С. С. Жихарев, Т. Л. Качанова, Н. Н. Ляшенко,
Г. И. Плеханова, Т. С. Разумовская, Т. М. Синицына,
В. И. Трофимов, Г. Б. Федосеев, Б. Ф. Фомин

МОДЕЛИ ДЛЯ ИССЛЕДОВАНИЯ БРОНХИАЛЬНОЙ АСТМЫ

1. Введение

Изучение бронхиальной астмы (БА) связано с проведением крупномасштабных исследований [7, 8], целью которых является: определение этиологии заболевания, выявление патогенетических механизмов БА, установление -их значимости в диагностике заболевания, апробация индивидуальной терапии ,в зависимости от характера внешних воздействий и особенностей реализующих систем на разных этапах болезни.

В целях получения данных о возможных этиопатогенетических вариантах БА и разработки концептуальной модели болезни были выполнены 546 комплексных исследований на 313 больных БА с определением при каждом исследовании до 1200 характеристик состояния больного.

В реализации программы исследований, получившей название Полного Исследовательского Комплекса (ПИК) БА, принимали участие 18 лабораторий семи различных медицинских институтов. В составе ПИК проводились исследования: общеклиническое, обычное лабораторное и рентгенологическое; цитологическое; микробиологическое; вирусологическое; микологическое; аллергологическое; иммунологическое (определение состояния клеточного и гуморального иммунитетов, определение циркулирующих в сыворотке крови бактериальных антигенов и антител); биохимическое (содержание биологически активных веществ в сыворотке крови); глюкокортикоидной активности надпочечников; адренергического дисбаланса на субклеточном, клеточном, органном и организменном уровнях; чувствительности и реактивности органов дыхания, выраженности бронхоспазма по реакции на фармакологические пробы; гемодинамики малого круга кровообращения; функционального состояния нервной системы с определением тонуса симпатического и парасимпатического отделов; психоневрологические; ЛОР-органов с определением ринобронхиального рефлекса по (реакции бронхов на механическое раздражение слизистой носа и ольфактивное раздражение.

2. Первичная обработка данных ПИК БА.

Начальный этап ПИК проводился в 1978—1983 гг. Множество всех данных за этот период можно условно представить в виде таблицы, имеющей 546 строк (количество выполненных исследований) и 1200 столбцов (число признаков состояния, определявшихся в каждом исследовании). В отдельных случаях полный набор признаков при выполнении ПИК получать не удавалось. Некоторые признаки по разным причинам вообще измерялись редко. В силу этого в таблице данных ПИК значительное число позиций оказались незаполненными.

Почти 900 признаков (74%) задавались на количественных шкалах, около 300 признаков — на порядковых и номинальных шкалах (19% и 7% соответственно). Нормальные распределения для большинства измеренных количественных признаков оказались нехарактерными. Практически для всех признаков было установлено наличие их взаимосвязей с другими признаками состояния. Линейные формы связей признаков, как правило, не подтверждались.

В ходе первичной обработки определялись: характеристики различных групп исследованных больных; параметры шкал признаков состояния, законы распределения; стандартные статистики (средние, дисперсии, квантили, коэффициенты корреляции); диаграммы рассеяния для пар признаков в специально подобранных группах исследований, парные и множественные регрессии, факторные нагрузки; статистики дисперсионного анализа; варианты кластеризации.

Число выборок при обработке данных ПИК БА варьировалось от нескольких десятков до сотен. При недостаточности данных некоторые качественные различия групп намеренно игнорировались.

В результате были получены семантические статистические структуры и интегрированные статические модели данных ПИК БА. Первичная обработка экспериментального материала позволила получить сведения о всевозможных группах больных БА в соответствии с запросами экспертов, при этом программные средства не накладывали каких-либо ограничений на характер запросов [2, 5].

3. Построение решающих правил индивидуальной диагностики БА

Исходными данными для дальнейших исследований были:

1) Число выделенных экспертами ПИК диагностических классов.
2) Разбиение всех 546 исследований ПИК по диагностическим классам.
3) Данные ПИК, применявшиеся в качестве обучающейся выборки.

Результаты каждого очередного завершившегося исследования рассматривались на семинаре ПИК. При постановке диагноза эксперты руководствовались эвристическими приемами. Они находили группы характерных признаков, свидетельствующих о наличии какого-то определенного клинико-этиопатогенетического механизма БА и на их основании разрабатывали модели-эталоны. В зависимости от количества совпадающих с моделью признаков больного включали в соответствующий диагностический класс.

Наши исследования, построенные не всестороннем анализе данных ПИК, должны были дать формальные подтверждения или же, напротив, опровержения предложенной семинаром экспертов классификации вариантов БА и привести в итоге к созданию эффективной автоматизированной процедуры индивидуальной диагностики больных астмой по результатам исследования их состояния.

Решающие правила индивидуальной диагностики больных БА создавались в три этапа: формирование пространства диагностических признаков, понижение размерности пространства диагностических признаков, разработка автоматизированной процедуры классификации состояния больных БА.

