

Получена: 21 ноября 2017 / Принята: 10 декабря 2017 / Опубликовано online: 31 декабря 2017

УДК 614.2 + 303.4

## **ПРИМЕНЕНИЕ РЕГРЕССИИ КОКСА В ЗДРАВООХРАНЕНИИ С ИСПОЛЬЗОВАНИЕМ ПАКЕТА СТАТИСТИЧЕСКИХ ПРОГРАММ SPSS**

**Екатерина Е. Шарашова** <sup>1</sup>,

**Камила К. Холматова** <sup>2</sup>,

**Мария А. Горбатова** <sup>2</sup>, <http://orcid.org/0000-0002-6363-9595>

**Андрей М. Гржибовский** <sup>2-5</sup>, <http://orcid.org/0000-0002-5464-0498>

<sup>1</sup> Арктический университет Норвегии, Тромсё, Норвегия;

<sup>2</sup> Северный Государственный Медицинский Университет, г. Архангельск, Россия;

<sup>3</sup> Национальный Институт Общественного Здоровоохранения, г. Осло, Норвегия;

<sup>4</sup> Международный Казахско-Турецкий Университет им. Х.А. Ясави, г. Туркестан, Казахстан;

<sup>5</sup> Северо-Восточный Федеральный Университет, г. Якутск, Россия.

### **Резюме**

В статье представлены теоретические основы проведения одного из наиболее популярных многомерных методов анализа дожития – регрессии Кокса. Данный метод анализа в большинстве случаев используется для определения независимого влияния потенциальных факторов риска на скорость наступления изучаемого события за изучаемый промежуток времени. На практическом примере описаны принципы проведения анализа пропорциональных рисков Кокса и основные принципы интерпретации полученной с помощью пакета прикладных статистических программ SPSS информации.

*Ключевые слова:* анализ дожития, выживаемость, регрессионный анализ Кокса, пропорциональность рисков, SPSS.

### **Abstract**

## **COX REGRESSION IN HEALTH SCIENCES USING SPSS SOFTWARE**

**Ekaterina E. Sharashova** <sup>1</sup>,

**Kamila K. Kholmatova** <sup>2</sup>,

**Maria A. Gorbatova** <sup>2</sup>, <http://orcid.org/0000-0002-6363-9595>

**Andrej M. Grjibovski** <sup>2-5</sup>, <http://orcid.org/0000-0002-5464-0498>

<sup>1</sup> Arctic University of Norway, Tromsø, Norway;

<sup>2</sup> Northern State Medical University, Arkhangelsk, Russia;

<sup>3</sup> Norwegian Institute of Public Health, Oslo, Norway;

<sup>4</sup> International Kazakh-Turkish University, Turkestan, Kazakhstan;

<sup>5</sup> North-Eastern Federal University, Yakutsk, Russia.

In this paper presents theoretical bases of the Cox regression as one of the most popular multidimensional methods of survival analysis. This method of analysis in most cases is used to determine the independent influence of potential risk factors on the rate of occurrence of the event under study during the time period studied. The practical example describes the principles of Cox proportional hazard analysis and the basic principles of interpreting the information received with the SPSS application statistical software package.

*Keywords:* survival analysis, survival, Cox regression analysis, proportionality of risks, SPSS.

Түйіндеме

## SPSS СТАТИСТИКАЛЫҚ БАҒДАРЛАМАЛАР ПАКЕТІН ПАЙДАЛАНУМЕН ДЕНСАУЛЫҚ САҚТАУДАҒЫ КОКС РЕГРЕССИЯСЫН ҚОЛДАНУ

**Екатерина Е. Шарашова**<sup>1</sup>,**Камила К. Холматова**<sup>2</sup>,**Мария А. Горбатова**<sup>2</sup>, <http://orcid.org/0000-0002-6363-9595>**Андрей М. Гржибовский**<sup>2-5</sup>, <http://orcid.org/0000-0002-5464-0498><sup>1</sup> Норвегия Арктикалық университеті, Тромсё, Норвегия;<sup>2</sup> Солтүстік Мемлекеттік Медициналық Университеті, Архангельск қ., Ресей;<sup>3</sup> Қоғамдық Денсаулық сақтау Ұлттық Институты, Осло қ., Норвегия;<sup>4</sup> Х.А. Ясави ат. Халықаралық Қазақ – Түрік Университеті, Туркестан, Қазақстан;<sup>5</sup> Солтүстік - Шығыс Федералдық Университеті, Якутск қ., Ресей;

Мақалада өмір сүру Кокс регрессиясын - талдаудың ең белгілі көпшамалы әдістерінің бірін өткізудің теориялық негіздері берілген. Көптеген жағдайларда талдаудың осы әдісі зерделенетін уақыт арасындағы зерттелетін жағдайдың орнауы жылдамдығына қауіптің әлеуетті факторларының тәуелсіз әсерін анықтау үшін қолданылады. Тәжірибелік үлгіде Кокс пропорционалдық қауіптеріне талдау өткізу принциптері және ақпараттың SPSS қолданбалы статистикалық бағдарламалары пакеті көмегімен алынған интерпретациялар негізгі принциптері.

**Негізгі сөздер:** өмір сүруге жетуді талдау, өміршеңдік, Кокс регрессиялық талдауы, қауіптердің пропорционалдығы, SPSS.

### Библиографическая ссылка:

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. С. 5-27.

Sharashova E.E., Kholmatova K.K., Gorbatova M.A., Grijbovski A.M. Cox regression in health sciences using spss software. *Nauka i Zdravookhranenie* [Science & Healthcare]. 2017, 6, pp. 5-27.

Шарашова Е.Е., Холматова К.К., Горбатова М.А., Гржибовский А.М. SPSS статистикалық бағдарламалар пакетін пайдаланумен денсаулық сақтаудағы Кокс регрессиясын қолдану // Ғылым және Денсаулық сақтау. 2017. №6. Б. 5-27.

В предыдущем выпуске журнала мы начали изучение принципов использования анализа выживаемости. Данные, изменяющиеся во времени, исследуются группой методов, обобщенно называемых анализом наступления события (event history analysis) или анализом дожития (survival analysis). Кратко напомним основные понятия, которые лежат в основе методов оценки выживаемости. Анализ дожития относится к статистическим подходам изучения зависимости возможного наступления события от времени. Исторически сложилось так, что термин «дожитие» подразумевал возможный летальный исход, однако, в медицине под

событием может рассматриваться также и выздоровление, и рецидив заболевания, и любые исходы, которые интересны для изучения относительно времени их возникновения [13].

Как уже упоминалось, наиболее распространенными методами являются таблицы выживаемости (life / mortality tables), метод Каплана–Мейера (Kaplan–Meier method), которые мы рассмотрели ранее, регрессия Кокса (Cox Regression) и регрессия Кокса с зависящими от времени переменными (Cox regression with time dependent covariates) [10, 19, 20]. В данной статье будет рассмотрен подробно регрессионный анализ Кокса и

упомянут регрессионный анализ Кокса с зависящими от времени предикторами.

Регрессия Кокса, которую также называют моделью пропорциональных рисков (Cox proportional hazards model), изучает зависимость времени дожития (survival time) от независимых переменных (predictor variables). Этот полупараметрический метод предполагает прогнозирование риска наступления события (hazard risk) для рассматриваемого объекта и оценивает влияние независимых переменных на этот риск. При этом риск наступления события является функцией, зависящей от времени, и выявляет вероятность наступления события для объектов, которые находятся в группе риска [15]. Никаких предположений о виде функции интенсивности/риска не делается, в этом состоит непараметрическая часть метода. Однако все переменные должны линейно влиять на логарифм функции риска наступления события, что составляет параметрическую компоненту метода.

Объектом исследования может быть индивид (пациент), для которого прогнозируется риск наступления события. По причине того, что объект уже попал под наблюдение и исследуется, он автоматически входит в группу риска, т.е. в любой промежуток времени с ним может случиться интересующее исследователя событие. Суть определения события состоит в том, чтобы выяснить, имеется ли риск того, что объект умрет, выздоровеет или с ним произойдет другое интересующее исследователя событие, в рассматриваемый период времени.

Также в модель включены независимые переменные (predictors) - характеристики объекта (например, возраст, пол пациента, сопутствующие заболевания и др.), которые могут влиять на риск наступления события.

Регрессия Кокса, как и все вышеперечисленные методы оценки дожития, имеет принципиальную особенность в контексте построения выборки: ее объем и структура до наступления события могут меняться. Это объясняется тем, что время не является определенным, вернее, оно определено только у тех объектов, у которых событие наступило. Для остальных участников выборки показатель времени остается неизвестным, т.к. событие может вовсе не произойти. Кроме того, объекты наблюдения

могут более не подлежать участию в исследовании в связи с изменением обстоятельств. Таким образом, с течением времени данные могут содержать неполную информацию – являться цензурированными (censored) [2]. Цензурирование является практически универсальным свойством данных анализа дожития, наиболее распространенной формой которого является правостороннее цензурирование - время наблюдения истекает или объект удаляется из исследования до того, как наступит событие (например, пациенты могут быть еще живы к концу исследования или выходят из-под наблюдения по разным причинам). Левосторонним цензурирование считается в том случае, если начало пребывания в группе риска неизвестно. Очень редко цензурирование может быть интервально-цензурированным, когда соблюдаются оба первых условия вместе. Последних двух вариантов исследования следует избегать в научной деятельности. Кроме того, в модели дожития цензурирование должно быть независимым от возможного значения риска для объекта, иначе результат будет искажен [11].

При анализе данных дожития две функции представляют фундаментальный интерес – это функция дожития (survivor function) и функция риска (hazard function) [1, 10, 12, 14, 20].

Пусть переменная  $T$  является временем дожития (survival time). Эта переменная отражает время, прошедшее от начала исследования (например, время от постановки диагноза) до события (например, смерти пациента по причине этого заболевания). Значения  $T$  можно рассматривать, как имеющие распределение вероятностей (probability distribution). Предположим, что функция плотности вероятности (probability density function) случайной величины  $T$  задается как  $f(T)$ . Тогда распределение вероятностей функции  $T$  в момент времени  $t$  выражается формулой [11, 16]:

$$F(T) = \Pr(T < t) = \int_0^t f(t) dt$$

Функция дожития,  $S(T)$  – это вероятность, что объект исследования выживет (событие не

наступит) по прошествии времени  $T$ . Тогда формула будет иметь вид:

$$S(T) = \Pr(T \geq t) \\ = 1 - F(T)$$

Функция риска (hazard function) – это вероятность, что событие наступит в течение ничтожно малого временного интервала, учитывая, что объект доживает до начала этого интервала. Математическим выражением для этой функции будет:

$$h(T) = \lim_{\Delta T \rightarrow 0} \frac{\Pr(T \leq t < (T + \Delta T) | T \leq t)}{F(T + \Delta T) - F(T)} \\ = \lim_{\Delta T \rightarrow 0} \frac{\Delta F(T)}{\Delta T} \\ = \frac{f(T)}{S(T)}$$

Кумулятивная функция риска (cumulative hazard function)  $H(T)$  – это сумма частот индивидуальных рисков от начала исследования до времени  $T$ :

$$H(T) = \int_0^T h(u) du$$

Таким образом, функция риска является производной от функции кумулятивного риска. Отношение функции кумулятивного риска к функции кумулятивного дожития выражается формулами:

$$S(T) = e^{-H(T)}$$

$$H(T) = -\ln(S(T))$$

Насколько мы можем судить по этим формулам, функция плотности вероятности, функция риска и функция дожития – математически взаимосвязаны [18].

