

Уважаемые читатели и авторы!

Мы благодарны всем Вам за участие в нашем журнале. Авторам – за ваши работы, читателям – за чтение, причем на самом деле сложно сказать, что из этого важнее и что больше мотивирует всех нас. Анализируя статьи, опубликованные в нашем журнале в прошлом году, обращает на себя внимание порой различный подход авторов к методологии, оценке полученных результатов, размеру выборок, выбывания и др. Оценка результатов исследования всегда базируется на трех вопросах:

- 1) кого оцениваем?
- 2) что и как оцениваем?
- 3) как долго оцениваем?

Сейчас 2020 г., эра big-data и искусственного интеллекта, но эти «банальные» вопросы по-прежнему актуальны. С другой стороны, мы не уникальны. В зарубежных высокорейтинговых журналах в 2000-х и первой половине 2010-х гг. была опубликована целая серия редакционных статей, в которых рассматривались типичные дефекты поступающих и публикуемых работ. Нам тоже придется пройти этот путь, хотя и с 10-летним опозданием.

Кого оцениваем? К этому вопросу можно отнести внутреннюю валидность (критерии включения и исключения), размер выборки и количество выбывших из исследования. На последнем остановимся подробнее. Пациенты выбывают из исследований почти всегда. О некорректном выбывании, снижающем ценность исследования, можно говорить в двух случаях: когда процент выбывших из сравниваемых групп весьма различается, либо когда выбывают пациенты, отличающиеся от тех, которые остаются. Среди выбывших пациентов могут быть не только те, которые асимптомны, полностью удовлетворены и просто не хотят вспоминать и возвращаться, но и, наоборот, абсолютно разочарованные. И мы никогда не знаем, каких пациентов в группе выбывших на самом деле больше.

Какой процент выбывших можно расценить как приемлемый? На этот вопрос нет четкого ответа [1]. Но чаще пользуются правилом, что выбывание <5% приводит к малой предвзятости, а выбывание >20% означает серьезную угрозу валидности [2].

В любом случае нужно понимать, что даже небольшой процент выбывания может серьезно

ограничивать исследование [3], а иногда и выбывание <20% может быть критичным. Есть методологическое правило подозревать в каждом выбывшем случае наихудший сценарий [2].

Говоря о количестве пациентов, включаемых в исследование, в наших работах в подавляющем большинстве случаев прослеживается пассивный подход: включают столько пациентов, сколько было на момент решения авторов подвести черту и начать анализ. В результате по некоторым сравниваемым параметрам получаются достоверные различия, а по другим – недостоверные. Но на самом деле ли они недостоверны (ошибка второго рода, β -ошибка), или при увеличении числа пациентов достоверность все-таки появится? Избежать этого позволяют элементарные онлайн-калькуляторы размера выборок (Sample Size Calculator). Но в поступающих в журнал работах авторы редко задаются вопросом: достаточно ли пациентов они включили в свое исследование и не публикуют ли его слишком рано?

Что и как оцениваем? К этому вопросу можно отнести проблему определения конечных точек, валидности используемых шкал и др.

Конечной точкой называют такой результат, при достижении которого участники, как правило, исключаются из дальнейшего исследования. «Суррогатная» конечная точка – это критерий, который сам по сути не является желаемым результатом лечения. «Суррогатной» конечной точкой может быть нормализация какого-то лабораторного показателя крови, который на самом деле может не иметь особого значения для пациентов, хоть и быть доказанным фактором риска чего-то.

Важно не ошибиться с выбором конечных точек. К примеру, авторы решили изучить результаты лечения диафизарных переломов и выбрали классические конечные точки – частоту сращений/несращений и амплитуду движений в смежных суставах, а результаты оценивали в течение полугода или года после операции. Такой подход нам был бы понятен 40–70 лет назад, но верно ли определили авторы конечную точку в своем исследовании? Что нам дает эта работа в 2020 г.? У диафизарных переломов, помимо этих конечных точек, есть еще одна важная конечная точка –

артроз смежных суставов. Конечно, для ее оценки нужно намного более длительное наблюдение, но что мешает включить в работу еще и «суррогатный» критерий отклонения механических осей? Каковы границы допустимых смещений? Конечно же, чем больше смещение, тем хуже, но нельзя сказать, что нам все известно в этом вопросе. Если бы в такую работу был включен «суррогатный» критерий отклонения осей, остаточного смещения, и было бы дано публичное обязательство на страницах журнала проследить реальное развитие артроза через 5–7 лет, да даже через 2–3 года, то такая работа была бы гораздо ценнее, а сами трудозатраты авторов на такую серию работ навряд ли были велики.

