

МЕТОДОЛОГИЯ НАУЧНЫХ ИССЛЕДОВАНИЙ

doi : 10.52485/19986173_2021_1_137

УДК: 519.23-614.2

Мудров В.А.

АЛГОРИТМЫ ПРИМЕНЕНИЯ АНАЛИЗА ВЫЖИВАЕМОСТИ
В БИОМЕДИЦИНСКИХ ИССЛЕДОВАНИЯХ
С ПОМОЩЬЮ ПАКЕТА ПРОГРАММ SPSS*Федеральное государственное бюджетное образовательное учреждение высшего образования «Читинская государственная медицинская академия» Министерства здравоохранения Российской Федерации, 672000, г. Чита, ул. Горького, 39а*

Цель исследования. Предметом исследования явилось изучение основ применения анализа выживаемости, позволяющего на основании независимых переменных прогнозировать время наступления конкретного события. Темой исследования явился анализ выживаемости в биомедицинских исследованиях. Целью работы явилось описание основных алгоритмов применения анализа выживаемости в биомедицинских исследованиях с помощью пакета программ SPSS.

Материалы и методы. Проведен научный обзор основных методов анализа выживаемости в биомедицинских исследованиях. Практические основы применения анализа выживаемости для определения времени наступления конкретного события рассмотрены на примере пакета программ «IBM SPSS Statistics Version 25.0» (International Business Machines Corporation, США).

Результаты. Определены оптимальные алгоритмы применения анализа выживаемости в биомедицинских исследованиях. Подробно описаны возможности применения анализа выживаемости в программе SPSS, даны рекомендации по интерпретации полученных результатов анализа.

Заключение. Использование описанных алгоритмов применения анализа выживаемости позволит повысить уровень представления результатов биомедицинских исследований.

Ключевые слова. Статистический анализ, алгоритм, выживаемость, дожитие, время наступления события, метод Каплана-Мейера, регрессионный анализ Кокса, пропорциональность рисков, SPSS.

Mudrov V.A.

SURVIVAL ANALYSIS ALGORITHMS IN BIOMEDICAL RESEARCH
USING SPSS SOFTWARE PACKAGE*Chita State Medical Academy, 39A Gorky str., Chita, Russia, 672000*

The aim of the research. The subject of the research was to study the basics of using survival analysis, which allows predicting the time of the event onset on the basis of independent variables. The research's topic was survival analysis in biomedical research. The aim of the study was to describe the main survival analysis algorithms in biomedical research using the SPSS software package.

Materials and methods. The scientific review of the main methods of survival analysis in biomedical research is carried out. The practical foundations of using survival analysis to determine the time of the event occurrence are considered on the example of the IBM SPSS Statistics Version 25.0 software package (International Business Machines Corporation, USA).

Results. The optimal algorithms for survival analysis application in biomedical research have been determined. The possibilities of survival analysis using in the SPSS program are described in detail, recommendations for the interpretation of the obtained analysis results are given.

Conclusion. The use of the described survival analysis algorithms will improve a presentation's level of biomedical research results.

Keywords. Statistical analysis, algorithm, survival, time of occurrence, Kaplan-Meier method, Cox regression analysis, proportionality of risks, SPSS.

В данной статье основное внимание уделено общим принципам применения анализа выживаемости/дожития для прогнозирования времени наступления конкретного события. Кроме того, автором рассмотрены основные методы анализа выживаемости, даны рекомендации по интерпретации полученных результатов, рассмотрены алгоритмы применения наиболее популярных методов анализа выживаемости: таблиц выживаемости,

метода Каплана-Майера и регрессии Кокса. Описание основных алгоритмов применения анализа выживаемости, представленных в доступном для читателя виде, является актуальным методом, позволяющим начинающим исследователям повысить уровень представления своих биомедицинских исследований [1-4].

Анализ выживаемости / дожития / времени наступления события представляет собой метод, позволяющий изучить закономерности наступления конкретного события у представителей наблюдаемой выборки с учетом времени до его наступления. Анализ выживаемости был назван так в связи с его первоначальным применением для анализа выживаемости людей по данным медицинской отчетности [5]. Однако с течением времени данный метод приобрел широкое распространение в медицинской статистике и стал трактоваться как «Анализ времени до наступления события». При этом в медицине под событием стали рассматривать не только летальный исход, но и выздоровление, рецидив или обострение заболевания и другие исходы, представляющие интерес и связанные со временем их наступления [6]. Для удобства работы, при заполнении данных, наличие изучаемого исхода следует кодировать единицей, а его отсутствие – нулем. Период времени от начального события до исследуемого исхода (контрольной точки) называется временем ожидания или временем до события [5]. Цензурированными (неполными) называются данные, которые содержат неполную информацию об исследуемом явлении. Цензурирование считается правосторонним, когда известен момент начала исследования, но не известен момент исхода (наблюдение прекратилось до момента наступления события по каким-либо причинам), и левосторонним, когда известен момент исхода исследования, но не известен момент начала заболевания (наблюдение началось только после наступления события). Левостороннего цензурирования в исследованиях следует избегать [5].