Отношение между риском наступления события и независимыми переменными выражается формулой:

$$h(T) = h_0(T) e^{\sum_{i=1}^p x_i \beta_i}$$

где  
 $T$  – время;

$x_1, \dots, x_p$  – независимые переменные (предикторы).

$h_0(T)$  – основной риск наступления события (baseline hazard rate), одинаковый для всех объектов наблюдения (при условии, что значения всех независимых переменных равны 0, то есть они никоим образом не влияют на исход);

$\beta_1, \dots, \beta_p$  – регрессионные коэффициенты (regression coefficients);

В этой модели регрессионные коэффициенты указывают на влияние каждого предиктора на функцию риска, и при увеличении значения предиктора на единицу, если значения остальных переменных неизменны, риск наступления события возрастает в  $e$  раз [11].

Таким образом, в регрессии Кокса нас будут интересовать три вида показателей: исход, период наблюдения, предикторы. Если бы нас интересовали только исход и предикторы, то мы могли бы воспользоваться для анализа методом логистической регрессии [8]. Если бы мы были заинтересованы в оценке исхода относительно времени, то можно было бы провести анализ методом Каплана-Мейера или построения таблиц дожития [3]. Только в регрессии Кокса мы можем оценить влияние множества предикторов на исход с учетом периода наблюдения.

По форме модель пропорциональных рисков схожа с моделью логистической регрессии, при оценке последней используется метод максимального правдоподобия (log-likelihood, -2LL в SPSS). Попытка использовать для оценки продолжительности жизни или других показателей критерий максимального правдоподобия, привела к необходимости решать интегральные уравнения (в связи с наличием цензурированных случаев), что представляет собой достаточно трудоемкий процесс. Для оценки модели пропорциональных рисков Д.Р. Коксом для восстановления зависимостей при анализе выживаемости предложен метод частичного правдоподобия (partial likelihood), который позволяет оценить значения компонент вектора коэффициентов модели, доставляющие максимум так называемой частичной функции максимального правдоподобия, и строить эффективные в

вычислительном отношении процедуры оценки, но по принципу идентичен методу максимального правдоподобия [8, 11]. Данный метод так же показывает, какая часть информации осталась не объясненной после применения модели для наших фактических данных. Модель обладает меньшей предсказательной способностью, если значения этого показателя возрастают [8]. В программе SPSS данный показатель модели с предикторами сравнивается с показателем базовой модели (без предикторов), которая приписывает всем объектам выборки значение исхода, которое чаще встречается в реальной выборке (в нашей выборке у 161 (58,1%) пациента ККТ не зарегистрирована).

В регрессии Кокса (как и в логистической регрессии) нулевая гипотеза (предиктор не имеет связи с зависимой переменной, т.е. его регрессионный коэффициент значимо не отличается от нуля) проверяется при помощи критерия Wald. Если регрессионный коэффициент значимо отличается от нуля, то независимая переменная вносит значимый вклад в предсказательную способность модели [13].

Показатель  $-2LL$  достаточно сложен для интерпретации, поэтому используется коэффициент  $\text{Exp}(B)$ , который показывает во сколько раз изменяется риск возникновения исхода, если значение предиктора изменяется на единицу.

Для применения регрессии Кокса необходимо соблюдение целого ряда критериев [17]: в основе регрессии Кокса имеется три базовых предположения относительно переменных, которые будут представлены первыми в списке. Следует также учитывать условия, которые являются значимыми для всех методов анализа дожития и были представлены в предыдущей статье [3]:

1. все предикторы независимы. Если выявлено взаимное влияние независимых переменных, то в модель необходимо включить функцию взаимодействия этих факторов;

2. все переменные линейно влияют на логарифм функции риска наступления события;

3. риск наступления события для любых двух объектов в любой интервал времени пропорционален. Например, если риск

развития инфаркта миокарда у мужчин в два раза выше, чем у женщин в возрасте пятидесяти лет, то он останется таким же в возрасте шестидесяти лет или любом другом возрасте. Условие подразумевает предположение о мультипликативной зависимости между функцией интенсивности и логлинейной функцией ковариат и является главным для построения модели в регрессии Кокса без зависимых от времени переменных. Данное предположение также называют гипотезой о пропорциональности рисков. Хотя естественно, что данное условие не всегда соблюдается, даже для представленного примера риск развития инфаркта миокарда у мужчин и женщин к возрасту 60-70 лет становится примерно одинаковым;

4. момент начала и окончания исследования (возникновения исхода или окончания периода наблюдения) или интервал наблюдения в единицах времени должны быть точно определены для каждого члена выборки;

5. определение исхода и момент его возникновения также должны быть четко зафиксированы;

6. цензурированные и нецензурированные наблюдения не должны отличаться по выживаемости друг от друга;

7. методы оценки выживаемости и определения исхода одинаковы на протяжении всего исследования;

8. условия, которые влияют на выживаемость, не меняются в ходе исследования;

Особенностью кодировки исхода в базе данных для удобства работы с программным обеспечением является то, что наличие исхода за период наблюдения должно быть закодировано единицей, а его отсутствие – нулем.

Рассмотрим регрессию Кокса на примере уже знакомого исследования, проведенного в Архангельске, главной целью которого было выявление предикторов осложнений и факторов, влияющих на прогноз пациентов с инфарктом миокарда при различных вариантах нарушения углеводного обмена. Сплошная выборка была сформирована из пациентов в возрасте до восьмидесяти лет, госпитализированных, в первые пять суток инфаркта миокарда. Данные 275 пациентов, которые прошли пероральный тест толерантности к глюкозе для диагностики

нарушений углеводного обмена и были выписаны из стационара, использовались для оценки отдаленного прогноза. В течение двух лет была оценена частота возникновения неблагоприятных сердечно-сосудистых исходов: смерть от сердечно-сосудистого заболевания, повторный инфаркт миокарда, нарушение мозгового кровообращения, госпитализация по поводу ишемической болезни сердца, реваскуляризация. Все перечисленные исходы затем были объединены и, конечным событием для анализа являлась комбинированная конечная точка (ККТ). В базе данных фиксировалась дата возникновения первого из данных исходов для пациента. Более подробно ознакомиться с основными результатами исследования можно в опубликованных ранее статьях [4-7].

Как мы уже упоминали, период наблюдения должен быть закодирован каким-либо интервалом времени (то есть количеством дней, месяцев и т.д.). Если в базу данных отдельно внесены даты возникновения

инфаркта миокарда и первого неблагоприятного сердечно-сосудистого события, то необходимо перекодировать их в одну переменную, обозначающую непосредственно интервал времени между двумя событиями, с помощью меню «Transform» на панели инструментов, далее подменю «Date and Time Wizard» или «Compute variable», что было подробно описано в предыдущей статье [3]. То есть время наблюдения пациента начиналось от даты возникновения инфаркта миокарда до момента возникновения любого первого из перечисленных сердечно-сосудистых событий, до выпадения пациента из-под наблюдения или до момента истечения двух лет после инфаркта миокарда (цензурирование). Присвоим имя новой переменной (duration days CEP, период наблюдения в днях для ККТ).

Для выполнения анализа дожития методом регрессии Кокса в SPSS на панели инструментов необходимо выбрать меню «Analyze», в нем – раздел «Survival», затем «Cox regression» (рис.1).

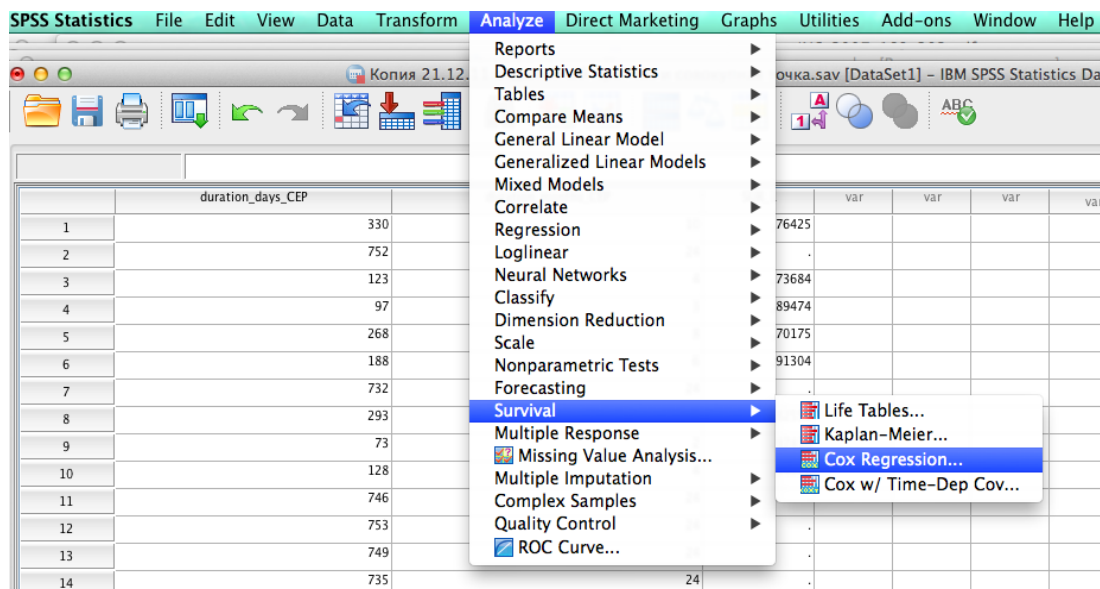


Рисунок 1. Выбор регрессии Кокса на панели инструментов.

В открывшемся диалоговом окне (рис. 2) в строку «Time» внесем переменную, обозначающую период времени наблюдения (период наблюдения в днях для ККТ), путем выделения самой переменной из списка в окне слева и нажатием на стрелку рядом со строкой, в строку «Status» внесем переменную, кодирующую изучаемый исход (в нашем случае, ККТ). Затем нажмем на «Define event» и внесем в появившемся окне с строку «Single value» значение переменной, которым мы

отмечали в базе наличие исхода (рис. 3). В нашем примере это значение переменной «комб\_без\_некоронарных»=1. Отметим, что переменная, обозначающая исход, не обязательно должна быть закодирована дихотомической (да/нет, 0/1). Ниже в этом окне представлены строки для внесения интервала значений непрерывной переменной (например, если в качестве исхода нас интересует развитие у пациента ожирения I степени, и у нас есть непрерывная переменная «индекс

массы тела» ( $\text{кг}/\text{м}^2$ ), то исходом для нас будут служить ее значения в интервале 30,0-34,99  $\text{кг}/\text{м}^2$  и для внесения нескольких значений из списка значений категориальной переменной (например, если значения того же индекса массы тела в базе зашифрованы в качестве категорий (0 – норма, 1 – избыточная масса

тела, 2 – ожирение I степени, 3 – ожирение II степени, 4 – ожирение III степени), а в качестве исхода нас будет интересовать сам факт наличия ожирения, то в строку «List of values» мы поочередно можем внести значения 2, 3, 4, добавляя их кнопкой Add в квадратное поле в центре окна.

