Объединение данных реальной клинической практики с результатами рандомизированных клинических исследований для лучшей информированности при принятии решений в онкологии

Усманова Т. А. , **Вербицкая Е. В.**
ФГБОУ ВО «Первый Санкт-Петербургский государственный медицинский университет имени И.П. Павлова», Санкт-Петербург, Россия

Аннотация

Рандомизированные контролируемые исследования (РКИ) являются «золотым стандартом» оценки действенности противоопухолевой терапии. Несмотря на высокую внутреннюю валидность таких исследований, их обобщаемость, то есть возможность перенести полученные результаты на широкую популяцию пациентов, ограничена, в связи с чем пользователи и работники системы здравоохранения могут столкнуться с более низкой эффективностью вмешательства в реальной практике, чем было заявлено в РКИ. Можно выделить множество причин формирования разрыва между действенностью и эффективностью (efficacy-effectiveness gap; EEG), то есть мерой воздействия в РКИ и в реальной клинической практике (РКП). Сюда относятся, например, различные характеристики пациентов в испытании и в РКП, приверженность терапии, особенности оказания медицинской помощи и другие. Для того чтобы проиллюстрировать эту проблему, представлен обзор ряда исследований, посвящённых оценке величины и анализу возможных причин указанного несоответствия. В большинстве приведённых исследований был выявлен EEG, предложены вероятные объяснения его наличия, а также произведены дополнительные оценки для установления вклада различных факторов в его величину. Авторы этих публикаций показывают, что, в отличие от участников РКИ, «реальные» пациенты старше, имеют худший функциональный статус, большее количество сопутствующих заболеваний, чаще являются женщинами, реже завершают начатое лечение или переход к следующей линии терапии. Кроме того, в данной статье предложены различные аналитические подходы к определению веса основных причинных факторов в формировании несоответствия между действенностью и эффективностью, что может использоваться при разработке методологии соответствующих исследований.

Имея в доступе информацию о размере EEG при использовании различных схем лечения в своём регионе и понимая, в какой степени тот или иной фактор может оказать влияние на величину этого разрыва, клиницист сможет прогнозировать эффективность лечения и выбирать оптимальную тактику для конкретного больного.

Ключевые слова: действенность; эффективность; несоответствие между действенностью и эффективностью; доказательства, собранные в реальной клинической практике; RWD; RWE

Для цитирования:
Усманова Т. А., Вербицкая Е. В. Объединение данных реальной клинической практики с результатами рандомизированных клинических исследований для лучшей информированности при принятии решений в онкологии. *Реальная клиническая практика: данные и доказательства.* 2022;2 (2):21 — 31. https://doi.org/10.37489/2782-3784-myrwd-14.

Поступила: 25 мая 2022 г. **Одобрена:** 06 июня 2022 г. **Опубликована:** 01 августа 2022 г.

Combining real-world data with randomized controlled trials results in better information oncology decision making

Usmanova T. A. , **Verbitskaya E. V.**
FSBEI HE I.P. Pavlov SPbSMU MOH Russia, St. Petersburg, Russia

Abstract

Randomized controlled trials (RCTs) are the gold standard for testing the efficacy of cancer therapy. Although the results of clinical trials have high internal validity, their generalizability, that is, the ability to transfer the results to a wide patient population, is limited. Therefore, users and health care workers may experience less effective intervention in real practice than stated in the RCT. There are many reasons for the formation of a gap between efficacy and effectiveness (efficacy-effectiveness gap; EEG), that is, the measure of impact on RCTs and the real-world. These reasons include, for example, different characteristics of patients in the trial and real practice, compliance to treatment, features of medical care, and others.

To illustrate this problem, a review of some studies on the estimation of the magnitude and analysis of the possible causes of this gap is presented. In most of the studies cited, EEG was identified, its probable explanations were proposed, and additional estimates were made to establish the contribution of various factors to its magnitude. These publications» authors show that real-world patients are older, have worse functional status, and have a greater number of comorbidities. They are women mostly and are less likely to complete the treatment they have started or move to the next line of therapy, in contrast to participants in RCTs. Additionally, this article proposes various analytical approaches to determine the weight of the main causal factors in the formation of a discrepancy between efficacy and effectiveness, which can be used in the development of the methodology of relevant studies.

Knowing the size of the EEG when using different treatment regimens in their region and understanding the extent to which one or another factor can influence the size of this gap, the clinician will be able to predict the effectiveness of treatment and choose the best therapy for a particular patient.

Keywords: efficacy; effectiveness; efficacy-effectiveness gap; EEG; real-world evidence; RWD; RWE

For citation:

Usmanova TA, Verbitskaya EV. Combining real-world data with randomized controlled trials results in better information oncology decision making. *Real-World Data & Evidence*. 2022;2 (2):21 — 31. https://doi.org/10.37489/2782-3784-myrwd-14.

Received: May 25, 2022. **Accepted:** June 06, 2022. **Published:** August 01, 2022.

Введение

Основным способом оценки действенности противоопухолевой терапии является проведение клинических исследований. Благодаря тщательно продуманному дизайну, строгим критериям отбора участников, максимальному исключению дополнительных воздействующих факторов мы получаем высокую внутреннюю валидность результатов таких исследований, то есть имеем право сделать вывод о том, что различия в исходах зависят от способа терапии. Однако перенести полученные результаты на широкую популяцию пациентов зачастую непросто [1]. Учитывая результаты соответствующих рандомизированных клинических исследований (РКИ) при принятии решений, клиницисты сталкиваются с низкой внешней валидностью (или обобщаемостью), то есть с возможностью обоснованно применить их результаты к определённой группе пациентов в конкретных клинических условиях в рутинной практике, особенно в случаях, когда исследования финансируются фармацевтическими компаниями [2]. Хотя критерии включения и исключения предназначены для того, чтобы обезопасить участников исследования путём исключения испытуемых, у которых ожидается низкая эффективность или высокая токсичность терапии, чрезмерно строгие критерии отбора приводят к тому, что исследуемая группа не отражает популяцию пациентов, получающих лечение в реальных условиях. Так, в США менее 3% онкологических больных участвуют в клинических исследованиях. Это связано как с недоступностью участия для некоторых пациентов, так и со строгостью критериев отбора, в связи с которыми в исследования зачастую попадают более молодые и здоровые пациенты [3]. Эти факторы наряду со многими другими, такими как практика назначения и применения лекарственных препаратов, приверженность лечению, влияние врача, доступ к медицинской помощи, разные способы измерения эффекта, приводят к формированию разрыва между «действенностью» (*англ.* efficacy), или величиной оценки воздействия в идеальных условиях, то есть при проведении РКИ, и «эффективностью» (*англ.* effectiveness), или величиной оценки воздействия при применении этого вмешательства в условиях реальной клинической практики [4-7]. Это несоответствие именуется «разрывом между действенностью и эффективностью» (*англ.* efficacy-effectiveness gap; EEG) и определяется как различия в соотношении «польза-риск», полученные на этапе клинических исследований и на этапе применения лекарственного средства в условиях реальной медицинской практики [5]. Наличие EEG, предположение о возможных причинах данного явления, а также пути его преодоления продемонстрированы в ряде публикаций, систематический обзор которых представили *Филиппова А.В.* и *Колбин А.С.* [6]. Среди 18 включённых в обзор статей авторы 16 пришли к выводу о том, что, по результатам их наблюдений, указанное несоответствие действительно существует, и авторы 2 статей заключили, что в их исследованиях несоответствия не обнаружено [6].