Формирование пространства диагностических признаков. Ценность признаков состояния для диагностики БА определялась с помощью информационных мер связи [3]. В качестве меры связи между признаком $\boldsymbol{x}$ из набора $\boldsymbol{X}$ признаков ПИК БА и диагностическим классом $\boldsymbol{y}$ из заданного множества $\boldsymbol{Y}$ диагнозов БА использовалась мера:

$M(x,y) = 1 - \frac{H(x,y)}{H(x)+H(y)}$, где

$H(x) = -\sum_{i \geq 1} p_{i*} \log_2 p_{i*}$ - энтропия $\boldsymbol{x}$, $H(y) = -\sum_{j \geq 1} p_{*,j} \log_2 p_{*,j}$, энтропия $\boldsymbol{y}$,

$H(x,y) = -\sum_{i \geq 1} \sum_{j \geq 1} p_{ij} \log_2 p_{ij}$ - энтропия совместного распределения $\boldsymbol{x}$ и $\boldsymbol{y}$.

В качестве оценок вероятностей p_{i*}, p_{*j}, p_{ij} принимались отношения частностей n_{i*}/n, n_{*j}/n, $n_{i,j}/n$, взятые из таблиц сопряженности (табл. 1).

Таблица 1

Сопряженности признака *x* с диагностическим классом *y*

Номер группы значений признака	Диагностические классы БА			Частотность *x* для *i*-й группы значений
	Аспириновая (1)	…	инфекция смешанная (7)	
1	n_{11}	…	n_{17}	n_{1*}
2	n_{21}	…	n_{27}	n_{2*}
…	…	…	…	…
r	n_{r1}	…	n_{r7}	n_{r*}
Частотность x для диагноза y	n_{*1}	…	n_{*7}	*n*

В случаях, когда x и y независимы, $M(x, y) = 0$. Условию полной связи x и y отвечает $M(x, y) = 0{,}5$. Найденные значения мер связи показали низкую информативность признаков для диагностики. Значения $M(x,y)$ изменялись в пределах от 0,010 до 0,100.

Селекция связей по значениям $M(x,y)$ на данном этапе была бы неоправданной. Практический опыт экспертов свидетельствовал о значимых взаимосвязях признаков и диагнозов из разработанной ими классификации. Оставалось предположить, что низкие значения меры $M(x,y)$ обусловлены влияниями факторов, не имеющих непосредственных связей с диагнозами БА, и большим общим уровнем шума, характерным для данных ПИК.

Для усиления информативности признаков применялся специально разработанный алгоритм фильтрации данных. По этому алгоритму каждая таблица сопряженности (см. табл. 1) приводилась к матричному представлению вида $P = \varepsilon C + (1 - \varepsilon)H$, где C и H означали соответственно зависимую и независимую от диагноза части матрицы P, $\varepsilon \in [0, 1]$. Алгоритм находил такие C и H, при которых значение ε становилось минимальным. Величина ε характеризовала отсев данных. В наших расчетах ε изменялась от 0,2 до 0,99. Признаки, для которых удовлетворялось неравенство $(1 - \varepsilon)n \geq N$ (N — минимальный допустимый объем выборки) в дальнейшем игнорировались.

Алгоритм фильтрации.

1. $k = 0,\ P = n_{ij}/n,\ C = (c_{ij}),\ H = (h_{ij}),\ i = \overline{1,r},\ j = \overline{1,s}$;
2. $C^0 = P$;
3. $C^{k+1} = A^T B,\ A = (a_{i^*}),\ B = (b_{*j})$;
4. $|c_{ij}^{k+1} - c_{ij}^{k}| < 1/n,\ i = \overline{1,r},\ j = \overline{1,s};\ k \leq 50$; на *7*.
5. $c_{ij}^{k+1} = \lambda \max(c_{ij}^{k}, a_{i^*}, b_{*j}) + (1 - \lambda)\min(c_{ij}^{k}, a_{i^*}, b_{*j})\ \ \lambda \in \{0;\ 0{,}25;\ 0{,}5\},\ i = \overline{1,r},\ j = \overline{1,s}$;
6. $c_{ij}^{k+1} = c_{ij}^{k+1} / \sum_{i,j} c_{ij}^{k+1},\ i = \overline{1,r},\ j = \overline{1,s}$; на 3.
7. $(p_{ij} = 0) \rightarrow (c_{ij}^{k+1} = 0),\ \varepsilon = \min_{i,j}(p_{ij} / c_{ij}^{k+1}),\ i = \overline{1,r},\ j = \overline{1,s}$;
8. $\overline{H} = P - \varepsilon C^{k+1},\ \sum_{i,j} \overline{h_{ij}} = 1,\ h_{ij} = (\overline{h_{ij}} / \sum_{i,j} \overline{h_{ij}}),\ i = \overline{1,r},\ j = \overline{1,s}$;
9. Конец

Данный алгоритм применялся для обработки таблиц сопряженности каждого из 1200 имеющихся признаков состояния (см. табл. 1). В итоге для последующих вычислений были получены зависимые от диагноза составляющие H.

Опыт использования алгоритма фильтрации показал его хорошую сходимость. Экспериментально было установлено, что наилучшей сходимости алгоритма отвечает значение $\lambda = 0{,}25$.

