

УДК: 519.7; 616-07

Оценка состояния здоровья популяции на основе энтропийного моделирования

Тырсин А.Н.^{*1}, Калев О.Ф.^{**2}, Яшин Д.А.^{***2}, Лебедева О.В.^{****3}

¹Уральский федеральный университет имени первого Президента России Б.Н. Ельцина,
Екатеринбург, 620002, Россия

²Южно-Уральский государственный медицинский университет,
Челябинск, 454092, Россия

³Челябинский государственный университет, Челябинск, 454001, Россия

Аннотация. В статье предпринята попытка на конкретном примере показать потенциальные возможности использования энтропийного моделирования многомерных стохастических систем в медицине. Проведена оценка состояния здоровья мужской сельской популяции по критериям «здоров», «практически здоров», «больной». Определены уровни энтропии параметров, характеризующих основные факторы риска хронических неинфекционных заболеваний, в зависимости от группы здоровья. В каждой группе выделен наибольший вклад отдельного параметра в энтропию популяции, определена степень зависимости этого параметра от других. Установлено, что с ухудшением состояния здоровья наблюдается рост энтропии в популяции. Этот рост вызван увеличением энтропии хаотичности.

Ключевые слова: энтропия, модель, многомерная случайная величина, дисперсия, индекс детерминации, фактор риска, популяция, неинфекционные заболевания, состояние здоровья.

ВВЕДЕНИЕ

Одной из мало изученных проблем в медицине является комплексная оценка популяционного здоровья одновременно по нескольким факторам риска, в их взаимосвязи. Это объясняется тем, что не ясно как учитывать вклад каждого фактора риска в общую оценку состояния здоровья. Обычно в таких случаях используются экспертные оценки, которые нельзя считать в полной мере объективными [1, 2]. Кроме того, биологические параметры людей характеризуются высокой вариабельностью, колеблемостью и, во многих случаях, негауссовским распределением, нелинейностью форм корреляционных зависимостей, что существенно усложняет проведение обычных процедур одномерного и многомерного статистического анализа.

Одним из перспективных альтернативных подходов для комплексного оценивания популяционного здоровья может оказаться энтропийное моделирование [3, 4]. Энтропия является фундаментальным свойством любых систем с неоднозначным, или вероятностным поведением [5]. Понятие энтропии является гибким и допускает четкую интерпретацию в терминах того раздела науки, где оно применяется. Оно все шире используется в современной науке для описания структурной организации и

*at2001@yandex.ru

**kalevang@mail.ru

***yashid.chel@mail.ru

****ya.olga.work@yandex.ru

дезорганизации, степени разрушения связей между элементами системы. Поэтому представляется, что энтропия может выступать в роли универсального параметра и идеально подходит для решения рассматриваемых задач о поведении сложных стохастических систем [6]. Любые объекты и явления живой и неживой природы без исключения содержат черты порядка и беспорядка, определенности и неопределенности, организованности и дезорганизованности.

Хронические неинфекционные заболевания (ХНИЗ) являются основными причинами смертности населения России. На их долю приходится 75% случаев смерти в общей структуре смертности [7]. Современная стратегия профилактики ХНИЗ основана на одновременном снижении приоритетных поведенческих (курение, пагубное потребление алкоголя, низкая физическая активность, нерациональное питание) и биологических факторов риска (артериальная гипертензия, дислипидемия, повышенный уровень глюкозы в крови, повышенная масса тела и ожирение) [7]. Особенностью биологических факторов риска является то, что они представляют собой параметры функционирования гомеостатических систем организма человека. Факторами риска развития ХНИЗ они становятся тогда, когда уровни их превышают значения, получившие значения «отрезных точек», которые ассоциируются с достоверным ростом смертности от разных причин и определяются в процессе эпидемиологических исследований. Это существенно снижает своевременность диагностирования ранних стадий заболеваний.

Задачей исследования является исследование изменения энтропии параметров, характеризующих биологические факторы риска ХНИЗ у мужчин с разными группами здоровья.

ОПИСАНИЕ ЭНТРОПИЙНОЙ МОДЕЛИ

Дифференциальная энтропия многомерных случайных величин

Дифференциальная энтропия многомерной непрерывной случайной величины $Y = (Y_1, Y_2, \dots, Y_m)$ была введена К. Шенноном в 1948 году [8]. Она определяется как

$$H(Y) = - \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f_Y(x_1, \dots, x_m) \ln f_Y(x_1, \dots, x_m) dx_1 \dots dx_m,$$

где $f_Y(x_1, \dots, x_m)$ – совместная плотность распределения случайных величин $Y_1, \dots, Y_m$.

Рассмотрим некоторые свойства дифференциальной энтропии.

Определение 1. Распределения вероятностей случайных величин X и Y называются *однотипными*, если существуют постоянные $a > 0$, b , такие что распределения случайных величин Y и $aX + b$ совпадают [9].

Из однотипности распределений вероятности случайных величин X и Y следует однотипность распределений вероятности случайных величин Y и X .

Определение 2. Параметр сдвига (в некоторых источниках его также называют параметром положения) – параметр μ , $\mu \in \Theta \subset R^k$, семейства функций $\{\varphi_\theta(\cdot)\}$, заданных на R^k по следующему правилу: $\varphi_\mu(\cdot) = \varphi(\cdot - \mu)$ для любого $\mu \in \Theta$, где $\varphi(\cdot)$ – некоторая заданная функция на R^k [9].

Определение 3. Параметр масштаба – один из параметров расположения, многообразием значений которых описывается семейство распределений вероятностей одного типа. Если распределение в R с функцией $F(x)$ принадлежит

тому же типу, что и фиксированное распределение с функцией $F_0(x)$, то $F(x) = F_0((x - \mu)/\lambda)$. Здесь $\lambda > 0$, μ – параметры масштаба и сдвига [10].