Онкологам особенно важно иметь хотя бы примерное представление о величине EEG не только для ответа на терзающий всех пациентов вопрос об их прогнозе, но и с целью выбора наиболее эффективной тактики терапии для конкретного больного. Стоит заметить, что самым надёжным способом установить причинно-следственную связь между участием в испытании и исходом была бы рандомизация пациентов на тех, которым будет предложено участие в РКИ, и тем, кому не будет. Но проведение такого исследования трудно осуществимо по практическим и этическим соображениям. Вместо этого сравнение основано на наблюдательных (обсервационных) исследованиях, в которых основное внимание уделяется представлению характеристик и абсолютных различий в выживаемости между пациентами, которые были

или не были включены в испытание [3]. Ввиду того, что величина этого разрыва до сих пор часто остаётся недооценённой, рассмотрим выводы нескольких авторов по этой проблеме.

Обзор публикаций, оценивающих несоответствие между действенностью и эффективностью

Группой *Unger JM et al.* проведено сравнение результатов выживания среди онкологических больных, получавших лечение в клинических исследованиях и вне их. Данные клинических исследований были получены от SWOG, национального консорциума клинических исследований, спонсируемого Национальным институтом рака, а данные реальной практики были взяты из базы данных Surveillance, Epidemiology, and End Results (SEER; Наблюдения, эпидемиология и конечные результаты). Анализируемые клинические исследования были поделены на обладающие хорошим и плохим прогнозом. В первую группу вошло 11 исследований, двухлетняя выживаемость в которых составила 50% и более, а во вторую — 10 исследований с двухлетней выживаемостью менее 50%. Показано, что наибольшая разница в выживаемости между онкологическими пациентами, получающими лечение в условиях клинического исследования и в реальной практике, наблюдается в первый год. Ни в одном из исследований с хорошим прогнозом выживаемость пациентов SWOG статистически значимо не отличалась от выживаемости пациентов SEER, тогда как в девяти из 10 исследований с плохим прогнозом у пациентов SWOG был статистически значимо более низкий риск смерти ($p < 0{,}001$). Авторы говорят, что несмотря на то, что в клинические исследования не попадают пациенты с плохим прогнозом или сопутствующими заболеваниями в анамнезе, обусловленные этим фактом различия в выживаемости могут не сохраняться с течением времени, если исследуемая когорта в остальном репрезентативна для общей онкологической популяции в отношении гистологии и стадии рака. Интересно также, что при анализе причин смерти смертность, не связанная с раком, была несколько ниже у пациентов SWOG на протяжении всех 5 лет наблюдения, а смертность, связанная с раком у пациентов SWOG, была заметно ниже в первый год, но аналогична пациентам SEER в последующие годы [3]. То есть, вероятно, меньшая смертность пациентов в РКИ по не связанным с раком причинам вызвана попаданием в исследования более здоровых пациентов без сопутствующей патологии в связи с критериями отбора, более низкая смертность от рака в первый год наблюдения обусловлена изначально лучшим прогнозом таких больных, а к выравниванию выживаемости приводят особенности опухоли.

Авторы анализа, проведённого при поддержке Канадского центра прикладных исследований в области борьбы с раком (ARCC), представили количественную оценку величины различий в выживаемости и токсичности между клиническими исследованиями и реальной практикой для современной системной терапии рака. Методология их исследования включала получение из данных РКИ сведений о выживаемости и частоте нежелательных явлений для каждого показания к приёму лекарственного препарата. Реальные данные по выживаемости и показателям госпитализации во время лечения были получены из канадских популяционных баз данных. При оценке медиан общей и годичной выживаемости EEG для каждого показания к приёму лекарственного препарата рассчитывали как разницу между данными реальной практики и РКИ. Анализ времени до события проводили с использованием анализа выживаемости Каплана — Мейера. На основе результатов клинических исследований были воссозданы псевдоиндивидуальные данные пациентов путём оцифровки опубликованных кривых общей выживаемости (ОВ). Различия в ОВ между доказательствами, собранными в реальной клинической практике (*англ.* Real-World Evidence; RWE), и данными клинических исследований были представлены в виде коэффициента риска (*англ.* hazard ratio; HR) с 95% доверительным интервалом (ДИ). Также авторы оценили различия в токсичности. Для этого они определяли разницу между частотой госпитализации пациентов в реальной практике и частотой серьёзных нежелательных явлений, наблюдавшихся в РКИ. В обзор вошло 20 вариантов системной терапии с 29 показаниями к её проведению. Авторы работы проанализировали 29 РКИ, в результатах которых либо были представлены медиана ОВ и годичная выживаемость, либо была опубликована кривая общей выживаемости Каплана — Мейера. Данные РКИ включали 8905 пациентов, а для RWE было выявлено в общей сложности 18240 пациентов, получавших парентеральную системную терапию. При сравнении ОВ из 29 проанализированных показаний в 28 (97%) была продемонстрирована более низкая ОВ в реальной практике по сравнению с данными клинических исследований. Более высокую эффективность в реальной практике продемонстрировал только пеметрексед при немелкоклеточном раке лёгкого с годичной выживаемостью 31% по сравнению с 30% в РКИ. Медиана средней разницы ОВ составила 5,2 месяца (межквартильный диапазон (IQR) 3,0-12,1 месяца), что свидетельствует об улучшении выживаемости в когорте клинических исследований. Средняя разница в годичной выживаемости между когортой клинических исследований и RWE составила 11% (IQR, 7-21%). Наибольшие расхождения в показа-

