

УДК 616.831—005.1—08

Ю. С. ТУНЯН, И. М. СЛИВНЯК

ДВУХПОРОГОВАЯ ЗАДАЧА В ПРОГНОЗИРОВАНИИ МОЗГОВЫХ ИНСУЛЬТОВ

Предложена методика определения степени предрасположенности обследуемых лиц к острым нарушениям мозгового кровообращения с целью рациональной организации профилактических мероприятий в амбулаторно-поликлинических условиях.

Предупреждение сердечно-сосудистых заболеваний, в частности, острых нарушений мозгового кровообращения, является весьма актуальной, но вместе с тем и сложной проблемой. В этом плане важное значение приобретает внесение коррективов в практику отбора контингента для диспансерного наблюдения. В данной работе предлагается математически обоснованная методика для выявления лиц, предрасположенных в той или иной степени к развитию острых нарушений мозгового кровообращения (ОНМК).

Условимся характеризовать состояние обследуемого лица по отношению к данному заболеванию принадлежностью его к одной из четырех групп: Γ_0 — группа здоровых, Γ_1 — группа умеренного риска, Γ_2 — группа повышенного риска и Γ_3 — группа больных. Обследование охватывает g признаков, результатом его является набор из g чисел $\vec{x} = (x_1, \dots, x_g)$. Требуется на основании полученного вектора \vec{x} отнести обследуемого к одной из категорий $\Gamma_0, \dots, \Gamma_3$. При этом, поскольку больные не нуждаются в первичном профилактическом обследовании, можно ограничиться тремя категориями $\Gamma_0, \Gamma_1, \Gamma_2$.

Исходными данными служат результаты обследования n больных (б) и m практически здоровых людей (з) — выборки $\vec{x}_6^{(1)}, \dots, \vec{x}_6^{(n)}$ и $\vec{x}_3^{(1)}, \dots, \vec{x}_3^{(m)}$. Для их различения используется дискриминатор

$$z = f(\vec{x}), \quad (1)$$

сопоставляющий каждому вектору \vec{x} определенное значение z . Таким образом, на языке величины z имеем две обучающие выборки

$$z_6^{(i)} = f(\vec{x}_6^{(i)}), \quad i = 1, \dots, n; \quad z_3^{(j)} = f(\vec{x}_3^{(j)}), \quad j = 1, \dots, m, \quad (2)$$

и нужно по результату обследования $z = f(\vec{x})$ определить, к какой из трех категорий $\Gamma_0, \Gamma_1, \Gamma_2$ принадлежит обследуемый.

Такая постановка является одним из вариантов известной задачи классификации при наличии обучающих выборок [1]. Однако в доступной литературе мы не нашли каких-либо относящихся к ней общих результатов. Близкая постановка задачи рассматривается в работе А. С. Мاستыкина с соавт. [2].

Разобьем ось z каким-либо способом на три области $\Delta_0, \Delta_1, \Delta_2$ и условимся при попадании значения z в область Δ_1 относить обследуемого к категории Γ_1 . При этом возможны разного рода ошибки, каждая из которых связана с некоторыми потерями. Пусть потери при отнесении здорового к категории Γ_2 и больного (если бы таковые проходили профилактическое обследование), а также лица из группы повышенного риска к категории Γ_0 равны 1. Потери же при отнесении лица, принадлежащего к любой из групп $\Gamma_0, \Gamma_2, \Gamma_3$, к группе Γ_1 будем считать равным θ , где $0 < \theta < 1$. Тогда средние потери в расчете на одного обследуемого, к которым приводит данное решающее правило, равны:

по группе здоровых —

$$R_3 \approx \frac{n(\Delta_2)}{n} + \theta \frac{n(\Delta_1)}{n},$$

по группе больных —

$$R_6 \approx \frac{m(\Delta_0)}{m} + \theta \frac{m(\Delta_1)}{m}.$$

Здесь $n(\Delta_1)$ и $m(\Delta_1)$ — число обследованных здоровых и соответственно больных, для которых результат обследования z попал в область Δ_1 .

