

© Маньяков Р.Р., Деменкова В.В., Агафонова Ю.В., Петряхин Д.Н., 2020
DOI 10.20310/2658-7688-2020-2-2(6)-20-32
УДК 614.2

Методика прогнозирования клинических исходов при помощи логистического регрессионного анализа в статистической программе IBM SPSS Statistics 20.0 (часть 1)

Рустам Ринатович МАНЬЯКОВ¹, Валерия Владимировна ДЕМЕНКОВА¹,
Юлия Владимировна АГАФОНОВА¹, Денис Николаевич ПЕТРЯХИН²

¹ФГБОУ ВО «Тамбовский государственный университет имени Г.Р. Державина»,
Медицинский институт

392000, Российская Федерация, г. Тамбов, ул. Интернациональная, 33

ORCID: <https://orcid.org/0000-0002-9935-8373>, e-mail: 8730241@mail.ru

ORCID: <https://orcid.org/0000-0002-2241-7048>, e-mail: vill20061@yandex.ru

ORCID: <https://orcid.org/0000-0003-2418-5129>, e-mail: Evamedical.ru@gmail.com

²ФГБВОУ ВПО «Военно-медицинская академия имени С.М. Кирова» МО РФ

194044, Российская Федерация, г. Санкт-Петербург, ул. Академика Лебедева, 6

ORCID: <http://orcid.org/0000-0001-6270-6509>, e-mail: Den-09-ra@yandex.ru

Methodology for prediction of clinical outcomes using logistic regression analysis in IBM SPSS Statistics 20.0 (part 1)

Rustam R. MANYAKOV¹, Valeriya V. DEMENKOVA¹,
Yuliya V. AGAFONOVA¹, Denis N. PETRYAKHIN²

¹Derzhavin Tambov State University, Medical Institute

33 Internatsionalnaya St., Tambov 392000, Russian Federation

ORCID: <https://orcid.org/0000-0002-9935-8373>, e-mail: 8730241@mail.ru

ORCID: <https://orcid.org/0000-0002-2241-7048>, e-mail: vill20061@yandex.ru

ORCID: <https://orcid.org/0000-0003-2418-5129>, e-mail: Evamedical.ru@gmail.com

²S.M. Kirov Military Medical Academy of the Ministry of Defense of the Russian Federation

6 Academician Lebedev St., St. Petersburg 194044, Russian Federation

ORCID: <https://orcid.org/0000-0001-6270-6509>, e-mail: Den-09-ra@yandex.ru

Аннотация. Представлено описание организации проведения методики построения математической модели прогноза риска клинического исхода по выявленным медико-социальным факторам. *Цель:* описание методики бинарного логистического регрессионного анализа в медицинских исследованиях. *Методы исследования.* Логистическая регрессия – это разновидность множественной регрессии, общее назначение которой состоит в анализе связи между несколькими независимыми переменными и зависимой переменной. Для применения бинарного логистического регрессионного анализа необходимо соблюдение следующих условий: результирующая переменная должна принимать два значения (благоприятный клинический исход/неблагоприятный клинический исход), независимые переменные могут быть как категориальными, так и интервальными, должны быть независимы друг от друга. *Результаты исследования.* На основе изученных медико-социальных факторов разработана математическая модель прогноза вероятности развития первичной инвалидности среди лиц, получивших черепно-мозговые травмы. Полученное уравнение регрессии содержит четыре переменных: возраст пациента, характер полученной черепно-мозговой травмы, уровень образования, тяжесть полученной черепно-мозговой травмы. Математическая модель предсказывает наступление неблагоприятного клинического исхода с долей вероятности 87,6 %. Полученные результаты

позволяют с допустимой долей чувствительности и специфичности рассчитать прогнозируемую вероятность изучаемого клинического исхода. *Вывод.* Представленная методика имеет практическое значение в плане совершенствования знаний начинающего врача-исследователя в области построения математических моделей прогнозирования вероятности изучаемого события.

Ключевые слова: логистическая регрессия; медико-социальные факторы; модель прогноза

Для цитирования: Маньяков Р.Р., Деменкова В.В., Агафонова Ю.В., Петряхин Д.Н. Методика прогнозирования клинических исходов при помощи логистического регрессионного анализа в статистической программе IBM SPSS Statistics 20.0 (часть 1). Медицина и физическая культура: наука и практика. 2020;2(6):20-32. DOI 10.20310/2658-7688-2020-2-2(6)-20-32.

Abstract. We presented a description of the organization of the method of constructing a mathematical model of clinical outcome risk prediction on the identified medical and social factors. *Purpose:* description of binary logistic regression analysis methodology in medical research. *Research methods.* Logistics regression is a form of multiple regression whose general purpose is to analyze the relationship between several independent variables and a dependent variable. For binary logistic regression analysis application, the following conditions must be met: the resulting variable must assume two values (favorable clinical outcome/adverse clinical outcome), the independent variables can be both categorical and interval, and must be independent of each other. *Results of a research.* On the basis of the studied medical and social factors, a mathematical model of the prediction of the probability of developing primary disability among persons who have suffered brain injuries has been developed. The obtained regression equation contains four variables: the patient's age, the nature of the brain injury, the level of formation, the severity of the brain injury. The mathematical model predicts the onset of an adverse clinical outcome with a probability of 87.6 %. The obtained results allow calculating the predicted probability of the studied clinical outcome with the permissible proportion of sensitivity and specificity. *Conclusion.* The presented technique is of practical importance in terms of improving the knowledge of the beginner researcher in the field of building mathematical models for predicting the probability of the event being studied.