телях выживаемости наблюдались для терапии рака молочной железы и опухолей мочеполовой системы, наименьшие — для рака лёгкого. При анализе подгрупп по линиям терапии большая разница в выживаемости наблюдалась для терапии второй линии (HR, 1,61 [95% ДИ, 1,50-1,74]) по сравнению с терапией первой линии (HR, 1,43 [95% ДИ, 1,35-1,51]). Также было обнаружено, что реальная когорта старше (средний возраст 66 лет [IQR, 62-70 лет]), чем когорта клинических исследований (средний возраст 61 год [IQR, 57-63 года]), а время до прекращения лечения в реальной практике короче в среднем на 1,4 месяца (IQR, 0,4-1,8 месяца). При анализе токсичности большая частота отмечена в реальной практике, средняя разница составила 14% (IQR, 9-22%). Авторы отмечают, что хотя более низкая продолжительность лечения может быть вторична и обусловлена особенностями пациента или токсичностью лечения, она также может отражать более низкую эффективность лечения в реальной практике. В данной публикации подчёркивается важность получения RWE для улучшения консультирования пациентов в отношении эффективности и токсичности системной терапии. Кроме того, выявление разрыва между действенностью и эффективностью подчёркивает необходимость сбора RWE после принятия результатов об эффективности, полученных в РКИ, для постоянной переоценки используемых технологий [8].

Группой *Cramer-van der Welle CM et al.* проведено исследование по оценке EEG для пациентов с мелкоклеточным раком лёгкого. Авторы исследования использовали данные голландского регистра онкологических заболеваний о 568 пациентах с мелкоклеточным раком лёгкого, которые получали лечение в семи крупных больницах. В ходе исследования был рассчитан «фактор Действенность-Эффективность (ДЭ)», представляющий собой отношение наблюдаемой ОВ каждого «реального» пациента (т. е. эффективность) к медиане общей выживаемости для пациентов, получающих аналогичное лечение в РКИ (т. е. действенность). Величина оцениваемого относительно 1,0 фактора ДЭ показывает степень различия ОВ в реальной практике по сравнению с результатами клинического исследования. Также в данном исследовании оценивали токсичность, с целью чего отбирали наблюдения с уменьшением дозы, ранним прекращением лечения и/или переходом на другое лечение в пределах данной линии. Обращает на себя внимание несколько выводов. Во-первых, общая выживаемость в реальных условиях на 21% ниже, чем в РКИ (медиана фактора ДЭ равна 0,79; 95% ДИ от 0,68 до 0,84; p<0,001), что соответствовало абсолютной разнице в медиане ОВ примерно в 2 месяца, причём медиана ОВ была короче для всех режимов терапии. Во-вторых, было показано, что больший EEG наблюдается для пациентов в возрасте ≥65 лет и с количеством баллов по шкале оценки общего состояния онкологического больного (шкала ECOG) ≥2: при анализе величины EEG в зависимости от возраста он составил 9% у пациентов <65 лет (фактор ДЭ 0,91, p=0,045) и 28% у пациентов в возрасте ≥65 лет (фактор ДЭ 0,72, p<0,001; n=319), а при анализе в зависимости от количества баллов по шкале ECOG величина несоответствия между действенностью и эффективностью у пациентов с ECOG 0-1 составила 14% (фактор ДЭ 0,86, p<0,001) против 38% у пациентов с ECOG ≥2 (фактор ДЭ 0,62, p<0,001; n=141). При совпадении обоих факторов (возраст ≥65 лет и количество баллов по шкале ECOG ≥ 2 (n=104)) EEG составил 43%. Как средний возраст, так и доля пациентов с ECOG ≥ 2 в реальном мире были больше, чем в РКИ (66 против 62 лет и 25% против 17%; оба p<0,001). Также показано, что более высокий возраст и ECOG-статус приводили к значительному увеличению токсичности [9]. Данное исследование не только продемонстрировало наличие разрыва между действенностью и эффективностью, но также частично объяснило его этиологию несоответствием характеристик пациентов, включаемых в РКИ и получающих лечение в условиях реального мира. Также авторы отмечают, что результаты их анализа согласуются с выполненным ранее *Povsic et al.* систематическим обзором реальной эффективности лечения мелкоклеточного рака лёгкого [10].

Тремя годами ранее упомянутая выше группа *Cramer-van der Welle C. M. et al.* провела аналогичное исследование для немелкоклеточного метастатического рака лёгкого, в котором были показаны сходные результаты — для всех схем терапии реальная ОВ была короче, чем ОВ, зафиксированная в РКИ. Медиана фактора ДЭ составила 0,77 (95% ДИ 0,70-0,85; p<0,001), причём EEG составил 19% у пациентов с количеством баллов по шкале ECOG 0-1 (фактор ДЭ 0,81, p<0,001) и 61% у пациентов с баллами по шкале ECOG ≥ 2 (фактор ДЭ 0,39, p<0,001; n=140, 12%). Также важно заметить, что в данном исследовании такие переменные, как возраст на момент постановки диагноза, пол, наличие сопутствующих заболеваний, гистологический тип опухоли и год постановки диагноза, не показали статистически значимой связи с EEG. Интересен и ещё один результат: вклад в величину несоответствия между действенностью и эффективностью, который вносит совокупность таких факторов, как более раннее завершение лечения, отсутствие перехода к другим линиям терапии, а также более высокий ECOG-статус пациентов реального мира, составляет лишь 35%, из которых

5,2% приходится на различия в ECOG-статусе. Одним из предположений, которое выдвигают авторы для объяснения EEG, является отбор в клиническое исследование более мотивированных пациентов [11].