Наиболее распространенными методами анализа дожития являются: таблицы выживаемости, метод Каплана-Мейера и регрессия Кокса [6]. Таблицы выживаемости и метод Каплана-Мейера предполагают учет не фактической даты начала или исхода / последнего контакта с пациентом, а анализ времени от начала исследования до интересующей контрольной точки (например, пятилетняя выживаемость пациентов с онкогинекологической патологией). Данный период времени исследователь может разделить на более короткие периоды, обосновав с теоретической точки зрения, при этом, необходимость выделения дополнительных контрольных точек (например, одно- и двухлетняя выживаемость). Таблицы дожития, как и метод Каплана-Мейера, позволяет оценивать любой дихотомический исход. Метод Каплана-Мейера представляет собой множительный непараметрический метод, позволяющий оценивать кумулятивную функцию выживания в момент возникновения каждого исхода. Использование таблиц дожития, являющихся по своей сути вариантом метода Каплана-Мейера для сгруппированных данных, позволяет оценивать выживаемость в фиксированные моменты времени от начала наблюдения, определенные самим исследователем. Между тем, использование таблиц дожития представляется более компактным в сравнении с методом Каплана-Мейера, что позволяет применять данный метод для обобщения результатов наблюдения за большими выборками [5].

Регрессия Кокса представляет собой модель пропорциональных рисков, которая изучает зависимость времени наступления события от независимых переменных. Являясь полупараметрическим методом, регрессия Кокса позволяет прогнозировать риск наступления события и оценивать влияние независимых переменных на этот риск. При этом риск является функцией, зависящей от времени наступления события. Основным условием применения регрессии Кокса является наличие линейной взаимосвязи между переменными и логарифмом функции риска наступления события. Объектом исследования может являться пациент, для которого прогнозируют риск наступления события. В качестве независимых переменных (предикторов) могут выступать характеристики объекта (например, срок гестации, пол плода, осложнения беременности и т.д.), которые гипотетически могут влиять на риск наступления конкретного события [6].

Целью работы явилось описание основных алгоритмов применения анализа выживаемости в биомедицинских исследованиях с помощью пакета программ SPSS.

Материалы и методы. Научный обзор наиболее популярных методов анализа выживаемости, используемых в биомедицинских исследованиях, проведен на основании анализа отечественных и зарубежных источников литературы. В качестве основополагающих рассмотрены общеизвестные источники, имеющие наибольший авторитет и широкое распространение [1, 3, 5, 6]. Практические основы применения анализа выживаемости для определения времени наступления конкретного события рассмотрены на примере пакета программ «IBM SPSS Statistics Version 25.0» (International Business Machines Corporation, США).

Результаты и их обсуждение. Цель анализа выживаемости состоит в получении статистических характеристик времени: от определенного момента до наступления некоторого интересующего исследователя события [7].

Согласно рекомендациям САМПЛ при описании анализа выживаемости следует указать [8, 9]:

- 1) цель применения анализа;
- 2) даты или события, соответствующие началу и концу анализируемого периода времени;
- 3) обстоятельства цензурирования данных;
- 4) статистические методы, использованные для оценки выживаемости;
- 5) обоснование применимости метода;
- 6) оценку вероятности дожития для определенных периодов наблюдения в каждой группе и их доверительные интервалы, а также число наблюдений;
- 7) накопленную вероятность событий в виде графика, особенно если события редкие;
- 8) медиану времени до события и ее доверительный интервал для сравнения с результатами других исследований;
- 9) результаты анализа в виде графика (например, Каплана-Мейера) или таблицы;
- 10) статистический метод, использованный для двух и более кривых выживаемости;
- 11) уровень значимости P , полученный при сравнении двух и более групп по времени до события;
- 12) регрессионную модель, использованную для оценки связи между независимыми признаками и выживаемостью или временем наступления события;
- 13) меру риска (например, отношение угроз) и доверительный интервал для каждого независимого признака.