Пусть X_1, X_2 – две непрерывные случайные величины, описываемые однотипными законами распределения с плотностями $f_{X_1}(x) = f(x; \mu_1, \lambda_1)$, $f_{X_2}(x) = f(x; \mu_2, \lambda_2)$ соответственно, где μ_1, μ_2 и λ_1, λ_2 – параметры сдвига и масштаба. Очевидно, что

$$f(x; \mu_2, \lambda_2) = \frac{\lambda_1}{\lambda_2} f\left(\frac{\lambda_1}{\lambda_2}x - \frac{\lambda_1}{\lambda_2}\mu_2 + \mu_1; \mu_1, \lambda_1\right). \quad (1)$$

Отсюда можно показать, что разность дифференциальных энтропий случайных величин X_1 и X_2 равна [4]

$$H(X_2) - H(X_1) = \ln\left(\frac{\lambda_2}{\lambda_1}\right). \quad (2)$$

Если непрерывные случайные величины X_1 и X_2 имеют конечные дисперсии $D[X_1] = \sigma_1^2$ и $D[X_2] = \sigma_2^2$, то формулу (2) можно записать в виде

$$H(X_2) - H(X_1) = \ln \frac{\sigma_2}{\sigma_1}.$$

В [11] приведен ряд непрерывных законов распределений, имеющих параметры масштаба и сдвига. К ним относятся нормальное распределение, распределение Лапласа, распределение Коши, логистическое распределение, показательное распределение, равномерное распределение и др. Там же приведены трехпараметрические распределения, имеющие параметры сдвига, масштаба и формы: гамма-распределение, распределение Вейбулла–Гнеденко, распределение Рэлея, логнормальное распределение, параболическое распределение и др. Поскольку параметр формы в результате линейного преобразования не изменяется, то эти распределения также удовлетворяют соотношению (1). Следует отметить, что указанное множество распределений является конечным. Но оно включает в себя практически все непрерывные распределения, используемые при моделировании реальных случайных явлений.

Теорема 1 [4]. Пусть имеется две системы непрерывных случайных величин $\mathbf{Y}^{(1)} = (Y_1^{(1)}, Y_2^{(1)}, \dots, Y_m^{(1)})$ и $\mathbf{Y}^{(2)} = (Y_1^{(2)}, Y_2^{(2)}, \dots, Y_m^{(2)})$, каждые соответствующие компоненты которых $Y_i^{(1)}, Y_i^{(2)}$ ($i = 1, 2, \dots, m$) имеют конечные дисперсии и описываются однотипными распределениями с некоторыми параметрами положения и масштаба. Тогда разность энтропий систем случайных величин $\mathbf{Y}^{(2)}$ и $\mathbf{Y}^{(1)}$ равна:

$$\Delta H(\mathbf{Y}) = H(\mathbf{Y}^{(2)}) - H(\mathbf{Y}^{(1)}) = \sum_{k=1}^m \ln \frac{\sigma_{Y_k}^{(2)}}{\sigma_{Y_k}^{(1)}} + \frac{1}{2} \sum_{k=2}^m \ln \frac{1 - R_{Y_k^{(2)}/Y_1^{(2)} \dots Y_{k-1}^{(2)}}^2}{1 - R_{Y_k^{(1)}/Y_1^{(1)} \dots Y_{k-1}^{(1)}}^2}, \quad (3)$$

где $\sigma_{Y_k^{(j)}/Y_1^{(j)} \dots Y_{k-1}^{(j)}} = \sigma_{Y_k^{(j)}} \sqrt{1 - R_{Y_k^{(j)}/Y_1^{(j)} \dots Y_{k-1}^{(j)}}^2}$; $R_{Y_k^{(j)}/Y_1^{(j)} \dots Y_{k-1}^{(j)}}^2$ – индексы детерминации соответствующих регрессионных зависимостей, $k = 2, 3, \dots, m, j = 1, 2$.

Обозначив

$$\Delta H(\mathbf{Y})_\Sigma = \sum_{k=1}^m \ln \frac{\sigma_{Y_k}^{(2)}}{\sigma_{Y_k}^{(1)}}; \Delta H(\mathbf{Y})_R = \frac{1}{2} \sum_{k=2}^m \ln \frac{1 - R_{Y_k^{(2)}/Y_1^{(2)} \dots Y_{k-1}^{(2)}}^2}{1 - R_{Y_k^{(1)}/Y_1^{(1)} \dots Y_{k-1}^{(1)}}^2},$$

представим формулу (3) как

$$\Delta H(\mathbf{Y}) = \Delta H(\mathbf{Y})_{\Sigma} + \Delta H(\mathbf{Y})_R, \quad (4)$$

где $\Delta H(\mathbf{Y})_{\Sigma}$ и $\Delta H(\mathbf{Y})_R$ – приращения энтропии за счет изменения дисперсий и индексов детерминаций случайных величин $Y_1, Y_2, \dots, Y_m$.

Если случайный вектор $\mathbf{Y}$ является гауссовским, то

$$\Delta H(\mathbf{Y}) = \sum_{k=1}^m \ln \frac{\sigma_{Y_k^{(2)}}}{\sigma_{Y_k^{(1)}}} + \frac{1}{2} \ln \frac{|\mathbf{R}_{\mathbf{Y}^{(2)}}|}{|\mathbf{R}_{\mathbf{Y}^{(1)}}|},$$

где $|\mathbf{R}_{\mathbf{Y}^{(j)}}|$ – определитель корреляционной матрицы $\mathbf{R}_{\mathbf{Y}^{(j)}}$ случайного вектора $\mathbf{Y}^{(j)}$, $j = 1, 2$.

Теорема 2 [4]. Пусть имеем систему непрерывных случайных величин $\mathbf{Y} = (Y_1, Y_2, \dots, Y_m)$, компоненты которой имеют конечные дисперсии. Ее энтропия

$$H(\mathbf{Y}) = \sum_{k=1}^m H(Y_k) + \frac{1}{2} \sum_{k=2}^m \ln(1 - R_{Y_k/Y_1 Y_2 \dots Y_{k-1}}^2). \quad (5)$$

Формула (5) является обобщением частного случая для гауссовских случайных векторов на произвольные непрерывные распределения, причем форма корреляционных зависимостей может иметь как линейный, так и нелинейный вид.