Понижение размерности пространства диагностических признаков. При выполнении ПИК каждый из 1200 признаков состояния определялся в какой-то одной лаборатории по специальной функционально-диагностической методике. С учетом данного обстоятельства все множество признаков состояния было разбито на непересекающиеся классы, в состав которых входили признаки, полученные по одной и той же методике исследований. На этой основе все полученные после применения алгоритма фильтрации матрицы H были рассортированы приблизительно на 40 групп по принадлежности определяющих их признаков к тем или иным методикам исследований. Работая далее отдельно с каждой такой группой матриц, нам удалось обойти ограничения на вычислительные ресурсы, естественным образом возникающие в случае обработки несгруппированных данных.

По каждой матрице H, входящей в какую-то определенную группу, вычислялась мера $M(x,y)$ связи признака состояния x с диагнозом y и оценивалась достоверность найденного значения меры. Экспериментальным путем были выявлены чувствительность меры к объему выборки n, изменению числа групп значений r признака x, характеру распределения частотности n_{ij} в таблицах сопряженности. В ходе вычислительных экспериментов выяснилось, что наиболее часто применяемая для оценки значимости информационной меры статистика [1] в нашем случае давала необоснованно завышенные значения.

Объяснением этому могли быть нарушения некоторых базовых предположений, влияющих на качество оценок достоверности. Таковыми могли, например, быть требования, предъявляемые к матрицам $\boldsymbol{H}$, а именно, $\boldsymbol{h_{ij} \geq 5}$, для всех $\boldsymbol{(i, j)}$, большие значения $\boldsymbol{r}$ и $\boldsymbol{s}$, большие объемы выборок $\boldsymbol{n}$. В действительности матрицы $\boldsymbol{H}$ имели нули, многие ненулевые их элементы имели значения, не превышающие 5, число столбцов у всех матриц $\boldsymbol{H}$ было постоянным и равнялось 7, почти во всех случаях объемы выборок не превосходили 100, а число строк изменялось от 2 до 10.

Достаточно хорошими получились оценки достоверности меры связи при использовании доверительного интервала:

$$\left[M_{max} = 1 - \frac{H(x,y)_{min}}{H(x)_{max} + H(y)_{max}}, M_{min} = 1 - \frac{H(x,y)_{max}}{H(x)_{min} + H(y)_{min}} \right],$$

где $\boldsymbol{H(x)_{max}}$, $\boldsymbol{H(y)_{max}}$, $\boldsymbol{H(x,y)_{max}}$, $\boldsymbol{H(x)_{min}}$, $\boldsymbol{H(y)_{min}}$, $\boldsymbol{H(x,y)_{min}}$ – означают максимальные и минимальные оценки энтропии. Для вычисления этих оценок использовались доверительные интервалы вероятностей $\boldsymbol{p_{i*}}$, $\boldsymbol{p_{*j}}$, $\boldsymbol{p_{ij}}$ взаимосвязи рассматриваемых признаков и диагнозов БА. В формальной записи такой доверительный интервал имел следующий вид [6]:

$$\left[max\left\{ p - t\sqrt{\frac{p(1-p)}{n}}, 0 \right\}; min\left\{ p + t\sqrt{\frac{p(1-p)}{n}}, 1 \right\} \right].$$

В приведенном выражении $\boldsymbol{p}$ означает вероятность взаимосвязи, $\boldsymbol{n}$ характеризует объем выборки, $\boldsymbol{t}$ задает коэффициент, соответствующий квантилям нормального распределения.

В результате обработки большого числа таблиц сопряженности была установлена возможность применения для оценки значимости меры связи критерия $\boldsymbol{M(x,y) \geq \alpha M_{min}}$. Все признаки состояния $\boldsymbol{x}$, не удовлетворяющие этому критерию, исключались из дальнейшей обработки. Экспериментально подтверждалось, что коэффициент $\boldsymbol{\alpha}$ в данном критерии должен иметь значение, равное 0,05. Оценки значимости меры $\boldsymbol{M(x,y)}$ получались при этом наилучшими.

В итоге для каждой группы матриц $\boldsymbol{H}$ были построены списки признаков состояния, имеющих значимые взаимосвязи с диагностическими классами БА. После объединения всех таких списков (общим числом около 40) образовался полный набор информативных признаков состояния. Число признаков в этом наборе не превышало 300.