Условия для применения метода Каплана-Мейера [5]:

1. Момент начала наблюдения должен быть четко определен (например, момент появления первых симптомов, установления диагноза и т.д.).
2. Момент исхода должен быть четко определен (например, антенатальная гибель плода, момент диагностики клинического узкого таза и т.д.).
3. Для нецензурированных случаев необходимо знать дату исхода или период времени от начала наблюдения до развития исхода.
4. Для цензурированных случаев необходимо знать дату последнего контакта или период времени от начала наблюдения до последнего контакта.
5. Цензурированные наблюдения не должны отличаться по выживаемости от нецензурированных.
6. Методы оценки выживаемости и определения исхода должны быть одинаковыми для объектов исследования, включенных как на ранних, так и на более поздних сроках исследования.
7. Условия исследования не должны меняться с течением времени и не должны влиять на выживаемость (например, лечение должно быть одинаковым на протяжении всего времени исследования для всех случаев).
8. Доли цензурированных случаев не должны статистически значимо различаться в исследуемых группах до момента окончания периода наблюдения.
9. В анализируемой выборке должно быть более 30 исследуемых объектов.

Алгоритм применения метода Каплана-Мейера в программе IBM SPSS Statistics Version 25.0 [5, 10]:

1. Необходимо внести исследуемые данные в исходную форму. Если период проведения исследования внесен в базу в виде конкретной единицы времени (минуты, часы, дни, месяцы, годы и т.д.), то проведение дополнительных расчетов не требуется. Если в базу данных внесены даты начала и окончания периода наблюдения, то необходимо выполнить преобразование указанных дат в единицы времени. Рассмотрим наиболее простой способ: в первую очередь следует обратить внимание на то, чтобы указанные даты имели тип переменной «Дата», далее в меню «Преобразование» необходимо выбрать раздел «Вычислить переменную», в результате чего откроется одноименное окно. В верхнем поле слева «Целевая переменная» необходимо указать имя новой переменной, в поле справа «Числовое выражение» необходимо перенести соответствующие переменные (дата начала и дата исхода), из даты исхода вычесть дату начала исследования, нажать на кнопку «ОК», в результате чего в базе данных появится новая переменная, отражающая исследуемый интервал времени в секундах. Если необходимо отразить интервал времени в минутах, то в поле «Числовое выражение» необходимо разность даты исхода и даты начала исследования разделить на 60, если в часах – разделить на 3600, если в днях – разделить на 86400 и т.д.
2. Для использования метода Каплана-Мейера необходимо в меню «Анализ» выбрать раздел «Анализ выживаемости», подраздел «Каплан-Майер». В открывшемся одноименном окне в строку «Время» необходимо перенести переменную, которая кодирует интересующий нас интервал времени (период исследования / наблюдения). В строку «Состояние» необходимо перенести переменную, которая кодирует интересующий нас исход (например, наличие разрыва промежути). Далее необходимо нажать кнопку «Задать событие» и в открывшемся окне «Каплан-Майер: задать событие для переменной состояния» указать цифру, которая отражает развитие исхода в нашем исследовании. После чего необходимо нажать кнопку «Продолжить». При нажатии в основном окне кнопки «Параметры» откроется окно «Каплан-Майер: Параметры», в котором в поле «Статистика» необходимо поставить галочки напротив «Таблицы дожития», «Среднее и медиана выживания», в поле «Графики» - «Анализ выживаемости». После чего необходимо нажать кнопку «Продолжить».
3. В строку «Фактор» следует перенести группировочную переменную, которая гипотетически влияет на исход, после чего станет активной кнопка «Сравнить уровни фактора», при нажатии на которую откроется одноименное окно. В открывшемся окне в поле «Статистические критерии» следует поставить галочки напротив «Лонг-ранг», «Бреслоу» и «Тарон-Уэр». Критерий «Лонг-ранг» принимает за нулевую гипотезу предположение, что выживаемость во всех группах равна. Данный критерий может выявить различия только в том случае, если кривые выживаемости не пересекаются, в сравнении с другими критериями в большей степени позволяет выявить поздние различия. Критерий «Бреслоу» оценивает количество наблюдаемых в группах риска развития исхода в каждый момент времени, в сравнении с другими критериями в большей степени позволяет выявить ранние межгрупповые различия. Критерий «Тарон-Уэр» оценивает квадратный корень из числа исследуемых в группе риска, в большей степени позволяет выявить поздние межгрупповые различия. Далее необходимо нажать кнопку «Продолжить», в главном окне «Каплан-Майер» – нажать кнопку «ОК».
4. В полученном заключении программы SPSS в первой таблице представлен сводный отчет по наблюдениям. В таблице дожития представлена накопленная доля дожития (отсутствия изучаемого исхода), рассчитываемая как произведение выживаемости в данный момент времени на выживаемость в предыдущий момент времени. Таблица «Средние значения и медианы для времени выживания» содержит среднее время, медиану ожидаемого времени и их 95% доверительный интервал без события. Данные значения необходимо искать на пересечении соответствующего столбца и строки, соответствующей изучаемому исходу. В