В общем случае согласно (5) энтропия многомерной случайной величины складывается из двух составляющих:

$$H(\mathbf{Y}) = H(\mathbf{Y})_{\Sigma} + H(\mathbf{Y})_R, \quad (6)$$

$$\text{где } H(\mathbf{Y})_{\Sigma} = \sum_{k=1}^m H(Y_k); \quad H(\mathbf{Y})_R = \frac{1}{2} \sum_{k=2}^m \ln(1 - R_{Y_k/Y_1 Y_2 \dots Y_{k-1}}^2).$$

Отметим, что формула (5) позволяет упростить задачу определения энтропии многомерного случайного вектора, так как состоит из более простых по сравнению с нахождением его плотности вероятности операций.

Энтропийная модель многомерной стохастической системы

Суть энтропийного подхода в медицине состоит в представлении человека в виде сложной системы, которая характеризуется многомерной случайной величиной $\mathbf{Y} = (Y_1, Y_2, \dots, Y_m)$, где каждый элемент Y_i – это показатель функционирования одной из подсистем организма, например, фактор риска. Это позволяет рассматривать проблему состояния здоровья человека с позиции системного анализа. При этом каждый фактор риска анализируется комплексно. С одной стороны, он рассматривается отдельно, а с другой – во взаимосвязи с остальными факторами.

Будем считать, что данное представление является адекватной математической моделью системы S . Каждый элемент Y_i вектора $\mathbf{Y}$ является одномерной случайной величиной, которая характеризует функционирование соответствующего элемента исследуемой системы. Элементы могут быть как линейно, так и нелинейно коррелированными.

Аддитивные представления (6) и (4), как самой энтропии, так и ее изменения показывают энтропийный дуализм. Изменение энтропии происходит аддитивным образом: с одной стороны – за счет изменения дисперсий, а с другой стороны – из-за изменения индексов детерминации случайных величин $Y_1, Y_2, \dots, Y_m$ (рис. 1).

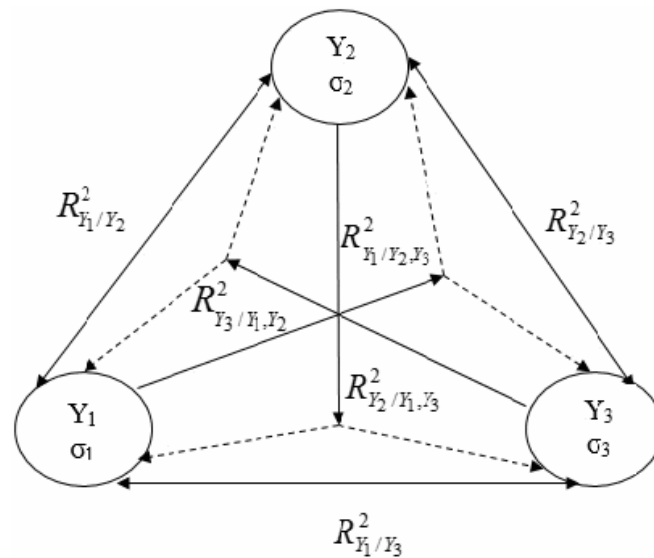


Рис. 1. Взаимозависимость энтропии случайного вектора.

Кроме того, формулы (6) и (4) имеют ясную интерпретацию через общесистемные закономерности аддитивности и целостности [12]. А именно, приращение $\Delta H(\mathbf{Y})_\Sigma$ определяет предельное изменение энтропии, соответствующее полной независимости элементов системы, и характеризует рассмотрение целостного объекта как состоящего из частей (аддитивность). Приращение $\Delta H(\mathbf{Y})_R$ отражает изменение степени взаимосвязей между элементами системы, характеризуя свойства системы как целого (целостность).

Таким образом, энтропийная модель (3), (5) объясняет причину энтропийного дуализма в развитии открытых стохастических систем. В [4] компоненту $H(\mathbf{Y})_\Sigma$ предложено условно называть энтропией хаотичности, а компоненту $H(\mathbf{Y})_R$ – энтропией самоорганизации. Следует заметить, что дуализм энтропии в том или ином виде отмечался в ряде публикаций [5, 13–15].

Предложенная энтропийная модель по сравнению с традиционным подходом, основанным на использовании информационной энтропии, обладает некоторыми существенными преимуществами. Во-первых, здесь не требуется оценивать вероятности пребывания многомерной системы в соответствующих состояниях. Это делает возможным ее применять на малых выборках данных.

Во-вторых, она легко реализуема на практике, поскольку требует проведения стандартных процедур оценивания дисперсий и индексов детерминации регрессионных моделей.

В-третьих, эта модель позволяет обнаружить причину изменения энтропии системы и оценить его количественно. Это делает возможным использование модели для различных задач. Например, к ним можно отнести рассмотренные в [3] оптимизационные задачи эффективного управления системами на основе увеличения или уменьшения ее энтропии, а также задачи диагностики и контроля состояния стохастических систем.

Для решения задач диагностики и контроля состояния системы используется представление энтропийной модели в виде (4). Будем считать, что две системы непрерывных случайных величин $\mathbf{Y}^{(1)}$ и $\mathbf{Y}^{(2)}$ соответствуют предыдущему и текущему периодам функционирования системы. Тогда, отслеживая изменение $\Delta H(\mathbf{Y})$ энтропии в целом и ее компонент $\Delta H(\mathbf{Y})_\Sigma$, $\Delta H(\mathbf{Y})_R$, можно сделать выводы о состоянии системы и происходящих в ней изменениях.