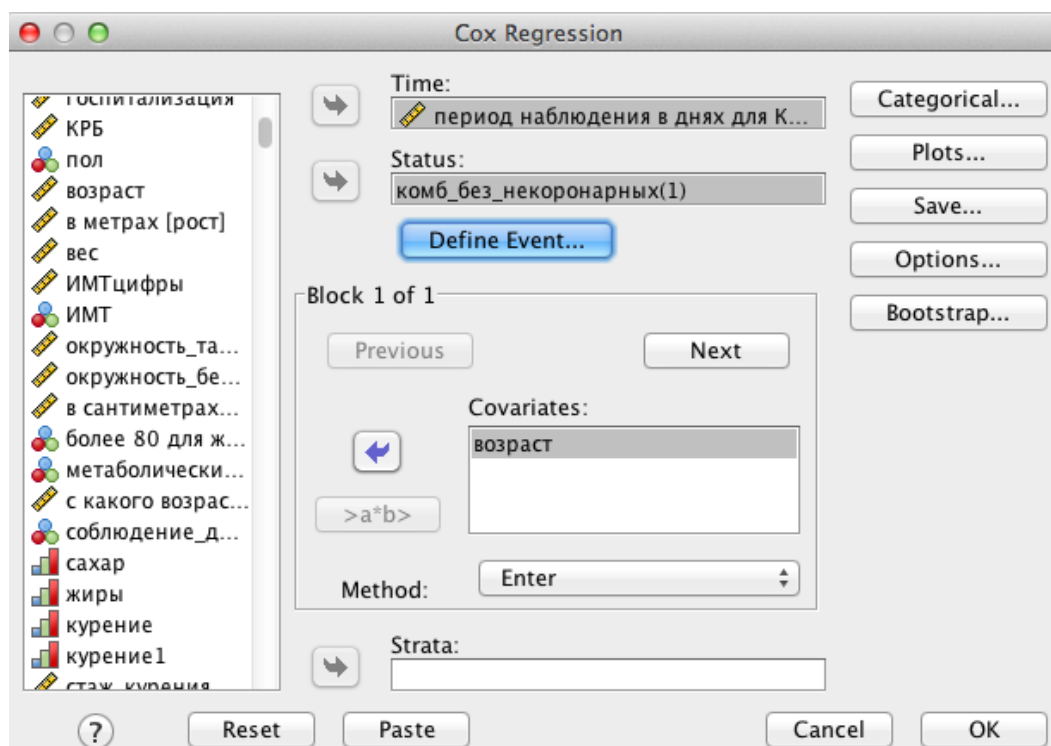


Рисунок 2. Диалоговое окно «Cox regression».

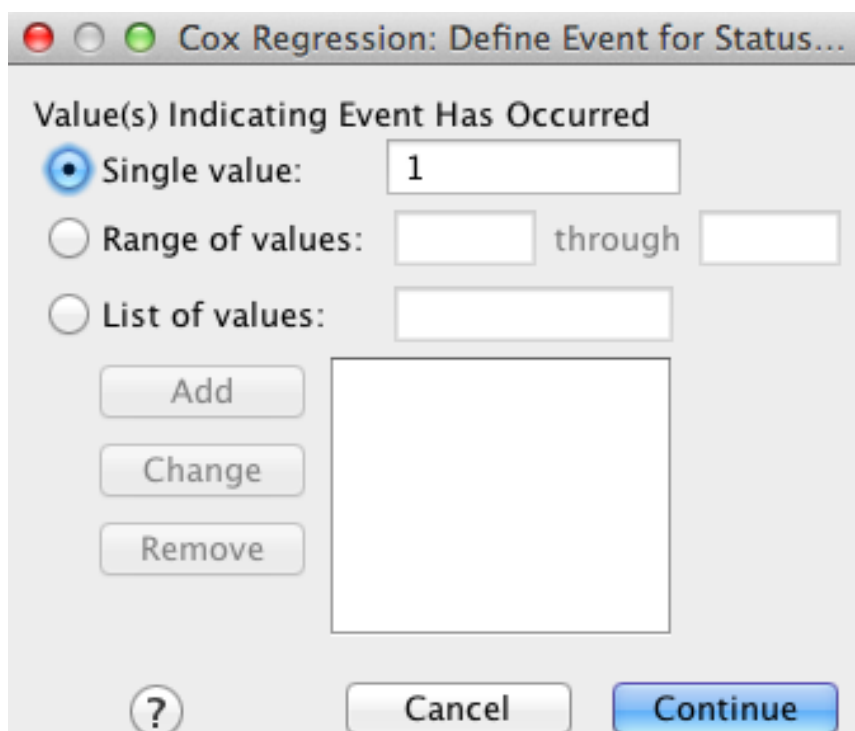


Рисунок 3. Диалоговое окно «Cox Regression: Define event».

Далее из списка переменных внесем в раздел «Covariates» интересующие нас независимые переменные (предикторы или ковариаты). Сначала проведем простой регрессионный анализ Кокса с учетом воздействия на зависимую переменную только одной независимой переменной «возраст». По умолчанию все независимые переменные считаются непрерывными. В нашем случае переменная также количественная, поэтому дополнительно ничего предпринимать не нужно.

Далее предлагается выбрать метод введения предикторов в модель. В связи с тем, что у нас пока всего один предиктор, то

выбор метода не имеет значения, оставим по умолчанию метод «Enter». Рассмотрим особенности методов введения на примере множественной регрессии Кокса.

Справа в окне построения графиков «Plots» (рис. 4) можно выбрать интересующие нас графики четырех возможных видов (отметим Survival). По умолчанию график будет построен по точкам, где независимая переменная имеет свое среднее значение, поэтому после названия переменной в скобках указывается слово «Mean». Если нас заинтересует какое-либо другое значение предиктора, то можно указать его в разделе «Change Value/Value».

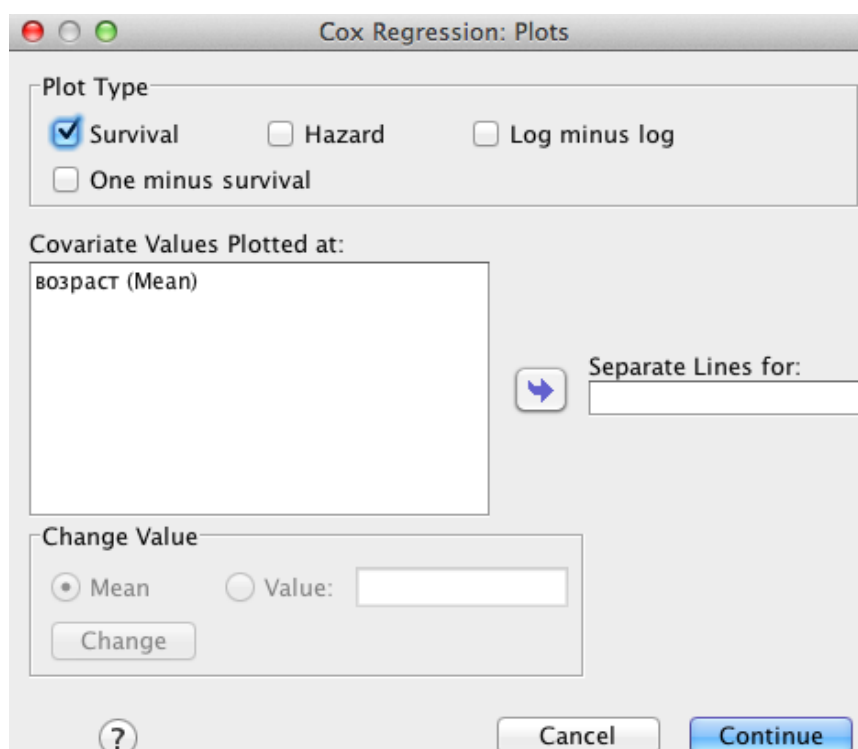


Рисунок 4. Диалоговое окно «Plots».

В разделе «Save» имеется возможность сохранить значения моделируемых функций (риска, вероятности дожития) для каждого члена выборки в качестве новой переменной в базе данных. Это позволит в дальнейшем работать отдельно с этой переменной, в том числе строить графики, отображающие данную переменную в других графических редакторах [1]. Отметим «Hazard function».

В окне «Options» должен быть отмечен пункт «доверительные интервалы для  $\exp(B)$ » (CI for  $\exp(B)$ ). Запустим анализ путем нажатия на кнопку ОК.

Оценка результатов анализа. Как и в любом регрессионном анализе первая часть

вывода содержит данные о базовой модели (Block 0), значение  $-2LL$  для модели без предикторов составляет 1205,451. Затем представлена модель с одной независимой переменной «возраст» (Block 1). В одной таблице представлены значения нескольких тестов для оценки качества построенной модели. Мы видим (рис. 5), что для модели с одной переменной значения  $-2LL$  уменьшилось (1200,597), значение  $\chi^2$  составляет 4,861, на уровне значимости 0,027 (overall (score) test), что не позволяет нам принять нулевую гипотезу, значит, предиктор улучшает предсказательную способность модели. В данной таблице также



представлены значения тестов likelihood ratio test,  $\chi^2$  при сравнении модели с таковыми из предыдущего шага ( $\chi^2$  при этом представляет собой разность между -2LL шагов 0 и 1) и блока (разность между значениями -2LL

блоков), но так как у нас модель одна и имеет единственный предиктор, то эти два показателя идентичны. Чаще всего показатели всех тестов имеют одну тенденцию и близки по значениям [17].

Omnibus Tests of Model Coefficients<sup>a</sup>

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
1200,597	4,861	1	,027	4,854	1	,028	4,854	1	,028

a. Beginning Block Number 1. Method = Enter

Рисунок 5. Оценка коэффициентов модели с одним предиктором.

Далее следует таблица с переменными, вошедшими в модель (рис. 6). В модели Кокса нет константы. Коэффициент регрессионного уравнения (B) для единственного фактора «возраст» составляет 0,021 (однако его сложно

интерпретировать, так как значение регрессионного коэффициента не дает информации о силе влияния переменной риск развития исхода).

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95,0% CI for Exp(B)	
							Lower	Upper
возраст	,021	,009	4,834	1	,028	1,021	1,002	1,040

Рисунок 6. Переменные в уравнении регрессии (Шаг 1).

Далее в таблице представлены стандартная ошибка коэффициента B (S.E.); критерий Вальда (Wald, критерий значимости коэффициента B для соответствующей независимой переменной, зависит от значения критерия и числа степеней свободы (df)); статистическая значимость критерия Вальда (Sig., при значениях <0,05 независимая переменная вносит значимый вклад в предсказательную способность модели); Exp(B) – экспонента B или  $e^B$ , отражает отношение рисков [2]. При увеличении возраста на один год риск более раннего наступления исхода увеличиваются в 1,021 раза или на 2,1%. Границы 95% ДИ (95,0% CI for Exp(B)) отражают, в каких пределах с вероятностью 95% находится значение этого коэффициента для популяции, из которой была сформирована изучаемая выборка. То есть ежегодно для членов популяции риск развития исхода может возрастать на 0,2-4,0%. Чем шире будут границы интервала, тем менее точно значение коэффициента b отражает его значение переменной для генеральной совокупности, и наоборот. Можно предложить другой вариант интерпретации: при увеличении возраста на десять лет риск развития исхода увеличивается в  $1,021^{10}$  (1,23) раз или на 23%. Тем не менее, это не

свидетельствует о том, что данная переменная линейно связана с логарифмом функции риска. На данном этапе можно лишь предположить, что в этой связи присутствует линейный компонент. Для подтверждения требуется отдельный анализ, который будет представлен далее.