Те же авторы провели и сравнение реальной эффективности с действенностью в РКИ для иммунотерапии пембролизумабом и ниволумабом при IV стадии немелкоклеточного рака лёгкого. Сравнивались ОВ и выживаемость без прогрессирования (ВБП) для первой линии иммунотерапии пембролизумабом и второй линии ниволумабом как наиболее часто назначаемых вариантов. Исследователи рассчитали не только фактор ДЭ, но и отношение рисков (HR). При расчёте фактора ДЭ медиана ОВ в реальной практике оказалась на 55% короче для пембролизумаба и на 35% короче для ниволумаба (коэффициент ДЭ 0,45 и 0,65 соответственно). Для пембролизумаба не выявлено существенной разницы в медиане ВБП (фактор ДЭ 0,85), а для ниволумаба медиана ВБП оказалась даже на 61% выше (фактор ДЭ 1,61), чем в РКИ. При расчёте отношения рисков оба варианта иммунотерапии показали сопоставимые результаты ВБП как в реальной практике, так и в исследованиях (HR 1,08 (95% ДИ 0,75-1,55) для пембролизумаба и HR 0,91 (95% ДИ 0,74-1,14) для ниволумаба), однако в реальных условиях ОВ была значительно короче для первой линии пембролизумаба (медиана ОВ 15,8 против 30 месяцев, HR 1,55; 95% ДИ 1,07-2,25). Частично этот разрыв можно объяснить тем, что в реальном мире меньшее количество пациентов переходят к следующей линии лечения (в два раза реже для первой линии и в 1,5 раза для второй линии терапии). Также при поиске объясняющих факторов для EEG установлено, что фактором, положительно коррелирующим с более длительными ВБП и ОВ, является развитие иммуноопосредованных нежелательных явлений, которые чаще наблюдались в условиях РКИ. Показана взаимосвязь более низких результатов лечения с более высоким ECOG-статусом. Это ещё раз подтверждает существующее мнение о том, что в исследования отбираются наиболее «сохранные» пациенты, в то время как пациенты с большими баллами по шкале ECOG в клинической практике также стремятся к получению иммунотерапии, что приводит к усугублению EEG. Кроме того, как полагают авторы, пациенты, получавшие иммунотерапию посредством участия в клиническом испытании, в случае неудачи имеют более высокую внутреннюю готовность искать дальнейшие варианты лечения, а клиники, занимающиеся включением в исследования, с готовностью сообщают им о других экспериментальных методах лечения [12]. До публикации результатов этой работы *Mamtani R, Lund J и Hubbard RA* [1] высказывали предположение, что для иммунотерапии будет наблюдаться выраженный EEG вследствие того, что она является привлекательным вариантом для пожилых и ослабленных пациентов, так как может давать потенциальную долгосрочную пользу, обладает благоприятным профилем токсичности, не метаболизируется в печени и почках, благодаря чему может назначаться пациентам с органной дисфункцией. Результаты описанного исследования не в полной мере подтверждают их прогноз. Вероятно, больший EEG наблюдался бы в случае анализа когорты, получающей исследуемую терапию по программе расширенного доступа. Такие программы предполагают использование экспериментального препарата у пациентов с серьёзными или угрожающими жизни заболеваниями, которые не могут быть включены в клиническое исследование, потому что не соответствуют критериям включения, и которые испробовали все стандартные режимы терапии. В случае проведения подобного анализа для иммунотерапии было бы полезно получить гипотетическое распределение выживания, которое наблюдалось бы, если бы пациенты из реального мира напоминали участников РКИ.

Известно, что две трети противоопухолевых лекарственных препаратов одобрены на основе суррогатных конечных точек, таких как выживаемость без прогрессирования или время до прогрессирования, измеренных в РКИ [13]. Возникают сложности при оценке на их основе ОВ в реальной практике. В качестве примера помощи врачам в преодолении этих трудностей можно привести систему ценностей Американского общества клинической онкологии (ASCO), целью которой является помощь врачам и пациентам в выборе предпочтительных методов лечения [14]. В состав такого его показателя, как «чистая польза для здоровья», входит компонент клинической пользы, в значительной мере зависящий от отношения рисков (HR), измеренного в РКИ. В обновлённой версии системы ценностей ASCO от мая 2016 г. компоненту клинической пользы назначается поправочный коэффициент 0,8 (т. е. снижение количества баллов на 20%) для РКИ с суррогатными конечными точками, такими как ВБП, по сравнению с РКИ, в которых конечной точкой является общая выживаемость [15]. Иными словами, ASCO предполагает, что реальное снижение смертности в результате лечения будет на 20% меньше по величине, чем это заявлено в клиническом испытании, где конечной точкой являлось ВБП [16]. В США при финансовой поддержке компании Novartis группой *Lakdwalla et al.* было проведено сравнение эффективности противоопухолевой терапии в реальном мире с результатами, полученными в условиях РКИ, для пяти наиболее

распространённых локализаций опухолей: рака молочной железы, колоректального рака, рака лёгких, яичников и поджелудочной железы [16]. Особенностью данной работы является то, что она не только продемонстрировала наличие разрыва между действенностью/эффективностью и высказала предположения о его причинах, но и оценила, как величина разрыва соотносится с тем, какие конечные точки (твёрдые или суррогатные) использованы при оценке результатов РКИ. В статье также оценивали, каков был бы этот разрыв, если бы пациенты в реальном мире соответствовали критериям отбора в клиническом испытании. Как и в описанной выше работе, исследователями проведён ретроспективный анализ базы данных SEER системы Medicare. Приведём основные, на наш взгляд, результаты этого масштабного и грамотно спланированного анализа. Для обзора был взят период получения пациентами лечения с 1991 по 2010 год. Обнаружено, что если пациенты в реальном мире подходят для участия в РКИ, то реальная ОВ оказывается сопоставима с ОВ, наблюдаемой в РКИ. Однако для этих пациентов эффективность, основанная на общей выживаемости в реальном мире, будет на 16% ниже, чем действенность, измеренная в РКИ на основе суррогатных конечных точек. Таким образом, для прогнозирования ОВ в реальном мире следует корректировать в меньшую сторону примерно на эту величину показатель действенности, основанный на суррогатных конечных точках. Также показано, что среди исследований, в которых в качестве конечной точки используется ОВ, более надёжно прогнозировали исходы в реальном мире те из них, в которых были большие размеры выборки, использовались более поздние линии терапии, исследование проводилось в США, а ожидаемая продолжительность жизни была меньше (рак поджелудочной железы и лёгкого). «Реальные» пациенты отличались от пациентов в РКИ: исходная когорта включала преимущественно пожилых пациентов со средним возрастом на момент лечения более 70 лет, более половины пациентов имели рак IV стадии на момент постановки диагноза, наблюдался больший процент женщин. В когорте, включавшей всех пациентов, а не только тех, которые имели право на участие в РКИ, коэффициент риска смерти для реального мира был на 8,2% выше (95% ДИ 5,7-10,8%), чем коэффициент риска смерти в РКИ при использовании ОВ в качестве конечной точки, и на 24,9% (95% ДИ 22,0-28,0%), чем в РКИ с использованием суррогатных конечных точек. Также авторы приводят поправочный коэффициент для системы ценностей ASCO, основанный на полученных ими эмпирических данных. Для различных схем лечения он различается как в большую, так и в меньшую сторону [16].