Оценку вклада произвольного l -го элемента в изменение энтропий хаотичности и самоорганизации находим через их предельные значения по формулам [4]

$$\Delta H(\mathbf{Y})_{\Sigma}^{(l)} = \ln \frac{\sigma_{Y_l^{(2)}}}{\sigma_{Y_l^{(1)}}}, \quad \Delta H(\mathbf{Y})_{\mathbf{R}}^{(l)} = \frac{1}{2} \ln \frac{1 - R_{Y_l^{(2)}/Y_1^{(2)} \dots Y_{l-1}^{(2)} Y_{l+1}^{(2)} \dots Y_m^{(2)}}^2}{1 - R_{Y_l^{(1)}/Y_1^{(1)} \dots Y_{l-1}^{(1)} Y_{l+1}^{(1)} \dots Y_m^{(1)}}^2}, \quad l = 1, 2, \dots, m.$$

Для диагностирования состояния многомерной стохастической системы будем придерживаться следующих этапов:

- 1) определение поведения системы (стабильная/нестабильная), поиск зависимостей поведения системы от времени, критических значений;
- 2) обнаружение характера изменения в системе («хаотичность» или «самоорганизация») в критических периодах;
- 3) на основании обнаруженной причины, проведение анализа: какой элемент системы (компонента системы $\mathbf{Y}$) оказался причиной изменения в системе;
- 4) формулирование вывода о влиянии изменения в системе с учетом выявленных критических моментов и причин их появления.

На этапе диагностики системы в некоторых случаях для интерпретации энтропийной модели полезно проводить анализ не только динамики энтропии (4), но ещё и статической энтропии (6). Это упрощает понимание процессов, происходящих в системе, так как интерпретация статической энтропии подтверждает интуитивное понимание системы и ее компонент.

Построение энтропийной модели по экспериментальным данным

Основная проблема использования моделей (4) и (6) состоит в оценивании индексов детерминации. Т.к. формы соответствующих регрессионных зависимостей не известны, поэтому необходимо использовать непараметрические методы.

Индекс детерминации показывает долю дисперсии результативной переменной Y , объясненной вариацией факторных переменных $X_1, \dots, X_m$, включённых в нелинейную модель регрессии [16]:

$$R_{Y/X_1 \dots X_m}^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (7)$$

где $\bar{y}$ – среднее значение результативной переменной, $\hat{y}_i$ – значения регрессии, y_i – фактические значения переменной Y .

Существует два основных подхода для построения непараметрической регрессии – это группировка и сглаживание данных.

Известные алгоритмы группировки данных [17] в многомерном случае, такие как FOREL, алгоритм k -средних, имеют ряд недостатков: 1) плохая применимость алгоритма при плохой разделимости выборки на кластеры; 2) необходимость априорных знаний о радиусе кластеров; 3) неустойчивость алгоритма (результат кластеризации сильно зависит от выбора начального объекта).

Известные методы сглаживания [18] (ядерное сглаживание, оценки k -ближайших соседей и др.) сталкиваются с проблемой выбора размера окрестности усреднения. Заниженное значение окрестности недостаточно сглаживает непараметрическую регрессию, что приводит к завышению индекса детерминации. А завышенное значение окрестности, наоборот, приводит к чрезмерному сглаживанию, при этом индекс детерминации занижается. Какого-либо универсального критерия выбора оптимальной окрестности сглаживания не предложено.

В [19] описан алгоритм, устраняющий данный недостаток. Он состоит в следующем. Пусть имеем многомерную выборку $(x_{i1}, \dots, x_{im}, y_i)$, $i = 1, \dots, n$. Формируем матрицу расстояний $\mathbf{D}$, элементы которой $d_{ij} = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2}$, $i, j = 1, \dots, n$.

Для каждого i -го наблюдения подбираем оптимальную выборку из L_i ближайших соседей согласно матрице $\mathbf{D}$ с номерами из соответствующего множества индексов $N(L_i) = \{i, i_1, i_2, \dots, i_{L_i-1}\}$ так, чтобы построенное уравнение линейной регрессии

$$\tilde{y}_l = a_{i0} + \sum_{k=1}^m a_{ik} x_{lk}, \quad l \in N(L_i) \quad (8)$$

имело минимальную дисперсию ошибок регрессии

$$s_{L_i}^2 = \frac{\sum_{l \in N(L_i)} (y_l - \tilde{y}_l)^2}{L_i - m - 1} \rightarrow \min_{3(m+1) < L_i \leq n}.$$

Для оптимальной локальной выборки из L_i ближайших соседей по формуле (8) формируем значение $\hat{y}_i$ непараметрической регрессии $\hat{y}_i = a_{i0} + \sum_{k=1}^m a_{ik} x_{ik}$.

По сформированному таким образом множеству значений $\hat{y}_i$, $i = 1, \dots, n$, по (7) найдем оценку индекса детерминации множественной регрессии.

Очевидно, что оптимальные размеры локальных выборок L_i в общем случае будут различными. Это позволит учесть изменение градиента теоретической функции регрессии в зависимости от значений факторных переменных и дисперсии случайной компоненты. Так как не использовалось никаких предположений о виде функции регрессии, то описанный алгоритм является непараметрическим.

Отметим, что оценку коэффициентов регрессии в (8) выполняют по-разному, в зависимости от особенностей исходных данных, например, можно использовать метод наименьших квадратов, метод наименьших модулей, а также робастные методы.

Энтропийная модель динамики, согласно теореме 1, может быть использована для данных, имеющих однотипные законы распределения. То есть изменения в системе не влекут за собой изменение типа закона распределения данных.

Чтобы на практике воспользоваться формулой (3), надо проверить соответствующие компоненты $Y_i^{(1)}$, $Y_i^{(2)}$ ($i = 1, 2, \dots, m$) на однотипность их эмпирических функций распределения $F_i^{(1)}(x)$, $F_i^{(2)}(x)$. Это можно сделать по-разному. Например, с помощью преобразования второй выборки путем приравнивания ее параметров масштаба и сдвига соответствующим параметрам первой выборки, и затем проверки статистической гипотезы однородности выборок $H_0: F_i^{(1)}(x) = F_i^{(2)}(x)$. При оценивании параметра сдвига (если он существует) будем использовать математическое ожидание, а при оценивании параметра масштаба – среднеквадратическое отклонение.

Для проверки статистической гипотезы однородности разработано много непараметрических критериев, наиболее распространенными среди которых являются двухвыборочные критерии Смирнова и омега-квадрат (Лемана–Розенבלата) [20].

