

Анализ выживаемости или событийный анализ: типовые ошибки ретроспективного метода

С. М. Куликов, Е. Н. Паровичникова, В. Г. Савченко

РЕФЕРАТ

Survival or time-to-event analysis: common pitfalls of retrospective approach

S. M. Kulikov, E. N. Parovitchnikova, V. G. Savchenko

SUMMARY

Retrospective analysis of the archival clinical data collected outside the controlled trials is the main source of false conclusions due to data manipulation. This flawed method produces common errors especially in case if life duration is used as an endpoint in survival analysis. It is pointed out that the mixing of searching and/or descriptive-promoting goals with the evidence aimed conclusions is incorrect and should be avoided. The main feature of evidence based study is the fact that the design of the study, inclusion/exclusion criteria, recruitment time and analysis plan are developed and exist before the experimental data will appear.

Special modeling program was written to generate the data and to simulate wrong procedures of data manipulation and analysis. In the paper the examples of late post factum selection of patients are demonstrated and the resulting bias of estimates is discussed.

The main conclusion is that retrospective time to event analysis of the passively collected data is not free of mistakes. Such type of the study and analysis should be used with limitation in discovering and descriptive studies.

Keywords

hematology, clinical trial, time-to-event analysis, survival analysis, overall survival, event-free survival, censoring, endpoints.

Hematology Research Center of RAMS, Moscow

Контакты: kulikov@blood.ru

Принято в печать: 4 июня 2010 г.

В статье обсуждаются недостатки исследовательских работ, основанных на ретроспективном анализе архивных клинических данных, собранных вне контролируемых планируемых клинических исследований. Описывается ряд методологических ошибок обработки и анализа данных, когда в исследовании основной конечной точкой (целевым критерием) является продолжительность жизни больного. Обосновывается некорректность совмещения поисковых методов и рекламно-описательных целей с доказательными выводами. В работе утверждается, что обязательной частью доказательных научно-медицинских исследований служит планирование и проспективность, для которых характерно то, что задачи и план исследования существуют до появления любых экспериментальных данных.

В работе приводятся примеры явной и неявной «постфактум» селекции данных и демонстрируются последствия такой обработки информации. Для того чтобы проиллюстрировать статью, были написаны специальные программы для генерации данных и моделирования типовых ошибок процедур ретроспективного событийного анализа.

Основной вывод данной статьи заключается в том, что ретроспективный событийный анализ (анализ выживаемости) пассивно собранных данных не свободен от ошибок, поэтому сфера его применения должна быть очень строго ограничена поисковыми и рекламно-описательными задачами.

Ключевые слова

гематология, клинические исследования, событийный анализ, анализ выживаемости, общая выживаемость, безрецидивная выживаемость, цензурирование, конечные точки.

ВВЕДЕНИЕ

С начала эры современной и прецизионной химиотерапии основным показателем эффективности лечения гемобластозов считался процент достигаемых полных ремиссий и их продолжительность. Но чем большему числу больных удавалось сохранить жизнь в результате совершенствования терапии, тем более очевидным становился тот факт, что эти параметры следует рассматривать как промежуточные, а основной целью противоопухолевого воздействия является увеличение

продолжительности жизни больного. Если ранее публикации ограничивались простыми процентными показателями, то в настоящее время во всех статьях, посвященных эффективности лечения гемобластозов, представлены результаты современного статистического анализа данных, иллюстрированные графиками с кривыми выживаемости Каплана—Мейера. Формальное и необдуманное использование строгих и правильных самих по себе процедур анализа может приводить к сомнительным или смещенным результатам. Ведь суть в том, что, к сожалению, методо-

логическая основа многих клинических работ изменилась за эти годы очень мало. В основном, это публикации-отчеты: вот так мы у себя лечим больных. Формально говоря, приводится ретроспективный анализ архивных данных за определенный период.

В целом статистический анализ экспериментальных данных может преследовать три кардинально различающиеся цели: 1) поисковая (например, поиск факторов риска, влияющих на выживаемость); 2) описательно-рекламная (например, оценка выживаемости по историческим данным); 3) доказательная (примером может служить анализ данных сравнительных проспективных клинических исследований).

В этой работе мы попытаемся продемонстрировать последствия подмены одной цели другой, когда ретроспективные исследования описательно-рекламного характера выдаются за доказательные. Особенно сильно такая подмена сказывается, когда в исследовании основной конечной точкой (целевым критерием) является продолжительность жизни больного.

Обязательная часть доказательных научно-медицинских исследований — планирование и проспективность. Для доказательных исследований характерно то, что задачи исследования, план (дизайн) проведения, даже план статистического анализа разрабатываются и существуют ДО (!) появления любых экспериментальных данных. Статистический анализ не может быть доказательным, если он базируется на данных, уже использованных в поисковых или описательно-рекламных целях. Это не недостаток статистического анализа, это методологическая ошибка проводящего анализ исследователя.

В настоящее время достаточно качественных учебных материалов по биостатистике, и в частности по анализу выживаемости [1, 2]. Конечно, совершенно очевидно, что для получения надежных клинических выводов необходимо использовать качественный программный статистический инструментарий, но это далеко не достаточное условие. В этой статье мы постараемся показать, от чего зависит достоверность заключений, сделанных на основе событийного анализа. В частности, мы продемонстрируем, к чему приводит расширенное и вольное толкование результатов ретроспективного анализа выживаемости.

Кроме того, необходимо подчеркнуть, что если экспериментальный материал собран неверно (вследствие неправильно спланированного исследования, селекции, отсутствия контроля и др.), то даже результаты правильного статистического анализа, выполненного с помощью качественных программ, не могут быть названы доказательными.