Keywords: logistics regression; medico-social factors; forecasting model

For citation: Manyakov R.R., Demenkova V.V., Agafonova Yu.V., Petryakhin D.N. Metodika prognozirovaniya klinicheskikh iskhodov pri pomoshchi logisticheskogo regressionnogo analiza v statisticheskoy programme IBM SPSS Statistics 20.0 (Chast' 1) [Methodology for prediction of clinical outcomes using logistic regression analysis in IBM SPSS Statistics 20.0 (part 1)]. *Meditsina i fizicheskaya kul'tura: nauka i praktika. – Medicine and Physical Education: Science and Practice.* 2020;2(6):20-32. DOI 10.20310/2658-7688-2020-2-2(6)-20-32. (In Russian, Abstr. in Engl.)

ВВЕДЕНИЕ

В настоящее время одним из широко распространенных подходов к прогнозированию является метод бинарного логистического регрессионного анализа, позволяющий по выявленным факторам (переменным) с известной степенью чувствительности и специфичности предсказать тот или иной клинический исход. Представленная статья раскрывает алгоритм построения математической модели прогноза риска изучаемого клинического исхода и анализа полученных результатов.

Целью настоящего исследования является описание методики бинарного логистического регрессионного анализа в медицинских исследованиях.

МАТЕРИАЛЫ И МЕТОДЫ

Логистическая регрессия (англ. *Logit model*) – это разновидность множественной регрессии, общее назначение которой состоит в анализе связи между несколькими независимыми переменными (называемыми также регрессорами или предикторами) и зависимой переменной. Бинарная логистическая регрессия применяется в случае, когда зависимая переменная является бернуллиевой (то есть принимает только два значения).

При помощи бинарной логистической регрессии можно оценивать вероятность того, что событие наступит для конкретного пациента (возникнут осложнения или не возникнут).

нут, будет установлена инвалидность или нет, заболеет или не заболеет, умрет или не умрет и т. д.).

Вероятность P отнесения пациента к одной из двух категорий вычисляется по формуле:

$$P = \frac{\exp(y)}{1 + \exp(y)}, \quad (1)$$

где значение зависимой переменной y линейно зависит от предикторов (независимых переменных) x_1, \dots, x_n , то есть выражается регрессионным уравнением вида:

$$y = b_0 + b_1x_1 + \dots + b_nx_n, \quad (2)$$

где b_0, b_1, \dots, b_n – регрессионные коэффициенты.

Для применения бинарного логистического регрессионного анализа необходимо соблюдение следующих условий: результирующая переменная должна принимать два значения (благоприятный клинический исход/неблагоприятный клинический исход), независимые переменные x_1, \dots, x_n могут быть как категориальные, так и интервальные, должны быть независимы друг от друга.

Для демонстрации процедуры построения математической модели прогноза клинического исхода использована русскоязычная версия программы IBM SPSS Statistics 20.0. В качестве исходной информации послужила база данных, включающая медико-социальные характеристики 177 пациентов, получивших черепно-мозговые травмы (ЧМТ).

Полученная математическая модель по результатам регрессионного анализа должна соответствовать следующим условиям: из полученных математических моделей отбиралось одно уравнение, имеющее предсказательную способность не менее 80 %, независимые переменные, вошедшие в состав уравнения прогноза, должны быть доступны при поступлении пациента на реабилитационное лечение и на всем протяжении катамнестического наблюдения за пациентом.

РЕЗУЛЬТАТЫ

Для построения математической модели прогноза клинического исхода необходима исходная информация.

Проще всего будет создание базы данных в программе Microsoft Excel. Для этого определим переменные, которые в дальнейшем могут быть использованы для процедуры построения математической модели прогноза риска первичной инвалидности: первичная инвалидность, характер полученных ЧМТ, тяжесть полученной травмы, госпитализация после получения ЧМТ, срок госпитализации, частота профилактических медицинских осмотров, занятие спортом не менее трех раз в неделю, возраст на момент обследования, возраст получения ЧМТ, повторная ЧМТ, образование. Назовем данный файл «База данных лиц с ЧМТ».

Импорт данных из Microsoft Excel в IBM SPSS Statistics 20.0. Запустите программу IBM SPSS Statistics 20.0. В открывшемся окне выберите пункт **создать новый запрос с помощью конструктора** → **Ок**. Откроется окно **Конструктор чтения баз данных** (рис. 1А), в котором необходимо выбрать пункт **Excel Files** → **Далее**.

В появившемся окне, нажав кнопку **Обзор**, необходимо выбрать исходный файл Microsoft Excel, в котором была создана база данных с названием «База данных лиц с ЧМТ», а затем нажать кнопку **Ок**.

В левом поле открывшегося окна (рис. 1Б) выделяем мышью **Лист № 1** и перетаскиваем его в правое окно → **Готово**.