При отклонении статистической гипотезы вместо модели (3) используем (5). При этом для каждой компоненты Y_k ($k = 1, 2, \dots, m$) энтропия вычисляется по формуле [8]

$$H(Y_k) = -h_k \sum_{j=1}^{L_k} p_j^{(k)} \ln p_j^{(k)},$$

где $p_j^{(k)} = n_j^{(k)} / n$ – оценка вероятности попадания в j -й интервал; L_k – количество интервалов; n_j^k – количество наблюдений, попавших в j -й интервал; n – объем выборки, h_k – длина интервала.

Расчет энтропии производился с помощью комплекса программ «Вычисление эмпирического корреляционного отношения», «Вычисление энтропии».

ЭКСПЕРИМЕНТАЛЬНЫЕ РЕЗУЛЬТАТЫ

Описание задачи исследования

Исследуем возможности энтропийного моделирования стохастических систем на примере анализа популяции в плане профилактики ХНИЗ по биологическим факторам риска. В качестве популяции рассмотрим группу мужчин с разными группами здоровья.

Статистический материал получен в результате комплексного сплошного углубленного клинико-эпидемиологического обследования мужской сельской популяции с гнездовой выборкой. Всего было обследовано 1402 мужчины одного из сел Челябинской области в возрасте от 18 до 85 лет, что составило 93% от списочного состава села. Для всех пациентов был проведен необходимый комплекс клинических, лабораторных и инструментальных методов обследования для квалифицированного заключения о состоянии здоровья. Работу проводила бригада специалистов, состоящая из сотрудников кафедры госпитальной терапии и семейной медицины Челябинской государственной медицинской академии и врачей Челябинской областной клинической больницы № 1.

Оценку и анализ состояния здоровья проводили в соответствии с детализированной классификацией групп для оценки состояния здоровья, предложенной А.Г. Марченко [21] по критериям здоров, практически здоров, больной:

I – здоровые:

а) не болели в период наблюдения острыми и хроническими заболеваниями, при осмотре не выявлено отклонений от нормы в состоянии здоровья;

б) болели редко острыми респираторными и др. острыми заболеваниями (не более 3 раз в течение года), при осмотре не выявлено отклонений от нормы в состоянии здоровья.

II – практически здоровые:

а) лица, имеющие такие факторы риска, как избыточный вес, гиперхолестеринемия (ГХС) (при отсутствии заболеваний, и если при обследовании не выявлено других отклонений в состоянии здоровья);

б) лица, которые часто (≥ 4 случаев в году) болели острыми заболеваниями, и при осмотре не выявлено отклонений от нормы в состоянии здоровья. К этой же группе относятся лица, имеющие одновременно такие факторы риска, как курение и низкая физическая активность;

в) лица, у которых установлены различные функциональные отклонения по результатам дополнительных исследований или симптомы без диагноза, предвестники болезни, состояния после перенесенных заболеваний и травм;

г) пациенты, у которых имеются заболевания, характеризующиеся потерей значимости при достижении определенного возраста, наличием длительного межприступного периода.

III – больные:

а) пациенты, у которых имеются хронические заболевания различной степени компенсации;

б) лица, имеющие физические недостатки и отдаленные последствия болезней и травм, характеризующиеся потерей трудоспособности или качества жизни.

ГХС устанавливалась при уровне общего холестерина крови (ОХС) не менее 5.17 ммоль/л, избыточный вес при уровне индекса массы тела 25–29.9, который определяли по формуле: масса тела (кг) / (рост (м))².

В качестве подсистем будем рассматривать показатели, характеризующие основные биологические измеримые факторы риска ХНИЗ. На начальном этапе энтропийного анализа необходимо определиться, какие параметры будут включены в расчет комплексной энтропии, чтобы элементы, которые не оказывают существенного влияния на поведение системы, не уменьшали удельный вес действительно значимых и имеющих прогностическое значение факторов. Цель исследования – анализ популяции в плане профилактики ХНИЗ, поэтому в анализ не включены данные иммунологических, биохимических и других лабораторных исследований. Для анализа популяционной энтропии были выбраны параметры, которые характеризуют факторы риска, которые имеют важное прогностическое значение и могут быть объективно и количественно измерены. К таким факторам отнесем уровень артериального давления, избыточную массу тела, высокий уровень глюкозы, дислипидемию.

Таким образом, сформировано четыре группы факторов риска ХНИЗ: артериальное давление (систолическое и диастолическое артериальное давление), масса тела (индекс массы тела), дислипидемия (уровень ОХС, триглицеридов, холестерина липопротеинов низкой плотности, холестерина липопротеинов высокой плотности), уровень глюкозы. Вследствие тесной корреляционной связи между систолическим и диастолическим артериальным давлением, а также между уровнями общего холестерина, холестерина липопротеинов низкой и высокой плотности, триглицеридов для дальнейшего энтропийного анализа были выделены четыре подсистемы «Уровень общего холестерина» (ОХС), «Систолическое артериальное давление» (САД), «Индекс массы тела» (ИМТ), «Уровень глюкозы» (Глюкоза).

Отметим, что на распространенность факторов риска и средние уровни параметров, их характеризующих, существенным образом влияет возраст. В частности, в обследованной нами популяции здоровых лиц в возрасте старше 34 лет их практически не оказалось. Поэтому для проведения анализа изменения популяционной энтропии в зависимости от статуса здоровья, в соответствии с рекомендациями для составления статистических таблиц для международного сравнения МКБ-10 были сформированы две возрастные группы: 18–24 лет, 25–34 лет.

Результаты исследований

Анализ энтропии параметров, характеризующих основные факторы риска ХНИЗ в группах здоровых, практически здоровых и больных был проведен у 593 человек.

Результаты анализа энтропии популяции в разных группах представлены в табл. 1.