Найденные признаки вполне могли оказаться взаимосвязанными. Требовалось отобрать их полное независимое подмножество. Для этой цели все признаки были упорядочены по убыванию значений меры $\boldsymbol{M(x,y)}$. Первый в списке признак, имеющий наибольшее значение меры, назначался «первым лидером» $\boldsymbol{x_л^1}$. Далее вычислялись значения меры взаимосвязи $\boldsymbol{M(x_л^1,x)}$ между $\boldsymbol{x_л^1}$ и каждым другим признаком $\boldsymbol{x}$ из рассматриваемого списка. Поскольку $\boldsymbol{x_л^1}$ — лидер, то $\boldsymbol{M(x_л^1,y) > M(x, \bar{y})}$, (здесь $\boldsymbol{y}$ и $\boldsymbol{\bar{y}}$ означают разные в общем случае диагностические классы). Если $\boldsymbol{M(x_л^1,x) \geq \beta\, M(x, \bar{y})}$, то признаки $\boldsymbol{x_л^1}$ и $\boldsymbol{x}$ считались взаимосвязанными, вследствие чего признак $\boldsymbol{x}$ исключался из списка информативных диагностических признаков состояния (коэффициент $\boldsymbol{\beta}$ определялся из эксперимента). Таким же в точности образом находились и исключались из списка все остальные взаимосвязанные с $\boldsymbol{x_л^1}$ признаки. После этого в качестве «второго лидера» $\boldsymbol{x_л^2}$ назначался ближайший в списке к $\boldsymbol{x_л^1}$ признак, и вся только что описанная для $\boldsymbol{x_л^1}$ процедура повторялась для $\boldsymbol{x_л^2}$. Процесс назначения лидеров, поиска и удаления из списка всех взаимосвязанных с ними признаков продолжался до тех пор, пока не были исчерпаны все возможности для выбора из числа оставшихся признаков очередного нового лидера. Список, состоящий из одних только лидеров $\boldsymbol{x_л^1}, \boldsymbol{x_л^2}, \ldots, \boldsymbol{x_л^m}$ рассматривался далее как полный набор независимых информативных диагностических признаков. Всего было отобрано 80 независимых диагностических признаков.

Автоматизированная процедура классификации состояний больных БА. Разработанная нами процедура классификации основана на байесовской модели [9]. В качестве исходных данных в этой процедуре использовались априорные вероятности диагностиче-

ских классов БА $p(y = j)$, $j \in \{1,\ldots,7\}$ и условные вероятности распределения множества применяемых при диагностике состояний больных астмой независимых признаков:

$$p(x_л^1,\ldots,x_л^m \mid y = j) = \prod_{l=1}^{m} p(x_л^l \mid y = j), \quad j \in \{1,\ldots,7\}.$$

В качестве оценок вероятностей $p(y = j)$ принимались отношения N_j / N, характеризующие частоты встречаемости в базе данных ПИК исследований, соответствующих по заключению семинара экспертов j-му варианту БА (N здесь равно 546, т. е. полному объему выполненных в рамках ПИК исследований). Оценки вероятностей $p(x_л^l \mid y = j)$ вычислялись по данным таблиц сопряженности как отношения p^l_{ij} / p^l_{*j}, $i = \overline{1,r}$; $j = \overline{1,7}$; $l = \overline{1,m}$, где p^l_{ij} – вероятность того, что признак $x_л^l$ получит значение из i – й группы, если был зафиксирован j – й вариант БА; $p^l_{*j} = \sum_i p^l_{ij}$ - вероятность взаимосвязи признака $x_л^l$ с j – м диагнозом.

Байесовская модель позволила найти условные вероятности каждого из возможных вариантов диагноза для заданных наборов значений диагностических признаков:

$$p(y = j \mid x^1_{iл},\ldots,x^m_{iл}) = \frac{\prod_{l=1}^{m} \frac{p^l_{ij}}{p^l_{*j}} p(y = j)}{\sum_{v=1}^{7} \prod_{l=1}^{m} \frac{p^l_{ij}}{p^l_{*j}} p(y = v)}$$

Выбор диагноза осуществлялся по максимальной условной вероятности:

$$\max_{j \in \{1,\ldots 7\}} (p(y = j \mid x^1_{iл},\ldots,x^m_{iл})) .$$

Проверка достоверности такого решения проводилась с помощью обучающей выборки из базы данных ПИК. Данные ПИК разбивались на группы по диагнозам, поставленным экспертами. Таким образом были построены 7 диагностических групп исследований, в каждой из которых отдельно взятые исследования задавались полными наборами значений независимых информативных признаков состояния и относились к одному и тому же диагнозу. Выполненная экспертами диагностика считалась точной.

Наборы данных каждой группы подвергались обработке с помощью байесовской процедуры классификации. В результате всякому исследованию были поставлены в соответствие условные вероятности диагностических классов, в том числе наиболее вероятный диагноз и условная вероятность класса, истинного для данного исследования. По итогам всех экспериментов (общим числом 546) был построен доверительный интервал $[A, B]$, где:

$$A = \min \frac{p(y = v)}{p(y = w)}, \quad B = \max \frac{p(y = v)}{p(y = w)}, \quad v, w \in \{1,\ldots,7\},$$

а $p(y = v)$ и $p(y = w)$ означают соответственно условные вероятности диагностических классов v и w, вычисленные по байесовской модели. При этом считались выполненными следующие условия:

- для левой границы - поставлен диагноз w, тогда как на самом деле имел место диагноз v;
- для правой границы - поставлен диагноз v, в то время как фактически имел место диагноз w.

Процесс автоматической классификации состояний больных БА завершался процедурой сравнения максимальной и ближайших к ней значений условных вероятностей классов. Если:

$$\frac{\max_{j \in \{1,\ldots 7\}} (p(y = j \mid x^1_{iл},\ldots,x^m_{iл}))}{p(y = j \mid x^1_{iл},\ldots,x^m_{iл})} \in [A, B],$$

то оба диагноза, удовлетворяющие этому условию, являются конкурирующими.

Упрощенный вариант диагностической процедуры, разработанный применительно к 25 признакам состояния, значения которых доступны для определения в условиях обычных поликлиник, обеспечивал точность диагностики в пределах 70 — 80%.

