

Машинный метод синтеза таблиц для диагностики и прогно-  
зирования течения болезни

Б.И. Готяпов, Г.В. Везнощено, Н.В. Орлова

Значительное повышение качества прогнозирования исходов, постановки диагноза заболевания может быть достигнуто с помощью использования во врачебной практике диагностических таблиц, применение которых реально в условиях любой клиники.

Известно несколько методов разработки диагностических таблиц:

- метод экспертных оценок [5],
- метод последовательной диагностической процедуры Вальда [2],
- дискриминантный функциональный анализ [6],

В связи с большой трудоемкостью разработки и составления таблиц с помощью этих методов и необходимостью осуществлять периодический пересмотр диагностических таблиц, т.к. клиника патологических процессов непрерывно изменяется и существует перемещение частоты признаков [1] становится актуальным вопрос о создании машинной системы

синтеза диагностических таблиц.

Данная работа посвящена созданию машинной системы синтеза таблиц диагностики и прогнозирования по методу последовательной диагностической процедуры Вальда.

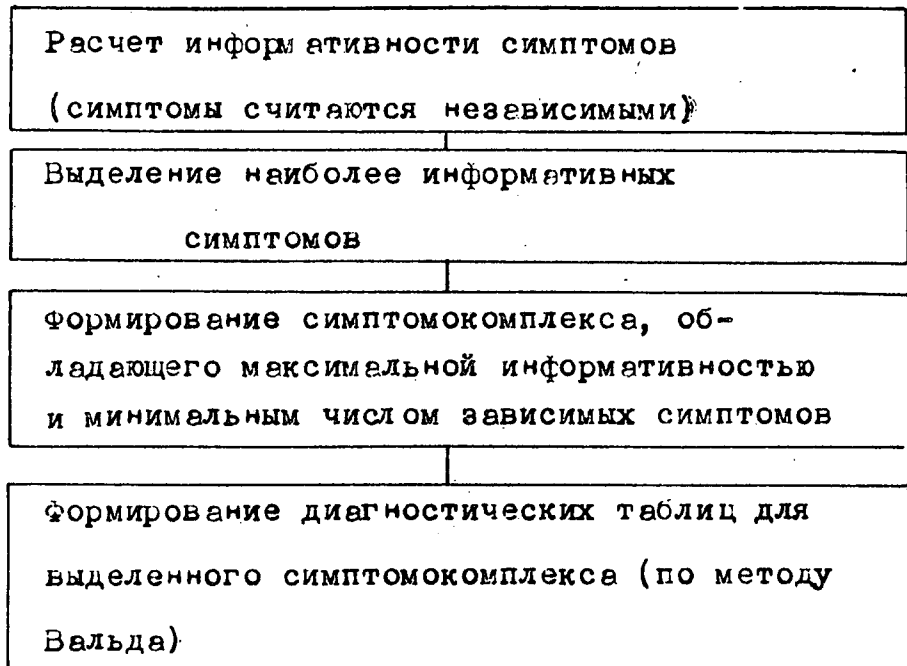
Самой сложной задачей, возникающей при использовании метода последовательной процедуры Вальда, а также при использовании остальных, указанных выше, методов формирования диагностических таблиц, является выделение симптомокомплекса. Существует несколько подходов к решению этого вопроса:

- метод экспертных оценок,
- выделение независимых признаков, характеризующих заболевание,
- выделение признаков, наиболее информативных по некоторому критерию.

Основным недостатком указанных подходов является то, что симптомы, входящие в выделенный симптомокомплекс, принимаются, как правило, независимыми.

В данной работе была осуществлена попытка создать интегрированный метод выделения симптомокомплекса и последующего формирования диагностических таблиц, учитыва-

ющий клинический опыт специалистов, зависимость и информативность исходных признаков. Алгоритм, соответствующий данному методу, представлен ниже.



Определение информативности симптомов осуществлялось по информационному критерию Кульбака [2]:

- для независимых симптомов

$$I(x_j) = \sum_i 10 \lg \frac{P(x_j^i/y^z)}{P(x_j^i/y)} [P(x_j^i/y^z) - P(x_j^i/y)],$$

где  $x_j^i$  -  $i$ -я

- градация  $j$ -го симптома,

$P(x_j^i / y^z)$  - вероятность появления  $x_j^i$  при условии, что прогнозируемый признак  $y$  примет значение  $y^z$ ,

$P(x_j^i / y)$  - вероятность появления  $x_j^i$  при условии, что прогнозируемый признак  $y$  не примет значение  $y^z$

- для групп зависимых симптомов  $X_1, \dots, X_m$ :

$$I(x_1, \dots, x_j, \dots, x_m) = \sum_{n_1=1}^{K_1} \dots \sum_{n_j=1}^{K_j} \dots \sum_{n_m=1}^{K_m} 10 \lg \frac{P(x_1^{n_1}, \dots, x_m^{n_m} / y^z)}{P(x_1^{n_1}, \dots, x_m^{n_m} / y)}$$

$$[P(x_1^{n_1}, \dots, x_m^{n_m} / y^z) - P(x_1^{n_1}, \dots, x_m^{n_m} / y)]$$

где  $K_j$  - число градаций  $j$ -го симптома.