Приведем еще один пример проблемы выбора конечной точки. Оценивая эффективность пероральных антикоагулянтов в качестве конечной точки, логично выбрать тромбоземболическое событие, и уже почти 15 лет так и происходит. Но еще в конце 2010-х гг. некоторые ортопеды отметили рост инфекционных осложнений после эндопротезирования на фоне пероральных антикоагулянтов, но достоверность роста частоты инфекций после эндопротезирования не была доказана ввиду одноцентрового характера исследований и небольшой частоты инфекций в целом. Понятно, что решить эту проблему можно было бы только в рамках очень крупного многоцентрового исследования, которое было начато только в 2016 г., спустя 10 лет после начала эры пероральных антикоагулянтов, и будет закончено в 2021 г. (PEPPER Trial, 25 тысяч пациентов [4]).

Проблема выбора конечных точек на самом деле не так проста, не нова, по-прежнему актуальна, и в ее признании нет ничего постыдного. Мы не уникальны в проблеме выбора конечных точек и проходим тот же путь, что и зарубежные коллеги: в официальном заявлении с изложением позиции 2012 г. AAOS отмечает: «...система оценки результатов и качества лечения должна развиваться и применяться шире. Текущее поколение критериев оценки больше оценивает сам процесс лечения или административные вещи... не доказана их корреляция с улучшением функциональных исходов, которые и интересуют пациентов в первую очередь» [5]. Например, почти все регистры эндопротезирования являются имплантат-ориентированными, а не пациенто-ориентированными, не позволяют оценить функциональный результат, и эта парадигма должна меняться [6]. Конечно же, это не может произойти быстро.

Скажем банальное – для оценки функциональных результатов используются шкалы. Некоторые больницы используют оценку по шкалам у всех пациентов, вне зависимости от того, включаются они в какое-либо исследование или нет. Такой подход

порой вызывает раздражение у практических врачей, например, американские ортопеды отмечают, что это «сложно, громоздко, долго и дорого» [7]. Да, это направление буксует и у нас, но цифровизация истории болезни когда-нибудь все равно будет «дружелюбной», и этот вопрос рутинной шкальной оценки решится.

Здесь важно уделить внимание тому, какую же шкалу (шкалы) мы будем использовать. Весьма отраднo все чаще видеть на страницах нашего журнала работы, посвященные кросс-культурной адаптации шкал. Работы по кросс-культурной адаптации шкал, несомненно, важны, но не всегда понятно, почему авторы выбрали именно эту шкалу. Хотелось бы прийти в отечественном ортопедическом сообществе к единому пониманию, какие именно шкалы стоит использовать. Так или иначе, мы приходим к единому цифровому контуру медицинской документации, и для «совместимости» оцениваемых в будущем данных эту работу по унификации шкал нужно всячески поддерживать и продолжать.

Как долго оцениваем? Выше этот вопрос уже затрагивался, когда мы говорили об оценке результатов лечения после диафизарных переломов. Совершенно нельзя принять, когда авторами выбирается недопустимо короткий срок наблюдения при том, что в других работах по аналогичной проблеме уже указывалось на важность отдаленных последствий.

С другой стороны, отсутствие других работ по изучению долгосрочных результатов не означает, что изучаемый вопрос имеет только краткосрочную актуальность. Например, после первой цементной вертебропластики, выполненной P. Galibert с коллегами во Франции в 1987 г. [8], возник энтузиазм, показания к вертебропластике были значительно расширены, и в течение долгих лет авторами публиковались впечатляющие положительные результаты со сроком наблюдения 1–2 года. Однако более отдаленные результаты выявили серьезные негативные последствия этой технологии и значительно уменьшили область ее применения.

Очень важным является вопрос преемственности в оценке результатов лечения ортопедической патологии. Формально в нашей стране «детская» и «взрослая» ортопедия являются одной специальностью, фактически только в ряде учреждений они существуют вместе, а большинству – живут отдельной жизнью. Отдаленные результаты лечения детей с ортопедической патологией в лучшем случае ограничиваются 18 годами, что абсолютно недостаточно для понимания эффективности используемых методов лечения при достаточно большой группе заболеваний. В этом может помочь неминуемая цифровизация медицинской документации, и через годы или десятилетия мы наконец-то

сможем получать доступ к детским рентгеновским снимкам взрослого пациента всегда, а не только в тех случаях, когда у ребенка есть грубая ортопедическая проблема и вся история снимков тщательно хранится родителями. Например, можно вспомнить о проблеме недиагностированных малых форм эпифизолиза головки бедренной кости и о его возможном влиянии на развитие САМ-импинджмента через десятилетия.