В результате откроется основное окно программы IBM SPSS Statistics 20.0 с исходными данными, которое нужно сохранить. Таким образом, мы создали файл с исходными данными в программе IBM SPSS Statistics 20.0 и сохранили его для последующей работы.

Для построения математической модели прогноза риска первичной инвалидности необходимо выявить все независимые переменные, имеющие статистически значимое влияние на изучаемый клинический исход. Такой анализ представляет собой сложную задачу, решению которой будет посвящена отдельная статья. Покажем лишь итоговую

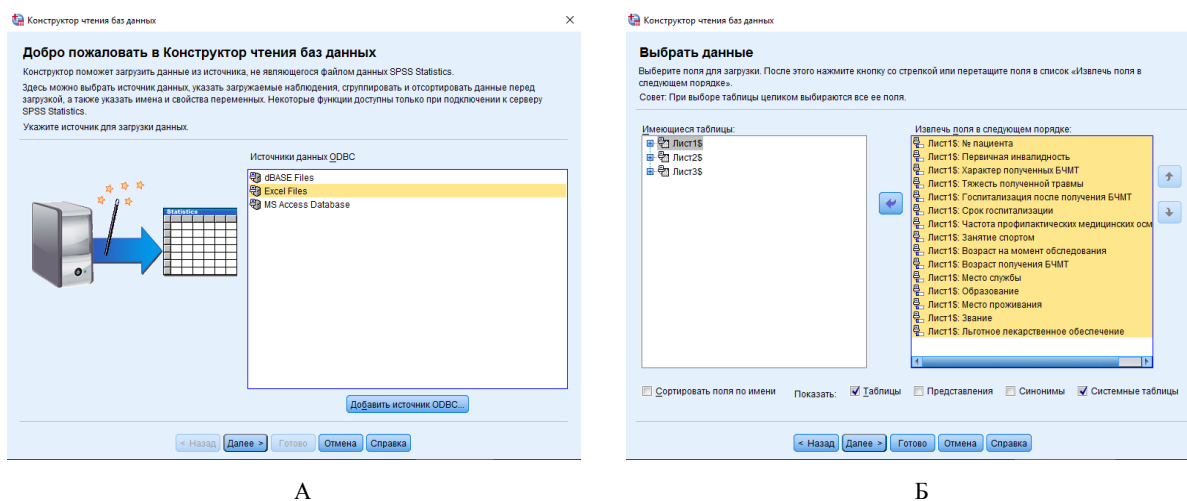


Рис. 1. А. Вид окна Конструктора чтения баз данных. Б. Вид окна выбора данных
Fig. 1. A. View of the Database Reader window. B. Data Selection Window View

**Результаты статистического анализа взаимосвязи
первичной инвалидности с независимыми переменными**

Таблица 1

Table 1

**Statistical analysis of the relationship between
primary disability and independent variables**

Переменная Variable	Кол-во наблюдений (n) The number of observations (n)	Статистический критерий Statistical criterion	Величина статистики Statistics value	p	Сила связи Binding force
Характер черепно-мозговой травмы The nature of brain injury	177	χ^2 Пирсона χ^2 Pearson	18,261	0,0001	0,3210
Тяжесть травмы Severity of injuries	132	χ^2 Пирсона χ^2 Pearson	49,197	0,0001	0,6100
Госпитализация после получения че- репно-мозговой травмы Hospitalized after receiving brain injury	106	χ^2 Пирсона χ^2 Pearson	3,803	0,0510	–
Срок госпитализации Hospitalization term	63	χ^2 Пирсона χ^2 Pearson	4,151	0,0420	–0,0257
Частота профосмотров Frequency of professional surveys	108	χ^2 Пирсона χ^2 Pearson	15,223	0,0001	–0,3750
Занятие спортом Doing sport	107	χ^2 Пирсона χ^2 Pearson	19,440	0,0001	0,4260
Возраст Age	177	χ^2 Вальда χ^2 Wald	24,222	0,0001	–
Возраст получения черепно-мозговой травмы Age of receiving brain injury	177	χ^2 Вальда χ^2 Wald	20,409	0,0001	–
Повторная боевая черепно-мозговой травмы Repeated fighting brain injury	177	χ^2 Пирсона χ^2 Pearson	2,404	0,1210	–
Образование Education	147	χ^2 Пирсона χ^2 Pearson	13,566	0,0001	0,3040

таблицу с переменными и достигнутым уровнем статистической значимости, используемым статистическим критерием и силой их взаимосвязи (табл. 1).

Более подробно об анализе статистической взаимосвязи переменных можно прочесть в литературе [1–7].

Таким образом, из 10 переменных, включенных в анализ, лишь 8 оказывают статистически значимое влияние на исследуемый клинический исход. Данные переменные могут использоваться в дальнейшем для построения математических моделей прогноза. Следует иметь в виду, что минимальное количество наблюдений на каждую переменную в итоговой математической модели должно быть не менее 10.

Настройка и запуск процедуры логистического регрессионного анализа. Для построения математической модели прогноза риска первичной инвалидности в основном окне программы IBM SPSS Statistics 20.0 необходимо выбрать в контекстном меню **Анализ** → **Регрессия** → **Логистическая** (рис. 2).

