ed. Psychology Press, 1996, 212 p.
13. Cureton E. E. Rank-biserial correlation. *Psychometrika*, 1956, vol. 21, no. 3, pp. 287–290.
14. Cureton E. E. Rank-biserial correlation when ties are present. *Educational and Psychological Measurement*, 1968, vol. 28, no. 1, pp. 77–79.
15. Glass G. V. Note on rank biserial correlation. *Educational and Psychological Measurement*, 1966, vol. 26, no. 3, pp. 623–631.
16. Willson V. L. Critical values of the rank-biserial correlation coefficient. *Educational and Psychological Measurement*, 1976, vol. 36, no. 2, pp. 297–300.
17. Sawilowsky S. S. New effect size rules of thumb. *Journal of Modern Applied Statistical Methods*, 2009, vol. 8, no. 2, pp. 597–599.
18. Lovakov A., Agadullina E. R. Empirically derived guidelines for effect size interpretation in social psychology. *European Journal of Social Psychology*, 2021, vol. 51, no. 3, pp. 485–504.
19. Gignac G. E., Szodorai E. T. Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, 2016, vol. 102, pp. 74–78.
20. Funder D. C., Ozer D. J. Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, 2019, vol. 2, no. 2, pp. 156–168.
21. Evans J. D. *Straightforward Statistics for the Behavioral Sciences*. Thomson Brooks/Cole Publishing Co, 1996, 624 p.
22. Hess M. R., Kromrey J. D. Robust confidence intervals for effect sizes: A comparative study of Cohen's d and Cliff's delta under non-normality and heterogeneous variances. *Annual meeting of the American Educational Research Association*, 2004, vol. 1.

23. Romano J., Kromrey J., Coraggio J., Skowronek J. Appropriate statistics for ordinal level data: Should we really be using t-test and Cohen's d for evaluating group differences on the NSSE and other surveys. *Annual meeting of the Florida Association of Institutional Research*, 2006, vol. 177, p. 34.

24. Marfo P., Okyere G. A. The accuracy of effect-size estimates under normals and contaminated normals in meta-analysis. *Heliyon*, 2019, vol. 5, no. 6, p. e01838.

25. Couper M. P., Tourangeau R., Conrad F. G., Singer E. Evaluating the effectiveness of visual analog scales: A web experiment. *Social Science Computer Review*, 2006, vol. 24, no. 2, pp. 227–245.

26. McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 1947, vol. 12, no. 2, pp. 153–157.

27. Anscombe F. J. The transformation of Poisson, binomial and negative-binomial data. *Biometrika*, 1948, vol. 35, no. 3/4, pp. 246–254.

28. Abadie A. Semiparametric difference-in-differences estimators. *The Review of Economic Studies*, 2005, vol. 72, no. 1, pp. 1–19.

29. Sant'Anna P. H. C., Zhao J. Doubly robust difference-in-differences estimators. *Journal of Econometrics*, 2020, vol. 219, no. 1, pp. 101–122.

30. Rosenbaum P. R., Rubin D. B. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 1983, vol. 70, no. 1, pp. 41–55.

Информация об авторе

Красько Ольга Владимировна, кандидат технических наук, доцент, ведущий научный сотрудник, Объединенный институт проблем информатики Национальной академии наук Беларуси.
E-mail: krasko@newman.bas-net.by
<https://orcid.org/0000-0002-4150-282X>

Information about the author

Olga V. Krasko, Ph. D. (Eng.), Assoc. Prof., Leading Researcher, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus.
E-mail: krasko@newman.bas-net.by
<https://orcid.org/0000-0002-4150-282X>