МЕТОДЫ

В статье рассматриваются типичные ошибки, встречающиеся в научной медицинской литературе, когда используются методы событийного анализа, в частности в онкогематологии. Хотя нетрудно найти конкретные примеры из реальных работ, мы не стали этого делать из этических соображений. Чтобы проиллюстрировать статью, были написаны специальные программы для генерации данных и моделирования ошибочных процедур. Программа генерирует случайным образом выборку данных, похожих на реальные. Далее эти данные подвергаются обработке и анализу, содержащему типичные ошибки. Выходные результаты такого рода анализа иллюстрируются графиками. Схожесть полученных графиков с рисунками из реальных статей намеренная и неслучайная. Модельная симуляция удобна и полезна тем, что мы знаем истинные характеристики выживаемости, такие как медиана продолжительности жизни или уровень выживаемости

на определенный момент, т. к. они заложены в модель. Поэтому можно сравнить оценки, полученные по выборке используемым некорректным методом анализа, и истинные.

Программы модельной генерации и обработки данных написаны на макроязыке статистического пакета SAS [3]. Типичное распределение продолжительности жизни онкогематологических больных хорошо описывается распределением, являющимся смесью двух экспоненциальных распределений [4]. Это значит, что группы исследуемых пациентов обычно гетерогенны и состоят по крайней мере из двух субпопуляций с разной прогнозируемой продолжительностью жизни. Визуально это реализуется в характерной форме кривой выживаемости. Она обычно состоит из двух участков: относительно крутого начального участка и более пологого участка, так называемого плато. Данная математико-статистическая модель и положена в основу симуляционной программы. Распределение точек цензурирования моделировалось как равномерное на отрезке времени наблюдения. Это соответствует равномерному во времени включению больных в исследование. В примере с активным/пассивным наблюдением за жизненным статусом больных генерировалось случайное «пропадание» информации о статусе больного в период после выписки из стационара.

ЧТО ТАКОЕ СОБЫТИЙНЫЙ АНАЛИЗ?

В классической биостатистике есть раздел «Анализ выживаемости» (англ. «Survival Analysis») [2]. Цель этого анализа — получить статистические характеристики времени: от какого-то определенного момента до наступления некоторого события (обычно это смерть). Отсюда и название. Основная отличительная черта анализа выживаемости как статистического метода заключается в том, что часть данных, представляемых для анализа, известна неточно. Например, о времени до события (смерти) для части объектов исследования известно лишь то, что оно не меньше некоторой величины. Эти больные либо еще живы на момент анализа, либо умерли, либо вышли из-под наблюдения, поэтому у них неизвестно время наступления события и нет информации, было ли событие вообще. В этих ситуациях объекты (пациенты) цензурируются на момент времени, когда о них было известно, что они живы. Цензурирование — это регистрация факта того, что исследуемое событие еще не наступило к моменту проведения анализа или последнего контакта с больным.

Как метод анализ выживаемости за последние 50 лет существенно обогатился новыми теоретическими статистическими моделями, задачами и вычислительными процедурами. И сейчас при решении конкретных медицинских прикладных задач выясняется, что измерения типа «время до изменения состояния объекта» встречаются значительно чаще, чем просто «время жизни». Значит, и сфера применения анализа выживаемости значительно шире, чем классического анализа данных о выживаемости. Например, такие события, как ремиссия, рецидив, осложнение и инфицирование, вполне могут быть подвергнуты статистическому анализу с использованием процедур анализа выживаемости.

Для того чтобы терминологически разделить метод и его приложение, вводится понятие «событийный анализ» (англоязычный аналог «time to event analysis» дословно переводится как «анализ времени до события»).

Событийный анализ — это статистический анализ, в котором исследуется время события, необязательно являющегося временем жизни объекта. По совокупности математико-статистических методов это классический анализ выживаемости. Слово «выживаемости» заменено на «событийный» для расширения сферы приложения.

СОБЫТИЯ

События — наиболее распространенные параметры изучения в медицине, демографии, социологии, основной из них — летальный исход. Конечность жизни человека, гетерогенность человеческой популяции, ограниченность времени клинических испытаний заставляют исследователей обращаться к изучению других событий в качестве целевых параметров исследования.

События, в отличие от других типов данных, записываются в виде пары значений: индикатор события и время. Индикатор события показывает, произошло ли событие за время исследования. Время — это промежуток времени от начала отсчета до события, если оно произошло, или до конца наблюдения за объектом, если оно не произошло.

Данные, собранные в виде временных рядов (например, температурный лист), часто можно редуцировать в данные типа события. Это оправдывается экономией объема собираемой и хранимой информации, обусловлено удобством анализа данных и их интерпретации. Редукция временного ряда до события может быть сделана, если исследователя интересует момент первого резкого изменения измеряемого параметра, которое и интерпретируется как событие или изменение статуса объекта. Например, динамическое измерение температуры тела может использоваться для обнаружения момента возникновения лихорадки, которая может быть вероятностно связана с другими событиями или результатом лечения. Если число регистрируемых осложнений велико, значителен объем другой собираемой информации, то запись, хранение и анализ всего температурного листа становятся избыточными. В этом случае информация о температурной динамике может быть сжата до двух чисел, описывающих событие «начало лихорадки».

РЕГИСТРАЦИЯ ВРЕМЕНИ НАСТУПЛЕНИЯ СОБЫТИЯ.
ТОЧНОСТЬ ИЗМЕРЕНИЯ ВРЕМЕНИ

Когда за объектом ведется непрерывное наблюдение, нет проблем с регистрацией времени наступления события. В этом случае время до наступления события измеряется с точностью, определяемой точностью записи в протоколе исследования, истории болезни и т.п. Например, время смерти известно обычно с точностью до дня (т.е. дата смерти). Есть, однако, события (например, рецидив заболевания), время (дата) наступления которого существенно зависит от периодичности измерения неких параметров, по их значению можно судить, наступило ли это событие или нет, от субъективности критериев оценки события и от объективности эксперта в оценке этих критериев. Точность измерения времени события может и должна учитываться при анализе. Периодичность измерения параметров, определяющих событие, должна быть по возможности одинаковой для изучаемых объектов и не изменяться за время наблюдения. Например, анализ крови для пациентов, включенных в исследование, должен выполняться регулярно и с одинаковой частотой, если по нему предполагается судить о наступлении ремиссии или рецидива. Фактически, когда регистрация события совершается эпизодически, мы всегда имеем дело не с точной фиксацией времени события, а с интервальным цензурированием.