Автоматизированная процедура индивидуальной диагностики по существу подтвердила только тот факт, что при постановке диагноза эксперты ПИК руководствовались неко-

торыми объективными непротиворечивыми приемами анализа симптомокомплексов, в которых в качестве исходных данных были задействованы значения измеряемых признаков состояния больных БА.

Широкое применение разработанных диагностических программ в практическом здравоохранении обосновано только в тех случаях, когда твердо установлены следующие факты:

- выделенные экспертами диагностические классы могут использоваться для характеризации в действительности различных форм течения БА;
- набор выделенных диагностических классов полный;
- схемы классификации всех возможных вариантов болезни аналогичны применявшимся экспертами ПИК схемам классификации индивидуальных наборов признаков состояния;
- данные исследований ПИК в действительности характеризуют состояние больных БА и являются определяющими при диагностике астмы.

Уточнение и исследование этих вопросов прежде всего связано с решением задач анализа структур причинных связей между признаками состояния.

4. Концептуальные модели структур причинных связей признаков состояния больных БА

При построении концептуальных моделей причинных связей с помощью экспертов отбирались системы представителей признаков состояния, которые согласно современным научным представлениям о природе возникновения, развития и течения БА должны оказывать сильное влияние на эндокринную, иммунную и нервную системы организма больного. Эксперты выделили 300 из 1200 признаков, на множестве которых требовалось выполнить структурный анализ причинных взаимосвязей основных внутрисистемных механизмов, ответственных за изменение реактивности бронхов.

Для каждой пары отобранных признаков вычислялись оценки степени тесноты связей. В результате была построена матрица таких оценок, каждая парная связь в которой описывалась значением меры информационной связи и ее доверительным интервалом. Элементы матрицы далее подвергались сортировке для обнуления тех ее элементов, которыми задавались слабые связи. Степень тесноты связи сравнивалась при этом с пороговым значением. Выбор пороговой величины осуществлялся на основе сопоставления значений меры связи с нижней границей ее доверительного интервала.

Для отобранных пар признаков с сильными связями, был получен граф связей, который предъявлялся экспертам для тщательного анализа. В виду больших размеров графа (300 вершин) анализу подвергались его различные частичные подграфы. Обсуждение графа связей на теоретическом семинаре экспертов завершилось выводом о возможности его использования в качестве первичной модели искомой структуры причинных связей признаков состояния. Используя этот граф, эксперты разработали описания подмножеств признаков, которые далее могли применяться в качестве носителей моделей основных физиологических механизмов, характерных при возникновении, течении и развитии БА. Например, на рис. 1 показан граф связей признаков, отображающий структуру модели влияния глюкокортикоидных гормонов на функцию внешнего дыхания.

Проверка подобных структур взаимосвязей могла бы быть построена на анализе законов преобразования веществ в организме человека. Однако эксперты не располагали необходимыми достоверными количественными описаниями процессов поведения организма в норме и патологии. Обсуждение всех зафиксированных связей проводилось экспертами исключительно на качественном уровне. Свои суждения эксперты чаще всего выражали в форме высказываний: «***X*** угнетает ***У***», «***X*** усиливает действие ***У***» и т. п. В таких условиях была опробована следующая схема оценки непротиворечивости структур связей признаков в отдельно исследуемых подграфах полного графа связей:

- эксперты называли признаки, которые в рассматриваемой структуре могли играть роль входных элементов;

- эксперты отмечали на графе возможные с их точки зрения направления действия и знаки каждой причинной связи. Кроме того, они описывали табличным или графическим способами известные им из клинической практики (либо теоретически допустимые) статические модели взаимосвязей типа «вход-выход»;
- аналогичные модели для каждой связи были построены по данным ПИК на основе методов линейной, полиномиальной и множественной регрессий;

По всем этим данным для каждого рассматриваемого подграфа составлялись алгебраические нелинейные уравнения взаимосвязей признаков. Например, для графа на рис. 1 одна из составленных нами систем уравнений имела вид:

- $КНГ = F_1(ИГА) = 5—0{,}1(0{,}1ИГА—12)^2$;
- $БИК = F_2(КНГ) = 1{,}6—0{,}1\ (КНГ—4)^2$;
- $ПК = F_3(БИК) = 10\sqrt{БИК}$;
- $ИГА = F_4(ПК, СВОКС) = 0{,}65(350/(ПК + 0{,}5) + 50) + 0{,}35(140/(СВОКС—0{,}5)+50)$;
- $ИПРА = F_5(КНГ) = 3{,}5—0{,}2(КНГ—4)^2$;
- $СТ = F_6(ИПРА) = 0{,}89/(ИПРА+1{,}33)—0{,}17$;
- $ОИУК = F_7(СТ, СВОКС) = 0{,}85(3{,}5СТ^2+0{,}05) +0{,}15(3—0{,}49СВОКС)$;
- $DL = F_8(VEMAX) = 35—0{,}25\ (VEMA\ X—\ 10)^2$;
- $VEMAX = F_9(ООЛ, СВОКС) = 0{,}85\ (0{,}75+360/(ООЛ—60)) + 0{,}15(1{,}87СВОКС—1)$;
- $DLVAK=F_{10}(DL, СТ, ЦЛАГО) = 0{,}60(0{,}2\ DL — 0{,}43) + 0{,}19(0{,}25/(СТ—0{,}01) +2) +0{,}21\ (4{,}5/ЦЛАГО+2)$;
- $ООЛ = F_{11}(КНГ)=0{,}9ООЛ+0{,}1(165/КНГ+150)$;
- $ВГО = F_{12}(ООЛ, ОИУК) = 0{,}9(12{,}4ООЛ + 1247) + 0{,}1(3666 + +254ОИУК—10ОИУК^2)$;
- $ЦЛАГС = F_{13}(БИК) = 0{,}82+0{,}62БИК$;
- $ЦЛАГО = F_{14}(ЦЛАГС) = 0{,}3+0{,}95ЦЛАГС$.