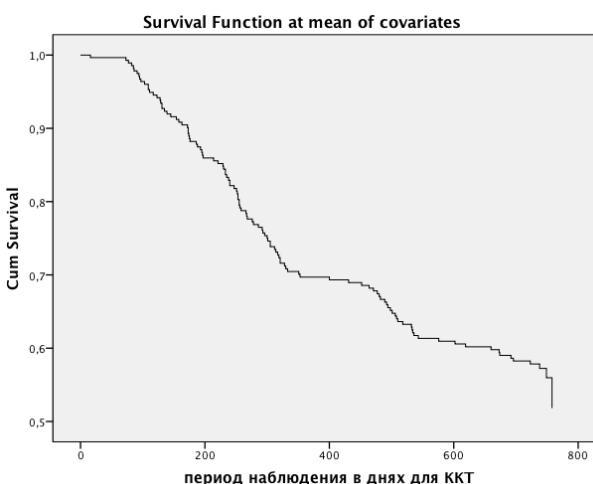


Рисунок 7. График кумулятивной выживаемости (вероятности, что ККТ не случится) в течение периода наблюдения при среднем значении переменной «возраст».

На рисунке 7 отражена графически вероятность того, что у пациента в возрасте 60,5 лет (средний возраст пациентов нашей выборки), не будет зарегистрирована ККТ в

течение периода наблюдения. Средний возраст в выводе представляется в отдельной таблице перед графиком. Каждая точка на кривой показывает, какова вероятность того, что усредненный пациент (то есть имеющий средние значения всех предикторов) нашей выборки не будет иметь ККТ, оставшись в выборке до момента времени, который мы выберем по оси абсцисс.

В регрессионном анализе Кокса можно оценивать независимый вклад нескольких предикторов одновременно.

Изучим влияние следующих переменных: возраст, пол, класс тяжести инфаркта миокарда, наличие в анамнезе артериальной гипертензии и мало изученный фактор лептин (адипокин, белок со свойствами гормона, вырабатываемый клетками жировой ткани и обладающий рядом функций в организме; рядом исследований доказано его неблагоприятное влияние на прогноз при инфаркте миокарда).

В диалоговом окне «Cox regression» в строках «Time» и «Status» оставим прежние переменные, в раздел «Covariates» внесем интересующие нас независимые переменные.

В подменю «Categorical» обозначим, что переменные пол, класс тяжести инфаркта миокарда, наличие артериальной гипертензии являются категориальными и отметим для них первую категорию как референтную.

Выберем метод введения предикторов в модель. Можно использовать метод форсированного ввода, Enter (все переменные одновременно вводятся в модель), метод последовательного ввода, forward, и метод последовательного исключения переменных, backward. При использовании двух последних пошаговых методов мы выбираем только ряд независимых переменных. Программа затем определяет порядок включения их в модель на основании предсказательной способности предикторов (наиболее значимо коррелирующий признак вводится первым). Включение предикторов в модель заканчивается, когда при добавлении очередного из них предсказательная способность модели статистически значимо не улучшается. Метод пошагового исключения Backward является противоположностью описанного метода. Подробно мы останавливались на методах введения в модель переменных при изучении

множественного линейного регрессионного анализа [9]. Используем Enter.

В окне построения графиков «Plots» выберем Survival и Hazard. Мы можем выбрать отдельные графики функции риска для значений категориальных или дихотомических переменных, внося интересующую нас переменную (выберем переменную «класс тяжести инфаркта») в окно «Separate Lines for».

В разделе «Save» отметим «Survival function» и «Hazard function», «DfBeta(s)», «Partial residuals».

В окне «Options» выберем «доверительные интервалы для  $\exp(B)$ » (CI for  $\exp(B)$ ) и «Correlation of estimates» (для проверки первого условия применения регрессии Кокса).

Оценка результатов анализа. На рисунке 8 представлены принципы кодирования категориальных переменных. Так как для всех переменных мы выбрали первую категорию в качестве категории сравнения, то нам будут посчитаны значения относительно женского пола, первого класса тяжести инфаркта миокарда и отсутствия артериальной гипертензии). Во втором столбце таблицы содержится информация о количестве объектов, которое принадлежит к каждой группе. В связи с тем, что определение уровня лептина было произведено только у 25% пациентов, в анализ отдаленного прогноза были включены 83 пациента.  $-2LL$  для базовой модели составило 360,223. Значение этого показателя для модели с предикторами значимо ниже ( $-2LL=332,409$ ,  $\chi^2=27,813$ ,  $p<0,001$ ), то есть наша модель обладает большей предсказательной способностью по сравнению с базовой.

Наибольший интерес для оценки независимого влияния предикторов представляют данные на рисунке 10. Если значение  $\exp(B)$  или отношения рисков больше единицы, то положительное значение данного фактора будет являться фактором, связанным с риском развития исхода, если меньше единицы – то будет ассоциировано с увеличением времени дожития (то есть выступать защитным фактором относительно исхода). Факторами, имеющими независимое влияние на риск развития ККТ, являются уровень лептина, наличие артериальной гипертензии и возраст пациента. Увеличение возраста, уровня лептина, факт наличия артериальной гипертензии связаны с

увеличением риска развития ККТ. Вариант оформления результатов регрессионного анализа Кокса представлен ниже в таблицах №1-2. Желательно отражать в таблице №1 все анализируемые переменные. Принцип построения таблицы №2 заключается в том, что простой регрессионный анализ проводится с учетом каждого предиктора, затем отбираются статистически значимые из них и вводятся в множественную модель. В модель

также можно включить общепризнанные факторы риска (в нашем примере «пол», «артериальная гипертензия»), независимо от его влияния в нашей выборке. Следует обратить внимание, если бы мы использовали способ отбора переменных для многофакторного анализа через простой анализ, то переменная «артериальная гипертензия» вообще могла бы не быть учтена.

**Categorical Variable Codings<sup>a,d,e</sup>**

		Frequency	(1) <sup>c</sup>	(2)	(3)
пол <sup>b</sup>	0=женский	28	1		
	1=мужской	55	0		
Класс_тяжести <sup>b</sup>	1=1	6	1	0	0
	2=2	32	0	1	0
	3=3	27	0	0	1
	4=4	18	0	0	0
АГ <sup>b</sup>	0=нет	13	1		
	1=есть	70	0		

- a. Category variable: пол
- b. Indicator Parameter Coding
- c. The (0,1) variable has been recoded, so its coefficients will not be the same as for indicator (0,1) coding.
- d. Category variable: Класс\_тяжести
- e. Category variable: АГ

**Рисунок 8. Кодирование категориальных переменных.**

**Omnibus Tests of Model Coefficients<sup>a</sup>**

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
332,409	24,290	7	,001	27,813	7	,000	27,813	7	,000

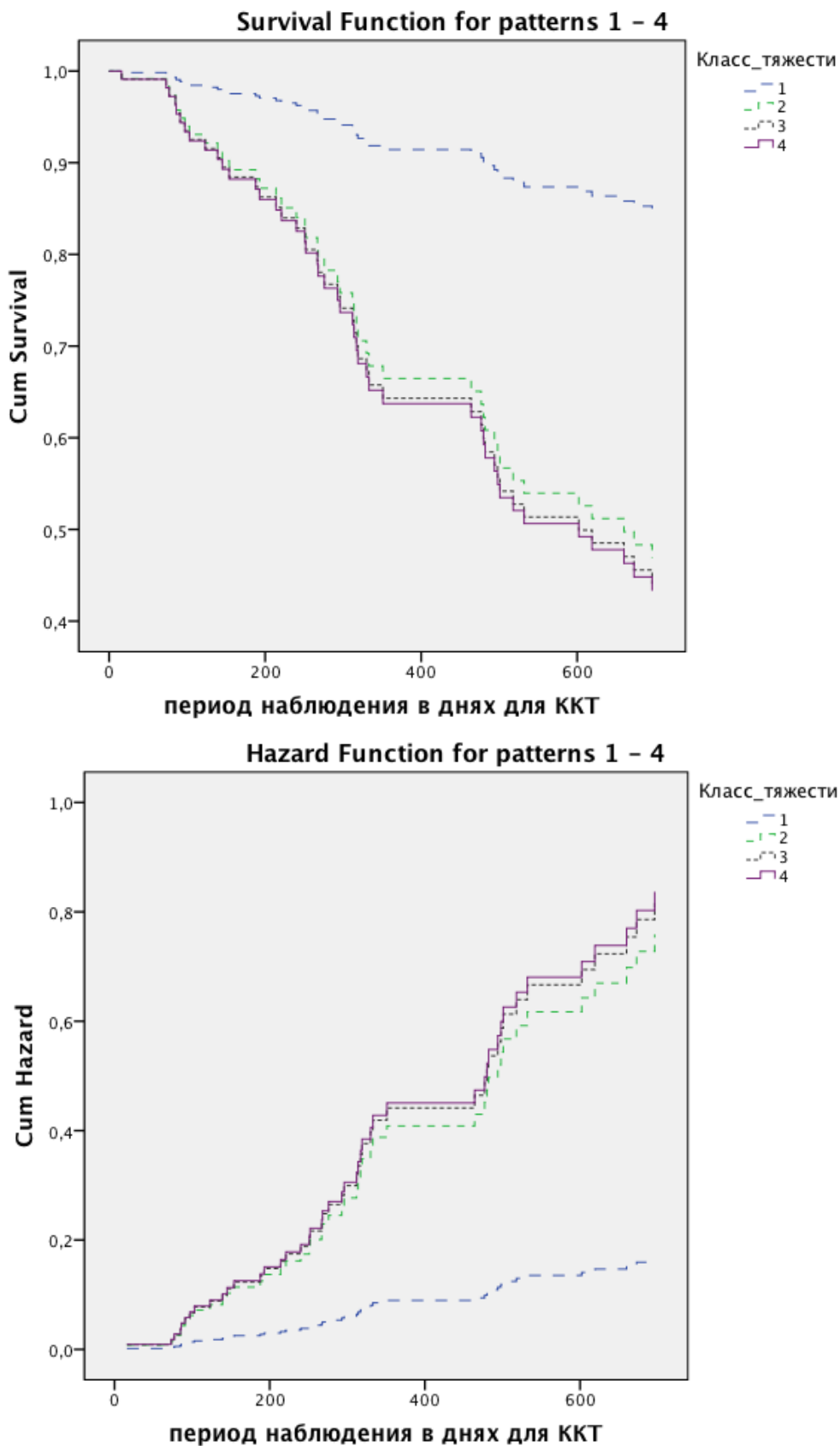
a. Beginning Block Number 1. Method = Enter

**Рисунок 9. Оценка коэффициентов модели с несколькими предикторами.**

**Variables in the Equation**

	B	SE	Wald	df	Sig.	Exp(B)	95,0% CI for Exp(B)	
							Lower	Upper
возраст	,034	,017	4,172	1	,041	1,034	1,001	1,068
пол	,080	,367	,048	1	,827	1,084	,528	2,225
АГ	1,230	,475	6,713	1	,010	3,423	1,349	8,681
лептин	,018	,005	14,811	1	,000	1,018	1,009	1,027
Класс_тяжести			2,429	3	,488			
Класс_тяжести(1)	-1,617	1,059	2,332	1	,127	,198	,025	1,581
Класс_тяжести(2)	-,097	,414	,055	1	,815	,907	,403	2,044
Класс_тяжести(3)	-,026	,407	,004	1	,950	,975	,439	2,166

**Рисунок 10. Переменные в уравнении регрессии (Шаг 1).**



**Рисунок 11. Графики кумулятивной выживаемости (вероятности, что ККТ не случится) и кумулятивного риска развития ККТ в течение периода наблюдения в зависимости от класса тяжести инфаркта миокарда (p=0,487).**

На графиках кумулятивной выживаемости и риска развития изучаемого исхода (являющихся зеркальным отображением друг

друга) построены четыре кривых согласно классу тяжести инфаркта миокарда (рис. 11). Визуально можно отметить, что прогноз

благоприятнее у пациентов с первым классом тяжести инфаркта, тогда как у пациентов с 2-4 классом прогноз практически не отличается. Однако, данные различия статистически не

значимы, данный предиктор не является значимым в модели (что может быть связано с небольшим объемом выборки).