Каждому способу выбора областей Δ_1 отвечают свои значения потерь R_3, R_6 . Естественно стремиться выбрать Δ_1 так, чтобы одновременно минимизировать обе эти величины. Это, однако, невозможно, так как R_3 и R_6 изменяются в противоположных направлениях. Поэтому будем минимизировать величину R_6 как более важную, а относительно потерь R_3 потребуем лишь, чтобы они были не слишком велики — не превосходили некоторой заданной границы α . Таким образом, оптимальное решающее правило обеспечивает $\min R_6$ при условии $R_3 \leq \alpha$.

Не останавливаясь на доказательстве, сформулируем способ отыскания оптимальных областей Δ_1 . Пусть $P_3(z), P_6(z)$ — плотности распределения величины z для групп здоровых и больных, задаваемые обучающими выборками (2). Тогда искомые области Δ_1 имеют вид

$$\begin{aligned} \Delta_0 &= \left\{ \frac{P_6(z)}{P_3(z)} < \lambda_1 \right\}, & \Delta_2 &= \left\{ \frac{P_6(z)}{P_3(z)} \geq \lambda_2 \right\}, \\ \Delta_1 &= \left\{ \lambda_2 \leq \frac{P_6(z)}{P_3(z)} < \lambda_1 \right\}, \end{aligned} \quad (3)$$

то есть состоят из тех значений z , для которых отношение $\frac{P_6(z)}{P_3(z)}$

соответственно меньше некоторого порога λ_1 , больше другого порога λ_2 или заключено между ними.

Способ отыскания λ_1, λ_2 зависит от значений параметров α, θ . Ограничимся наиболее важным для практики случаем

$$\alpha \leq \theta < \frac{1}{2}. \quad (4)$$

При этом пороги λ_1, λ_2 связаны соотношением

$$\lambda_2 = \left(\frac{1-\theta}{\theta} \right)^2 \lambda_1,$$

а λ_1 — наименьшее число, для которого

$$\frac{1}{n} [\theta n(\Delta_1) + n(\Delta_2)] \leq \alpha,$$

где области Δ_1, Δ_2 определены формулами (3).

Остановимся специально на случае нормальных плотностей распределения

$$P_3(z) = N(a_3, \sigma_3^2), \quad P_6(z) = N(a_6, \sigma_6^2). \quad (5)$$

Для упрощения записей примем предположения (4)

$$a_3 < a_6, \quad \sigma_3 < \sigma_6.$$

Обозначим

$$U = \frac{\sigma_6^2}{\sigma_3^2}, \quad v = \frac{a_3 - a_6}{\sigma_3}, \quad \varphi(\lambda) = \sqrt{\frac{u}{u-1} \left[\ln(\lambda^2 u) + \frac{v^2}{u-1} \right]},$$

$$F(\lambda) = \Phi \left[\frac{v}{u-1} + \varphi(\lambda) \right] - \Phi \left[\frac{v}{u-1} - \varphi(\lambda) \right],$$

где $\Phi(x)$ — функция Лапласа [3]. Тогда области Δ_i имеют вид

$$\Delta_0 = \{z < z_1\}, \quad \Delta_1 = \{z_1 \leq z < z_2\}, \quad \Delta_2 = \{z \geq z_2\},$$

где пороги z_1, z_2 связаны соотношениями

$$z_1 = a_3 + \sigma_3 \cdot \left[\frac{v}{u-1} + \varphi(\lambda_1) \right],$$

а величины λ_i определяются из уравнений

$$\begin{cases} \lambda_2 = \left(\frac{1-\theta}{\theta} \right)^2 \lambda_1, \\ \theta F(\lambda_1) + (1-\theta) F(\lambda_2) = 1 - \alpha. \end{cases}$$