В открывшемся окне процедуры логистической регрессии (рис. 3А) в поле, где перечислены все переменные, необходимо, выделив зависимую переменную, кодирующую изучаемый исход (в нашем примере это **первичная инвалидность**), перенести в поле **Зависимая переменная**. Независимые переменные необходимо перенести в поле ковариаты.

В раскрывающемся списке **Метод** выбираем метод построения уравнения регрессии Включение ОП: Вальд, при выборе которого в итоговую модель регрессионного анализа войдут переменные, значение статистики Вальда которых достигает уровня статистической значимости.

Также в диалоговом окне Логистическая регрессия справа имеются четыре пункта: Категориальные, Сохранить, Параметры и Бутстреп. Пункты Категориальные и Бутстреп не рассматриваются, более подробно об этом можно прочесть в литературе¹.

¹IBM SPSS Statistics Core System Users Guide: руководство пользователя по базовой системе IBM SPSS Statistics 20. URL: ftp://public.dhe.ibm.com/software/analytics/spss/documentation-/statistics/20.0/ru/client/Manuals/IBM_SPSS_Statistics_Core_System_Users_Guide.pdf (дата обращения: 24.02.2015).

При нажатии пункта **Сохранить** отобразится диалоговое окно запроса параметров сохранения процедуры **Логистическая регрессия: Сохранить** (рис. 3Б). В диалоговом окне запроса параметров сохранения на вкладке **Предсказанные значения** отмечаем галочкой **Вероятности, Стьюдентизированные** на вкладке **Остатки**. Остальные настройки параметров сохранения оставляем по умолчанию. После нажатия кнопки **Продолжить** возвращаемся в главное диалоговое окно Логистической регрессии.

При нажатии пункта **Параметры** на экране появится окно запроса параметров вывода результатов логистической регрессии (рис. 4), в котором на вкладке **Статистики и графики** отметим галочкой **Критерий согласия Хосмера–Лемешова, Поточный вывод остатков, ДИ для $e^{хр(B)}$** , остальные настройки параметров вывода оставляем по умолчанию.

После настройки параметров, необходимых для построения математической модели, в главном диалоговом окне Логистической регрессии необходимо нажать кнопку ОК, в результате чего будет выведен вывод с расчетами.

Интерпретация полученных результатов. Результаты проведенных математических расчетов процедуры бинарного логистического регрессионного анализа представлены в табл. 2–8.

В табл. 2 представлены данные о количестве наблюдений. Так, включенных наблюдений для анализа – 122 (55,2 %) и 99 (44,8 %) наблюдений, которые не использовались для построения модели прогноза.

Зависимая переменная, в нашем случае – это первичная инвалидность, принимает два значения: отсутствие инвалидности/наличие инвалидности. В базе данных представленная переменная кодируется значениями «0» – отсутствие инвалидности, «1» – наличие инвалидности.

Все независимые переменные, которые были перенесены в поле **Ковариаты** основного диалогового окна настройки процедуры логистической регрессии, представлены в табл. 3 итоговых результатов с уровнем статистической значимости. Так, переменная

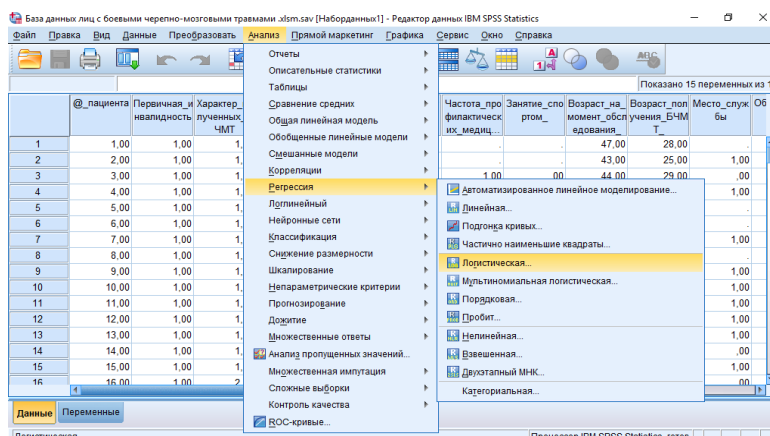
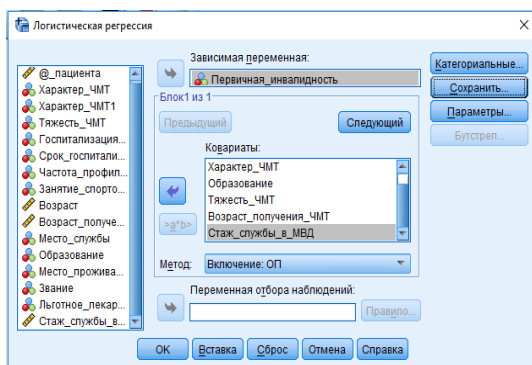
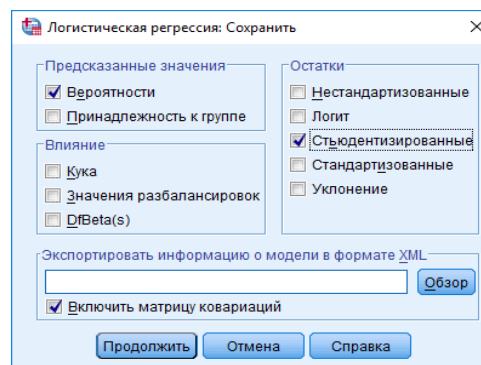