Таблица 1.

**Результаты множественного регрессионного анализа Кокса для оценки факторов риска развития комбинированной конечной точки у пациентов с инфарктом миокарда.**

Показатель	B (SE)	95% ДИ для exp B		
		Нижний	Exp B	Верхний
Возраст, г.	0,03 (0,02)*	1,00	1,03	1,07
Пол, муж	0,08(0,37)	0,53	1,08	2,23
АГ ранее	1,23 (0,48)**	1,35	3,42	8,68
Лептин, нг/мл	0,02 (0,01)***	1,01	1,02	1,03
2 КТ ИМ	-1,62 (1,06)	0,03	0,20	1,58
3 КТ ИМ	-0,10 (0,41)	0,40	0,91	2,04
4 КТ ИМ	-0,03 (0,41)	0,44	0,98	2,17

Примечание. Хи-квадрат модели 24.29, p=0.001. \* <0,05, \*\* =0,01, \*\*\* <0,001.

B – регрессионный коэффициент, SE – стандартная ошибка, ДИ – доверительный интервал, АГ – артериальная гипертензия в анамнезе, КТ ИМ – класс тяжести инфаркта миокарда (референсная категория – 1 класс тяжести ИМ)

Таблица 2.

**Результаты однофакторного и многофакторного регрессионного анализа Кокса для оценки факторов риска развития комбинированной конечной точки у пациентов с инфарктом миокарда.**

Показатель	Однофакторный анализ Кокса			Многофакторный анализ Кокса		
	ОР	95% ДИ	p	ОР	95% ДИ	p
Лептин, нг/мл	1,01	1,01-1,02	0,001	1,02	1,01-1,03	<0,001
Возраст, г.	1,02	1,00-1,04	0,028	1,03	1,00-1,07	0,041
Пол, муж	1,36	0,93-1,99	0,115	1,08	0,53-2,23	0,827
АГ ранее	1,12	0,68-1,85	0,662	3,42	1,35-8,68	0,010

Примечание. ОР – отношение рисков (Hazard ratio), ДИ – доверительный интервал, АГ – артериальная гипертензия в анамнезе.

Перейдем к проверке необходимых условий проведения регрессионного анализа Кокса.

Для проверки первого условия, что все предикторы тесно не коррелируют друг с другом (условие отсутствия коллинеарности),

мы отмечаем «Correlation of estimates» в окне «Options» (построим корреляционную матрицу). Корреляции между регрессионными коэффициентами переменными представлены на рисунке 12.

**Correlation Matrix of Regression Coefficients**

	возраст	пол	АГ	лептин	Класс_тяжес ти(1)	Класс_тяжес ти(2)
пол	-,340					
АГ	,246	,054				
лептин	,300	-,378	,455			
Класс_тяжести(1)	,091	-,041	,051	-,020		
Класс_тяжести(2)	,088	,105	,140	,033	,242	
Класс_тяжести(3)	-,087	,073	,145	,003	,237	,620

**Рисунок 12. Корреляционный матрикс регрессионных коэффициентов.**

Согласно корреляционному матриксу выявлены корреляции очень слабой (до 0,2),

слабой (0,2-0,5) и средней (0,5-0,7) силы, что не мешает применению модели.

Только сильные корреляционные связи (>0,7 или >0,9 по данным различных источников) могут повлиять на результаты анализа [1, 8]. Если мы выявили сильные связи или хотим проверить связи средней силы, то в регрессионный анализ следует включить переменную, отображающую взаимосвязь предикторов, путем выделения двух переменных и нажатия на кнопку >a\*b>. Еще одним вариантом исключения взаимодействия является включение в модель

только одного из сильно коррелирующих между собой предикторов (если такой вариант возможен).

В качестве примера введем в модель переменную, отражающую взаимодействие предикторов «артериальная гипертензия» и «лептин». Данная переменная значимо не влияет на риск развития ККТ, поэтому в данном случае взаимодействие переменных, способное оказать влияние на модель, мы исключаем (рис. 13).

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95,0% CI for Exp(B)	
							Lower	Upper
пол	,088	,368	,057	1	,811	1,092	,530	2,248
возраст	,034	,016	4,297	1	,038	1,034	1,002	1,068
Класс_тяжести			2,434	3	,487			
Класс_тяжести(1)	-1,616	1,063	2,312	1	,128	,199	,025	1,595
Класс_тяжести(2)	-,097	,434	,050	1	,822	,907	,388	2,122
Класс_тяжести(3)	-,020	,422	,002	1	,962	,980	,429	2,240
АГ	1,237	1,120	1,219	1	,269	3,445	,383	30,961
лептин	,018	,005	11,755	1	,001	1,018	1,008	1,028
АГ*лептин	,000	,010	,000	1	,996	1,000	,981	1,019

Рисунок 13. Переменные в уравнении регрессии (Шаг 1) с дополнительной переменной, отражающей взаимодействие предикторов.

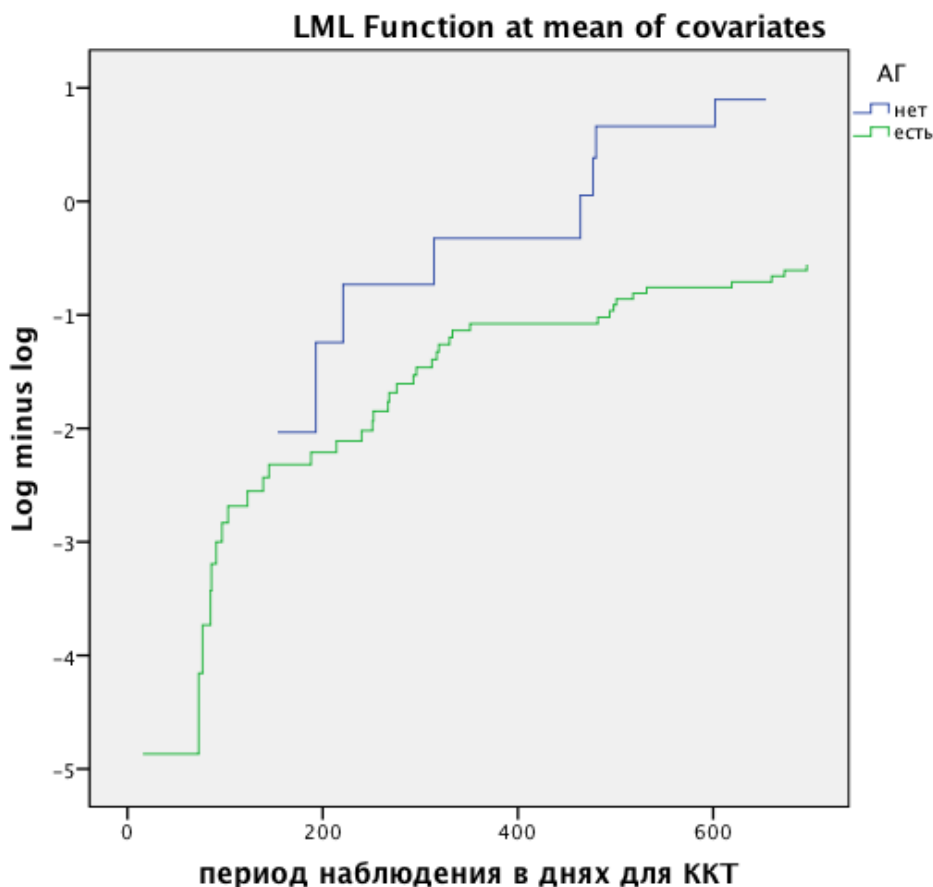


Рисунок 14. График «Log minus log» в зависимости от наличия/отсутствия артериальной гипертензии.

Затем мы проверим третье условие. Для тестирования того, что риск наступления события для любых двух объектов в любой интервал времени пропорционален, возможны следующие способы: если необходимо проверить категориальную переменную, то можно оценить и сравнить базальные функции риска для каждой категории этой переменной; для категориальных и количественных переменных, мы можем создать переменную взаимодействия между изучаемым предиктором и временем и оценить, отличается ли значение коэффициента новой переменной от нуля в уравнении регрессии Кокса.

Например, если сравнить двух пациентов одинакового возраста, пола, класса тяжести инфаркта и с равным уровнем лептина, но при этом у одного будет артериальная гипертензия, а у другого нет, то функции риска могут не быть пропорциональны в любой момент времени на протяжении периода наблюдения и их отношение может иметь зависимость от времени. Для оценки гипотезы следует провести регрессионный анализ Кокса со стратификацией по артериальной гипертензии. Откроем диалоговое окно «Cox regression», все оставляем так же, за исключением того, что переменную «АГ» из раздела «Covariates» перенесем в раздел «Strata» (то есть, стратифицируя по наличию гипертензии, мы не сможем оценить этот

показатель в модели). В разделе Plots отметим «Log minus log», так как для оценки пропорциональности рисков необходимо оценить именно этот тип графика (рис. 14).

Если исходные функции риска пропорциональны, то кривые на графике будут параллельными. Мы видим, что линии более или менее параллельны, однако немного расходятся, поэтому для точного вывода о соблюдении данного условия проверим эту переменную вторым способом. Зайдем в «Analyze», в раздел «Survival», затем «Cox regression with time dependent covariates». В появившемся окне «Expression for T\_COV\_» необходимо выбрать период времени, который будет отражать время в новой переменной «Time (T)», которая автоматически появится в начале списка переменных в окне слева. Нам нужно создать переменную времени, эквивалентную зависимой переменной. Выберем переменную «период наблюдения в днях», разделим ее на 30, то есть она будет эквивалентна зависимой переменной и будет отражать период наблюдения в нашем исследовании в месяцах. После нажатия на кнопку «Model» откроется окно, аналогичное таковому в простой регрессии Кокса (рис. 15), построим модель с учетом только переменной «артериальная гипертензия».

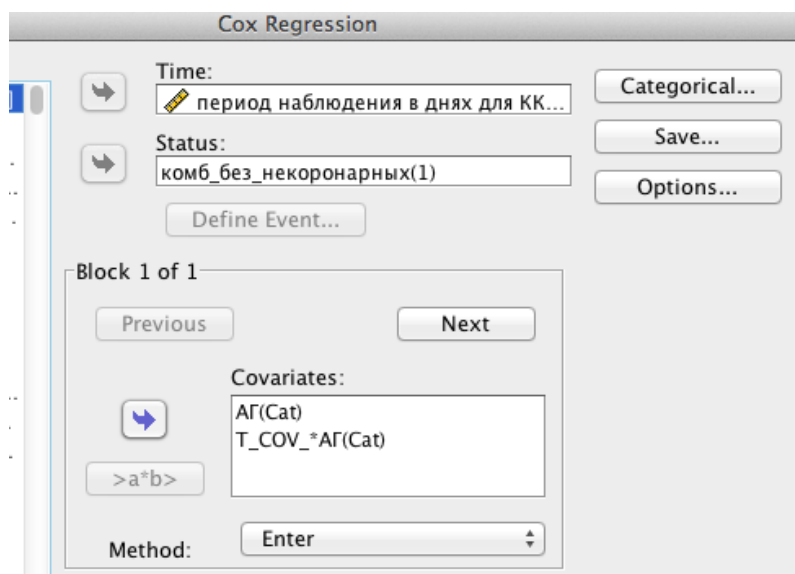


Рисунок 15. Диалоговое окно «Cox regression with time-dependent covariates».