Важно, чтобы исследователи всегда четко понимали, насколько изучаемые ими вопросы соотносятся с продолжительностью их наблюдения, и в этом вопросе полезнее излишняя критичность, которая будет опровергнута временем, чем преждевременная уверенность.

Как интерпретировать результаты? В этом вопросе мы хотим уделить внимание статистической и клинической значимости, которые, к сожалению, не всегда понимаются авторами.

Если получена статистическая значимость, то перед выводом нужно задать себе вопрос: оценивалась истинная конечная точка, или достоверные различия были получены по суррогатному критерию? Например, авторы изучали частоту вывихов после эндопротезирования различными доступами и обнаружили, что при одном из доступов антеверсия чашки статистически значимо отличается от антеверсии при другом доступе. В данном случае конечная точка – именно вывих, а не антеверсия, хотя она (точнее ее отсутствие) считается достоверным фактором риска вывиха. Вывод о том, что в группе с таким-то доступом повышается риск вывиха, будет ложным. Или другой пример – асимптомные малые тромбозы вен голени после операции, погоня за которыми на самом деле не так и важна для пациента, а попытки исправить их ведут к росту тех же геморрагических осложнений. К сожалению, эта банальная ошибка выводов на основе «суррогатного» критерия в той или иной альтерации встречается нередко.

Еще один важный момент – а важны ли конкретному пациенту достоверные преимущества, полученные в результате исследования? Здесь вспоминается крупное исследование с несколькими тысячами пациентов, у которых изучалось использование пробиотиков для коррекции гастроинтестинальных симптомов при антибиотикотерапии. В результате оказалось, что длительность симптомов в группе пробиотиков статистически значимо меньше на 0,4 суток. Насколько важно это пациенту на самом деле? Авторы того исследования считали, что важно.

Закономерен вопрос: что следует из приведенных примеров, и каковы будут пожелания для дальнейшей работы. Следует заметить, что эра локального изучения нескольких историй болезней в архиве уверенно уходит в прошлое и остается актуальной только для очень узких вопросов. Решение – за регистрами и цифровизацией, которые позволят более достоверно оценить результаты лечения и выявить их слабые стороны. В то же время на пути «регистравости» и цифровизации нам нужно решать проблему оценки функциональных результатов и важных для пациента результатов вообще. Формулируя выводы, каждый исследователь должен задать себе вопрос – насколько клинически значимы полученные результаты, действительно ли этот эффект стоил затрат, и не строится ли клинический вывод на основе «суррогатной конечной точки». Помимо этого, авторы должны отдавать себе отчет, что полученные результаты – это результаты сегодняшнего дня, и они могут измениться. При планировании исследования всегда нужно задаваться вопросом: достаточны ли сроки наблюдения для конкретного изучаемого параметра? С другой стороны, не нужно бояться изложить результаты как предварительные с обязательством изучить и более отдаленные исходы.

Литература

1. Sacket D.L., Richardson W.S., Rosenberg W. Evidence-Based Medicine: How to Practice and Teach EBM. New York: Churchill Livingstone; 2000. 261 p.
2. Dettori J.R. Loss to follow-up. *Evid Based Spine Care J.* 2011 Feb; 2(1): 7–10. doi: 10.1055/s-0030-1267080.
3. Bhandari M., Guyatt G.H., Swiontkowski M.F. User's guide to the orthopaedic literature: how to use an article about a surgical therapy. *J Bone Joint Surg Am.* 2001;83(6):916–926. doi: 10.2106/00004623-200106000-00015.
4. Comparative Effectiveness of Pulmonary Embolism Prevention After Hip and Knee Replacement (PEPPER). <https://clinicaltrials.gov/ct2/show/NCT02810704>
5. American Academy of Orthopaedic Surgeons. Position statement. Public reporting of provider performance. <https://www.aaos.org/globalassets/about/position-statements/1183-public-reporting-of-provider-performance.pdf>.
6. Franklin P.D., Allison J.J., Ayers D.C. Beyond joint implant registries: a patient-centered research consortium for comparative effectiveness in total joint replacement. *JAMA.* 2012;308:1217-1218. doi: 10.1001/jama.2012.12568.
7. Ayers D.C., Bozic K.J. The Importance of Outcome Measurement in Orthopaedics. *Clin Orthop Relat Res.* 2013; 471(11):3409-3411. doi: 10.1007/s11999-013-3224-z.
8. Galibert P., Deramond H., Rosat P., Le Gars D. [Preliminary note on the treatment of vertebral angioma by percutaneous acrylic vertebroplasty]. *Neurochirurgie.* 1987;33(2):166-168. (In French).

член редколлегии д-р мед. наук А.П. Середя

главный редактор профессор Р.М. Тихилов

заместитель главного редактора д-р мед. наук И.И. Шубняков