ЦЕНЗУРИРОВАНИЕ

Говоря математическим языком, точка цензурирования — это левая граница временного интервала вероятного наступления случая, т.е. это последний момент времени, для которого точно известно, что исследуемое событие еще не

произошло. Самая распространенная ситуация — это цензурирование текущей датой (сегодняшним днем). Когда у объекта событие еще не произошло и это известно точно, то продолжительность жизни для этого объекта рассчитывается как разница между текущей датой и датой начала исследования, в этом случае наблюдение считается цензурированным. Продолжительность жизни для объекта, у которого произошло событие (смерть), рассчитывается как разница между известной датой события (смерти) и датой начала исследования, и тогда наблюдение считается случаем. Существуют ситуации, когда исследуемое событие не происходит по причине наступления другого конкурирующего события.

Примеры цензурирования, вызванного конкурирующим событием, и связанные с этим ошибки событийного анализа.

1. Старое цензурирование. Точкой цензурирования во многих исследованиях является дата последнего контакта с пациентом. Если больной вышел из-под наблюдения или поддерживает очень редкие контакты с исследователями, при этом последний контакт был давно, то это с большой вероятностью свидетельствует о низкой активности в отслеживании статуса больного. Такое, не очень тщательное наблюдение, отличающееся относительно редкими контактами с пациентом, — одна из самых распространенных причин вероятного смещения оценок выживаемости в сторону их улучшения.

2. Смерть как цензурирование. При оценке вероятности сохранения ремиссии или вероятности развития рецидива длительность ремиссии у больных, умерших в полной ремиссии, цензурируется датой смерти. Здесь очень часто могут возникать так называемые ошибки пропуска, НЕрегистрации целевого события (рецидива), когда случай-событие (рецидив) и случай-цензурирование (смерть) следуют один за другим в коротком временном интервале. Так, момент рецидива просто может быть пропущен ввиду относительной редкости лабораторных измерений. В этом случае возможна недооценка вероятности наступления рецидива, т.к. фактически точка события ошибочно рассматривается как точка цензурирования.

3. Резистентность как событие и как цензурирование. Недостижение полной ремиссии или эффекта терапии, т.е. резистентность, вообще сложно трактовать как событие, независимо от того, фигурирует ли оно в списке конечных точек (бессобытийная выживаемость) или является точкой цензурирования в оценке вероятности ответа на терапию. Если резистентность оценивается косвенно, как длительное ожидание ответа или отсутствие ответа на терапию к какому-то определенному моменту времени, это вообще не событие. В этом случае резистентность — это состояние, и регистрация этого состояния — это не объективное, а субъективное событие, т.е. принятое врачом-исследователем решение о смене терапии. В любом случае момент возникновения этого состояния неизвестен и отличается от времени регистрации этого состояния. При анализе резистентности как события необходима осторожность и четкость дефиниций, при этом точность времени ее регистрации может быть высокой при достаточной частоте измерений объективных лабораторных индикаторов резистентности.

4. Выполнение трансплантации костного мозга как конечное постоянное цензурирование. Это один из очень эффективных способов изменить показатели выживаемости у исследуемой когорты больных в сторону значительного их увеличения, поскольку все последующие за процедурой события уже не будут регистрироваться.

КТО, КАК И ЗАЧЕМ СТРОИТ КРИВЫЕ ВЫЖИВАЕМОСТИ?

Строго говоря, кривая выживаемости — это график математической функции. Эта функция характеризует вероятность того, что событие не произойдет до указанного времени. Если событие — летальный исход, то вертикальная координата точки на кривой выживаемости — это численное значение **вероятности** объекта дожить до момента времени, соответствующего горизонтальной координате этой точки. То, что мы видим в статьях и презентациях, — это экспериментальная **оценка** этой математической функции, полученной в результате анализа данных о продолжительности жизни группы исследуемых пациентов. Эта оценка сделана с некоторой точностью. Точность оценки функции выживаемости зависит от количества объектов, по которым вычислена эта статистика, и процентного соотношения числа пациентов с известной продолжительностью (умерших) и цензурированных пациентов (живых). Чем больше в процентном соотношении в исследуемой группе умерших больных, тем точнее оценка.

Оценки выживаемости, полученные по 10–20 пациентам при 1–5 числе случаев, очень неточны и имеют незначительную научную и практическую ценность. Нередко можно видеть даже в крупнейших международных журналах кривые выживаемости, построенные для двух или трех больных. Такой метод представления данных методически некорректен и бессодержателен.

Важно понимать также, что основное предназначение событийного анализа вообще и кривых выживаемости в частности — сравнительный анализ, например сравнение результатов лечения больных по разным протоколам или сравнение результатов лечения разных вариантов одного заболевания. Использование кривых выживаемости для индивидуального прогноза продолжительности жизни крайне неоднозначно и должно применяться с осторожностью.

РЕТРОСПЕКТИВНЫЙ АНАЛИЗ КАК ОСНОВА «МЕМУАРНОЙ СТАТИСТИКИ»

Отчеты о ретроспективных исследованиях еще пока очень распространены в научной литературе. Использование событийного анализа в таких работах наиболее подвержено риску получения ненадежных и смещенных результатов и, как следствие, риску ошибочного толкования. Две главные причины этих ошибок — это селекция объектов исследования задним числом и неоднородность условий наблюдения за больным в течение исследования.