В таблице оценки регрессионных коэффициентов (рис. 16) мы видим, что нет эффекта взаимодействия переменной «артериальная гипертензия» с переменной

времени, поэтому третье условие соблюдено. Аналогичным образом мы проверили соблюдение этого условия для всех предикторов.

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95,0% CI for Exp(B)	
							Lower	Upper
АГ	,104	,257	,165	1	,685	1,110	,671	1,837
T_COV_*АГ	-5,095	201,946	,001	1	,980	,006	,000	4,833E+169

**Рисунок 16. Переменные в уравнении регрессии (Шаг 1) с дополнительной переменной, отражающей взаимодействие предиктора и переменной времени.**

Условие пропорциональности рисков также может нарушать наличие нетипичных случаев. Предложено несколько способов их выявления: the Schoenfeld, the Martingale, the Deviance, the Score, DfBETAs, Leverage Displacement, LMAX и др. [16, 17]. Рассмотрим наиболее широко применяемые из них.

«Выскакивающие» случаи (outliers) – случаи, для которых модель работает плохо, в линейном и логистическом регрессионном анализе оценивались с помощью анализа остатков (residuals). Остатки представляют собой разность между реальным и предсказанным значением зависимой переменной для каждого объекта выборки. В регрессии Кокса невозможно подсчитать остатки, так как не определяется предсказанное значение исхода, поэтому используются не фактические остатки, а их эквивалент, где вместо предсказанного используется «аппроксимированное» (fitted) значение (хотя их традиционно называют «остатками»). Для выявления выскакивающих случаев можно рассчитать несколько видов количественных эквивалентов остатков: частичные остатки Шёнфельда (partial Schoenfeld residuals), мартингальные остатки (Martingale residuals), генерализованные остатки отклонений (deviance residuals), остатки Кокса-Снелла (Cox-Snell). В результате анализа будут выявлены два варианта объектов: те, кто живет долго, несмотря на высокий предсказанный риск; те, кто получает исход рано, будучи в группе низкого риска его развития. Подробно разберем анализ остатков Шёнфельда, представленный в SPSS, и изучим метод оценки мартингальных остатков, в связи с тем, что он может быть использован для оценки как третьего, так и второго условий применения регрессии Кокса.

Частичные остатки Шёнфельда рассчитываются для каждого предиктора

модели с учетом нецензурированных объектов выборки и представляют собой разность фактического значения ковариаты и условно ожидаемого взвешенного среднего, где веса определяются согласно шкалам риска (risk scores) [16, 17]. Наибольшую диагностическую ценность для анализа условия пропорциональности рисков представляет график значений остатков, отложенный относительно времени (Hosmer и Lemeshow (1999); Therneau и Grambsch (2000)). Если условие пропорциональности соблюдается, то остатки будут распределены случайным образом вокруг горизонтальной линии относительно ноля, если не соблюдается, то в их расположении будет прослеживаться какой-либо тренд. Частичные остатки будут рассчитаны для всех переменных, если в разделе «Save» мы отметим «Partial residuals». Мы видим, что для переменной «лептин» остатки распределены равномерно, а для переменной «класс тяжести инфаркта миокарда 1» выявлен один случай, который дальше всех отстоит от ноля, тем не менее его значение располагается в пределах [1], что не является нарушением условия (рис. 17).

Маргинальные остатки измеряют разность между вероятностью того, что у объекта в группе риска случится изучаемый исход и предсказанным количеством случаев согласно модели. Максимальное значение такого остатка равно единице, минимальное – минус бесконечность. Высокое отрицательное значение остатков будет у случаев из группы высокого риска развития исхода, которые при этом имеют длительный период выживаемости. Мартингальные остатки не представлены в выводе SPSS автоматически, но их можно рассчитать через меню Transform, Compute variable по формуле  $martingale = event(KKT) - haz\_1$ . Последний показатель мы сохранили через меню «Save»/«Hazard function» при проведении регрессионного анализа Кокса.



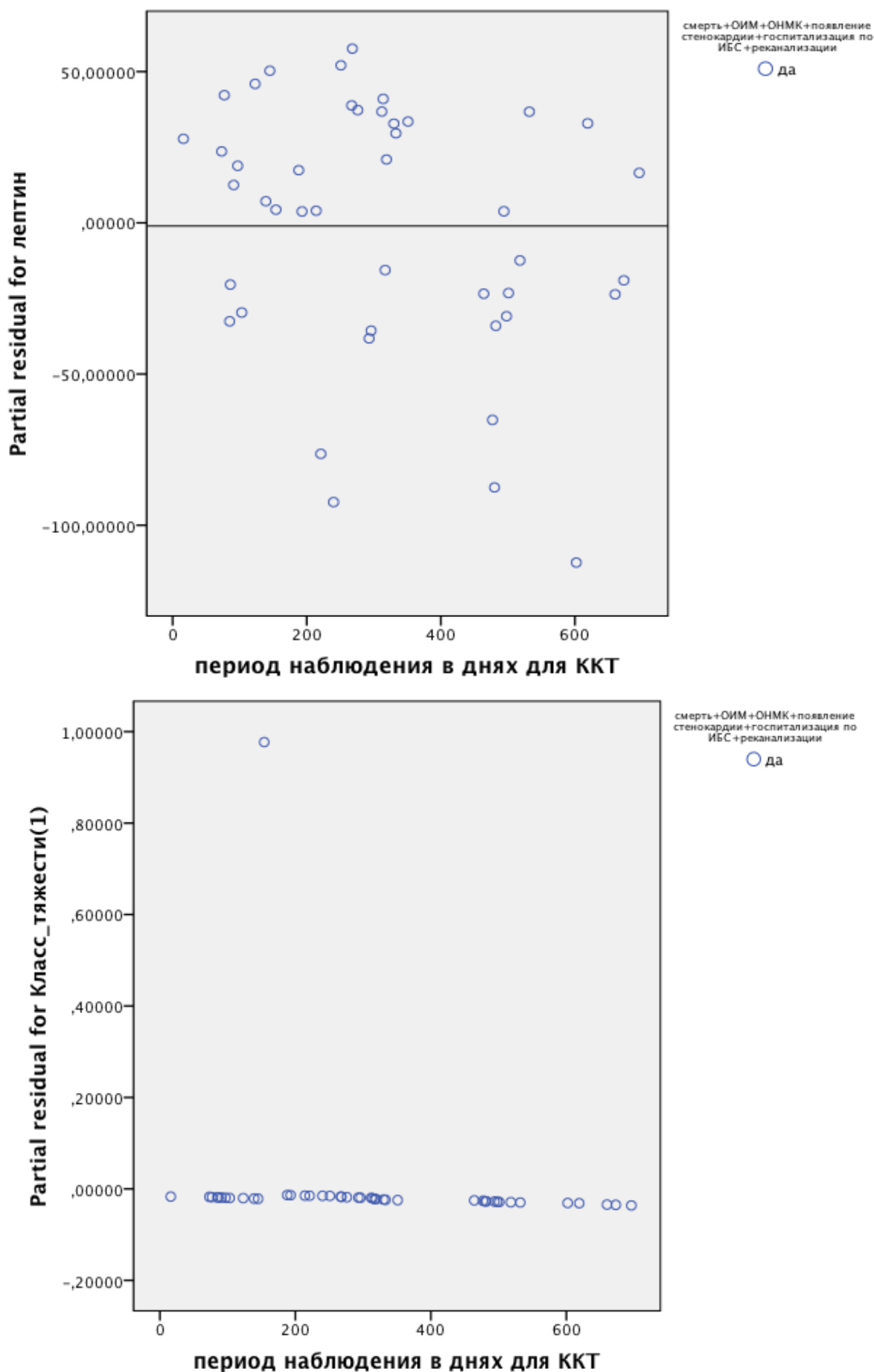


Рисунок 17. Скатерограммы распределения частичных остатков Шёнфельда для переменных «лептин», «класс тяжести инфаркта миокарда 1» согласно периоду наблюдения.

Проверим данное условие на количественной переменной «лептин». Построим скаттерограмму зависимости значений мартингаловых остатков от уровней лептина (меню «Graphs», «Legacy dialogs»,

«Scatter/dot», «Simple scatter») (рис. 18). Остатки расположены равномерным облаком около нуля, значения нескольких из них расположены в пределах  $|2|$ , что также не является нарушением условия (рис. 19).

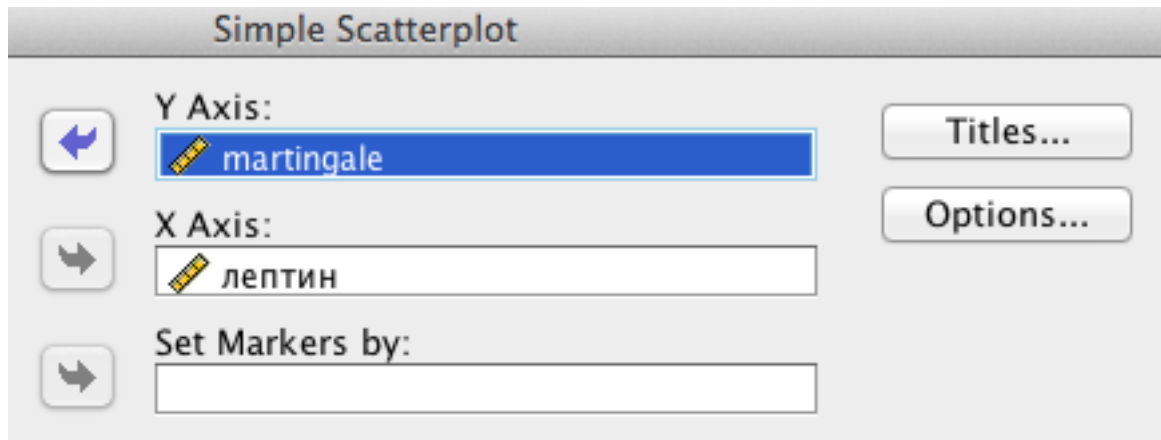


Рисунок 18. Диалоговое окно «Graphs: Simple Scatterplot».

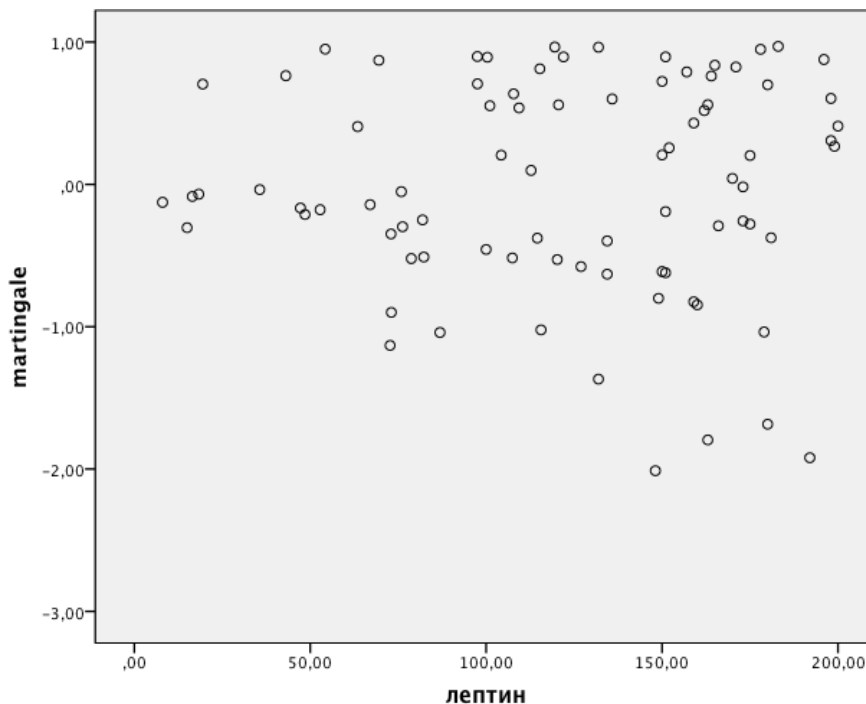


Рисунок 19. Скатерограмма зависимости мартингаловых остатков от значений лептина.