Рис. 2. Вид стартового окна программы IBM SPSS Statistics 20.0 и запуск процедуры логистической регрессии
Fig. 2. View of IBM SPSS Statistics 20.0 startup window and start of logistics regression procedure



A



B

Рис. 3. А. Вид основного диалогового окна настройки процедуры логистической регрессии.
 Б. Вид окна запроса параметров сохранения
Fig. 3. A. View of the main dialog box for setting up the logistics regression procedure.
 B. Save options query window view

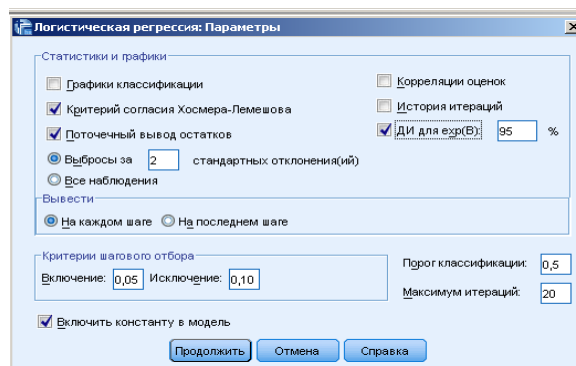


Рис. 4. Вид окна запроса параметров вывода результатов процедуры логистического анализа
Fig. 4. View of query window for output parameters of logistics analysis procedure results

Таблица 2

Сводка обработки наблюдений

Table 2

Report of processing of observations

Отобранные наблюдения Selected observations	включенные в анализ included in the analysis	122	55,2
	пропущенные наблюдения the missed observations	99	44,8
	всего total	221	100,0
Неотобранные наблюдения Not selected observations		0	0,0
Всего Total		221	100,0
а. Если включены веса, общее число наблюдений находится в таблице классификации			
a. If weights are included, the total number of observations is in the classification table			

Таблица 3

Переменные, не включенные в уравнение

Table 3

Variables not included in equation

Исследуемые переменные Studied variables		Значение критерия Value criterion	Степень свободы Degree freedoms	Уровень значимости (p) Level importance (p)
Шаг 0 Step 0	Возраст Age	22,158	1	0,000
	Характер_ЧМТ The nature of TBI	13,110	1	0,000
	Образование Appearing	7,284	1	0,007
	Тяжесть_травмы Severity of injury	42,323	1	0,000
	Возраст_получ_ЧМТ Age of receiving TBI	13,338	1	0,000
	Стаж_работы Work experience	0,026	1	0,872
Обобщенные статистики Summary statistics		71,645	6	0,000

Стаж работы статистически значимо не связана с исследуемым исходом ($p = 0,872$) и, соответственно, не будет использована для построения математической модели. Остальные независимые переменные имеют статистически значимую связь с клиническим исходом среди исследуемого контингента лиц.

Критерий Хи-квадрат (χ^2) – это критерий статистической значимости совокупного влияния всех независимых переменных в модели на зависимую переменную. На каждом шаге происходит добавление одной независимой переменной (табл. 4), в результате

происходит увеличение значения критерия χ^2 , что свидетельствует об улучшении качества модели.

Процедура останавливается, когда значение критерия χ^2 перестает изменяться. На шаге 4 значение критерия χ^2 составляет 91,461 с 4 степенью свободы и достигнутым уровнем статистической значимости, равной 0,0001.

В табл. 5 представлены величины, характеризующие качество модели. Так, величина – 2 Log Правдоподобие (отрицательное удвоенное значение логарифма функции) показывает, насколько хорошо модель соответствует

Таблица 4

Объединенные тесты для коэффициентов модели

Table 4

Combined tests for model coefficients

	Шаг Step	Хи-квадрат Chi square	Степень свободы Degree freedoms	Значение Value
Шаг 1 Step 1	Шаг Step	47,777	1	0,000
	Блок Block	47,777	1	0,000
	Модель Model	47,777	1	0,000
Шаг 2 Step 2	Шаг Step	17,749	1	0,000
	Блок Block	65,526	2	0,000
	Модель Model	65,526	2	0,000
Шаг 3 Step 3	Шаг Step	16,667	1	0,000
	Блок Block	82,193	3	0,000
	Модель Model	82,193	3	0,000
Шаг 4 Step 4	Шаг Step	9,268	1	0,002
	Блок Block	91,461	4	0,000
	Модель Model	91,461	4	0,000

Таблица 5

Сводка для модели

Table 5

The report for model

Шаг Step	-2 Log Правдоподобие -2 Log Credibility	R квадрат Кокса и Снелла Cox & Snell R Square	R квадрат Нэйджелкерка Nigelkerk R Square
1	118,058 ^a	0,324	0,436
2	100,308 ^a	0,416	0,559
3	83,641 ^a	0,490	0,660
4	74,373 ^a	0,527	0,710

a. Оценивание закончено на итерации номер 6, потому что оценки параметра изменились менее чем на 0,001
 a. The evaluation is finished on iteration number 6 because the parameter estimates have changed by less than 0.001

Таблица 6

Критерий Хосмера–Лемешова

Table 6

Hosmer–Lemeshow Criterion

Шаг Step	Хи-квадрат Chi square	Степень свободы Degree freedoms	Значение Value
1	0,000	0	.
2	6,003	8	0,647
3	10,076	8	0,260
4	4,465	8	0,813

исходным данным, при этом снижение данной величины соответствует улучшению качества математической модели. Показатели R^2 Кокса и Снелла и R^2 Нэйджелкерка показывают, насколько изменение зависимой переменной объясняется изменением совокупности независимых переменных. Значения этих показателей меняются от 0 до 1, однако значение показателя R^2 Кокса и Снелла, равное 1, является теоретически недостижимым, этого недостатка лишен показатель R^2 Нэйджелкерка.