Формально говоря, любой статистический анализ ретроспективен, т. к. проводится на данных, уже полученных. Но характер исследования и способ сбора экспериментального материала могут быть как ретроспективными, так и проспективными. В чем же заключается главное, существенное, отличие проспективного метода от ретроспективного? Оно не в правилах формирования целей, задач и планов, а в том, что в проспективном исследовании экспериментальных данных попросту **НЕТ** на момент формулирования цели, а главное, выборка объектов (пациентов) формируется ДО появления результатов воздействия на объекты, т. е. результатов лечения больных. В этом смысле проспективные исследования являются слепыми по отношению к будущим результатам, следовательно, свободными от осознанных или неосознанных манипуляций с данными и селекции объектов.

Особенно сильно искажение истинных результатов происходит при выполнении **ретроспективного** событийного анализа. Когда исследователь анализирует свой архивный

материал, свои «клинические мемуары», он формирует выборку и группы сравнения, уже зная исходы, конечные точки пациентов. Избежать субъективности при этом очень сложно.

Например, сформулирована задача: оценить выживаемость пациентов с конкретным заболеванием, лечение которым было проведено в определенной клинике за какой-то определенный период, скажем, за 10 лет, и сравнить ее либо с литературными данными, либо с данными какой-то контрольной выборки. Естественно, конечной целью, даже если она и не сформулирована, является демонстрация высокой эффективности терапии в указанной клинике в отчетный период. Даже искреннего и честного исследователя подстерегают минимум три типа ошибок, приводящих к смещенным статистическим оценкам характеристик выживаемости и связанным с этим выводами.

Ошибка номер один. Это последовательный, пошаговый анализ, выполняемый путем последовательной оптимизации критериев включения. На первом этапе строится кривая выживаемости по всем пациентам, госпитализированным в клинику за 10 лет. Затем смотрят на «ступеньки» — события. Врач чаще всего помнит поименно больных, с которым связана конкретная «ступенька» (смерть) на кривой выживания. Замечает, например, что многие из них — это пациенты, лечение которым уже было начато на этапе до госпитализации в конкретную клинику. Вот вам и готовый первый пункт в критериях селекции — предшествующее лечение. Как правило, список потенциальных причин для не включения больных в анализ достаточно широк. Выбор всегда есть: неточно доказанный диагноз, низкая эффективность предшествующего воздействия, неблагоприятный социальный статус, сопутствующие тяжелые заболевания и т. д. Тот факт, что критерии включения формируются не до, а после получения данных и зависят от них, — грубейшая ошибка анализа.

Ошибка номер два. Когда возможности селекции задним числом исчерпаны или исследователь хочет выглядеть честным, он пытается рассортировать пациентов с целью выделить подгруппы с относительно удачными результатами лечения. При этом происходит подмена анализа или поиска факторов риска категоризацией, т. е. разбиением на группы и «открытием» новых нозологий. Один известный врач, глядя на кривую выживаемости, сказал, что все неудачи должны быть тщательно проанализированы. Если результатом такого «тщательного» анализа задним числом будет «уточнение» диагноза у больных с неудачным исходом лечения, то в следующих презентациях кривая выживаемости для этой группы, конечно, будет выглядеть значительно лучше, поскольку больные с «уточненным диагнозом» будут просто исключены из анализа.

Как поступить по-другому? Можно разделить всю группу на две: с удачами и неудачами лечения; «покопаться» и найти признаки, которые дифференцируют эти подгруппы лучше всего. В принципе это нормальная поисковая эвристическая процедура. Но в корректной научной методологии это всего лишь метод генерации гипотезы, которая в дальнейшем должна быть подтверждена на независимой группе пациентов в ходе нового проспективного исследования. В распространенной практике «мемуарного» статистического анализа этим этапом пренебрегают: **что нашел, то и доказал.**

На рис. 1 приведен смоделированный пример такой ситуации. Параметры модели подобраны таким образом, что 5-летняя общая выживаемость составляет 50%. На рис. 1, А представлена истинная кривая (кривая 2) и кривая выживаемости, построенная по случайно сге-