В случае, если  $i$ -я градация независимого симптома или определенная совокупность градаций группы зависимых симптомов однозначно определяют исход (диагноз), т.е.

если одна из условных вероятностей равна нулю, то отношение условных вероятностей принималось равным  $10^5$ .

Введение этого усложнения было вызвано тем, что в случае возникновения такой ситуации информативность симптома (группы зависимых симптомов) равнялась бы бесконечности при любой априорной вероятности  $i$ -ой градации независи-

мого симптома или определенной совокупности градаций групп зависимых симптомов. Определение зависимых симптомов осуществлялось сравнением коэффициентов корреляции с пороговой величиной.

Разработанный алгоритм формирования симптомокомплекса, обладающего максимальной информативностью и состоящего из наименьшего числа зависимых симптомов, может быть описан следующим образом.

Среди симптомов, имеющих наибольшую информативность выделяются зависимые. Если в выбранном наборе симптомов есть признаки, зависящие более чем от  $K$  симптомов, входящих в этот набор, то определяется признак, от которого зависит наибольшее число других симптомов, и этот признак исключается. Далее опять осуществляется поиск симптомов, зависящих более чем от  $K$  симптомов и т.д.

После формирования набора признаков, в котором нет ни одной группы из  $(K+1)$ го зависимых симптомов, производится оценка информативности таких групп и упорядочивание групп зависимых и независимых симптомов по информативности.

Выбор параметра  $K$  осуществляется из условия

необходимой простоты организации диагностической таблицы, в данной работе К применяется равным 2.

Описанный алгоритм был реализован на языке Фортран-IV для ЭВМ ЕС-1020.

В связи с ориентацией использования программы непосредственно практическим врачом организация входных и выходных данных осуществлялась на основе следующих принципов.

Исходная информация кодировалась следующим образом: "Симптомы, которые могут принимать два значения "присутствует" или "отсутствует" кодируются условными символами "1" и "0" соответственно".

"Признаки, которые могут иметь любое значение из некоторого интервала, разбиваются на градации (не больше 8) и кодируются номером подинтервала, в который попадает истинное значение признака". Таким образом, вся исходная информация о каждом больном кодируется на одной перфокарте, каждый столбец которой соответствует определенному симптому (максимально возможное число анализируемых симптомов равняется 80).

Такая организация входной информации значительно

упрощает процесс ее подготовки и контроля.

В выдаваемую машиной информацию входят сформированный симптомокомплекс и диагностическая таблица.

Примером организации выходной информации являются приведенные ниже выделенный симптомокомплекс и диагностическая таблица исхода родов при бактериурии:

комплекс II 2 14 23 48 6 20 3 7

ДИАГНОСТИЧЕСКАЯ ТАБЛИЦА

Симпто- мов	их градации	диагностический коэффициент
II 2 14		-0,601E 01
	1 1 3	-0,810E 00
	1 2 2	0,207E 01
	1 2 3	0,207E 01
	1 3 2	0,612E 01
	2 1 2	-0,810E 00
	2 1 3	-0,486E 01
	2 2 2	0,458E 01
	2 2 3	-0,159E 02
	2 3 2	-0,369E 01
	2 3 3	0,240E 02
	2 4 2	-0,641E 01
	3 1 2	0,131E 02

	3 2 2	-0,997E 01
	3 3 2	0,131E 02
	3 3 3	-0,810E 00
	3 4 2	0,612E 01
23	I	0,131E 02
	2	-0,478E 01
	3	-0,179E 00
48	2	0,503E 01
	4	0,619E 01
6		
	I	0,601E 01
	2	0,163E 00
	3	-0,544E 01
20		
	I	0,122E 01
	2	-0,894E 01
	3	0,322 01
3		
	I	0,435E 00
	3	-0,114E 01
7		
	I	0,487E 00
	2	-0,444E 01



В выделенный симптомокомплекс вошли 9 наиболее информативных симптомов, среди которых три (II,2,14) являются зависимыми.

В таблице симптомы располагаются по убыванию информативности. Таким образом, выделенная группа зависимых симптомов является наиболее информативной.

В диагностической таблице слева направо перечисляются номера симптомов, соответствующие им градации и диагностические коэффициенты.

Разработанная библиотека программ была использована для формирования таблиц прогноза и диагноза заболеваний почек у беременных. Известно, что при заболеваниях почек у беременных довольно сложно дифференцировать пиелонефрит беременных, хронический пиелонефрит и сочетанный токсикоз, поскольку характерные признаки этих заболеваний нивелируются. В связи с этим нередко постановка диагноза, а следовательно, и выбор индивидуального комплексного лечения затруднены и запаздывают во времени. Дифференцирование диагноза а также прогнозирование исходов течения заболевания для матери и плода в плане ближайших и отдаленных последствий представляет большую сложность и порою проблематичность для врача акушера-гинеколога.