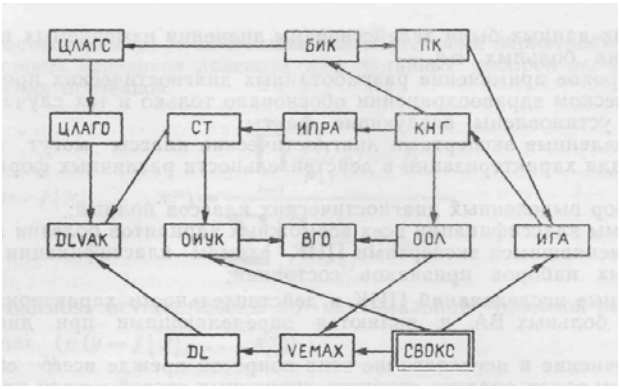


Рис. 1. Модель гормональной регуляции системы внешнего дыхания:

ЦЛАГС — циркулирующий антиген легочной ткани (свободный), БИК — быстродействующий ингибитор каллекреина, ПК — прекаллекреин, ЦЛАГО — циркулирующий антиген легочной ткани (общий), СТ — серотонин тромбоцитарный, ИПРА — исходная протоминрасщепляющая активность, КНГ — кининоген, DLVAK — отношение диффузионной способности к альвеолярной вентиляции по крови, ОИУК — 5 оксинин долуксусная кислота, ВГО — внутригрудной объем, ООЛ — остаточный объем легких; ИГА — иммуноглоб>лин А, DL—диффузионная способность, VEMAX — максимальный поток, СВОКС—свободные 11—ОКС

Решение системы уравнений подобного вида проводится итерационными методами [4]. Для рассмотренного примера начальные значения выбирались из базы данных ПИК. Переменная ***СВОКС***, представленная на графе (рис. 1) вершиной без входных дуг, использовалась в качестве входа модели. Ее изменение описывалось в виде линейной убывающей зависимости. Значения остальных переменных последовательно пересчитывались. Помимо значений переменных на каждом шаге вычислялись разности первого и второго порядков, а также их сигнум-функции. Результаты выводились в форме графиков (см. рис. 2) и передавались экспертам для интерпретации. Прогоны моделей осуществлялись при разных вариантах структур, направлений, знаков и функциональных описаний связей, подсказанных экспертами. Попытки интерпретации полученных из экспериментов данных в большинстве случаев успеха не имели.

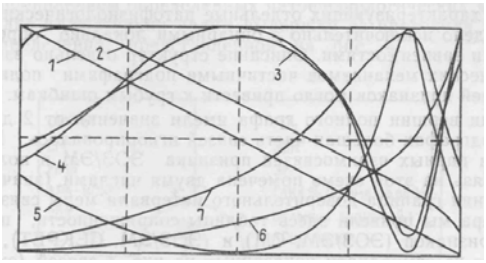


Рис. 2. Решение системы уравнений модели при изменении параметра ***СВОКС***:
1 - ***СВОКС***; 2 - ***DL***; 3 - ***БИК***; 4 - ***DLVAK***; 5 - ***СТ***; 6 - ***ИГА***; 7 - ***КНГ***.

Заметим, что в рассуждениях о структурах взаимосвязей признаков состояния, характеризующих отдельные патофизиологические процессы, мы имели дело исключительно с бинарными локально определенными причинными зависимостями. Описание структур отдельно взятых патофизиологических механизмов частичными подграфами полного графа взаимосвязей признаков могло привести к грубым ошибкам.

