

## МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ВЕРОЯТНОСТИ СОБЫТИЯ «ИНСУЛЬТ – ФАКТОРЫ РИСКА»

**Кадырова И. А.**

*Карагандинский государственный медицинский университет, Караганда, Казахстан (100000, Караганда, ул. Гоголя, 40), e-mail: [Iriska441@yandex.ru](mailto:Iriska441@yandex.ru)*

---

В статье приведен материал по математическому моделированию вероятности заболевания инсультом у отдельного пациента. Методика может применяться также и для определения этих параметров у социальных групп и граждан. Был проведен большой объем лабораторных исследований в одном из медучреждений г. Караганды, затем была определена корреляция между событием «инсульт» и 35 факторами риска. Следующим этапом являлась разработка регрессивной модели «инсульт – факторы риска». Модель разрабатывалась методом логистической регрессии. Полученные уравнения позволили определять вероятность наступления инсульта. Так как уравнение регрессии близко по своей сути к детерминированному, то возможно его дифференцирование. Первая производная от вероятности по каждому из факторов риска позволила определить скорость нарастания вероятности. Это положение является новым.

---

Ключевые слова: инсульт, факторы риска, уравнение регрессии, вероятность.

## MATHEMATICAL MODEL OF THE PROBABILITY OF THE EVENT "STROKE – RISK FACTORS"

**Kadyrova I. A.**

*Karaganda State Medical University, Karaganda, Kazakhstan(100000, Karaganda,street Gogolya, 40),e-mail: [Iriska441@yandex.ru](mailto:Iriska441@yandex.ru)*

---

The article contains data on the mathematical modeling of the stroke probability in the individual patient. This method can also be applied for determining these parameters for social groups and individuals. A large amount of laboratory research was performed in hospital of Karaganda city, then the correlation was determined between the event "stroke" and 35 risk factors. Next stage was to develop a regression model "stroke-risk factors." This model was developed by logistic regression method. The obtained equations allowed to determine the probability of occurrence of a stroke. Since the regression equation is inherently close to the determined, it can be differentiated. First derivative of the probability of each of the risk factors allowed us to determine the rate of increase of probability. This provision is new.

---

Key words: stroke, risk factors, regression equations, probability.

Введение. Инсульт является одной из главных причин смертности в мире [2]. Сахарный диабет 2 типа (далее СД) значительно увеличивает риск развития инсульта [4]. Для осуществления скрининга состояния здоровья пациентов необходимо разработать методику, прогнозирующую вероятность возникновения инсульта у пациентов с СД и у пациентов без него, учитывая возможные факторы риска.

Целью исследования явилась разработка математической модели «инсульт – факторы риска» для определения вероятности возникновения инсульта в социальных группах и у отдельных граждан.

Материалы и методы исследования. В исследование были включены 153 участника в возрасте от 40 до 83 лет с равным включением мужчин и женщин.

Первая группа – 32 пациента с нарушениями мозгового кровообращения. За текущий год 3 пациента перенесли инсульт.

В группу больных с сахарным диабетом включено 94 пациента, имеющих сахарный диабет в компенсаторной стадии, инсульт за текущий год – у 9 человек. Контрольную группу составили 30 практически здоровых людей, сопоставимых по возрасту и полу. Критериями включения в контрольную группу были возраст от 40–80 лет, нормальное артериальное давление, ИМТ в пределах 18,5–25,0, показатели биохимического анализа крови и аппаратных методов диагностики, соответствующих относительной норме.

У всех участников исследования проводилось анкетирование. В анкете представлены вопросы на выявление факторов риска.

Затем были проведены следующие измерения: индекс массы тела (ИМТ), систолическое артериальное давление (САД) и диастолическое артериальное давление (ДАД), частота дыхательных движений (ЧДД) и частота сердечных сокращений (ЧСС). Определение уровня глюкозы крови (в этом и других анализах) утром натощак в стандартных условиях. Гликозилированный гемоглобин определялся иммунологическим методом при помощи реагентов Vital и спектрофотометра DR 2800 с длиной волны 443 нм. Исследования свертывающей системы крови (ПТИ, количество фибриногена, агрегация тромбоцитов, АПТВ) проводилось на одноканальном анализаторе параметров гемостаза Clot – 1. Для определения биохимических показателей крови (холестерин, триглицериды, АЛАТ, АСАТ, общий билирубин, прямой билирубин, мочевины, креатинин, общий белок) использовались реагенты фирмы Vital с биохимическим анализатором BioSystem A-15. ЭКГ исследование проводилось в 12 отведениях на электрокардиографе BTL-088D, Великобритания 2011, а ЦДК б/ц ствола проводилось сканером MEDISON SONOACE X8. Линейный датчик 5–12 мГц.