В представленной математической модели показатель R^2 Нэйджелкерка составил 0,710, то есть изменение зависимой переменной на 71,0 % объясняется включенными в модель переменными.

Критерий согласия Хосмера–Лемешова показывает (табл. 6), насколько хорошо представленная модель согласуется с исходными

данными. Значение критерия должно быть не менее 0,5–0,6. Величина значения критерия составила 0,813, что соответствует высокому качеству модели.

Процент корректных предсказаний представлен в табл. 7, при этом наилучшее значение прогноза можем наблюдать на шаге 4. Так, процент корректных предсказаний отсутствия первичной инвалидности составил 93,0 %, наличия первичной инвалидности – 78,4 %, общий прогноз – 86,9 %, что соответствует условиям, оговоренным выше. В примечании указано значение порога отсечения модели.

В табл. 8 представлены данные, необходимые для построения уравнения регрессии прогноза риска первичной инвалидности, и оценка значимости каждой переменной, входящей в состав уравнения.

Таблица 7

Таблица классификации

Table 7

Таблица классификации

Шаг Step	Наблюдаемые Observed		Предсказанные Predicted		
			Первичная_инвалидность Primary disablement		Процент корректных The percentage of correct
			0,00	1,00	
Шаг 1 Step 1	Первичная_инвалидность Primary disablement	0,00	70	1	98,6
		1,00	25	26	51,0
	Общий процент General percent				78,7
Шаг 2 Step 2	Первичная_инвалидность Primary disablement	0,00	65	6	91,5
		1,00	17	34	66,7
	Общий процент General percent				81,1
Шаг 3 Step 3	Первичная_инвалидность Primary disablement	0,00	66	5	93,0
		1,00	8	43	84,3
	Общий процент General percent				89,3
Шаг 4 Step 4	Первичная_инвалидность Primary disablement	0,00	66	5	93,0
		1,00	11	40	78,4
	Общий процент General percent				86,9
а. Разделяющее значение = 0,500					
а. Separating value = 0.500					

Таблица 8

Переменные в уравнении

Table 8

Variables in the equation

Показатели Indicators	B	Стандартная ошибка Standard error	Вальд Wald	Степень свободы Degree freedoms	Значение Value	exp(B)	95 % Доверительный интервал для exp(B) 95 % Confidential in- terval for exp(B)		
							нижняя lower	верхняя upper	
Шаг 1 ^a Step 1 ^a	Тяжесть травмы Severity of injury	4,288	1,045	16,824	1	0,000	72,800	9,383	564,847
	Константа Constant term	-5,317	1,121	22,518	1	0,000	0,005		
Шаг 2 ^b Step 2 ^b	Возраст Age	0,172	0,046	14,181	1	0,000	1,187	1,086	1,298
	Тяжесть травмы Severity of injury	4,297	1,065	16,279	1	0,000	73,515	9,115	592,920
	Константа Constant term	-12,813	2,399	28,523	1	0,000	0,000		
Шаг 3 ^c Step 3 ^c	Возраст Age	0,183	0,048	14,620	1	0,000	1,201	1,093	1,319
	Характер черепно- мозговой травмы The nature of brain injury	2,426	0,636	14,558	1	0,000	11,318	3,254	39,359
	Тяжесть травмы Severity of injury	4,694	1,105	18,038	1	0,000	109,240	12,523	952,912
	Константа Constant term	-16,866	2,938	32,963	1	0,000	0,000		
Шаг 4 ^d Step 4 ^d	Возраст Age	0,243	0,061	15,761	1	0,000	1,275	1,131	1,438
	Характер черепно- мозговой травмы The nature of brain injury	2,389	0,667	12,844	1	0,000	10,903	2,952	40,270
	Образование Appearing	-1,980	0,704	7,903	1	0,005	0,138	0,035	0,549
	Тяжесть травмы Severity of injury	4,634	1,131	16,777	1	0,000	102,941	11,209	945,409
	Константа Constant term	-18,211	3,345	29,632	1	0,000	0,000		
a.	Переменные, включенные на шаге 1: Тяжесть травмы								
a.	Variables included in step 1: Severity of injury								
b.	Переменные, включенные на шаге 2: Возраст								
b.	Variables included in step 2: Age								
c.	Переменные, включенные на шаге 3: Характер черепно-мозговой травмы								
c.	Variables included in step 3: The nature of brain injury								
d.	Переменные, включенные на шаге 4: Образование								
d.	Variables included in step 4: Appearing								

Так, в уравнение вошли четыре независимых переменных, включенные в анализ на шаге 4: возраст, характер полученной ЧМТ, тяжесть полученной травмы, образование, а также свободный член, называемый константой.