Хотя представленные выше данные не оставляют сомнений в существовании EEG в отношении лечения онкологических заболеваний со стабильно более низкой эффективностью и большей токсичностью, наблюдаемыми в реальном мире, они не в полной мере помогают установить этиологию этого разрыва. Большинство из проанализированных исследований на основании сравнения пациентов обеих групп позволяют заключить, что пациенты реального мира старше, имеют худший ECOG-статус, большее количество сопутствующих заболеваний, чаще досрочно прекращают начатое лечение или получают более низкую дозу лекарственного препарата, реже переходят к другим линиям терапии, среди них выше процентное соотношение женщин, чем среди участников РКИ. Также можно предположить, что эти больные менее комплаентны, не полностью соответствуют показаниям к назначаемой терапии, чаще проживают далеко от крупных медицинских центров с соответствующими ограничениями для получения регулярного наблюдения и медицинской помощи. Кроме того, пациенты в реальном мире, помимо сравниваемых линий лечения, могут получать предшествующую или последующую терапию, возможно отсутствующую у пациентов в РКИ, иметь специфичные для заболевания факторы, наличие которых ограничило бы их участие в клиническом испытании, вследствие чего они могут демонстрировать более низкую эффективность лечения. Требуется дальнейшее изучение этиологии EEG для оценки вклада каждого из возможных факторов в его величину.

Стратегии по улучшению обобщаемости

Хотя этиология EEG до сих пор не вполне ясна, однако уже сейчас, благодаря, в частности, описанным выше исследованиям и полученным в них результатам, можно выделить более значимые компоненты EEG, а также структурировать используемые подходы для оценки вклада некоторых из них. Как показать, какая часть EEG обусловлена различиями характеристик «реальных» пациентов и пациентов — участников РКИ, и оценить долю, приходящуюся на остальные факторы? С этой целью можно использовать методы, позволяющие согласовать характеристики пациентов, включённых в РКИ и оцениваемых в исследованиях реальной практики. Одним из них является метод согласованного скорректированного непрямого сравнения (*англ.* matching adjusted indirect comparison). Изначально данный метод использовался для осуществления непрямых сравнений на основании результатов РКИ двух вариантов вмешательства с идентичным контролем, когда для одного из них доступны индивидуальные данные пациентов, а второй содержит только агрегированные данные. Для реализации этого подхода

используются первичные индивидуальные данные пациентов для одного из РКИ, включённых в сеть доказательств непрямого сравнения, для приведения их в соответствие с характеристиками пациентов, включённых в другое РКИ [17], и затем с помощью процедуры взвешивания сравнивается ответ между двумя РКИ, но уже в идеально сбалансированных популяциях [17, 18]. Применение данной методики позволяет уменьшить гетерогенность популяций. Применительно к оценке вклада характеристик пациентов в клинических исследованиях и реальной практике суть подхода выглядит следующим образом: так как данные по конкретным субъектам в РКИ зачастую недоступны, но у нас есть доступ к данным пациентов реального мира (например, из эпидемиологических баз данных, реестров пациентов или медицинских карт), возможно приведение анализируемой когорты пациентов реального мира в соответствие с когортой РКИ, а затем выполнение расчёта EEG на сбалансированных таким образом группах. В результате выполнения этого действия мы получим величину EEG, которая наблюдалась бы, если бы реальная популяция соответствовала популяции участников РКИ. Имея величину EEG, рассчитанную для исходного состава реальной популяции, не составляет труда вычислить разницу между этими показателями, благодаря чему можно оценить, какая часть EEG обусловлена различиями характеристик «реальных» пациентов и пациентов — участников РКИ, то есть какая часть разрыва связана с критериями отбора в РКИ. Также можно предположить, что оставшаяся часть разрыва во многом опосредована преимуществами, получаемыми в результате участия в испытании, например большей приверженностью лечению или тщательному мониторингу токсичности.

Альтернативой для преодоления недоступности информации о конкретном испытуемом являются методы воссоздания данных о выживаемости на основе опубликованных в результатах РКИ кривых Каплана — Мейера [19]. На основе этих графиков получают псевдоиндивидуальные данные пациентов с последующим прямым сравнением результатов терапии в реальной практике и в исследованиях с использованием метода регрессии выживаемости для оценки величины «испытательного» эффекта. Разница в выживаемости между пациентами в РКИ и в реальной практике после поправки на характеристики пациентов представляет собой часть EEG, которая не может быть объяснена различиями больных [20].

Но так ли необходимо оценивать EEG? Возможно, было бы лучше повысить обобщаемость результатов РКИ? Рассмотрим несколько подходов для приближения к этой цели с присущими им ограничениями. Конечно, лучшим способом повысить внешнюю валидность исследований было бы создание репрезентативной выборки субъектов в исследовании [21]. Однако для этого требуется, во-первых, заранее знать точный состав интересующей популяции, а во-вторых, иметь доступ ко всем представителям данной популяции. Даже если бы эти критерии были соблюдены, создание случайной выборки из всей популяции было бы дорогостоящим и оказалось бы возможным лишь в том случае, если бы все выбранные субъекты были готовы к участию. В противном случае преимущества случайной выборки теряются [22], так как типы субъектов, которые соглашаются на участие, могут значительно отличаться от тех, кто в итоге будет получать вмешательство в реальной практике. Также возможно создание целенаправленной выборки на основании набора характеристик, присущих целевой популяции. Но и тут есть нюанс: если исследуемое вмешательство разрабатывается для одной целевой популяции, впоследствии лица, принимающие решения, могут быть заинтересованы в распространении результатов данного исследования на другие популяции [23]. То есть создание идеально обобщаемой выборки в РКИ изначально представляется невыполнимой задачей.

Помимо проведения исследований по оценке EEG, существует ряд аналитических стратегий, разработанных для прогнозирования эффектов вмешательства на популяции после проведения исследования (или, чаще, серии исследований). Некоторые из этих стратегий состоят в простом документировании различных факторов, связанных с обобщаемостью. Другие, в том числе постстратификационные методы, методы, связанные с метаанализом, и метод повторного взвешивания, на самом деле нацелены на выполнение поставленной задачи. Постстратификация является наиболее простым и распространённым способом оценки эффектов лечения в популяции посредством разбиения выборки на подгруппы по результатам рандомизированного исследования. Она оценивает эффекты лечения для конкретных подгрупп в исследовании, а затем усредняет их, используя пропорции населения, чтобы оценить средний эффект для реальной популяции. К примеру, если в исследовании участвовали 20% женщин и 80% мужчин, а в интересующей популяции соотношение составляло бы 50/50, последующая стратификация брала бы равновзвешенное среднее специфических оценок эффекта для мужчин и женщин из исследования для ответа на вопрос, каким будет эффект в популяции. Этот метод очень прост, но ограничен количеством переменных, для которых могут быть сделаны корректировки. Постстратификационные ячейки (подгруппы) могут оказаться очень маленькими даже при попытке постстратификации по основным демографическим данным,