таблице «Полные сравнения» представлены результаты применения трех критериев («Лонг-ранг», «Бреслоу» и «Тарон-Уэр»): необходимо обратить внимание на столбец «Значимость», который позволяет судить о наличии ($p < 0,05$) или отсутствии ($p > 0,05$) значимости различий. На основании наличия значимости двух критериев можно сделать вывод, что между группами имеются значимые различия. График «Функции выживания» содержит графическое отображение происходящих событий в каждой группе исследования, при этом каждое событие отражается на графике «ступенькой» / «шагом вниз». Если кривые выживаемости расходятся сразу и не пересекаются, то также говорят о наличии различий. Момент расхождения кривых отражает момент наблюдения, когда между изучаемыми группами появляются статистически значимые различия.

5. Если имеется более двух групп исследования, указанные критерии («Лонг-ранг», «Бреслоу» и «Тарон-Уэр») свидетельствуют лишь о том, что различия имеются только между некоторыми группами. Чтобы выделить группы, статистически значимо различающиеся между собой, необходимо при нажатии кнопки «Сравнить уровни фактора» в открывшемся окне поставить галочки напротив «Лонг-ранг», «Бреслоу» и «Тарон-Уэр», а также выбрать вариант сравнения «Попарно по все стратам», а не «Сразу по всем стратам». В полученном заключении программы SPSS необходимо обратить внимание на таблицу «Парные значения», в частности на значимость, располагающуюся на пересечении соответствующих групп. Значимость $p < 0,05$ свидетельствует о наличии различий между соответствующими группами.
6. Если исходные данные содержат стратификационную переменную, которая обуславливает их расслоение (например, класс тяжести заболевания), то в основном окне «Каплан-Майер» следует перенести ее в строку «Страты», что позволит проводить анализ выживаемости для каждой стратифицированной переменной по отдельности.
7. Для изучения одновременного влияния большего количества предикторов изучаемого исхода следует применять регрессионный анализ Кокса.

Условия применения таблиц дожития аналогичны таковым для метода Каплана-Мейера [5].

Алгоритм применения таблиц дожития в программе IBM SPSS Statistics Version 25.0 [5, 10]:

1. Необходимо внести исследуемые данные в исходную форму. Если период проведения исследования внесен в базу в виде конкретной единицы времени (минуты, часы, дни, месяцы, годы и т.д.), то проведение дополнительных расчетов не требуется. Если в базу данных внесены даты начала и окончания периода наблюдения, то необходимо выполнить преобразование указанных дат в единицы времени. Рассмотрим наиболее простой способ: в первую очередь следует обратить внимание на то, чтобы указанные даты имели тип переменной «Дата», далее в меню «Преобразование» необходимо выбрать раздел «Вычислить переменную», в результате чего откроется одноименное окно. В верхнем поле слева «Целевая переменная» необходимо указать имя новой переменной, в поле справа «Числовое выражение» необходимо перенести соответствующие переменные (дата начала и дата исхода), из даты исхода вычесть дату начала исследования, нажать на кнопку «ОК», в результате чего в базе данных появится новая переменная, отражающая исследуемый интервал времени в секундах. Если необходимо отразить интервал времени в минутах, то в поле «Числовое выражение» необходимо разность даты исхода и даты начала исследования разделить на 60, если в часах – разделить на 3600, если в днях – разделить на 86400 и т.д.
2. Для применения таблиц дожития необходимо в меню «Анализ» выбрать раздел «Анализ выживаемости», подраздел «Таблицы дожития». В открывшемся одноименном окне в строку «Время» необходимо перенести переменную, которая кодирует интересующий нас интервал времени (период исследования / наблюдения). В поле «Показывать интервалы времени» необходимо внести значение изучаемого периода времени и шаг (интервал времени, гипотетически играющий значимую роль). В строку «Состояние» необходимо

перенести переменную, которая кодирует интересующий нас исход (например, наличие разрыва промежности). Далее необходимо нажать кнопку «Задать событие» и в открывшемся окне «Таблицы дожития: задать событие» указать цифру, которая отражает развитие исхода в нашем исследовании. После чего необходимо нажать кнопку «Продолжить». При нажатии в основном окне кнопки «Параметры» откроется окно «Таблицы дожития: Параметры», в котором в поле «График» необходимо поставить галочку напротив «Анализ выживаемости», в поле «Сравнить уровни первого фактора» выбрать «Все». После чего необходимо нажать кнопку «Продолжить».