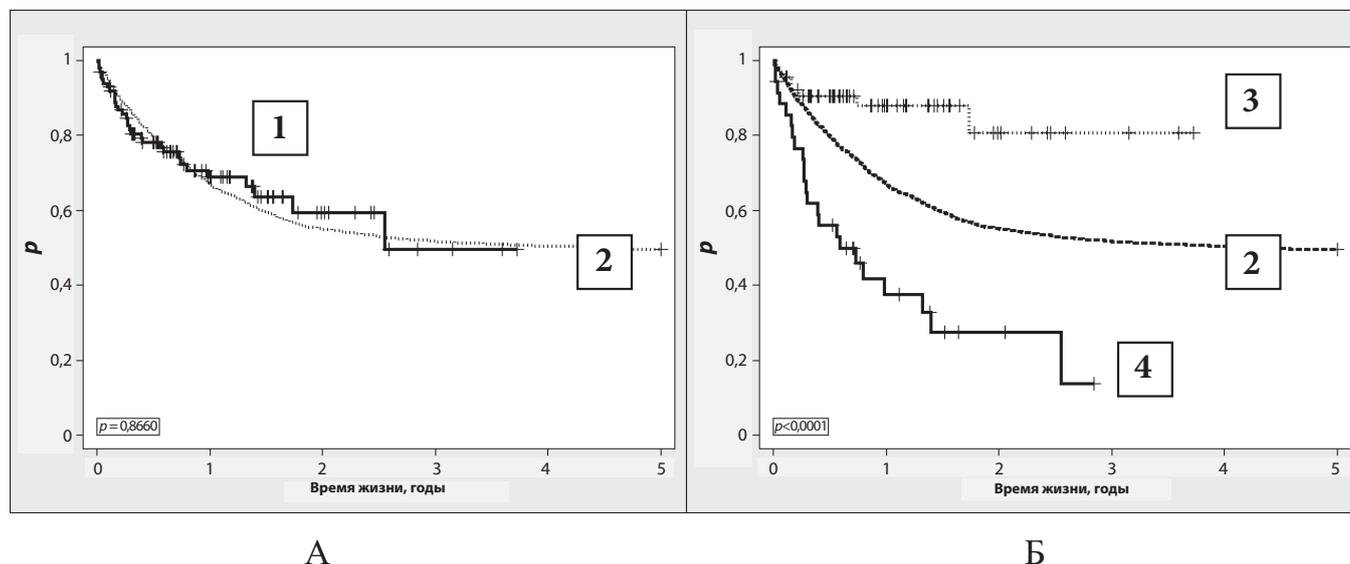


Рис. 1. Результат ретроспективной категоризации группы больных (модельная симуляция):

А — выживаемость в общей группе больных; Б — выживаемость после разбиения на подгруппы; кривая 1 — оценка общей выживаемости по всей группе из 100 больных; кривая 2 — истинная выживаемость; кривая 3 — оценка общей выживаемости 65 больных, отобранных после «анализа неудач» в подгруппу 1-й категории (удачная терапия); кривая 4 — оценка общей выживаемости 35 больных, отобранных в группу 2-й категории (неудачная терапия)

нерированным данным о продолжительности жизни для 100 больных (кривая 1). Предполагается, что исследователь поработал с данными и постфактум разделил выборку на две категории с целью выделить подгруппы больных, удачно и неудачно пролеченных. При этом исследователь, естественно, не был «слеп» и знал конечные результаты терапии всех больных. Поэтому умерших пациентов (случаи) с большей вероятностью (80 %) он относил в первую категорию, а с вероятностью 20 % — во вторую, живых больных он с большей вероятностью (80 %) относил во вторую категорию, а с меньшей (20 %) — в первую. Компьютерная программа смоделировала эту сортировку и разделила группу на две подгруппы в соответствии с этим правилом, в группу «удачная терапия» попало 65 пациентов, в группу «неудачная терапия» — 35. Затем было проведено обычное сравнение оценок выживаемости в этих подгруппах, как будто бы мы имеем дело с обычным классифицирующим фактором. В результате получились кривые выживаемости, изображенные на рис. 1, Б, где выделены две категории: «удачная терапия» (кривая 3) и «неудачная терапия» (кривая 4).

Как видно на рис. 1, выживаемость больных из группы «удачной терапии» (первая категория) составила около 80 % при истинной 50 %. Приведенный анализ — пример исследовательского лукавства или, если хотите, лжи, когда больные с плохой эффективностью исследуемого метода лечения исключаются из общей анализируемой популяции (мол, если у них не получен эффект, значит, болезнь у них другая), а вывод об эффективности исследуемого метода лечения делается на основании выживаемости больных из группы «удачной терапии».

Ошибка номер три. Пассивный мониторинг статуса пациента. Очевидно, что наблюдение за больным во время госпитализации и после нее существенно отличается. Строго говоря, событийный анализ может и должен распространяться только на тот отрезок времени, в течение которого не изменяются условия получения информации о статусе пациента. После выписки из стационара, во-первых, сильно изменяется временной регламент получения информации о жизненном статусе больного: информация поступает значительно реже. Интервал между контактами с пациентом или его родственниками может достигать многих меся-

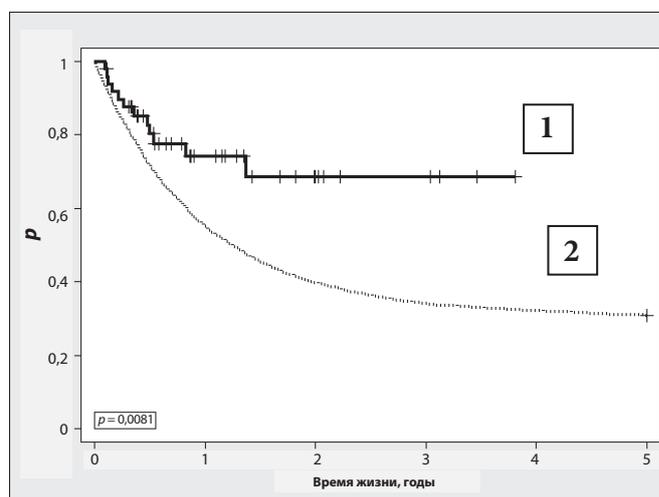


Рис. 2. Результат ретроспективного анализа общей выживаемости больных, у которых есть фазы активного и пассивного наблюдения за жизненным статусом (модельная симуляция):

1 — общая выживаемость 100 больных в условиях вероятной потери информации о летальных исходах; 2 — истинная выживаемость при постоянном эффективном наблюдении

цев. Во-вторых, если не проводится специального клинического исследования, после выписки активное наблюдение за статусом больного часто просто прекращается. При этом существует вероятность, что информация о конечной точке (смерти больного) не доходит до врача и больной в информационной базе остается «вечно живым» на момент последнего контакта. Чем выше вероятность потери информации о смерти, тем, естественно, больше смещение оценки выживаемости.

На рис. 2 приведен модельный пример такой ситуации. В модели предполагается, что истинная 4-летняя выживаемость у данной популяции больных составляет 32 %. Если предположить, что активное наблюдение за больным осуществляют только в течение 4 мес., после которых информация о смерти пациента не приходит (случайно) в среднем в 50 % случаев, то в результате моделирования указанной ситуации 4-летняя выживаемость, полученная в условиях вероятной потери информации о летальных исходах, будет сильно завышена (68 %).