таким как пол, раса, этническая принадлежность и возрастные группы. Другой набор методов полезен, когда по определённой теме доступно несколько исследований. К ним относятся метаанализ, подходы к синтезу исследований и методы взвешивания. Целью метаанализа является объединение результатов нескольких сходных исследований, однако он уделяет мало внимания тому, насколько пациенты в исследованиях похожи на «реальных». Исследовательский синтез отвечает той же цели, что и метаанализ, но потенциально способен решить проблему обобщаемости [23]. Он позволяет объединить результаты различных типов исследований, например информацию об эффектах программы из рандомизированного исследования с информацией из наблюдательного (обсервационного) исследования, которое может содержать более репрезентативную выборку [24]. То есть использование этого метода может приблизить нас к поставленной цели — объединить данные реальной практики и результаты РКИ. А методы взвешивания могут помочь оценить эффекты терапии на уровне популяции на основе данных РКИ. При этом подходе испытуемые взвешиваются, чтобы они походили на популяцию. Также можно использовать коэффициенты склонности (*англ.* propensity score) [25] для обобщения и оценки сходства рандомизированной пробной выборки и целевой популяции. Этот аналитический подход обеспечивает суммарную оценку различий в наблюдаемых характеристиках индивидуумов в исследовании и в популяции. Затем можно использовать методы коррекции коэффициентов склонности, чтобы, во-первых, изучить, насколько хорошо эти характеристики отражают различия между выборкой и генеральной совокупностью, а затем оценить, каковы будут эффекты для генеральной совокупности [23]. Данный метод похож на постстратификацию, но позволяет учесть больше ковариат. Как мы видим, описанные подходы к улучшению обобщаемости результатов контролируемых исследований достаточно сложны, не всегда применимы и, в конце концов, даже при их использовании полученная информация будет теоретической и потребует проверки в исследованиях реальной клинической практики.

Однако на основании описанных аналитических стратегий по прогнозированию эффективности на целевой популяции можно было бы разработать методы для детальной оценки вклада различных факторов в величину EEG. Например, упомянутый выше подход, основанный на применении коэффициентов склонности, отлично подходит для сравнения результатов терапии в двух наблюдательных исследованиях. Кратко опишем суть данного метода: при сравнении двух способов лечения в эксперименте путём рандомизации мы делим испытуемых на две группы, каждая из которых получает свой вариант терапии. Поскольку можно ожидать, что состав двух групп будет идентичен, имеет смысл проведение прямого сравнения ответа на терапию. В наблюдательных же исследованиях результаты, полученные в ходе прямого сравнения, будут вводить в заблуждение, так как состав групп неслучаен и может существенно различаться (под коэффициентом склонности в нашей ситуации понимается условная вероятность назначения индивидууму определённого варианта лечения на основании присутствия у него тех или иных ковариат). Чтобы преодолеть эту трудность, *Rosenbaum & Rubin* (1983 г.) [25] определили балансирующие баллы, которые могут быть использованы для построения подгрупп похожих людей в каждой группе лечения. Ожидается, что сравнения подгрупп дадут хорошую информацию о разнице между методами лечения [26]. Кроме того, методы взвешивания можно использовать для количественной оценки соотношения между действенностью и эффективностью, связанного с конкретными характеристиками пациента, такими как возраст или количество баллов по шкале ECOG. Путём взвешивания может быть скорректировано распределение времени выживания, наблюдаемое для реальных пациентов, для получения гипотетического распределения выживания, которое бы наблюдалось, если бы пациенты из реального мира напоминали участников клинических исследований [1]. Таким образом, взвешивание распределения времени выживания, наблюдаемого у реальных пациентов, можно использовать для согласования распределения характеристик пациентов на основе представленных в результатах клинических исследований средних значений с использованием описанного выше метода согласованного скорректированного непрямого сравнения [18].

Вклад в величину EEG, вероятно, может вносить и влияние вмешивающихся факторов (или конфаундинг-факторов). Данным термином описывается связь между двумя и более причинными факторами, наблюдаемыми в наборе данных, в котором логически невозможно разделить вклады отдельных причинных факторов на изучаемый исход [5]. Влияние конфаундинг-факторов — это ситуация, когда мера связи между воздействующим фактором и исходом искажена из-за взаимосвязи воздействующего (изучаемого) фактора с другим фактором, иногда неизвестным исследователю, также влияющим на исход. Например, когда на эффект фармакологического лечения могут оказывать влияние раса или генетические факторы, которые могут быть известны или нет [27]. Применительно к пациентам, получающим противоопухолевую терапию в реальных условиях, можно привести следующие примеры: пациент мог досрочно прекратить лечение, как вследствие низкой приверженности, так

и из-за развития нежелательных реакций; исчезнуть из системы медицинского наблюдения по причине как прекращения обращения за медицинской помощью, так и смерти или смены места жительства; перейти на приём другого лекарственного препарата, как из-за неэффективности терапии, так и вследствие недоступности принимаемого ранее лекарственного средства в его регионе или развития нежелательных реакций. При необходимости оценить вклад факторов, имеющих конфаундеры, при разработке методологии исследования можно использовать подход ограничения на включение пациентов или подбор сходных пар [5]. До тех пор пока не будет произведена корректировка или поправка на конфаундер, его влияние невозможно отличить от влияния изучаемого фактора воздействия.

Благодаря полученным с помощью вышеописанных методов данным можно оценить, какая часть EEG обусловлена различиями характеристик «реальных» пациентов и пациентов — участников РКИ, то есть какая часть разрыва связана с критериями отбора в РКИ, какая приходится на преимущества, получаемые в результате участия в испытании, а какая опосредована другими факторами.

В дополнение к методам, которые учитывают различия в характеристиках пациентов, необходимы соответствующие методы для устранения различий в качестве и полноте данных. Сбор данных в клинических исследованиях осуществляется очень тщательно, в то время как в электронных медицинских картах или иных источниках данных реального мира ключевые показатели, например ECOG-статус, могут отсутствовать или регистрироваться не с таким постоянством, как в РКИ. Для преодоления ограничений, связанных с качеством данных, необходимы методы, которые учитывают отсутствующие или подверженные ошибкам ковариаты [1]. Для достижения этой цели возможно проведение анализа чувствительности, то есть оценки влияния возможных неточностей на исход основного анализа [5]. В электронных медицинских картах могут отсутствовать данные вследствие ухода пациентов из системы здравоохранения или отсутствия регулярной оценки (например, для выявления прогрессирования). Систематическое отсутствие сбора данных о смертности приводит к переоценке выживаемости по сравнению с реальными данными [28], что может привести к недооценке величины EEG. Это не имеет практического значения при онкологических заболеваниях с низкой выживаемостью, но если выживаемость более высокая, отсутствие данных может привести к необъективности оценок вследствие цензурирования пациентов. Во избежание этой систематической ошибки, следует тщательно продумывать, когда подвергать цензуре пациентов, вышедших из-под наблюдения [29].