Коэффициенты регрессии b_0, b_1, \dots, b_n позволяют определить размер вклада каждой переменной в модель прогноза, при этом коэффициент с положительным знаком повышает вероятность принадлежности к группе, отрицательный – уменьшает эту вероятность. Чем больше значение коэффициента $b_j, j=0,1,\dots,n$, тем более значительно данная переменная влияет на рассчитанный результат вероятности. Коэффициент b_j каждой переменной используется для построения уравнения регрессии.

Величина статистики Вальда в совокупности с числом степеней свободы показывает значимость коэффициента b_j . В представленных расчетах статистика Вальда каждого коэффициента имеет достаточно высокий уровень статистической значимости.

В табл. 8 для $B = b_j$ приводится значение $\exp(B)$ – отношение шансов (ОШ), равное отношению шанса того, что событие произойдет в одной группе, к шансу того, что оно произойдет в другой группе. Это значение показывает, на сколько изменится шанс при изменении значения переменной на единицу.

Доверительный интервал для $\exp(B)$ – интервал, который покрывает неизвестный параметр с заданной надежностью 95 %.

Для построения уравнения регрессии необходимо подставить в уравнение (2) полученное на четвертом шаге значение коэффициента $b_j, j=0,1,\dots,n$. Таким образом получаем:

$$y = -18,211 + (0,243x_1) + (2,389x_2) - (1,980x_3) + (4,634x_4), \quad (3)$$

где x_1 – возраст; x_2 – характер ЧМТ; x_3 – образование; x_4 – тяжесть травмы.

Подставив полученное значение y в формулу (1), рассчитываем прогнозируемый риск P наступления неблагоприятного кли-

нического исхода в виде первичной инвалидности.

В случае, если рассчитанный показатель P составляет более 0,5 (порога отсечения), прогнозируется высокая вероятность неблагоприятного клинического исхода, если менее 0,5 – низкая вероятность.

В математическую модель прогноза риска первичной инвалидности вошли четыре переменные, три из которых имеют положительное значение коэффициента регрессии (b_1, b_2, b_4), при увеличении значения которых на единицу повышается вероятность первичной инвалидности. Их называют факторами риска. К ним относятся возраст, характер ЧМТ, тяжесть ЧМТ. Одна переменная – образование, имеет отрицательное значение коэффициента регрессии b_3 , при уменьшении значения которой на единицу понижается вероятность неблагоприятного исхода. Такие переменные называются протективными.

По величине значения ОШ можно судить о степени влияния независимой переменной на изменение зависимой переменной. Чем выше значение ОШ, тем сильнее влияние независимой переменной.

Например, рассмотрим вариант, при котором пациент А характеризуется следующими значениями переменных: возраст составил 35 лет ($x_1 = 35$), по характеру полученная травма – изолированная ($x_2 = 1$), образование среднее ($x_3 = 0$), получена тяжелая травма ($x_4 = 2$). Подставив соответствующие значения в уравнение регрессии математической модели прогноза риска первичной инвалидности (3) значения переменных, получаем:

$$y = -18,211 + (0,243 \times 35) + (2,389 \times 1) - (1,98 \times 0) + (4,634 \times 2) = 1,951. \quad (4)$$

При этом прогнозируемый риск первичной инвалидности согласно модели составит:

$$P = \frac{\exp(1,951)}{1 + \exp(1,951)} = 0,875555641. \quad (5)$$

Таким образом, значение P математической модели выше значения 0,5, следовательно, у пациента имеется высокий риск первичной инвалидности.

ВЫВОДЫ

Итак, представленный в рекомендациях алгоритм расчетов и интерпретация результатов бинарного логистического регрессионного анализа может помочь исследователю построить собственную математическую модель прогноза вероятности любого клинического собы-

тия, в результате которого возможно численно оценить риск того или иного события.

Представленная методика имеет практическое значение в плане совершенствования знаний начинающего врача-исследователя в области построения математических моделей прогнозирования вероятности изучаемого события.

СПИСОК ЛИТЕРАТУРЫ

1. Гланц С. Медико-биологическая статистика. М.: Практика; 1998. 459 с.
2. Зайцев В.М., Лифляндский В.Г., Маринкин В.И. Прикладная медицинская статистика. СПб.: ООО «Издательство ФОЛИАНТ»; 2003. 432 с.
3. Леонов В.П. Логистическая регрессия в медицине и биологии. URL: http://www.biometrica-tomsk.ru/logit_1.htm (дата обращения: 20.05.2017).
4. Наследов А. SPSS 19: профессиональный статистический анализ данных. СПб.: Питер; 2011. 400 с.
5. Петри А., Сэбин С. Наглядная медицинская статистика. М.: ГЭОТАР-Медиа; 2009. 168 с.
6. Реброва О.Ю. Статистический анализ медицинских данных. Применение пакета прикладных программ STATISTICA. М.: Медиа Сфера; 2006. 312 с.
7. Юнкеров В.И., Григорьев С.Г. Математико-статистическая обработка данных медицинских исследований. СПб.: ВМедА; 2002. 266 с.