Следующим этапом явилась статистическая обработка измерений, которую проводили согласно общепринятым методам в программе Statistica 6.1. Для получения коэффициентов корреляции была создана квадратная корреляционная матрица. Полученная зависимость подчинялась нормальному закону распределения [3].

Поскольку данные исследования включали как качественные, так и количественные признаки, появилась необходимость создания матрицы с закодированными значениями. Определялся максимум и минимум изучаемого фактора, полученный интервал разбивался на необходимое количество пронумерованных интервалов (количество и размер интервалов определялся исследователем согласно нормальным значениям показателей, а также частоте появления значений в интервале).

Событие (инсульт) в матрице – зависимая переменная и кодировалось 1 и 0. Остальные же факторы риска явились независимыми переменными и кодировались в зависимости от количества интервалов у данного фактора.

Для ЦДК выделили следующие наиболее часто встречающиеся критерии у участников исследования: атеросклероз сонной артерии, повышение периферического сопротивления в сонных артериях, деформация позвоночных артерий, компрессия позвоночных артерий, ускорение кровотока в бассейне позвоночных артерий, норма. Для ЭКГ критериями выступили следующие данные: аритмия, тахикардия, экстрасистолия, нарушение проводимости, нарушение реполяризации, норма. При наличии признака ячейке присваивалось значение 1, при отсутствии – 2.

Результаты и обсуждение. Анализ корреляционных связей проводился для двух групп: 1 – участники с произошедшим инсультом и участники контрольной группы; 2 – пациенты с СД с относительно нормальными показателями для обозначения минимальной вероятности возникновения инсульта и пациенты с СД с произошедшим инсультом.

В первой группе мы наблюдаем тесную зависимость инсульта от следующих факторов: повышенный индекс массы тела, длительность АГ, повышение САД, ДАД, ЧСС, изменение показателей мочевины и креатинина. Самым тесно коррелируемым признаком из всех рассмотренных является количество фибриногена. Интересно, что длительность АГ тесно коррелирует с гликозилированным гемоглобином.

Во второй группе значимая теснота корреляции присутствует между событием – инсульт и длительностью АГ, и САД. Между факторами длительность заболевания СД и длительность АГ существует значимый коэффициент корреляции. Столь большая разница в количестве коррелируемых факторов в разных группах обусловлена разными путями развития нарушения мозгового кровообращения у больных СД и пациентами, не имеющих данной патологии.

Таблица. Обобщенная таблица данных

№Х	Фактор	Коэффициент корреляции фактора с событием инсульт		Диапазон интервала	Коды интервала	Коэффициент регрессии	
		1 гр.	2 гр.			1 гр.	2 гр.
X1	Возраст	0,16	0,19	Меньше 60–90	1–4	- 2,45497	- 0,2401 33
X2	ИМТ	0,38	-0,05	16–31 и выше	1–4	10,56	- 1,9328 9
X3	Длительность АГ	0,28	0,23	0–21 и выше	1–6	2,7	- 4,1692 9
X5	САД	0,28	0,34	110–191 и выше	1–9	5,05695 9E+00	2

X6	ДАД	0,31	0,08	60–110 и выше	1–5	1,9506	- 1,2411 3
X7	ЧСС	0,36	0,14	Ниже 60–81 и выше	1–4	1,12763 9	3,9
X8	ЧДД	0,13	-0,01	16–21 и выше	1–3	3,57748	5,31
X9	Глюкоза крови	0,01	-0,02	3,3–12 и выше	1–4	- 5,67782	- 3,2460 4
X10	Холестерин	0,03	-0,04	Ниже 5.2–8 и выше	1–4	- 4,47648	- 4,4364 6
X11	Триглицериды	-0,08	0,13	0,14–4,5 и выше	1–3	- 11,1834	7
X12	ПТИ	-0,07	0,11	77–101 и выше	1–4	3,56	- 4,4534 4
X13	Фибриноген	0,44	-0,02	2,3–7,5	1–3	- 3,07065	- 4,5787 4
X14	Агрегация тромбоцитов	-0,22	0,04	13–19 и выше	1–3	2,31037	18
X15	АПТВ	-0,18	-0,03	23–41 и выше	1–3	- 0,75682 1	- 4,0882 4
X16	Алат	-0,14	0,14	0–61 и выше	1–3	- 1,28629	- 4,0114 4
X17	Асат	-0,07	0,07	0–61 и выше	1–3	- 8,35471	- 2,4751 7
X18	Билирубин общий	0,24	0,14	8,6–20,6 и выше	1–3	1,99058	- 1,2984 7
X19	Билирубин прямой	0,07	0,15	3,4–12 и выше	1–3	- 2,20222	- 0,6665 65
X20	Мочевина	0,26	0,07	25–51 и выше	1–4	6	9,6200 53E+00
X21	Креатинин	0,30	0,07	40–116 и выше	1–5	- 7,56844	- 7,4919 3
X22	Общий белок	-0,14	-0,10	Ниже 65–86 и выше	1–4	1,38996	1,4229 0
X23	Тахикардия	–	–	Да / нет	1–2	- 5,99556	0,2294 91
X24	Нарушение проводимости миокарда	–	–	Да / нет	1–2	- 1,58150	1,7763 81