Таблица 1. Уровни энтропии в группах здоровых, практически здоровых и больных лиц

Возраст, лет	Статус здоровья	Энтропия хаотичности	Энтропия самоорганизации	Энтропия общая
18–24	Здоровые	7.023	–0.391	6.63
	Практически здоровые	7.362	–0.174	7.19
	Больные	8.11	–0.34	7.77
25–34	Здоровые	5.71	–0.49	5.22
	Практически здоровые	7.77	–0.4	7.37
	Больные	8.49	–0.41	8.08

По мере ухудшения состояния здоровья популяции происходит увеличение общей популяционной энтропии, объединенной по всем факторам риска. Это заключение имеет практическое значение, поскольку, если у нескольких обследованных популяций

в какой-то одной будет выше уровень энтропии, то без проведения дорогостоящих и продолжительных исследований предварительно можно сказать о худшем состоянии здоровья данной группы и выделить ее для продолжения обследования.

Рост энтропии от группы «Здоровые» к группе «Практически здоровые» можно объяснить неблагоприятным воздействием таких патологических факторов как курение и низкая физическая активность, заболеваемость частыми простудными и другими острыми заболеваниями, наличием заболеваний в состоянии длительной ремиссии и др.

Энтропия в группе «Больные» выше, чем в группе «Практически здоровые» и значительно выше, чем в группе «Здоровые». Это можно объяснить тем, что к патологическому влиянию факторов риска на человеческий организм в отдельности и на всю популяцию в целом добавляется дополнительное повреждающее воздействие самих ХНИЗ.

Следует отметить, что энтропия у здоровых лиц в более молодой по возрасту группе (18–24 лет в сравнении с 25–34 лет) выше. По всей видимости, это происходит из-за того, что окончательный рост и формирование человеческого организма происходит до 25 лет и, поэтому у молодых лиц наблюдается больший разброс биологических показателей.

Энтропия самоорганизации в группе «Практически здоровые» выше, чем в группах «Здоровые» и «Больные» в обеих возрастных категориях. Это означает, что взаимодействие между подсистемами организма в группе «Практически здоровые» слабее, так как в этом состоянии организм, с одной стороны, оказывает меньшее сопротивление болезням в сравнении с группой «Здоровые». С другой стороны, пациенты в группе «Больные» уже имеют хронические заболевания, к которым организм приспособился, следовательно, взаимодействие между подсистемами в группе «Больные» сильнее, чем в группе «Практически здоровые».

Для более детального анализа в отдельности подсистемы «ОХС», «САД», «ИМТ», «Глюкоза» приведены результаты расчета энтропии хаотичности подсистем «ОХС», «САД», «ИМТ», «Глюкоза» (рис. 2 и 3) и результаты расчета индексов детерминации для регрессии каждого фактора риска на остальные факторы (рис. 4 и 5).

Далее рассмотрим динамику энтропии во всех возрастных группах при ухудшении состояния здоровья. Для того чтобы использовать энтропийную модель динамики (3), предварительно была проверена однотипность данных. Для каждой подсистемы была проведена проверка гипотезы однородности с помощью критерия Смирнова на уровне значимости $\alpha = 0.05$. Вероятность ошибочного отклонения нулевой гипотезы на первом этапе во всех статистических тестах $p > 0.1$, что значительно больше уровня значимости $\alpha = 0.05$. Следовательно, гипотеза H_0 не отклоняется.

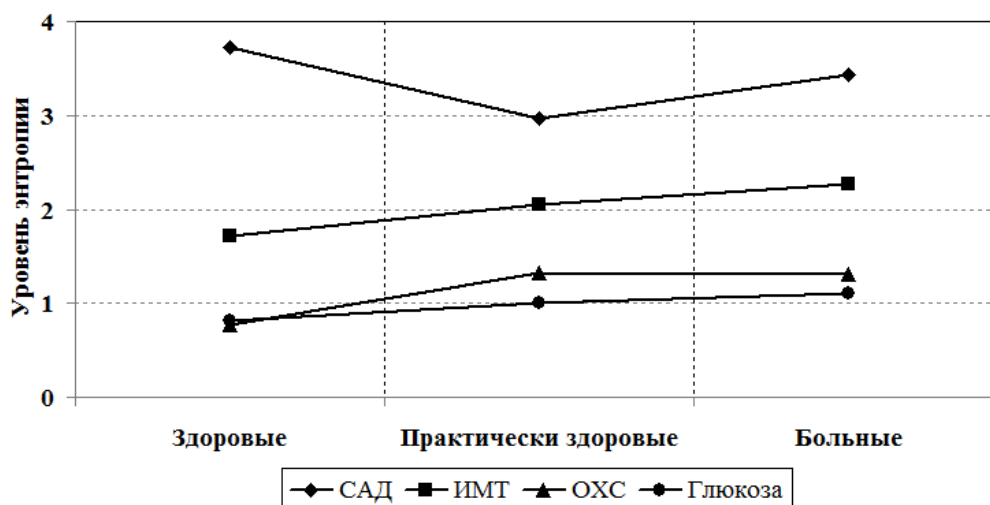


Рис. 2. Энтропия хаотичности по каждому фактору риска для возраста 18–24 лет.

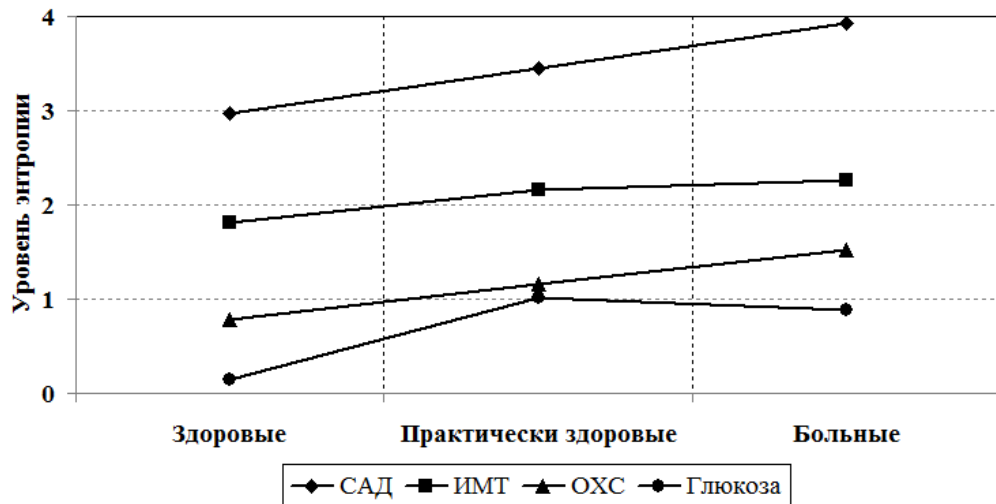


Рис. 3. Энтропия хаотичности по каждому фактору риска для возраста 25–34 лет.