ЗАМЕНИТЕЛИ ОБЩЕЙ ВЫЖИВАЕМОСТИ

Общая выживаемость считается основным критерием эффективности в клинических исследованиях в онкологии и гематологии [5, 6]. Однако ввиду ограниченности времени наблюдения в клинических исследованиях для увеличения вероятности наступления событий в рассмотрение вводят другие конечные точки, которые можно считать аналогами или заместителями (анг. *surrogates*) общей выживаемости.

Характеристики ответа на терапию, такие как ремиссия или потеря ответа (рецидив), часто применяют в качестве дополнительных или вторичных точек. При их использовании в качестве критериев эффективности главной проблемой становится вопрос об их корреляции с отдаленными результатами — общей выживаемостью.

Много методологических трудностей и ошибок связано с использованием и интерпретацией таких параметров анализа, как безрецидивная или бессобытийная выживаемость. Если для построения кривых общей выживаемости применяется одна конечная точка — смерть больного, то для построения безрецидивной или бессобытийной выживаемости используют смешанные конечные точки: в список исследуемых событий кроме летального исхода вводят другие неблагоприятные события, такие как рецидив, прогрессия, развитие другого онкологического заболевания, которые, несомненно, имеют корреляцию с продолжительностью жизни больного, но не сиюминутную и не всегда однозначную. Кроме того, следует подчеркнуть, что уж если используется анализ выживаемости со смешанными конечными точками (безрецидивная и/или бессобытийная выживаемость), необходимо всегда давать четкие определения этим конечным точкам и времени начала отсчета. И конечно, эти определения должны быть сформулированы до начала исследования, т. е. до появления первых данных.

В значительной части работ по онкогематологии приводится и общая, и безрецидивная выживаемость. Можно спорить о целесообразности и информативности одновременного представления обеих этих оценок. Гораздо более информативно, по нашему мнению, вместо кривых безрецидивной выживаемости представить анализ вероятности достижения ремиссии и/или развития рецидива. Однако многие авторы, по-видимому, для экономии места изображают общую и безрецидивную выживаемость на одном графике. Это уже большая методологическая ошибка по нескольким причинам.

Причина первая и главная. Общая выживаемость оценивается для **всей** исследуемой группы, и точкой начала отсчета считается дата начала лечения (например, дата трансплантации). Безрецидивная выживаемость оценивается для подгруппы больных, у которых была достигнута ремиссия, и точкой начала отсчета считается дата ремиссии. На одном графике не могут без специальных ухищрений изображаться функции, у которых аргументы разные, т. е. то, что откладывается по горизонтальной оси.

Причина вторая. Некорректно сравнивать две разные характеристики, полученные на разных группах: на всей группе и подгруппе. При этом часто возникает наивно глупый вопрос: «Какая из выживаемостей должна быть выше: общая или безрецидивная?» Правильного ответа на этот вопрос нет. С одной стороны, т. к. наблюдаются два типа событий — смерть и рецидив, кривая безрецидивной выживаемости должна идти ниже. С другой стороны, группа пациентов, у которых получен ответ (ремиссия), — это отобранная более благополучная подгруппа. В ней любой неблагоприятный исход может быть менее вероятным, чем смерть в общей группе. Поэтому не следует изображать кривые общей

и безрецидивной (бессобытийной) выживаемости на одном графике, чтобы не провоцировать бессмысленные сравнения и спорные интерпретации.

Вообще ретроспективный событийный анализ пассивно собранных данных, т. е. данных, собранных вне контролируемого планируемого исследования, должен быть исключен из списка корректных научных методов исследования и рекомендован к использованию в сильно ограниченных задачах поискового анализа или рекламно-описательных целях. Селекция задним числом, когда результаты терапии известны, может быть вполне искренней и необязательно сознательной фальсификацией. Несмотря на это, такое использование событийного анализа является грубой ошибкой и значительно искажает истинную картину.

Еще одна опасность ретроспективного анализа архивов — это не только искажение реальной картины для читателя, потребителя суррогатной научной информации. Сами исследователи могут вполне искренно верить в свои результаты и выводы и продолжать двигаться к недостижимым целям.

«КВАДРАТНЫЕ» КРИВЫЕ ВЫЖИВАЕМОСТИ, ИЛИ КРИВЫЕ «КАТАСТРОФ»

Если просматривать статьи с «картинками» выживаемости в таких уважаемых журналах, как «Blood» или «Leukemia», то редко можно найти кривые с ярко выраженным изломом. Излом, или, выражаясь математическим языком, резкое изменение производной на кривой выживаемости, свидетельствует о резком изменении функции риска. Функция риска характеризует интенсивность наступления случаев, т. е. вероятность наступления события в ближайшем к данному моменту малом интервале времени, при условии что события раньше не было. Если говорить о выживаемости, то функция риска — это вероятность больного умереть в течение данного дня при условии, что он до этого дожил. Скачок функции риска свидетельствует либо о резком изменении объективного состояния пациента (например, окончание рискованного периода агранулоцитоза), либо о значительном изменении состава группы (естественном выбывании пациентов с высоким риском ранней летальности, т. е. об исходной гетерогенности исследуемой группы). Анализ скрытой гетерогенности и переменного во времени риска — сложная математико-статистическая задача [7]. Но хотелось бы остановиться на странном, необычном виде кривых выживаемости, имеющем не объективные природные причины, а являющемся следствием методологических особенностей и погрешностей анализа данных. На следующем рисунке изображена типичная картинка, которую любят приводить в своих работах некоторые врачи-исследователи. Обычно это следствие «прецедентной» практики некоторых научных журналов. Авторы рисуют картинки в одной манере для простоты сравнения результатов.