Пример. Для рассматриваемой задачи профилактики заболевания ОНМК был разработан набор из 19 признаков (табл. 1, 2). Величина x_{19} имеет 4 возможных значения (табл. 2), а каждая из величин x_1, \dots, x_{18} — два значения: $x_i = 1$ — верхний уровень и $x_i = 0$ — нижний

i	Признак*	U _i
1	Неблагополучное семейное положение**	0
2	Длительные психогенные травмы (в анамнезе)	6
3	Черепно-мозговая травма (в анамнезе)	0
4	Малоподвижный образ жизни	6
5	Неиспользование ежегодного отпуска	0
6	Склонность к перееданию	7
7	Избыточное употребление жирной пищи	3
8	Курение сигарет (папирос) более 20 штук в день	1
9	Стаж курения 30 лет и более	7
10	Злоупотребление алкоголем (в анамнезе)	7
11	Повышенная метеочувствительность	7
12	Сердечно-сосудистые заболевания у родителей и (или) спсбсов	11
13	Сердечно-сосудистые заболевания у остальных кровных родственников	6
14	Повышенный вес	2
15	Гиперстеническая конституция	1
16	Повышение артериального давления	17
17	Повышение содержания сахара в крови	4
18	Повышение содержания холестерина в крови	3
19	Возраст	1

* В этом столбце при $i=1, \dots, 18$ указаны верхние уровни признаков, отвечающие значениям $x_i = 1$.

** Имеются в виду одинокие, разведенные, вдовцы (вдовы).

уровень i -го признака. Исходными данными послужили результаты проведенного в 1974 г. обследования 100 больных и 100 практически здоровых мужчин — жителей г. Еревана. В качестве (1) был выбран линейный дискриминатор Фишера

$$z = \sum_{i=1}^{19} u_i x_i \quad (6)$$

(по [4]). Значения коэффициентов u_i , подсчитанные на основании указанной выборки и нормированные при помощи условия $\sum u_i^2 = 100$, приведены в табл. 1. Как видно из этой таблицы, признаки с номерами $i = 1, 3, 5$ с точки зрения разделения обучающих выборок (2) неинформативны.

Выборки (2) привели к нормальным законам (5) с параметрами $a_3 = 20,7$, $\sigma_3 = 11,2$, $a_6 = 44,5$, $\sigma_6 = 14,7$.

Для $\theta = 0,25$ и различных значений $\alpha \leq \theta$ были подсчитаны пороги z_1 , z_2 и средние потери R_6 (табл. 3). Как и следовало ожидать, с возрастанием α величина R_6 убывает, а пороги z_1 , z_2 смещаются влево.

В частности, при $\alpha = 0,12$ получаем

$$z_1 = 25, \quad z_2 = 40 \quad (7)$$

и $R_6 = 0,16$. Практически это означает следующее. Для обследуемого лица вычисляем по формуле (6) величину z и сравниваем ее с порогами (7). Согласно (6), величина z равна сумме значения x_1 (табл. 2) и тех из коэффициентов u_1, \dots, u_{18} (табл. 1), для которых соответствующие признаки оказались у обследуемого лица на верхнем уровне. При $z < 25$ относим данное лицо к группе здоровых, при $25 < z < 40$ — к группе умеренного и при $z > 40$ — повышенного риска. Указанная

Таблица 2

Возраст	x_{19}
40—49 лет	1
50—59 лет	4
60—69 лет	8
70 лет и старше	12

Таблица 3

α	Z_1	Z_2	R_6
0,24	17,4	34,6	0,085
0,15	22,3	37,9	0,135
0,12	25,0	39,7	0,160
0,10	26,9	41,4	0,195

процедура обеспечивает минимум потерь $R_6 = 0,16$ по группе больных (если бы таковая была подвергнута обследованию) среди всех процедур, для которых средние потери по группе здоровых $R_3 < 0,12$.

Интересно сравнить при порогах (7) теоретические значения величин R_3, R_6 с выборочными. Для использованной при получении (7) выборки мужчин — жителей г. Еревана число значений z для группы больных, попавших в области $\Delta_1 = \{25 < z < 40\}$ и $\Delta_0 = \{z < 25\}$, составило 29 и 11; для группы здоровых — в области Δ_1 и $\Delta_2 = \{z > 40\}$ — соответственно 25 и 7. Это приводит к средним потерям

$$R_6 = \frac{1}{100} \left(11 + \frac{29}{4} \right) = 0,18, \quad R_3 = \frac{1}{100} \left(7 + \frac{25}{4} \right) = 0,13$$

в соответствии с табл. 3.