Нам также необходимо оценить случаи, оказывающие влияние на модель (influential cases). При выполнении анализа с исключением случая, оказывающего влияние на модель (influential case), регрессионные коэффициенты модели изменяются. По этой разнице, которая рассчитывается в SPSS и сохраняется в базе под именем DfBeta для каждого объекта, можно оценить степени влияния отдельного случая на предсказательную способность модели. Однако, размер показателей будет зависеть от единиц измерения каждого из предикторов,

поэтому выявить универсальное критическое значение показателя, выше которого случаи будут считаться «оказывающими сильное влияние на модель», достаточно проблематично. Рекомендуется использовать значения  $>|1|$  или  $>|2|$  [9]. Проверку можно провести или через меню «Frequencies», где внесем в окно для анализа все сохраненные в базе DfBetas, а в меню «Statistics» выберем максимальное и минимальное значение; или построив графики (рис. 20) зависимости значений DfBetas от любой из переменных (выберем DfBetas переменной «КТ ИМ 1» от

значений переменной «порядковый номер», выделив разными маркерами варианты исхода). Число случаев, остатки которых выходят за пределы 2 или 3 стандартных отклонений, не должно превышать 5% или 1% от объема изучаемой выборки, иначе построенная модель будет обладать низкой предсказательной точностью и/или плохо соответствовать имеющимся данным. Визуализация поможет нам также оценить направление отклонения от нуля значений

DfBetas, что значения DfBetas расположены хаотичным облаком около нуля независимо от исхода. Значение только одного уже выявленного ранее случая незначительно превышает пределы |1|, но не превышает пределы |2| (рис. 21). Таким образом, мы не выявили случаев, оказывающих сильное влияние на модель. В сомнительных случаях следует провести регрессионный анализ повторно с исключением выявленных случаев и оценить, насколько изменится итоговая модель.

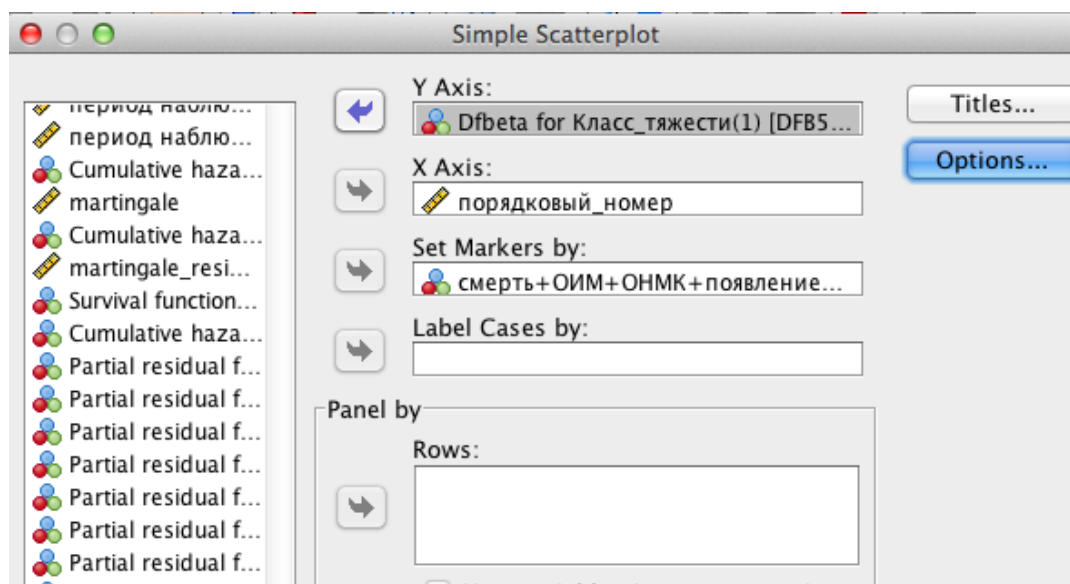


Рисунок 20. Диалоговое окно «Scatter/Dot» для построения скаттерограммы для DfBetas переменной «класс тяжести инфаркта миокарда 1» с разделением относительно исхода.

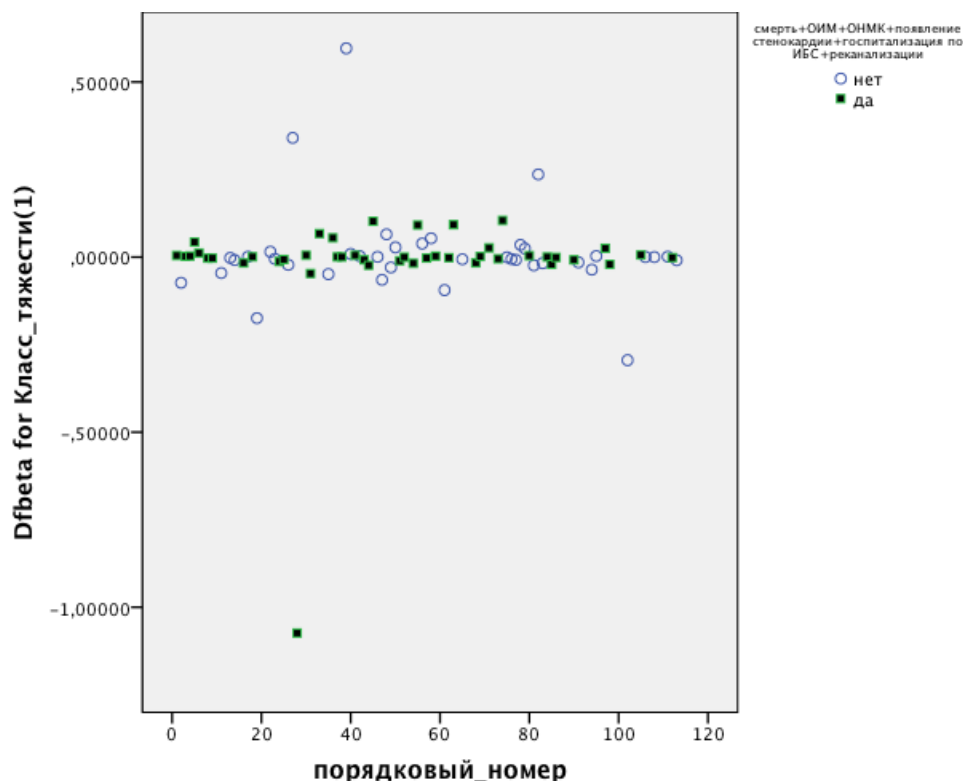
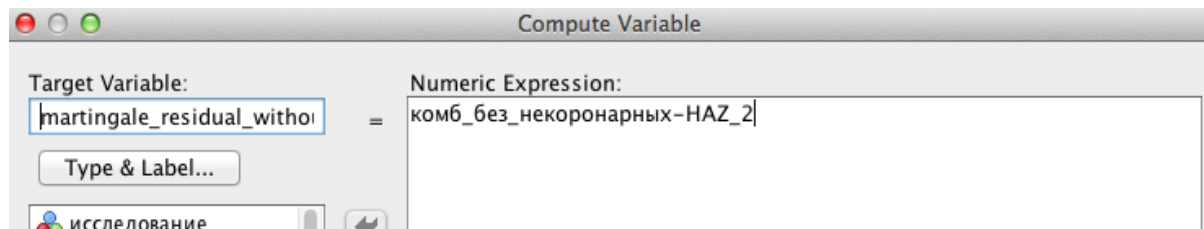


Рисунок 21. Скаттерограмма для DfBetas переменной «класс тяжести инфаркта миокарда 1» с разделением относительно исхода.

Для проверки второго условия, что все переменные линейно влияют на логарифм функции риска наступления события, необходимо оценить мартингальные остатки (martingale residuals). Для оценки линейного влияния переменной на логарифм функции

риска надо построить базовую модель (без включения этого предиктора), при этом сохранить значения Hazard и снова рассчитать мартингальные остатки без значения лептина (рис. 22, 23).



**Рисунок 22. Диалоговое окно «Compute variable» для расчета новой переменной «мартингальные остатки без учета лептина».**

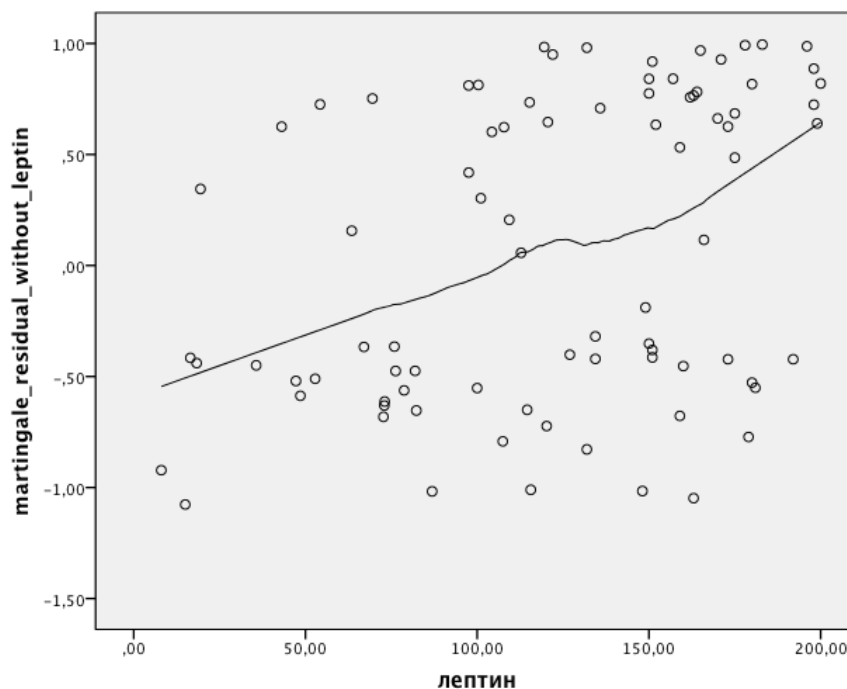
	HAZ_1	martingale	XBE_1	HAZ_2	XBE_2	martingale_residual_without_leptin
1	,44039	,56	,26678	,23461	-,37813	,77
2	1,03698	-1,04	,39631	,77213	,29030	-,77
3	,39597	,60	1,60684	,11330	,68567	,89
4	,17500	,83	1,06492	,07210	,65525	,93
5	,69170	,31	1,19409	,27567	,11276	,72
6	,23856	,76	,67217	,21832	,53360	,78

**Рисунок 23. Новые переменные «hazard function», «martingale residuals» в базе данных.**

Построить график зависимости последних остатков от значений лептина. Затем нажать дважды на сам график, чтобы можно было с ним работать, выбрать в строке над графиком значок «Add fit line and total», затем «прямая Лоэсса» Loess smooth (рис. 24). Зависимость должна быть более или менее линейной

(возможны небольшие колебания относительно прямой), что, как мы уже упоминаю, представляет собой параметрическую часть функции риска Кокса.