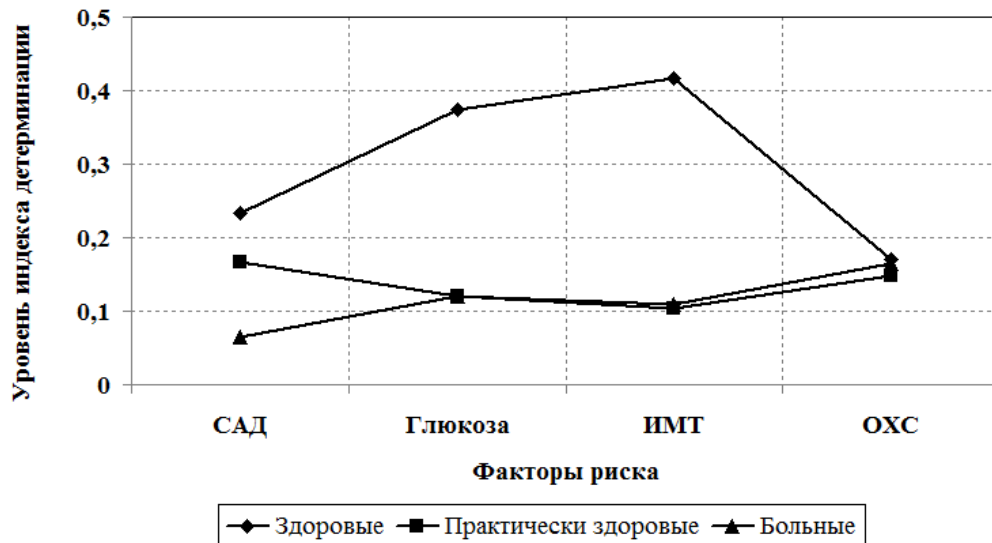


Рис. 4. Индексы детерминации для регрессии каждого фактора риска на остальные факторы для возраста 18–24 лет.

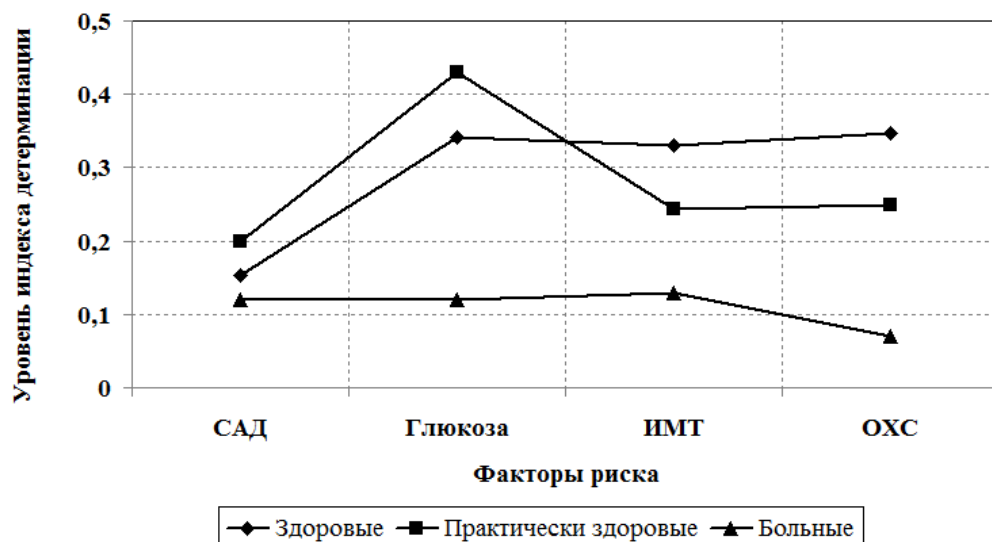


Рис. 5. Индексы детерминации для регрессии каждого фактора риска на остальные факторы для возраста 25–34 лет.

Максимальное увеличение энтропии наблюдается при переходе от здорового состояния к практически здоровому в возрастной группе 25–34 лет. Из табл. 2 видно, что в наибольшей степени на это изменение повлияла энтропия хаотичности, следовательно, увеличение энтропии произошло в результате увеличения степени разброса данных.

Таблица 2. Динамика энтропии при ухудшении состояния здоровья

Возраст, лет	Статус здоровья	Динамика энтропии хаотичности	Динамика энтропии самоорганизации	Динамика энтропии
18–24	«Здоровые» → «Практически здоровые»	0.339	0.217	0.56
	«Практически здоровые» → «Больные»	0.748	–0.166	0.58
25–34	«Здоровые» → «Практически здоровые»	2.06	0.09	2.15
	«Практически здоровые» → «Больные»	0.72	–0.01	0.71

Согласно табл. 2 видим, что наибольшая динамика наблюдается у энтропии хаотичности в возрастной категории 25–34 лет при переходе от здорового состояния к практически здоровому. Найдем предельные значения изменения энтропии хаотичности для каждого фактора риска: $\Delta H_{\Sigma}^{CAD} = 0.480$, $\Delta H_{\Sigma}^{ИМТ} = 0.339$, $\Delta H_{\Sigma}^{ОХС} = 0.364$, $\Delta H_{\Sigma}^{Глюкоза} = 0.874$. Следовательно, наибольшее влияние на повышение энтропии оказала подсистема «Глюкоза».