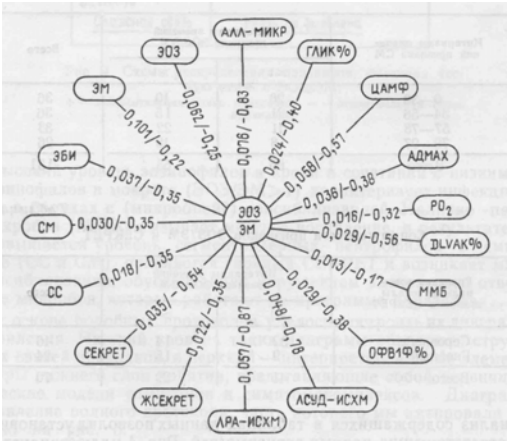


Рис. 3. Схема парных взаимосвязей признака ЭОЗ/ЭМ с остальными признаками, представленными в графе взаимосвязей:
AR.MAX — максимальное артериальное давление; АЛЛ-МИКР — аллергены микробного происхождения; ГЛИК% — гликогенолиз; ЖСЕКРЕТ — характер мокроты; ЛРА-ИСХМ — бронхиальное сопротивление (ринобронхиальный рефлекс); ЛСУД-ИСХМ — удельная проводимость (ринобронхи-альный рефлекс); 1ММБ — патогенные микробы; ОФВ1Ф% — изменение объема форсированного выдоха за одну секунду после адреналина (в %); СЕКРЕТ — характер секрета (бронхоскопия); СМ — сегментоядерные нейтрофилы (мокрота); СС — сегментоядерные нейтрофилы (бронхосмыв); ЦАМФ — циклический аденозинмонофосфат; ЭБИ — клетки эпителия бронхов измененные (бронхосмыв); ЭМ — эозинофилы (мокрота); ЭОЗ — эозинофилы (кровь); DLVAK% — изменение удельной диффузионной способности по крови после бронхолитика (в %); PO2 — парциальное напряжение кислорода (кровь).

Степени вершин полного графа имели значения от 2 до 70. В частичных подграфах большая часть связей игнорировалась. На рис. 3 дана схема парных взаимосвязей признака ***ЭОЗ/ЭМ*** в полном графе. Каждая связь на этой схеме помечена двумя числами (значение меры связи, нижняя граница доверительного интервала меры связи). В качестве примера мы привели здесь таблицы сопряженности, построенные для пар признаков (***ЭОЗ/ЭМ, СМ***) и (***ЭОЗ/ЭМ, СЕКРЕТ***), характеризующие две из семнадцати показанных на рис. 3 связей (см. табл. 2 и 3).

Таблица 2

Сопряженности признаков ЭОЗ/ЭМ и СМ

Интервалы значений признака СМ	Интервалы значений признака ЭОЗ/ЭМ		Всего
	≤ 1	> 1	
0 - 33	26	10	36
34 - 56	21	15	36
57 - 78	11	22	33
79 - 97	8	18	26
Всего	**66**	**65**	**131**

Таблица 3

Сопряженности признаков ЭОЗ/ЭМ и СЕКРЕТ

Градации номинальной шкалы признака СЕКРЕТ	Интервалы значений признака ЭОЗ/ЭМ		Всего
	≤ 1	> 1	
Серозный	32	14	46
Гнойный	9	15	24
Всего	**41**	**29**	**70**

Анализ содержащихся в таблицах данных позволил установить знаки соответствующих парных зависимостей. Рис. 4 иллюстрирует характер распределения связей признаков, образующих пары с ***ЭОЗ/ЭМ***. Аналогичные рис. 4 представления были получены для каждой из 300 вершин-признаков полного графа причинных связей. Схемы распределения признаков (каждая в отдельности, парами или тройками, но не более) предъявлялись экспертам для интерпретации. При этом они использовали сведения о всех выявленных типах связей и качественных оценках уровней признаков (например «высокий», «средний» или «низкий» уровни для признаков, определенных на интервальных шкалах или шкалах отношений). Также они запрашивали дополнительные данные из хранимых в памяти ЭВМ историй болезни пациентов, результаты исследований которых были использованы при составлении таблиц сопряженности. Ответы на эти вопросы оперативно предоставлялись системой накопления, хранения и обработки информации, обслуживающей вычислительные эксперименты с данными ПИК [5].

Итоги работы экспертов над схемами распределения признаков оформлялись в виде протоколов, объясняющих роль каждого признака, участвующего в формировании симптомов и симптомокомплексов, обусловленных проявлением определенных патофизиологических механизмов БА. В порядке иллюстрации дадим выдержку из протокола экспертизы схемы распределения, представленной на рис. 4.

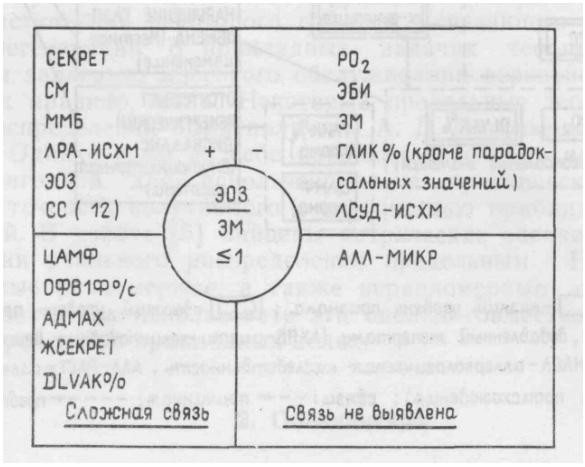


Рис. 4. Схема распределения признаков, имеющих тесную связь с ***ЭОЗ/ЭМ***:
" + " - положительная связь признаков; " — " - отрицательная связь признаков.