Например, в некоторых сообществах исследователей есть обычай изображать безрецидивную выживаемость, включая в список конечных точек не только смерть и рецидив, но и раннюю летальность, т. е. в анализ включаются все пациенты (а не только те, у кого получена полная ремиссия), и отсчет времени ведется от начала лечения (а не от момента достижения ремиссии). В принципе, если не искажается временная шкала (что тоже бывает, когда ранняя летальность регистрируется для всех больных одним днем — точкой О на горизонтальной оси времени) и отсчет времени действительно ведется от начала терапии, а не от момента ремиссии, то формально с точки зрения статистики все правильно. На графике изображается оценка вероятности на-

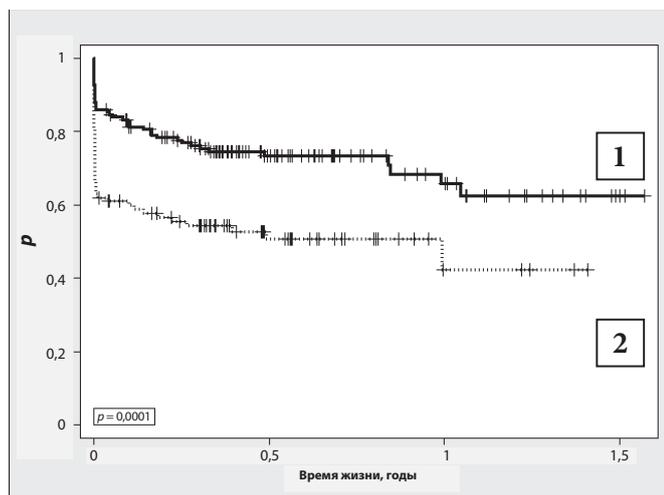


Рис. 3. Результаты симуляционного моделирования анализа безрецидивной выживаемости больных в двух группах, при котором отсчет идет от начала лечения:

1 — группа с ранней летальностью 15%; 2 — группа с ранней летальностью 30%

бора случаев: смерть и рецидив. Некорректность подхода в том, что в анализ объединяются разные по статистической природе случаи — рецидив и смерть. Если смерть может наступить в любое время, то рецидивы в начальный период (до ремиссии) по определению не могут произойти. Эта некорректность может приводить к явной ошибке, когда сравниваются кривые. Примером могут служить кривые выживаемости, приведенные на рис. 3.

Как видно на представленном графике, вероятность наступления случаев (смерть + рецидив) в постремиссионный период в двух группах абсолютно одинакова — кривые параллельны. При этом оценки безрецидивной выживаемости статистически сильно различаются ($p = 0,0001$), что и фиксируется в выводах. Но на самом деле отличаются они только за счет разной ранней летальности (15 и 30% — так заложено в модели), в то же время вероятность сохранения ремиссии в двух группах совершенно одинаковая.

Идеологически безрецидивная выживаемость — это характеристика безрецидивного течения постремиссионного периода заболевания, и включение в анализ эффекта ранней летальности (события, предшествующего возможному наступлению ремиссии) может привести к ошибочной интерпретации результатов, что и продемонстрировано в этом модельном примере.

Есть бесспорный математический принцип: кривая выживаемости должна начинаться из точки на графике с координатами $t = 0, P = 100\%$ (где t — время отсчета, P — доля больных), а для вероятности наступления события — из точки с характеристикой $t = 0, P = 0\%$. Это означает, что плотность вероятности в начальный момент — величина больше нуля. Из этого следует, что на кривой выживаемости нет вертикального или горизонтального продолжительного участка в начальной точке. Если на графике есть начальный вертикальный участок, т. е. кривая начинается на уровне ниже 100%, то это значит, что в начальный момент происходит какая-то катастрофа, приводящая к **одномоментной** потере части пациентов. Если же кривая какое-то время идет горизонтально на уровне 100%, а затем уходит вниз, это выглядит как 100%-я искусственно обеспеченная гарантия отсутствия летальных исходов в этот конкретный промежуток времени и отмена гарантии после него.

Таким образом, если график выживаемости или других конечных точек имеет необычную форму (резкие изменения наклона, горизонтальные и вертикальные продолжительные

участки), это всегда должно вызывать подозрения. Причины этого, скорее всего, искусственные и связаны с методологией сбора и обработки данных.

АНАЛИЗ «РЕСПОНДЕР-НЕРЕСПОНДЕР». ПРОДЛЕВАЕТ ЛИ ЖИЗНЬ ВЫИГРЫШ В ЛОТЕРЕЮ?

В свое время был достаточно популярным сравнительный анализ выживаемости тех, у кого был получен ответ на лечение (респондеры), и тех, у кого его не было (нереспондеры). О некорректности данного вида анализа еще четверть века назад в авторитетном медицинском журнале «Journal of Clinical Oncology» была написана специальная методологическая статья [8, 9]. В статье однозначно показано, что постфактум разбиение исследуемой группы на ответивших и не ответивших на лечение и сравнение их оценок выживаемости с отсчетом от начала лечения является некорректным методом анализа. Статья J. Anderson и соавт. [8] процитирована за эти 20 лет около 400 раз, и данный тип анализа почти исчез со страниц авторитетных журналов. Несмотря на это, статьи, в которых этот псевдометод используется, появляются с завидным постоянством, поскольку он очень привлекателен своей кажущейся простотой и наглядностью. По сути, этот метод вполне соответствует принципам ретроспективной «мемуарной» статистики, когда сортировка пациентов осуществляется после получения информации о результате воздействия.