3. В строку «Фактор» следует перенести группировочную переменную, которая гипотетически влияет на исход, после чего станет активной кнопка «Задать диапазон», при нажатии на которую откроется одноименное окно. В открывшемся окне необходимо внести интересующие исследователя минимальное и максимальное значение группировочной переменной. После чего необходимо нажать кнопку «Продолжить», в главном окне «Таблицы дожития» – нажать кнопку «ОК».
4. В полученном заключении программы SPSS в таблице дожития представлен сводный отчет каждой группы исследования. В первом столбце указано начало интервала времени, который продолжается до времени следующего интервала. Во втором столбце указано количество обследованных, достигших определенного интервала времени. Соответственно, в первой строке данного столбца указано общее число пациентов, включенных в исследование. В третьем столбце указано число выбывающих в течение соответствующего интервала времени, в четвертом – число подвергающихся риску в этом интервале времени. В пятом столбце представлено число терминальных случаев (случаев развития исхода), а в шестом – их доля для соответствующего интервала времени. «Доля терминальных случаев» отражает вероятность того, что событие произойдет у члена выборки, который вошел в состав выборки в соответствующем интервале времени. «Доля выживших», входящая в седьмой столбец таблицы, отражает вероятность того, что событие не произойдет у члена выборки, который вошел в состав выборки в соответствующем интервале времени. Доля выживших равна разности единицы и доли терминальных случаев. Наибольший интерес для исследователей представляет восьмой столбец таблицы «Накопленная доля выживших к концу интервала», который позволяет оценить вероятность выживаемости (отсутствия исследуемого исхода) в зависимости от интервала времени (1,0 соответствует 100% выживаемости). В девятом столбце представлена стандартная ошибка функции дожития, с помощью которой оценивают вариабельность функции дожития. Столбец «Плотность вероятности» позволяет оценить вероятность наступления изучаемого события (например, разрыва промежности) в изучаемом интервале времени. Стандартная ошибка плотности вероятности, располагающаяся в одиннадцатом столбце данной таблицы, позволяет оценить вариабельность показателя плотности вероятности. Предпоследний столбец «Коэффициент смертности» отражает вероятность того, что в течение соответствующего интервала времени произойдет событие у того объекта, у которого оно еще не произошло на момент начала этого временного интервала. Последний столбец «Стандартная ошибка коэффициента смертности» отражает вариабельность коэффициента смертности.
5. В таблице «Среднее время выживания» представлено среднее значение ожидаемого времени жизни (отсутствия изучаемого исхода) в каждой исследуемой группе.
6. График «Функции выживания» содержит графическое отображение происходящих событий в каждой группе исследования, при этом каждое событие отражается на графике «ступенькой» / «шагом вниз». Если кривые выживаемости расходятся сразу и не пересекаются, то говорят о наличии различий. Момент расхождения кривых отражает момент наблюдения, когда между изучаемыми группами появляются статистически значимые различия.
7. Таблица «Полные сравнения» содержит значение обобщенного теста Уилкоксона (Гехана), степень свободы и значимость. Значимость $p < 0,05$ свидетельствует о наличии

статистически значимых различий между исследуемыми группами. Уровень значимости $p > 0,05$ не позволяет опровергнуть нулевую гипотезу об отсутствии различий в функции дожития между исследуемыми группами.

8. Для изучения одновременного влияния большего количества предикторов изучаемого исхода следует применять регрессионный анализ Кокса.

Регрессия Кокса в большинстве случаев используется для определения независимого влияния потенциальных факторов риска на скорость наступления изучаемого события за изучаемый промежуток времени. Регрессия Кокса отличается от логистической регрессии тем, что кроме исхода и предикторов, нас еще интересует период наблюдения. В сравнении с методом Каплана-Мейера и таблицами дожития регрессия Кокса позволяет оценить влияние множества предикторов на исход с учетом периода времени [6].

Условия для применения регрессии Кокса [6]:

1. Отсутствие коллинеарности – отсутствие высокой корреляционной связи между независимыми переменными (предикторами), в противном случае в модель необходимо включить функцию взаимодействия зависимых предикторов.
2. Все переменные должны линейно влиять на логарифм функции риска наступления изучаемого события.
3. Риск наступления изучаемого события должен быть пропорционален для любых двух объектов в любой интервал времени.
4. Момент начала наблюдения должен быть четко определен (например, момент появления первых симптомов, установления диагноза и т.д.).
5. Момент исхода должен быть четко определен (например, антенатальная гибель плода, момент диагностики клинического узкого таза и т.д.).
6. Для нецензурированных случаев необходимо знать дату исхода или период времени от начала наблюдения до развития исхода.
7. Для цензурированных случаев необходимо знать дату последнего контакта или период времени от начала наблюдения до последнего контакта.
8. Цензурированные наблюдения не должны отличаться по выживаемости от нецензурированных.
9. Методы оценки выживаемости и определения исхода должны быть одинаковыми для объектов исследования, включенных как на ранних, так и на более поздних сроках исследования.
10. Условия исследования не должны меняться с течением времени и не должны влиять на выживаемость (например, лечение должно быть одинаковым на протяжении всего времени исследования для всех случаев).
11. Доли цензурированных случаев не должны статистически значимо различаться в исследуемых группах до момента окончания периода наблюдения.
12. В анализируемой выборке должно быть более 30 объектов исследования.

Алгоритм применения регрессии Кокса в программе IBM SPSS Statistics Version 25.0 [6, 10]:

1. Необходимо внести исследуемые данные в исходную форму. Если период проведения исследования внесен в базу в виде конкретной единицы времени (минуты, часы, дни, месяцы, годы и т.д.), то проведение дополнительных расчетов не требуется. Если в базу данных внесены даты начала и окончания периода наблюдения, то необходимо выполнить преобразование указанных дат в единицы времени. Рассмотрим наиболее простой способ: в первую очередь следует обратить внимание на то, чтобы указанные даты имели тип переменной «Дата», далее в меню «Преобразование» необходимо выбрать раздел «Вычислить переменную», в результате чего откроется одноименное окно. В верхнем поле слева «Целевая переменная» необходимо указать имя новой переменной, в поле справа «Числовое выражение» необходимо перенести соответствующие переменные (дата начала и дата исхода), из даты исхода вычесть дату начала исследования, нажать на кнопку «ОК», в результате чего в базе данных появится новая переменная, отражающая исследуемый

интервал времени в секундах. Если необходимо отразить интервал времени в минутах, то в поле «Числовое выражение» необходимо разность даты исхода и даты начала исследования разделить на 60, если в часах – разделить на 3600, если в днях – разделить на 86400 и т.д.

2. Для применения регрессии Кокса необходимо в меню «Анализ» выбрать раздел «Анализ выживаемости», подраздел «Регрессия Кокса». В открывшемся одноименном окне в строку «Время» необходимо перенести переменную, которая кодирует интересующий нас интервал времени (период исследования / наблюдения). В строку «Состояние» необходимо перенести переменную, которая кодирует интересующий нас исход (например, наличие разрыва промежности). Далее необходимо нажать кнопку «Задать событие» и в открывшемся окне «Регрессия Кокса: задать событие для переменной состояния» указать цифру, которая отражает развитие изучаемого исхода. Следует сказать, что переменная, отражающая исход, не обязательно должна быть дихотомической, в данном окне имеется возможность введения интересующего нас диапазона значений непрерывной переменной, выступающего в данной ситуации в качестве исхода. Кроме того, данное окно позволяет выбрать список значений (например, все три степени ожирения матери) в качестве исхода, добавив их поочередно в квадратное поле в центре окна. После чего необходимо нажать кнопку «Продолжить».
3. В поле «Ковариаты» необходимо перенести переменные, интересующие исследователя в качестве предикторов. Нажатие кнопки «Категориальные» позволит выбрать категориальные переменные. В качестве опорной следует выбрать первую категорию. Далее следует выбрать метод введения, по умолчанию установлен метод форсированного ввода «Enter», при котором все переменные вводятся в модель одновременно. Начинаящим исследователям изменять тип ввода нецелесообразно. При нажатии кнопки «Графики» откроется одноименное окно, в котором следует поставить галочки напротив «Анализ выживаемости» и «Риск». Для значений категориальных или дихотомических переменных можно выбрать отдельные графики функции риска, перенеся их в строку «Отдельные линии для». При нажатии в основном окне кнопки «Сохранить» откроется окно «Регрессия Кокса: Сохранить», в котором в поле «Сохранить переменные модели» необходимо поставить галочки напротив «Функция выживания», «Функция риска», «DfBeta(s)» и «Частные остатки». После чего необходимо нажать кнопку «Продолжить». В окне «Регрессия Кокса: Параметры», которое появляется при нажатии кнопки «Параметры», следует выбрать «ДИ для $\exp(B)$ » и «Корреляция оценок» (для проверки первого условия применения регрессии Кокса). После чего в текущем окне необходимо нажать кнопку «Продолжить», а в главном окне «Регрессия Кокса» – нажать кнопку «ОК».
4. В полученном заключении программы SPSS в первой таблице представлен сводный отчет по наблюдениям. В таблице «Категориальное кодирование переменных» отражены принципы кодирования категориальных переменных. Во втором столбце данной таблицы содержится информация о частоте встречаемости объектов исследования в каждой группе. Блок 0 содержит данные о базовой модели без предикторов, следует обратить внимание на значение «-2Log-правдоподобие». Таблица «Универсальные критерии коэффициентов модели» в Блоке 1 содержит значения «-2Log-правдоподобие», «Хи-квадрат» и «Значимость». Уровень значимости $p < 0,05$ свидетельствует о том, что предикторы улучшают прогностическую способность модели. Кроме того, следует иметь в виду, что модель обладает большей прогностической способностью в сравнении с базовой только в том случае, когда «-2Log-правдоподобие» в Блоке 1 меньше, чем в Блоке 0. В данной таблице также представлены значения соответствующих критериев при сравнении модели с таковыми из предыдущего шага, что также отражает прогностическую способность изучаемой модели.
5. Наибольший интерес для исследователя представляет таблица «Переменные в уравнении», в которой особое внимание следует обратить на столбцы «Значимость» и «ДИ 95.0% для $\exp(B)$ ». Если статистическая значимость (критерия Вальда) $p < 0,05$, то можно сделать