Таким образом, при анализе энтропии по основным измеримым модифицируемым ФР ХНИЗ (подсистемы артериальное давление, уровень общего холестерина, индекс массы тела, глюкоза крови) в комплексе с оценкой популяционного здоровья по критериям «здоров», «практически здоров», «больной» установлено:

1) у больных лиц в сравнении с практически здоровыми, а у практически здоровых в сравнении со здоровыми выше уровень популяционной энтропии;

2) взаимодействие между подсистемами (артериальное давление, уровень общего холестерина, индекс массы тела, глюкоза крови) наиболее слабо выражено в группе практически здоровых лиц;

3) максимальное увеличение популяционной энтропии происходит у возрастной категории 25–34 лет при переходе от здорового состояния к практически здоровому, в результате действия подсистемы «Глюкоза». Это говорит о том, что в этой возрастной группе необходимо особое внимание уделять уровню глюкозы в крови, так как изменение именно этого фактора риска ХНИЗ оказало наибольшее влияние на переход от здорового состояния к практически здоровому.

ЗАКЛЮЧЕНИЕ

В статье были рассмотрены возможности использования энтропийного моделирования стохастических систем в медицинской диагностике. В качестве задачи исследования был проведен анализ изменения энтропии параметров, характеризующих

биологические факторы риска хронических неинфекционных заболеваний у мужчин с разными группами здоровья.

Использование энтропийного анализа при профилактике хронических неинфекционных заболеваний в комплексе с оценкой популяционного здоровья позволяет определить наиболее критичное изменение состояния здоровья популяции и выяснить, какая подсистема организма человека повлияла на это изменение. Результаты исследования показали адекватность построенной энтропийной модели.

Представляет интерес развитие энтропийного моделирования для оценки изменения состояния здоровья индивидуальных пациентов.

Работа выполнена при финансовой поддержке Российского научного фонда, проект № 14-18-00574.

СПИСОК ЛИТЕРАТУРЫ

1. Кирьянов Б.Ф., Токмачёв М.С. *Математические модели в здравоохранении*. Великий Новгород: НовГУ им. Ярослава Мудрого, 2009. 279 с.
2. Цинкер М.Ю., Кирьяков Д.А., Камалтдинов М.Р. Применение комплексного индекса нарушения здоровья населения для оценки популяционного здоровья в Пермском крае. *Известия Самарского научного центра РАН*. 2013. Т. 15. № 3(6). С. 1988–1992.
3. Тырсин А.Н., Соколова И.С. Энтропийно-вероятностное моделирование гауссовских стохастических систем. *Математическое моделирование*. 2012. Т. 24. № 1. С. 88–103.
4. Тырсин А.Н., Ворфоломеева О.В. Исследование динамики многомерных стохастических систем на основе энтропийного моделирования. *Информатика и ее применения*. 2013. Т. 7. В. 4. С. 3–10.
5. Климонтович Ю.Л. *Введение в физику открытых систем*. М.: Янус-К, 2002. 284 с.
6. Timashev S.A., Tyrsin A.N. Entropy Approach to Risk-Analysis of Critical Infrastructures Systems. In: *Vulnerability, Uncertainty, and Risk. Analysis, Modeling, and Management: Proceedings of the ICVRAM 2011 and ISUMA 2011 Conferences*. Ed. Ayyub B.M. Reston: ASCE, 2011. P. 147–154.
7. *Профилактика хронических неинфекционных заболеваний. Рекомендации*. Москва, 2013. 128 с. URL: <http://www.webmed.irkutsk.ru/doc/pdf/prevent.pdf> (дата обращения: 03.04.2015).
8. Shannon C.E. A Mathematical Theory of Communication. *The Bell System Technical Journal*. 1948. V. 27. P. 379–423, 623–656.
9. *Математическая энциклопедия*. Под ред. Виноградова И.М. М.: Советская энциклопедия, 1984. Т. 4. 1216 с.
10. *Математическая энциклопедия*. Под ред. Виноградова И.М. М.: Советская энциклопедия. 1982. Т. 3. 1184 с.
11. Вадзинский Р.Н. *Справочник по вероятностным распределениям*. СПб.: Наука, 2001. 295 с.
12. *Теория систем и системный анализ в управлении организациями: справочник*. Под ред. Волковой В.Н., Емельянова А.А. М.: Финансы и статистика, 2006. 848 с.
13. Вильсон А.Д. *Энтропийные методы моделирования сложных систем*. М.: Наука, 1978. 248 с.
14. Pena D., Van der Linde A. Dimensionless measures of variability and dependence for multivariate continuous distributions. *Communications in Statistics: Theory and Methods*. 2007. V. 36. № 10. P. 1845–1854.

15. Николис Г., Пригожин И. *Самоорганизация в неравновесных системах: от диссипативных структур к упорядоченности через флуктуации*. М.: Мир, 1979. 512 с.
16. *Эконометрика*. Под ред. Елисеевой И.И. М.: Финансы и статистика, 2007. 576 с.
17. Загоруйко Н.Г. *Прикладные методы анализа данных и знаний*. Новосибирск: Изд-во Института математики, 1999. 270 с.
18. Хардле В. *Прикладная непараметрическая регрессия*. М.: Мир, 1993. 349 с.
19. Лебедева О.В., Тырсин А.Н. Построение энтропийной модели динамики многомерной стохастической системы по экспериментальным данным. В: *Актуальные вопросы технических наук в современных условиях: сборник научных трудов по итогам международной научно-практической конференции*. Выпуск II. СПб.: ИЦРОН, 2015. С. 53–60. URL: <http://www.izron.ru/upload/iblock/f7e/1..pdf> (дата обращения: 03.04.2015).
20. Орлов А.И. О проверке однородности двух независимых выборок. *Заводская лаборатория*. 2003. Т. 69. № 1. С. 55–60.
21. Марченко А.Г. Критерии оценки состояния здоровья населения при его комплексном изучении. *Советское здравоохранение*. 1979. № 2. С. 23–28.

Материал поступил в редакцию 27.04.2015, опубликован 08.06.2015.