Конкретными целями прикладной части работы явились:

1. Изучение особенностей течения беременности, родов и послеродового периода при пиелонефрите беременных в сравнении с другими группами заболеваний почек у беременных (сочетанных токсикоз, хронический пиелонефрит, бактериария).

2. Прогнозирование исхода беременности, родов и восстановление вероятностного прогноза для матери и плода при пиелонефрите в ближайшем и отдаленном времени.

3. Выявление существенных закономерностей течения беременности при пиелонефрите, создание модели течения беременности и родов при этом заболевании.

Основным материалом для синтеза таблиц послужили статистические данные, полученные при обработке архивного материала (1220 историй болезни с верифицированным диагнозом).

В качестве дифференциальных прогностических тестов первоначально были выбраны 88 признаков, характеризующих срок беременности, анемию жизни и заболевания, симптоматику в момент осмотра врачом, данные лабораторных тестов и эффект проведенной терапии.

Мы сочли необходимым уточнить, какое влияние на

прогноз родов для матери и плода оказывается перенесенная в прошлом соматическая патология, особенности течения данной беременности при клиническом и лабораторном обследовании.

При анализе течения пиелонефрита беременных были выделены критерии, подчеркивающие тяжесть патологического процесса в почках: температурная реакция, данные исследования крови и мочи. Например, из 88 изученных признаков при составлении приведенной диагностической таблицы в нее вошли только 8, наиболее информативных. Остальные признаки были исключены в связи с тем, что различия некоторых из них при благоприятном исходе не были статистически значимыми. К числу таких признаков были отнесены: исход предшествующей беременности, длительность течения заболевания, одно- или двустороннее поражение почек и др. ✓

Формируемая диагностическая таблица проста в применении, используется в качестве бланка, который после заполнения его врачом, становится источником данных для проверки эффективности таблицы и ее корректировки.

Процедура заполнения таблицы и ее использование состоит в следующем. Для каждого признака врач подчеркивает соответствующий диапазон, отмечающийся у больной.

Затем выбирает относящиеся к этим диапазонам приведенные алгебраическую сумму до достижения величины  $\pm P$  ( $P$ -принятая пороговая величина,  $+$  означает благоприятный прогноз,  $-$  неблагоприятный). Если выполнялось, что алгебраическая сумма  $S$  для всех признаков лежит в пределах  $-P < S < P$ , то принимается решение "неопределенный диагноз" т.е.

имеющейся информации недостаточно для прогнозирования с выбранным уровнем надежности.

В большинстве случаев прогнозирования по приведенной таблице определенный ответ был получен после использования первых 5-7 признаков, что обусловлено большой информативностью этих признаков: срок беременности, СОЭ, лейкоцитоз, пиурия, температура, симптом Пастернацкого, время начала заболевания.

Прогностически наиболее информативными неблагоприятными предрасполагающими факторами оказались болезни мочевыделительной системы в анамнезе, раннее начало заболевания, высокое СОЭ и пиурия. Использование перечисленных признаков не может полностью заменить использование таблицы, которая содержит более полную информацию о признаках.

Клиническая апробация составленных таблиц показала правильность прогнозирования исхода заболевания в 85%

случаев, в 5% случаев прогноз был ошибочным, в остальных случаях — неопределенным. Таким образом, в связи с трудностью постановки диагноза и решения вопроса об исходе заболевания почек у беременных для матери и плода применение составленных диагностических и прогностических таблиц является перспективным в акушерско-гинекологической практике.

#### Л и т е р а т у р а

- [1] Амосов Н. М., Минцер О. П., Палец Б. Л.: "Кардиология", 1977, 7, с. 19-25.
- [2] Гублер Е. В.: "Вычислительные методы распознавания патологических процессов", Л., 1970.
- [3] Минцер О. П., Чепский Л. П., Цыганый А. А., Заславский С. Я.: "Проблемы медицинской кибернетики", М., 1972.
- [4] Головня Л. Д., Губерман Ш. А.: "Кардиология", 1977, 7, с. 138-143.
- [5] Мисюк Н. С., Довнар И. А.: "Дифференциальная диагностика острых нарушений мозгового кровообращения при помощи таблицы-матрицы", сб. "ЭВМ в невропатологии", Минск, 1968.

- 6 Thurmayer R., Otte M., Thurmayer G. R.: Computer aid for diagnosing pancreatic, hepatic and gastric diseases by pancreatic function test. "MEDINFO 74. Proc. 1st World Conf. Med. Inform." Stockholm, 1974, Vol. 2., Amsterdam e. a., 1974, 607-612.