X25	Нарушение процессов реполяризации миокарда	–	–	Да/нет	1–2	- 2,31624	- 0,340166
X26	Аритмия	–	–	Да/нет	1–2	6,2482	- 24,3798
X27	Экстрасистолия	–	–	Да/нет	1–2	- 11,0942	- 27,0057
X28	Норма ЭКГ	–	–	Да/нет	1–2	1,149431	1,142954
X29	Атеросклероз сонной артерии	–	–	Да/нет	1–2	- 0,581736	6,0955
X30	Повышение периферического сопротивления	–	–	Да/нет	1–2	- 4,18548	- 5,64649
X31	Деформация ПА	–	–	Да/нет	1–2	0,658318	- 19,4154
X32	Компрессия ПА	–	–	Да/нет	1–2	- 4,66957	16
X33	Повышение ускорение кровотока в бассейне ПА	–	–	Да/нет	1–2	9,98	16
X34	Норма ЦДК	–	–	Да/нет	1-2	- 1,63149	2,914050E+01
X35	Гликозилированный гемоглобин	0,13	0,02	4-12 и выше	1-9	- 7,53494	0,917
X36	Длительность заболевания СД	-	0,18	0-21 и выше	1-6	1,785463	- 1,38249

Следующим этапом явилось математическое моделирование, которое основывалось на методе логистической регрессии.

Математическая модель позволяет: изучить характер изменения показателя при изменении действующих на систему факторов, оценить степень влияния факторов на величину

показателя-отклика, прогнозирование показателя-отклика для заданных уровней факторов, определение оптимальных уровней факторов, для обозначения желаемых или требуемых значений показателей состояния системы [5].

Эта модель имеет вид:

$$y = \exp(b_0 + b_1 \cdot x_1 + \dots + b_i \cdot x_i) / \{1 + \exp(b_0 + b_1 \cdot x_1 + \dots + b_i \cdot x_i)\};$$

$$0 < y < 1,$$

где:  $y$  – вероятность возникновения инсульта;

$b_0$  – свободный член;

$b_1 \dots b_i$  – коэффициенты регрессии факторов  $x_1 \dots x_i$ .

В результате обработки получили коэффициенты регрессии, сведенные в таблицу 1.

В качестве примера приведем уравнение регрессии для пациентов 1 группы:  $Y = \text{EXP}(0.394929 - 2.45497 \cdot X_1 + 10.56 \cdot X_2 + 2.7 \cdot X_3 + 5.056959E+00 \cdot X_5 + 1.9506 \cdot X_6 + 1.127639 \cdot X_7 + 3.57748 \cdot X_8 - 5.67782 \cdot X_9 - 4.47648 \cdot X_{10} - 11.1834 \cdot X_{11} + 3.56 \cdot X_{12} - 3.07065 \cdot X_{13} + 2.31037 \cdot X_{14} - 0.756821 \cdot X_{15} - 1.28629 \cdot X_{16} - 8.35471 \cdot X_{17} + 1.99058 \cdot X_{18} - 2.20222 \cdot X_{19} + 6 \cdot X_{20} - 7.56844 \cdot X_{21} + 1.38996 \cdot X_{22} - 5.99556 \cdot X_{23} - 1.58150 \cdot X_{24} - 2.31824 \cdot X_{25} + 6.2482 \cdot X_{26} - 11.0942 \cdot X_{27} + 1.149431 \cdot X_{28} - 0.581736 \cdot X_{29} - 4.18548 \cdot X_{30} + 0.658318 \cdot X_{31} - 4.66957 \cdot X_{32} + 9.98 \cdot X_{33} - 1.63149 \cdot X_{34} - 7.53494 \cdot X_{35} + 1.785463 \cdot X_{36}) / (1 + \text{EXP}(0.394929 - 2.45497 \cdot X_1 + 10.56 \cdot X_2 + 2.7 \cdot X_3 + 5.056959E+00 \cdot X_5 + 1.9506 \cdot X_6 + 1.127639 \cdot X_7 + 3.57748 \cdot X_8 - 5.67782 \cdot X_9 - 4.47648 \cdot X_{10} - 11.1834 \cdot X_{11} + 3.56 \cdot X_{12} - 3.07065 \cdot X_{13} + 2.31037 \cdot X_{14} - 0.756821 \cdot X_{15} - 1.28629 \cdot X_{16} - 8.35471 \cdot X_{17} + 1.99058 \cdot X_{18} - 2.20222 \cdot X_{19} + 6 \cdot X_{20} - 7.56844 \cdot X_{21} + 1.38996 \cdot X_{22} - 5.99556 \cdot X_{23} - 1.58150 \cdot X_{24} - 2.31824 \cdot X_{25} + 6.2482 \cdot X_{26} - 11.0942 \cdot X_{27} + 1.149431 \cdot X_{28} - 0.581736 \cdot X_{29} - 4.18548 \cdot X_{30} + 0.658318 \cdot X_{31} - 4.66957 \cdot X_{32} + 9.98 \cdot X_{33} - 1.63149 \cdot X_{34} - 7.53494 \cdot X_{35} + 1.785463 \cdot X_{36}))$ ;