Идея этого псевдоанализа проста: пациентов делят на две подгруппы в зависимости от того, был ли достигнут эффект при лечении или нет; затем строят и сравнивают кривые выживаемости; продолжительность жизни отсчитывается от начала терапии. Математическое доказательство ошибочности метода приведено в упомянутой статье [8]. Основная идеологическая некорректность метода в том, что на момент начала терапии, от которой ведется наблюдение за жизненным статусом пациента, информация о возможном эффекте у конкретного пациента, конечно, отсутствует. Для анализа выживаемости или других конечных точек у больных в зависимости от эффективности проведенного им лечения используются более сложные математико-статистические методы анализа с применением регрессионных моделей с переменными во времени ковариатами, описание которых выходит за рамки данной статьи.

Проиллюстрируем некорректность метода «респондер-нереспондер» на следующем модельном примере. Возьмем группу пожилых пациентов с онкологическим заболеванием. Для 80% больных средняя продолжительность жизни 1 год, а для остальных 20% — 10 лет. Предположим, что разыгрывается некая социальная лотерея для таких пациентов. Тиражи проводятся регулярно и часто, скажем, каждый день. Средняя длительность ожидания пациентом выигрыша лотереи — 1 год. Спрашивается, продлевает ли жизнь больного выигрыш в лотерею? Ответ очевиден: поскольку события (смерть и выигрыш) независимы, то ответ — нет, не продлевает. В модели генерируется два типа событий: смерть и выигрыш лотереи. Пациенты, которые выиграли, относились в категорию «выигравших», те же, у кого смерть наступила раньше выигрыша, — в другую категорию. Для тех и других построены оценки выживаемости, изображенные на рис. 4.

В результате анализа выживаемости мы «неожиданно» получаем вывод, что выигравшие живут дольше... Разве не очевидно? Медиана продолжительности жизни больных в группе выигравших составляет 1,3 года, а в группе проигравших — 0,68 года (почти в 2 раза меньше!). При этом истинная медиана для всей группы составляет 0,9 года. Результат «анализа» — сильно значимый, судя по графикам

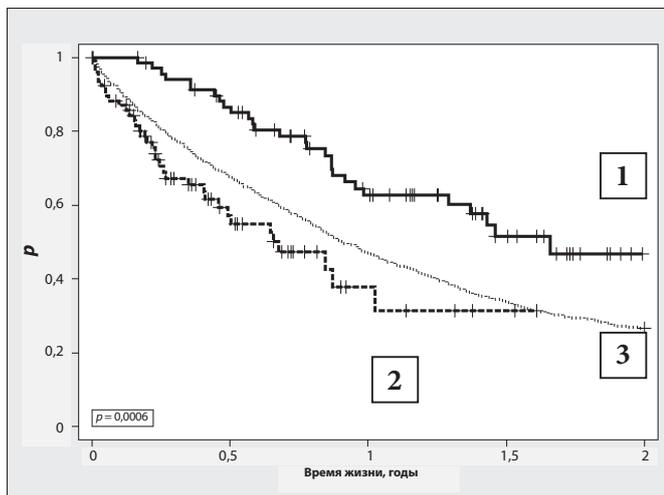


Рис. 4. Выживаемость пожилых больных, участвующих в розыгрыше лотереи (модельная симуляция). Результат ретроспективного псевдоанализа по сравнению группы выигравших в лотерею (1) и невыигравших (2); 3 — истинная выживаемость

и цифрам, и, очевидно, абсурдный по сути. Суть ошибки в распространенной путанице причины и следствия: выигрывают (дожидаются выигрыша) те, кто живет дольше.

ВЫВОДЫ

Основной вывод данной статьи заключается в том, что ретроспективный событийный анализ (анализ выживаемости) пассивно собранных данных не свободен от ошибок как результат поздней селекции и манипуляции данными, поэтому сфера его применения должна быть очень строго ограничена поисковыми и рекламно-описательными задачами.

Для получения фактов, которые могут быть использованы в доказательных целях, необходимы проспективные контролируемые исследования. Дизайн исследования, селекция объектов, технология и инструменты сбора данных, план анализа должны разрабатываться и фиксироваться до появления экспериментальных данных. Это обязательные условия получения несмещенных результатов и объективных выводов.

Кроме того, пусть ни у кого не вызывает сомнения, что для профессионалов представляемые в статьях и презентациях кривые выживаемости или событийного анализа являются своего рода детектором лжи — по ним всегда можно судить о чистоте эксперимента и экспериментатора.

ЛИТЕРАТУРА

1. Электронный учебник по статистике Statsoft. Анализ выживаемости. <http://www.statsoft.ru/home/textbook/default.htm>.
2. Lee E.T. Statistical methods for survival data analysis. Belmont, CA: Lifetime Learning, 1980.
3. SAS Institute Inc. 2004. SAS® 9.1.3, Cary, NC: SAS Institute Inc.
4. Lambert P., Thompson J. Estimating and modeling the cure fraction in population-based cancer survival analysis. *Biostatistics* 2007; 8(3): 576–94.
5. Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics. Guidance for Industry U.S. Department of Health and Human Services Food and Drug Administration, May 2007.
6. Куликов С., Савченко В., Маслова Е., Паровичникова Е., Гудилина Ю. Клинические экспериментальные исследования в гематологии: планирование, управление данными, представление результатов. *Тер. арх.* 1996; 7: 65–72.
7. Wienke A. Frailty Models, Max Planck Institute for Demographic Research, MPIDR Working Paper WP 2003–032, September 2003, <http://www.demogr.mpg.de/>.
8. Anderson J.R., Cain K.C., Gelber R.D. Analysis of survival by tumor response. *J. Clin. Oncol.* 1983; 1: 710–9.
9. Anderson J.R., Cain K.C., Gelber R.D. Analysis of Survival by Tumor Response and Other Comparisons of Time-to-Event by Outcome Variables. *J. Clin. Oncol.* 2008; 26(24): 3913–5.