Остальные условия также были соблюдены (проверены в предыдущей статье) [3].



**Рисунок 24. Скатерограмма зависимости мартингальных остатков без учета лептина от значений лептина.**

Таким образом, модель пропорциональных рисков Кокса является удобным инструментом для биомедицинских исследований, так как ее использование предоставляет возможность произвести относительно точный анализ дожития по сравнению с другими методами в связи с включением набора независимых переменных, влияющих на риск наступления события. Именно поэтому данный метод рекомендуется для оценки любых исходов относительно времени и наиболее часто встречается в публикациях, главной задачей которых является оценка событий во временном интервале.

### Литература:

1. Бурдяк А.Я. Применение анализа «Анализ наступления события (Event history analysis) с помощью пакета SPSS». SPERO. Социальная политика: экспертиза, рекомендации, обзоры. 2007. №6. С. 189-202.
2. Румянцев П.О., Саенко В.А., Румянцева У.В., Чекин С.Ю. Статистические методы анализа в клинической практике. Часть. 2. Анализ выживаемости и многомерная статистика // Проблемы Эндокринологии. 2009. №6. С. 48-56.
3. Холматова К.К., Гржибовский А.М. Применение анализа дожития в здравоохранении // Наука и Здоровоохранение 2017. №5. С. 5-28.
4. Холматова К.К., Дворяшина И.В. Прогностическое значение уровней гликемии, зарегистрированных при поступлении, у пациентов с инфарктом миокарда // Архивъ внутренней медицины. 2014. №1. С. 25-29.
5. Холматова К.К., Дворяшина И.В., Супрядкина Т.В. Влияние гликемии на ранний прогноз пациентов с инфарктом миокарда без сахарного диабета 2 типа в анамнезе. // Кардиоваскулярная терапия и профилактика. 2014. №2. С. 29-34.
6. Холматова К.К., Дворяшина И.В., Супрядкина Т.В. Различные варианты нарушений углеводного обмена и их влияние на течение инфаркта миокарда у пациентов г. Архангельска // Экология человека. 2013. №10. С.14-22.
7. Холматова К.К., Дворяшина И.В., Фомкина И.А., Супрядкина Т.В. Прогностическое значение содержания адипоцитокинов у пациентов с инфарктом миокарда и различными вариантами

нарушений углеводного обмена // Сахарный диабет. 2014. №3. С. 90-95.

8. Холматова К.К., Шарашова Е.Е., Горбатова М.А., Гржибовский А.М. Применение множественного логистического регрессионного анализа в здравоохранении // Наука и Здоровоохранение 2017. №4. С. 5-26.
9. Шарашова Е.Е., Холматова К.К., Горбатова М.А., Гржибовский А.М. Применение множественного линейного регрессионного анализа в здравоохранении с использованием пакета статистических программ SPSS // Наука и Здоровоохранение 2017. №3. С. 5-31.
10. Юнкеров В.И., Григорьев С.Г. Математико-статистическая обработка данных медицинских исследований. СПб.: ВМедА, 2002. 266 с.
11. Cox D.R. Regression models and life tables (with discussion) // J. R. Statist. Soc., Series B. 1972. N 2. P. 187- 220.
12. Foster J. Understanding and using advanced statistics. Foster J., Barkus M., Yavorsky C. London: SAGE Publications Ltd., 2006. 178 p.
13. Hosmer JR. DW, Lemeshow S. Applied Survival Analysis; Regression Modeling of Time to Event Data. New York: John Wiley & Sons, 1999. 416 p.
14. Kleinbaum D.G., Klein M. Survival analysis: a self-learning text (3<sup>rd</sup> ed.). New York, 2012. 591 p.
15. Machin D., Cheung Y., Parmar M. Survival analysis: a practical approach (2<sup>nd</sup> ed.). New York, 2006. 267 p.
16. NCSS statistical software. Cox regression: [site]. URL: [http://ncss.wpengine.netdna-cdn.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Cox\\_Regression.pdf](http://ncss.wpengine.netdna-cdn.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Cox_Regression.pdf)
17. Norusis M.J. SPSS 15.0 advanced statistical procedures companion. New Jersey, 2007. 418 p.
18. Peat J., Barton B. Medical statistics: a guide to data analysis and critical appraisal (1<sup>st</sup> ed.) // NY: Blackwell Publishing, 2005. 324 p.
19. Rao S.R., Schoenfeld D.A. Survival methods // Circulation. 2007. N 115. P. 109-113.
20. Statsoft. Survival/Failure Time Analysis: [site]. URL: <http://www.statsoft.com/Textbook/Survival-Failure-Time-Analysis>

**References:**

1. Burdyak A.Y. *Primenenie analiza «Analyze nastupleniya sobytiya (Event history analysis) s pomoshju paketa SPSS»*. [The application of the «event history analysis in SPSS»] SPERO. Sotsialnaya politika: ekspertiza, rekomendacii, obzory. [SPERO. Social politics: expertise, recommendations, reviews]. 2007. N 6. P. 189-202. [in Russian]
2. Rumyansev P.O., Saenko V.A., Rumyanseva U.V., Chekin S.U. Statisticheskie metody analiza v klinicheskoi praktike. Chast 2: Analiz vyzhivaemosti i mnogomernaya statistika. [Statistical methods of analysis in clinical practice & Part 2: Survival analysis and multiple statistics]. *Problemy endokrinologii* [Problems of endocrinology]. 2009. N 6. P. 48-56. [in Russian]
3. Kholmatova K.K., Grijbovski A.M. *Primenenie analiza dozhitia v zdravoochranenii* [The application of survival analysis in health sciences]. *Nauka i zdravoochranenie* [Science and Health Care] 2016. №5. pp. 5-28. [in Russian]
4. Kholmatova K.K., Dvoryashina I.V. Prognosticheskoe znachenie urovnei glikemii, zaregistrovannykh pri postuplenii, u pacientov s infarctom miokarda. [Prognostic value of admission glycaemia levels in patients with myocardial infarction]. *Arhiv vnutrennei mediciny* [Archive of internal medicine]. 2014. N 1. P. 25-29. [in Russian]
5. Kholmatova K.K., Dvoryashina I.V., Supryadkina T.V. Vliyanie glikemii na rannij prognoz pacientov s infarctom miokarda bez saharnogo diabeta 2 tipa v anamneze [Influence of glycaemia on short-term prognosis of patients with myocardial infarction and without diabetes mellitus type 2 in anamnesis]. *Kardiovaskulyarnaya terapiya i profilaktika* [Cardiovascular therapy and prevention]. 2014. N 2. P. 29-34. [in Russian]
6. Kholmatova K.K., Dvoryashina I.V., Supryadkina T.V. Razlichnye varianty narushenij uglevodnogo obmena i ih vliyanie na techenie infarcta miokarda u pacientov g. Arkhangel'ska. [Different glucose metabolism disorders and its influence on the myocardial infarction course in patients in Arkhangel'sk]. *Ekologiya cheloveka* [Human ecology]. 2013. N 10. P.14-22. [in Russian]
7. Kholmatova K.K., Dvoryashina I.V., Fomkina I.A., Supryadkina T.V. Prognosticheskoe znachenie sodержaniya adiptyscitokinov u ptsientov s infarctom miokarda i razlichnymi variantami narushenii uglevodnogo obmena. [Prognostic value of adipokines' levels in patients with myocardial infarction and glucose metabolism disorders]. *Sakharnyi diabet* [Diabetes mellitus]. 2014. N 3. P. 90-95. [in Russian]
8. Kholmatova K.K., Sharashova E.E., Gorbatova M.A., Grijbovski A.M. *Primenenie mnozhestvennogo logisticheskogo regressionnogo analiza v zdravoochranenii s ispol'zovaniem paketa statisticheskikh programm SPSS* [The application of multiple logistic regression analysis in health sciences using SPSS software]. *Nauka i zdravoochranenie* [Science & Healthcare] 2017. №4. C. 5-26. [in Russian]
9. Sharashova E.E., Kholmatova K.K., Gorbatova M.A., Grijbovski A.M. *Primenenie mnozhestvennogo lineinogo regressionnogo analiza v zdravoochranenii s ispol'zovaniem paketa statisticheskikh programm SPSS* [The application of multiple logistic regression analysis in health sciences using SPSS software]. *Nauka i zdravoochranenie* [Science & Healthcare] 2017. №3. C. 5-31. [in Russian]
10. Junkerov V.I., Grigoriev S.G. *Matematiko-statisticheskaya obrabotka dannykh meditsinskikh issledovaniy* [Mathematical and statistical analysis of the medical research data]. SPb: VMedA, 2002. 266 p. [in Russian]
11. Cox D. R. Regression models and life tables (with discussion). *J. R. Statist. Soc., Series B*. 1972. N 2. P. 187-220.
12. Foster J. Understanding and using advanced statistics. Foster J., Barkus M., Yavorsky C. London: SAGE Publications Ltd., 2006. 178 p.
13. Hosmer JR. DW, Lemeshow S. *Applied Survival Analysis; Regression Modeling of Time to Event Data*. New York: John Wiley & Sons, 1999. 416 p.
14. Kleinbaum D.G., Klein M. *Survival analysis: a self-learning text (3<sup>rd</sup> ed.)*. New York, 2012. 591 p.
15. Machin D., Cheung Y., Parmar M. *Survival analysis: a practical approach (2<sup>nd</sup> ed.)*. New York, 2006. 267 p.
16. NCSS statistical software. Cox regression: [site]. URL: [http://ncss.wpage.netdna-cdn.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Cox\\_Regression.pdf](http://ncss.wpage.netdna-cdn.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Cox_Regression.pdf)

17. Norusis M.J. SPSS 15.0 advanced statistical procedures companion. New Jersey, 2007. 418 p.

18. Peat J., Barton B. Medical statistics: a guide to data analysis and critical appraisal (1<sup>st</sup> ed.). NY: Blackwell Publishing, 2005. 324 p.

19. Rao S.R., Schoenfeld D.A. Survival methods. *Circulation*. 2007. N 115. P. 109-113.

20. Statsoft. Survival / Failure Time Analysis: [site]. URL:

<http://www.statsoft.com/Textbook/Survival-Failure-Time-Analysis>

#### **Контактная информация:**

**Гржибовский Андрей Мечиславович** – доктор медицины, магистр международного общественного здравоохранения, Старший советник Национального Института Общественного Здравоохранения, г. Осло, Норвегия; Заведующий ЦНИЛ СГМУ, г. Архангельск, Россия; Профессор Северо-Восточного Федерального Университета, г. Якутск, Россия; Почетный доктор Международного Казахско-Турецкого Университета им. Х.А. Ясяви, г. Туркестан, Казахстан; Почетный профессор ГМУ г. Семей, Казахстан.

**Почтовый адрес:** INFA, Nasjonalt folkehelseinstitutt, Postboks 4404 Nydalen, 0403 Oslo, Norway.

**Email:** Andrej.Grijibovski@gmail.com

**Телефон:** +4745268913 (Норвегия), +79214717053 (Россия), +77471262965 (Казахстан).