Уравнение (1)

Для определения вероятности наступления инсульта необходимо вместо  $X_1 \dots X_{36}$  подставить код интервала, которому принадлежат показатели. При отрицательном коэффициенте необходимо подставлять значения кода в обратном порядке. Также мы определили, что шансы развития инсульта в группе с СД в 2,08 раза выше, чем у участников без СД.

Критерии значимости составляли  $p=0,02038$ , при  $\chi^2=54,174$  для второй группы, а для первой группы  $p=0,03683$ , при  $\chi^2=51,352$ . Данные критерии значимости подтверждают работоспособность модели [1].

Уравнение регрессии по своей математической сущности приближается к детерминированному. Дифференцирование функции вероятности не дает нам точные

расчетные данные, но позволяет определить знак скорости нарастания вероятности заболевания инсультом.

Производная от вероятности имеет вид:  $\frac{dy}{dx_i} = \frac{az}{(1+z)^2}$ ,

где:  $a = b_0 + b_1 + \dots + b_i - 1$ ;

$z = \exp b_0 + b_1 + \dots + b_i * x$ .

При увеличении  $x_i \frac{dy}{dx_i} \rightarrow 0$

Таким образом, если при увеличении факторов риска вероятность события увеличивается, то скорость возрастания вероятности уменьшается.

Функция вероятности от любого фактора риска представляет собой гиперболу, асимптотически приближающуюся к 1, а функция изменения скорости вероятности – гиперболу, стремящуюся к 0.

Закключение. Проведенные исследования позволили сделать следующие выводы:

1. Установлены коэффициенты корреляции. Наибольшая теснота связи в первой группе у события «инсульт» с факторами: повышенный индекс массы тела, длительность АГ, повышение САД, ДАД, ЧСС, изменение показателей мочевины и креатинина. Самым тесным коррелируемым признаком из всех рассмотренных является количество фибриногена. Во второй группе выявлена значимая теснота корреляции между событием «инсульт» и длительностью АГ и САД.
2. Разработана математическая модель на основе логистической регрессии, позволяющая определять вероятность наступления инсульта у больных СД и пациентов, не имеющих СД.
3. Анализ математической модели позволил установить скорость нарастания инсульта в выборке, что можно спроецировать на генеральную совокупность населения.
4. Проведенное исследование явилось базой для разработки методики прогнозирования вероятности возникновения инсульта у различных социальных групп и граждан.

### Список литературы

1. Гланц С. Медико-биологическая статистика учеб. для вузов. – М.: Практика, 1998. – С. 459 с.
2. Официальное периодическое издание: Всемирная организация здравоохранения: Десять ведущих причин смерти // Информационный бюллетень. – 2011. – № 310. URL: [http://www.who.int/governance/eb/who\\_constitution\\_ru.pdf](http://www.who.int/governance/eb/who_constitution_ru.pdf).

3. Реброва О. Ю. Статистический анализ медицинских данных. – М.: Медиа-Сфера, 2006. – 305 с.
4. Щепанкевич Л. А., Вострикова Е. В., Пилипенко П. И. Клинические особенности ишемического инсульта, развившегося на фоне сахарного диабета 2 типа // Медицина и образование в Сибири. – 2012. – № 3. URL:[http://www.ngmu.ru/cozo/mos/article/text\\_full.php?id=73](http://www.ngmu.ru/cozo/mos/article/text_full.php?id=73).
5. Юнкеров В. И., Григорьев С. Г. Математико-статистическая обработка данных медицинских исследований. – СПб.: ВМедА, 2002. – 266 с.

**Рецензенты:**

Жаутикова Сауле Базарбаевна, д-р мед. наук, профессор, заведующая кафедрой патологической физиологии «Карагандинский государственный медицинский университет», г. Караганда.

Грузин В. В., д-р техн. наук, профессор, заведующий кафедрой «Карагандинский государственный технический университет», г. Караганда.