REFERENCES

1. Glants S. *Mediko-biologicheskaya statistika* [Medical and Biological Statistics]. Moscow: Practice; 1998. 459 p. (In Russian).
2. Zajcev V.M., Lifyandskij V.G., Marinkin V.I. *Prikladnaya medicinskaya statistika* [Applied Medical Statistics]. St. Petersburg: Ltd. "FOLIANT Publishing"; 2003. 432 p. (In Russian).
3. Leonov V.P. *Logisticheskaya regressiya v medicine i biologii* [Logistics Regression in Medicine and Biology]. (In Russian). Available at: http://www.biometrica.tomsk.ru/logit_1.htm (accessed: 20.05.2017).
4. Nasledov A. *SPSS 19: professional'nyj statisticheskij analiz dannyh* [SPSS 19: Professional Statistical Analysis of Data]. St. Petersburg: Piter; 2011. 400 p. (In Russian).
5. Petri A., Sebin S. *Naglyadnaya medicinskaya statistika* [Evident Medical Statistics]. Moscow: GEOTAR-Media; 2009. 168 p. (In Russian).
6. Rebrova O.Yu. *Statisticheskij analiz meditsinskih dannyh. Primeneniye paketa prikladnyh programm STATISTICA* [Statistical Analysis of Medical Data. Application of the STATISTICA Application Package]. Moscow: Media Sfera; 2006. 312 p. (In Russian).
7. Yunkerov V.I., Grigorev S.G. *Matematiko-statisticheskaya obrabotka dannyh meditsinskih issledovanij* [Mathematical and Statistical Processing of Medical Research Data]. St. Petersburg: VMedA; 2002. 266 p. (In Russian).

Информация об авторах

Маньяков Рустам Ринатович, кандидат медицинских наук, доцент кафедры патологии Медицинского института. Тамбовский государственный университет имени Г.Р. Державина, г. Тамбов, Российская Федерация. E-mail: 8730241@mail.ru

Вклад в статью: идея и дизайн исследования, написание части текста, редактирование.

ORCID: <https://orcid.org/0000-0002-9935-8373>

Деменкова Валерия Владимировна, аспирант, кафедра общественного здоровья и здравоохранения Медицинского института. Тамбовский государственный университет имени Г.Р. Державина, г. Тамбов, Российская Федерация. E-mail: vill20061@yandex.ru

Вклад в статью: написание части текста, редактирование.

ORCID: <https://orcid.org/0000-0002-2241-7048>

Агафонова Юлия Владимировна, аспирант, кафедра общественного здоровья и здравоохранения Медицинского института. Тамбовский государственный университет им. Г.Р. Державина, г. Тамбов, Российская Федерация. E-mail: Evamedical.ru@gmail.com

Вклад в статью: анализ литературы, дизайн исследования, написание части текста, редактирование.

ORCID: <https://orcid.org/0000-0003-2418-5129>

Петряхин Денис Николаевич, адъюнкт, кафедра автоматизации управления медицинской службой с военно-медицинской статистикой. Военно-медицинская академия имени С.М. Кирова МО РФ, г. Санкт-Петербург, Российская Федерация. E-mail: Den-09-ra@yandex.ru

Вклад в статью: обработка материала, написание части текста статьи.

ORCID: <http://orcid.org/0000-0001-6270-6509>

Конфликт интересов отсутствует.

Для контактов:

Маньяков Рустам Ринатович
E-mail: 8730241@mail.ru

Поступила в редакцию 11.03.2020 г.
Поступила после рецензирования 29.04.2020 г.
Принята к публикации 15.05.2020 г.

Information about the authors

Rustam R. Manyakov, Candidate of Medicine, Associate Professor of Pathology Department of Medical Institute. Derzhavin Tambov State University, Tambov, Russian Federation. E-mail: 8730241@mail.ru

Contribution: idea and design of the study, writing part of the text, editing.

ORCID: <https://orcid.org/0000-0002-9935-8373>

Valeriya V. Demenkova, Post-Graduate Student, Public Health and Health Care Department of Medical Institute. Derzhavin Tambov State University, Tambov, Russian Federation. E-mail: vill20061@yandex.ru

Contribution: writing part of text, editing.

ORCID: <https://orcid.org/0000-0002-2241-7048>

Yuliya V. Agafonova, Post-Graduate Student, Public Health and Health Care Department of Medical Institute. Derzhavin Tambov State University, Tambov, Russian Federation. E-mail: Evamedical.ru@gmail.com

Contribution: literature analysis, design, writing part of text, editing.

ORCID: <https://orcid.org/0000-0003-2418-5129>

Denis N. Petryakhin, Post-Graduate Student, Department of Automation Management of the Medical Service with Military Medical Statistics (AUMS). S.M. Kirov Military Medical Academy of the Ministry of Defense of the Russian Federation, St. Petersburg, Russian Federation. E-mail: Den-09-ra@yandex.ru

Contribution: material processing, writing part of the text.

ORCID: <https://orcid.org/0000-0001-6270-6509>

There is no conflict of interests.

Corresponding author:

Rustam R. Manyakov
E-mail: 8730241@mail.ru

Received 11 March 2020
Reviewed 29 April 2020
Accepted for press 15 May 2020