вывод, что соответствующая независимая переменная вносит значимый вклад в прогностическую способность модели. Если 95% доверительный интервал для $\text{Exp}(B)$ включает в себя единицу, то можно сделать вывод, что соответствующая независимая переменная не вносит значимый вклад в прогностическую значимость модели. Столбец « $\text{Exp}(B)$ » отражает отношение рисков: если значение $\text{Exp}(B)$ больше 1, то положительное значение соответствующего предиктора является фактором риска развития изучаемого исхода, если меньше 1 – протективным фактором.

6. Анализ таблицы «Корреляционная матрица коэффициентов регрессии» позволяет проверить первое условие применения регрессии Кокса: отсутствие коллинеарности между исследуемыми предикторами, в противном случае в модель необходимо включить функцию взаимодействия зависимых предикторов (например, произведения соответствующих предикторов) или исключить из модели один из сильно коррелирующих факторов. Следует обратить внимание, что только сильные корреляционные связи (более 0,7) могут влиять на результаты анализа.
7. Графики «Накопленное выживание» и «Накопленный риск» позволяют визуальным способом определить наличие статистической значимости различий между исследуемыми категориями. Если указанные кривые сразу расходятся и не пересекаются, то говорят о наличии различий, в противном случае – об отсутствии. Момент расхождения кривых отражает момент наблюдения, когда между изучаемыми категориями появляются статистически значимые различия.
8. Для проверки третьего условия возможности применения регрессии Кокса, можно использовать стратификационную переменную, которую в основном окне «Регрессия Кокса» следует перенести в строку «Страты», что позволит проводить анализ выживаемости для каждой стратифицированной переменной по отдельности. Если исходные функции риска пропорциональны, то кривые на графиках будут параллельными. Между тем, визуальная оценка графиков позволяет оценить соблюдение третьего условия только ориентировочно.
9. Для проверки указанного третьего следует применять регрессию Кокса с ковариатами, зависящими от времени.

Алгоритм применения регрессии Кокса с ковариатами, зависящими от времени, в программе IBM SPSS Statistics Version 25.0 [6, 10]:

1. Для применения регрессии Кокса с ковариатами, зависящими от времени, необходимо в меню «Анализ» выбрать раздел «Анализ выживаемости», подраздел «Кокс с ковариатами, завис.от времени». В открывшемся окне «Вычислить зависящую от времени ковариату» в строку «Выражение для T_COV_» необходимо перенести исследуемую временную переменную, которую необходимо разделить на гипотетически значимый интервал времени, новая эквивалентная изучаемой переменной создается автоматически.
2. При нажатии кнопки «Модель» открывается окно, аналогичное таковому в простой регрессии Кокса. Для проверки соблюдения третьего условия необходимо поочередно построить модели с учетом только одного предиктора.
3. Для оценки взаимодействия между исследуемым предиктором с переменной времени необходимо обратить внимание на таблицу «Переменные в уравнении». Значимость $p > 0,05$ в соответствующей строке свидетельствует о соблюдении третьего условия.
4. Другие таблицы в данном заключении аналогичны заключению при применении простой регрессии Кокса.

Для проверки второго условия необходимо оценить мартингальные остатки путем построения базовой модели без включения соответствующего предиктора. После чего необходимо построить график (скатерограмму) зависимости мартингальных остатков новой базовой модели от исключенного предиктора, построить на данном графике прямую Лоесса и визуальным способом оценить линейность данной прямой [6].