Если постоянно проводить анализ данных реального мира, сравнения действенности и эффективности, а также влияния на результаты факторов, удастся накопить большой массив данных, охватывающий всевозможные целевые популяции в различных регионах. На основании постоянно обновляющихся сведений возможно будет создание статистической модели, непрерывно обучающейся на вновь появляющихся данных. С её помощью можно будет мгновенно рассчитать предполагаемую эффективность лечения в зависимости от характеристик пациентов. Если заключить полученную модель в регулярно обновляющуюся компьютерную программу с понятным интерфейсом, это поможет клиницистам в принятии решений и, возможно, внесёт больше определённости в жизнь пациентов.

Заключение

Хотя клинические исследования остаются «золотым стандартом» оценки действенности терапии, их дизайн не позволяет учесть обстоятельства применения лечебных схем в условиях реальной клинической практики, в связи с чем для широких групп населения эффективность лечения зачастую сложно предугадать. Несмотря на положительные результаты, продемонстрированные в исследованиях, возможна ситуация, когда применение вмешательства в реальной клинической практике не приведёт к улучшению исходов для «реальных» пациентов, особенно если они сильно отличаются от испытуемых по ряду характеристик. В статье описаны аналитические стратегии, которые можно использовать для оценки EEG и вклада различных факторов в его величину. На наш взгляд, улучшить информированность клиницистов при принятии решений можно путём применения следующих последовательных шагов:

1) Проводить надлежащим образом спланированные пострегистрационные исследования на реальных популяциях пациентов в отношении как можно большего числа подходов к терапии онкологических заболеваний в различных регионах.
2) На основании данных, полученных в этих исследованиях, а также в РКИ, производить оценку разрыва между действенностью и эффективностью.
3) Оценивать, какая доля от общего показателя разрыва приходится на различия в характеристиках пациентов, и выдвигать предположения о том, какие факторы вносят наибольший вклад в долю разрыва, оставшуюся после исключения их несоответствия.
4) Разрабатывать, по возможности, методологические подходы для оценки вклада этих факторов в общий уровень разрыва.

5) Разрабатывать и применять методы для учёта отсутствующих данных.
6) Использовать методы, позволяющие нивелировать возможные ошибки, например анализ чувствительности.
7) Учитывать возможность наличия конфаундинг-факторов и использовать методы борьбы с ними.
8) На основе постоянно обновляющихся данных составить модель для расчёта вероятности ответа на терапию.

Таким образом, имея перед глазами вышеперечисленную совокупность данных, а также оценивая особенности пациента и другие вероятные влияющие факторы в каждом конкретном случае, онколог сможет предугадать, насколько эффективным окажется лечение у конкретного больного.

ДОПОЛНИТЕЛЬНАЯ ИНФОРМАЦИЯ

Конфликт интересов. Авторы декларируют отсутствие конфликта интересов.

Участие авторов. Усманова Т. А. — поиск литературы, анализ, написание статьи; Вербицкая Е. В. — редактирование статьи.

СВЕДЕНИЯ ОБ АВТОРАХ

Усманова Татьяна Андреевна — клинический ординатор кафедры клинической фармакологии и доказательной медицины, ФГБОУ ВО ПСПбГМУ им. И. П. Павлова Минздрава России, Санкт-Петербург, Россия

Автор, ответственный за переписку
email: smirnovatatyana01@gmail.com,
https://orcid.org/0000-0003-2827-1382

Вербицкая Елена Владимировна — доцент кафедры клинической фармакологии и доказательной медицины, зав. отделом биомедицинской статистики, ФГБОУ ВО ПСПбГМУ им. И. П. Павлова Минздрава России, Санкт-Петербург, Россия

email: elena.verbitskaya@gmail.com,
https://orcid.org/0000-0003-3770-993X

Литература

1. Mamtani R, Lund J, Hubbard RA. Considering the totality of evidence: Combining real-world data with clinical trial results to better inform decision-making. *Pharmacoepidemiol. Drug Saf.* Jun. 01, 2021;30 (6):814-6. doi: 10.1002/pds. 5218
2. Rothwell PM. External validity of randomised controlled trials: «to whom do the results of this trial apply? *Lancet* (London, England). Jan. 2005;365 (9453):82-93. doi: 10.1016/S0140–6736 (04) 17670-8
3. Unger JM et al. Comparison of survival outcomes among cancer patients treated in and out of clinical trials. *J. Natl. Cancer Inst*. 2014;106 (3). doi: 10.1093/JNCI/DJU002
4. European Commission. Enterprise and Industry Directorate General for Health & Consumers. High level pharmaceutical forum 2005-2008: conclusions and recommendations. 2008.
5. Исследования реальной клинической практики / А. С. Колбин, Д. Ю. Белоусов, С. К. Зырянов, В. В. Омельяновский, Д. А. Сычёв, А. Л. Хохлов, С. Л. Плавинский, Б. К. Романов, М. В. Журавлёва, Е. В. Вербицкая, Е. А. Вольская, Д. А. Рождественский, С. В. Глаголев, М. Ю. Фролов, А. В. Рудакова, А. В. Павлыш, Ю. Е. Балыкина, А. А. Курылёв, А. В. Иванов, Д. С. Козлов, С. Б. Васильченко, Т. И. Галимов, О. А. Логиновская, С. А. Мишинова, Т. А. Гольдина, В. А. Булатов, Е. В. Лаврентьева, В. В. Горин, Л. А. Худова, В. В. Секачёв, И. В. Ефименко, Л. И. Карпенко, А. А. Поликарова. — М.: Издательство ОКИ: Буки Веди, 2020. — 208 с.: ил. — ISBN 978-5-4465-2902-5. — Доступно по: https://izdat-oki.ru/issledovaniya-realnoj-klinicheskoj-praktiki. [Kolbin AS et al. Real-world data study. — M.: Publishing House OKI: Buki Vedi, 2020. — 208 p.: ill. — ISBN 978-5-4465-2902-5].
6. Филиппова А. В., Колбин А. С. Соответствие результатов клинических исследований и реальной медицинской практики. *Учёные записки СПбГМУ им. И. П. Павлова*. May 2018;25 (1):7-14. doi: 10.24884/1607-4181-2018-25-1-7-14
7. Eichler H-G et al. Bridging the efficacy — effectiveness gap: a regulator»s perspective on addressing variability of drug response. *Nat. Rev. Drug Discov*. Jul. 2011;10 (7):495-506. doi: 10.1038/nrd3501
8. Phillips CM et al. Assessing the efficacy-effectiveness gap for cancer therapies: A comparison of overall survival and toxicity between clinical trial and population-based, real-world data for contemporary parenteral cancer therapeutics. *Cancer.* Apr. 2020;126 (8):1717-26. doi: 10.1002/CNCR. 32697
9. Cramer-van der Welle CM et al. Systematic evaluation of the efficacy-effectiveness gap of systemic treatments in extensive disease small cell lung cancer. *Pharmacoepidemiol. Drug Saf.* Apr. 2021;30 (4):445-50. doi: 10.1002/pds. 5179
10. Povsic M, Enstone A, Wyn R, Kornalska K, Penrod JR, Yuan Y. Real-world effectiveness and tolerability of small-cell lung cancer (SCLC) treatments: A systematic literature review (SLR). *PLoS*