*«Высокий уровень эозинофилов в крови в сочетании с низким уровнем эозинофилов в мокроте (**ЭОЗ/ЭМ>1**) характеризует инфекционный процесс в бронхах с (микробной?) сенсибилизацией. Наличие патогенных микробов (**ММБ**) вызывает местное воспаление, в результате которого повышается уровень сегментоядерных нейтрофилов в смыве и мокроте (**СС и СМ**), появляется гнойный **СЕКРЕТ** и возникает микробная сенсибилизация, обусловленная нарушением адекватного ответа на наличие микробов, которые работают по программе паразитов» ...*

На основе подобных протоколов удалось построить их диаграммные представления. Нижний уровень таких диаграмм образуют структуры ***n***-арных связей признаков, а верхний - интерпретированные элементами структуры нижнего слоя понятия, представляющие собой первичные семантические модели симптомов и симптомокомплексов. Диаграммное представление полного протокола, часть которого мы цитировали выше, показано на рис. 5.

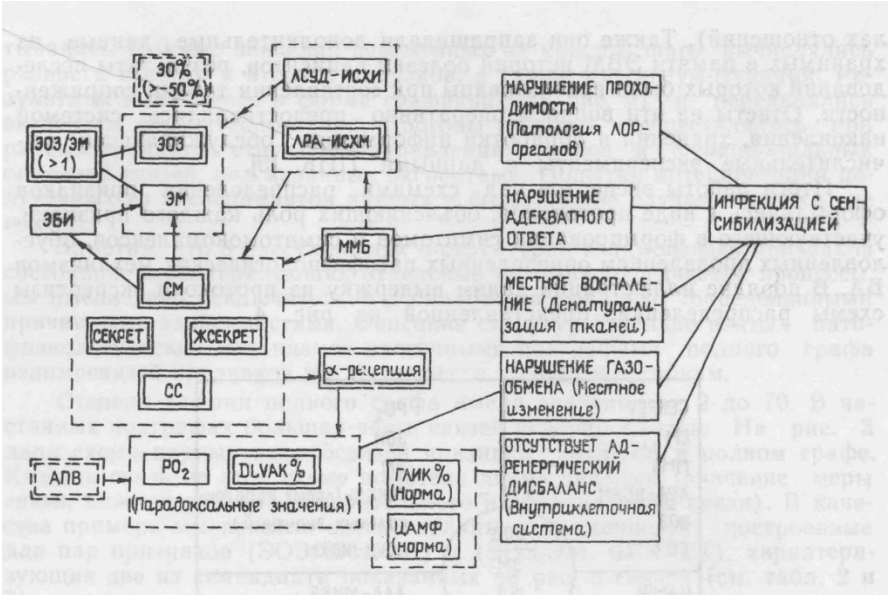


Рис 5. Экспертная морфологическая модель структур и причинных связей симптомов этиопатогенетических механизмов БА, обусловленных внутренними связями признака ***ЭОЗ/ЭМ>1***

Обозначение: ▭ - низкий уровень признака, ▣ - Высокий уровень признака;
▭ - признак, добавленный экспертами (АХПВ – ацетилхолинподобные вещества, Г - гистамин, НАСЛ - аллергологическая наследственность, АЛЛРАСТ - аллергены растительного происхождения);
Связи: ⟶ - причинная; - - - → - предполагаемая причинная.

Диаграммные представления подобного вида (общим числом около 300) использовались в качестве исходных данных при составлении экспертных морфологических дескриптивных моделей патофизиологических механизмов БА. Процедура построения таких моделей должна выполняться машинным способом с привлечением специальных средств представления знаний или, например, языка ПРОЛОГ.

ЛИТЕРАТУРА

1. *Айвазян С.А., Енюков И.С., Мешалкин Л.Д.* Прикладная статистика. Основы моделирования и первичная обработка данных. - М.: Финансы и статистика, 1983. - 471 с.
2. *Аврамчук Е.Ф., Бачурин А.С., Исаев А.Е., Качанова Т.Л. и др.* Базовая система языковых и программных средств моделирующего центра // Программно-математическое обеспечение ЭВМ. - Л.: ЦНИИ «Румб», 1986 - Вып. 42. - 184 с.
3. *Елисеева И.И., Рукавишников В.О.* Группировка, корреляция, распознавание образов. - М.: Статистика, 1977. - 143 с.
4. *Корн Г., Корн Т.* Справочник по математике для научных работников и инженеров / Пер. с англ, под общ. ред. И. Г. Арамановича. - М.: Наука, 1984. - 832 с.
5. *Бачурин А.С., Бачурина Т.И., Качанова Т.Л. и др.* Система накопления, хранения и обработки информации. // Программно-математическое обеспечение ЭВМ. - Л.: ЦНИИ «Румб», 1987. Вып. 44. - 158 С.
6. Статистические методы анализа информации в социологических исследованиях / Отв. ред. Г. В. Осипов. - М.: Наука, 1979. - 319 с.
7. *Федосеев Г.Б.* Современные представления о причинах возникновения, особенностях течения и лечения бронхиальной астмы: Актовая речь / 1-й Ленинградский мед, ин-т им. академика И. П. Павлова. - Л., 1982. - 28 с.
8. *Федосеев Г.Б., Жихарев С.С., Лаврова Т.Р. и др.* Физиологические и патофизиологические механизмы проходимости бронхов. - Л.: Наука, 1984. - 280 с.
9. *Хант Э.* Искусственный интеллект. - М.: Мир, 1978. - 558 с.