Таким образом, наиболее универсальным методом из вышеперечисленных является регрессия Кокса, так как она позволяет не только произвести относительно точный анализ

выживаемости, но и позволяет включить в модель значительное число независимых переменных, влияющих на риск развития конкретного события [6].

Заключение. Использование описанных алгоритмов применения анализа выживаемости позволит повысить уровень представления результатов биомедицинских исследований. Пакет программ «IBM SPSS Statistics Version 25.0» позволяет достаточно просто и наглядно применять основные методы анализа выживаемости для анализа времени наступления изучаемых событий [10].

Список литературы:

1. Бююль А., Цёфель П. SPSS: Искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей. СПб. ДиаСофтЮП. 2005.
2. Левин И.А., Манухин И.Б., Пономарева Ю.Н., Шуметов В.Г. Методология и практика анализа данных в медицине: монография. М. Тель-Авив. АПЛИТ. 2010.
3. Реброва О.Ю. Статистический анализ медицинских данных. Применение пакета прикладных программ STATISTICA. М. МедиаСфера. 2006.
4. Peacock J.L., Peacock P.J. Oxford Handbook of Medical Statistics. Oxford University Press. 2011.
5. Шарашова Е.Е., Холматова К.К., Горбатова М.А., Гржибовский А.М. Применение анализа выживаемости в здравоохранении с использованием статистических программ SPSS. Наука и здравоохранение. 2017. 5. 5-28.
6. Шарашова Е.Е., Холматова К.К., Горбатова М.А., Гржибовский А.М. Применение регрессии Кокса в здравоохранении с использованием пакета статистических программ SPSS. Наука и здравоохранение. 2017. 6. 5-27.
7. Куликов С.М., Паровичникова Е.Н., Савченко В.Г. Анализ выживаемости или событийный анализ: типовые ошибки ретроспективного метода. Клиническая онкогематология. 2010. 3(2). 176-183.
8. Lang T.A., Altman D.G. Basic statistical reporting for articles published in Biomedical Journals: The “Statistical Analyses and Methods in the Published Literature” or the SAMPL Guidelines. International Journal of Nursing Studies. 2015. 52(1). 5-9. doi:10.1016/j.ijnurstu.2014.09.006.
9. Lang T.A., Altman D.G. Statistical analyses and methods in the published literature: The SAMPL guidelines. Medical Writing. 2016. 25(3). 31–36. doi: 10.18243/eon/2016.9.7.4.
10. Наследов А. IBMSPPSSStatistics 20 и AMOS: профессиональный статистический анализ данных. Санкт-Петербург. Питер. 2013.

References:

1. Buyul A., Cefel P. SPSS: the art of information processing. Moscow. DiaSoft. 2005. in Russian.
2. Levin I.A., Manukhin I.B., Ponomareva Yu.N., Shumetov V.G. Methodology and practice of data analysis in medicine: monograph. Moscow. Tel-Aviv. APLIT. 2010. in Russian.
3. Rebrova O.Yu. Statistical analysis of medical data. Application of the STATISTICA application software package. Moscow. Media Sphere. 2006. in Russian.
4. Peacock J.L., Peacock P.J. Oxford Handbook of Medical Statistics. Oxford University Press. 2011.
5. Sharashova E.E., Kholmatova K.K., Gorbatova M.A., Grjibovski A.M. Survival analysis in health sciences using SPSS software. Science & Healthcare. 2017. 5. 5-28. in Russian.
6. Sharashova E.E., Kholmatova K.K., Gorbatova M.A., Grjibovski A.M. Cox regression in health sciences using SPSS software. Science & Healthcare. 2017. 6. 5-27. in Russian.
7. Kulikov S.M., Parovitchnikova E.N., Savchenko V.G. Survival or time-to-event analysis: common pitfalls of retrospective approach. Clinical oncohematology. 2010. 3(2). 176-183.
8. Lang T.A., Altman D.G. Basic statistical reporting for articles published in Biomedical Journals: The “Statistical Analyses and Methods in the Published Literature” or the SAMPL Guidelines. International Journal of Nursing Studies. 2015. 52(1). 5-9. doi:10.1016/j.ijnurstu.2014.09.006.

9. Lang T.A., Altman D.G. Statistical analyses and methods in the published literature: The SAMPL guidelines. Medical Writing. 2016. 25(3). 31–36. doi: 10.18243/eon/2016.9.7.4.
10. Nasledov A. SPSS Statistics 20 and AMOS: professional statistical analysis of data. Saint-Petersburg. Peter. 2013.in Russian.