One. Jul. 2019;14 (7). doi: 10.1371/JOURNAL.PONE. 0219622

11. Cramer-van der Welle CM, Peters BJM, Schramel FMNH, Klungel OH, Groen HJM, Van De Garde EMW. Systematic evaluation of the efficacy-effectiveness gap of systemic treatments in metastatic nonsmall cell lung cancer. *Eur. Respir. J.* 2018;52 (6). doi: 10.1183/13993003.01100-2018
12. Cramer-van der Welle CM et al. Real-world outcomes versus clinical trial results of immuno-therapy in stage IV non-small cell lung cancer (NSCLC) in the Netherlands. *Sci. Rep*. Dec. 2021;11 (1). doi: 10.1038/S41598-021-85696-3
13. Kim C, Prasad V. Cancer Drugs Approved on the Basis of a Surrogate End Point and Subsequent Overall Survival: An Analysis of 5 Years of US Food and Drug Administration Approvals. *JAMA Intern. Med.* Dec. 2015;175 (12):1992-4. doi: 10.1001/JAMAINTERNMED. 2015.5868
14. Schnipper LE et al. American Society of Clinical Oncology Statement: A Conceptual Framework to Assess the Value of Cancer Treatment Options. *J. Clin. Oncol*. Aug. 2015;33 (23):2563-77. doi: 10.1200/JCO. 2015.61.6706
15. Schnipper LE et al. Updating the American Society of Clinical Oncology Value Framework: Revisions and Reflections in Response to Comments Received. *J. Clin. Oncol.* Aug. 2016;34 (24):2925-33. doi: 10.1200/JCO. 2016.68.2518
16. Lakdawalla DN et al. Predicting Real-World Effectiveness of Cancer Therapies Using Overall Survival and Progression-Free Survival from Clinical Trials: Empirical Evidence for the ASCO Value Framework. *Value Health.* Jul. 2017;20 (7):866-75. doi: 10.1016/J. JVAL. 2017.04.003
17. Signorovitch JE et al. Comparative effectiveness without head-to-head trials: A method for matching-adjusted indirect comparisons applied to psoriasis treatment with adalimumab or etanercept. *Pharmacoeconomics.* 2010;28 (10):935-45. doi: 10.2165/11538370-000000000-00000
18. Signorovitch JE et al. Matching-adjusted indirect comparisons: a new tool for timely comparative effectiveness research. *Value Health.* Sep. 2012;15 (6):940-7. doi: 10.1016/J. JVAL. 2012.05.004
19. Guyot P, Ades AE, Ouwens MJNM, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. *BMC Med. Res. Methodol*. Feb. 2012;12. doi: 10.1186/1471-2288-12-9
20. Webster-Clark MA, Sanoff HK, Stürmer T, Peacock Hinton S, Lund JL. Diagnostic Assessment of Assumptions for External Validity: An Example Using Data in Metastatic Colorectal Cancer. *Epidemiology.* Jan. 2019;30 (1):103-11. doi: 10.1097/EDE. 0000000000000926
21. Braslow JT, Duan N, Starks SL, Polo A, Bromley E, Wells KB. Generalizability of studies on mental health treatment and outcomes, 1981 to 1996. *Psychiatr. Serv*. Oct. 2005;56 (10):1261-8. doi: 10.1176/APPI. PS. 56.10.1261
22. Shadish WR. The logic of generalization: Five principles common to experiments and ethnographies. *Am. J. Community Psychol*. Jun. 1995;23 (3):419-28. doi: 10.1007/BF02506951
23. Stuart EA, Bradshaw CP, Leaf PJ. Assessing the generalizability of randomized trial results to target populations. *Prev. Sci*. Apr. 2015;16 (3):475-85. doi: 10.1007/S11121-014-0513-Z
24. Imai K, King G, Stuart EA. Misunderstandings between experimentalists and observationalists about causal inference. *J. R. Stat. Soc. Ser. A.* Apr. 2008;171 (2):481-502. doi: 10.1111/J. 1467-985X. 2007.00527. X
25. Rosenbaum PR. Model-Based Direct Adjustment. *J. Am. Stat. Assoc*. Jun. 1987;82 (398):387. doi: 10.2307/2289440
26. Danielsson S. The propensity score and estimation in nonrandom surveys-an overview. *Mod. Stat. Surv. Methods.* 2002;1 (2000):1-8. Accessed: May 11, 2022. [Online]. Available: https://www.semanticscholar.org/paper/The-propensity-score-and-estimation-in-nonrandom-Danielsson/9904c874c40eb5b90a018ad2ccf93be58af8f65c.
27. Miquel S, Porta OUP. International Epidemiological Association, Dictionary of Epidemiology — Oxford Reference. 2008.
28. Carrigan G et al. An evaluation of the impact of missing deaths on overall survival analyses of advanced non-small cell lung cancer patients conducted in an electronic health records database. *Pharmacoepidemiol. Drug Saf.* May 2019;28 (5):572-81. doi: 10.1002/PDS. 4758
29. Lesko CR, Edwards JK, Cole SR, Moore RD, Lau B. When to Censor? *Am. J. Epidemiol*. Mar. 2018;187 (3):623–32. doi: 10.1093/AJE/KWX281 MyRWD