Еще больший интерес представляет сравнение теоретических значений $R_6 = 0,16, R_3 = 0,12$ с их выборочными значениями, подсчитанными по новым группам больных и здоровых, не использованным при составлении табл. 3. С этой целью в 1977 г. было проведено дополнительное обследование 55 больных и 70 здоровых мужчин — жителей г. Еревана. Для группы больных число значений z , попавших в области Δ_1 и Δ_0 , составило 12 и 4; для группы здоровых — в области Δ_1 и Δ_2 — соответственно 22 и 4. Средние потери равны

$$R_6 = 0,13, \quad R_3 = 0,135,$$

что близко к их теоретическим значениям.

Аналогичная проверка была произведена при тех же порогах z_1, z_2 для выборки из 100 больных и 50 здоровых мужчин — жителей г. Моск-

вы. Число больных, попавших в области Δ_1 и Δ_0 , составило 28 и 10; число здоровых — в области Δ_1 и Δ_2 — соответственно 16 и 3. Отсюда $R_6 = 0,17$, $R_3 = 0,14$, что снова близко к теоретическим значениям R_6 , R_3 , полученным на основании исходной выборки мужчин — жителей г. Еревана. Таким образом, по-видимому, есть основания считать последнюю представительной для генеральной совокупности мужчин — жителей крупных городов.

Описанная выше методика установления степени предрасположенности обследуемых лиц к развитию ОНМК может в определенной мере способствовать уточнению и расширению контингента лиц, охваченных диспансерным наблюдением, благодаря выделению трех градаций (группы здоровых, умеренного и повышенного риска) вместо обычно рассматриваемых двух. Она достаточно проста для применения в амбулаторно-поликлинических учреждениях и может быть использована при организации рациональных профилактических мероприятий для больных с рассматриваемой патологией.

Научная группа по изучению сосудистой патологии головного мозга при ЦНИЛ Ереванского медицинского института

Поступила 15/XI 1977 г.

Յալ. Ս. ԹՈՒՆԱՆ, Ի. Մ. ՍԼԻՎՆԱԿ

ԵՐԿՇԵՄՔԱՅԻՆ ԽՆԻԻՐԸ ՈՒՂԵՂԻ ԱՐՅԱՆ ՇՐՋԱՆԱՌՈՒԹՅԱՆ ՍՈՒՐ ԽԱՆԳԱՐՈՒՄՆԵՐԻ ԿԱՆԽԱԳՈՒՇԱԿՄԱՆ ԺԱՄԱՆԱԿ

Հոդվածում քննարկվում է հետադոտման ենթակա անձի դասակարգումը հնարավոր 3 խմբերից մեկին. առողջների, ուղեղի արյան շրջանառության սուր խանգարման նկատմամբ շփոթվոր վտանգի և մեծացած վտանգի ենթակաների: Տրված են Երևան և Սոսկվա քաղաքների տղամարդկանց հետադոտման նյութերը: Առաջարկված մեթոդիկան կարող է որոշակի հիմք ծառայել ուղեղի արյան շրջանառության սուր խանգարումների ուսցիոնալ պրոֆիլակտիկայի կազմակերպման համար:

YU. S. TUNIAN, I. M. SLIVNIAK

BILIMINAL PROBLEM IN PROGNOSIS OF CEREBRAL INSULTS

Methods of determination of the degree of predisposition of examined persons to acute disorders of cerebral circulation are given for rational organization of prophylactic arrangements in ambulatory—polyclinical conditions.

Л И Т Е Р А Т У Р А

1. Айвазян С. А., Бежаева З. И., Староверов О. В. Классификация многомерных наблюдений. М., 1974.
2. Мастыкин А. С., Семак А. Е., Гришков Е. Г., Рябцева Т. Д., Крылова Е. А. Кибернетика в невропатологии и психиатрии (тезисы докладов респ. симп., июнь 1976 г.), Минск, 1976.
3. Смирнов Н. В., Большев Л. Н. Таблицы математической статистики М., 1965.
4. Урбах В. Ю. Статистический анализ в биологических и медицинских исследованиях. М., 1975.