УДК 616.988.25-002.954.2-037

РАЗРАБОТКА МАТЕМАТИЧЕСКОЙ МОДЕЛИ ПРОГНОЗИРОВАНИЯ ЗАБОЛЕВАЕМОСТИ КЛЕЩЕВЫМ ЭНЦЕФАЛИТОМ

Цокова Т.Н., Козлов Л.Б.

ГОУ ВПО Тюменская государственная медицинская академия Росздрава, Тюмень

Подробная информация об авторах размещена на сайте «Учёные России» - http://www.famous-scientists.ru

Разработана математическая модель прогнозирования инфекционной заболеваемости на модели природно-очаговой инфекции, возбудителем которой является вирус клещевого энцефалита. Математическая модель представлена в виде аддитивного временного ряда, включающая тренд, случайные компоненты и сезонные составляющие, имеющие разную периодичность: менее года, 3 года и многолетнюю.

Математическое моделирование эпидемических процессов (ЭП) позволяет осуществлять качественный долгосрочный прогноз инфекционной заболеваемости, выявлять факторы, влияющие на динамику ЭП. В 70-х годах XX столетия Л.Н. Большев с соавторами [2, 3] разработали математическую модель прогноза заболеваемости клещевым энцефалитом (КЭ). Для построения модели авторы провели предварительный анализ влияния частоты присасывания клещей и иммунной прослойки населения всех возрастов на частоту заболеваний КЭ. Е.И.Болотин с соавторами [1] разработали методику факторного временного прогнозирования эпидемических проявлений очагов КЭ с использованием выявленных критических уровней заболеваемости.

Нами для разработки математического моделирования ЭП использована многолетняя динамика заболеваемости КЭ на юге Тюменской области и составлены временные ряды в климатических подзонах. Математическая модель позволила без предварительного анализа факторов, влияющих на ЭП, осуществлять прогноз заболеваемости КЭ.

По анализу временных рядов можно предсказать будущие значения временного ряда на основании предыдущих фактических данных [5]. В процессе составления временного ряда необходимо идентифицировать и формально описать его. Как только математическая модель будет определена, ее можно экстраполировать, не обращая внимания на процессы, влияющие на изменение показателей временного ряда.

Данный метод использован нами для изучения динамики ЭП, вызванного вирусом КЭ. Сложность возникла при идентификации временного ряда. При монотонном тренде происходит устойчивое нарастание или убывание значений временного ряда. Анализировать такой ряд обычно нетрудно. В качестве экстраполирующей функции ряда чаще всего выбирают известные математические зависимости – линейные, параболические, экспоненциальные и др., которые часто не отражают динамику ЭП. Прогнозируемые значения такого ряда будут являться оценочными и не обеспечивают точности расчётов.

Реже используют полиномиальное преобразование временного ряда. Сглаживание ряда с помощью полиномов связано с проведением трудоёмких вычислений, что не позволяет широко использовать данный метод для прогноза инфекционной заболеваемости, и в научной литературе сведения о применении полиномиального метода для этих целей отсутствуют.

Для прогнозирования заболеваемости КЭ нами использована аддитивная модель и временной ряд был представлен функциональной зависимостью:

$$y(t) = f(t) + s(t) + \varepsilon(t), \quad (1)$$

где y(t) – значение временного ряда в момент времени t; f(t) – основная составляющая ряда (тренд); s(t) – сезонная составляющая (отражает повторяемость показателей заболеваемости в течение сезона года); ε(t) – случайная компонента (отражает неучтённые и случайные факторы, влияющие на ЭП). Временные ряды могут содержать одновременно все перечисленные компоненты или их различные комбинации. Такой ряд называют временным с сезонной составляющей.

Для получения тренда использовано уравнение полинома Чебышева высокого порядка, что позволило снизить погрешности в прогнозе заболевания. Аппроксимирующая функция f(t) теоретически может быть выражена многочленом любой степени m, например:

$$\overline{y}_x = a_0 + a_1 x + a_2 x^2 + \ldots + a_m x^m \quad (2)$$

При каждом повышении порядка полинома требуется определение не только нового параметра, α_{m+1}, но ввиду изменения системы “нормальных уравнений”, проводят пересчет всех остальных параметров: от α_o до α_m.

Рассмотрев общий случай использования метода наименьших квадратов, П. Л. Чебышев разработал метод вычисления уравнения регрессии, позволяющий определять добавляемый параметр без пересчета найденных ранее параметров, ограничиваясь лишь вычислением нового параметра. Добавляемый член имеет вид: $a_{m+1}P_{m+1}(t)$, где $P_{m+1}(t)$ определяется по общей формуле:

$$P_{m+1}(t) = P_1(t) * P_m(t) - \frac{m^2(n^2 - m^2)}{Ч(2m-1)(2m+1)} * P_{m-1}(t) \quad (3)$$

В этом случае уравнение регрессии принимает вид [4]:

$$\overline{y_t} = a_0 P_0(t) + a_1 P_1(t) + a_2 P_2(t) + \ldots + a_m P_m(t)$$

$$a_0 = \frac{\sum y_i}{n};$$

....

$$a_m = \frac{\sum y_i P_m(t)}{\sum P_m^2(t)}. \quad (4)$$

После нахождения уравнения тренда вычисляют остаточную дисперсию по формуле:

$$D_m = \frac{S_m}{n - m - 1}, \quad (5)$$

где n-m-1 – степень свободы, m – степень полинома, n – объём выборки.

Для получения S_m применяют формулу:

$$S_m = \sum_{i=1}^{n} (y_i - y_i^m)^2, \quad (6)$$

где y_i^m - значение ординаты, рассчитанное по уравнению полинома m – степени. Переход к многочленам более высокого порядка производят до тех пор, пока остаточная дисперсия продолжает уменьшаться. Если остаточная дисперсия при выравнивании по многочлену m+1 порядка по сравнению с остаточной дисперсией, полученной для уравнения порядка m, уменьшается незначительно, переход к уравнениям более высокого порядка сле-

дует прекратить и аппроксимацию считать достаточной.

Ошибка расчёта (среднеквадратическое отклонение полученной функции от экспериментальных точек) должно быть одного порядка с погрешностью введённых табличных данных, так как среднеквадратическое отклонение зависит от y, n и вида выбранной функции y^*. Вычисляют среднеквадратическое отклонение d_n полученной теоретической кривой y^* от экспериментальной y:

$$d_n = \sqrt{\frac{1}{(n-1)} \cdot \sum_{i=1}^{n} \left[y_i - y_i^*\right]^2} \qquad (7)$$

и сравнивают с погрешностью эксперимента e. При этом возможно:

1) $d_n > e$ - аппроксимация слишком грубая, степень полинома необходимо увеличить;

2) $d_n < e$ - аппроксимация физически недостоверна, истинная функция “сплющена”, старшие степени полинома лишние и, следовательно, надо уменьшить степень полинома;

3) $d_n \approx e$ - степень полинома оптимальна.

Погрешность эксперимента рассчитывали по формуле:

$$e^2 = \frac{\sum_{i=1}^{n} (y_i - \overline{y})^2}{n-1}, \qquad (8)$$

где y_i - наблюдаемые значения, $\overline{y}$ - среднее значение введённых параметров, n – объём выборки.

Аппроксимирующая функция для строго периодической сезонной составляющей s(t) в уравнении (1) находили, используя тригонометрическую функцию, приподнятую над осью t и сдвинутую вдоль неё, в виде:

$$s(t) = a_o + a_1 \cos\left(\frac{pt}{\mathbf{l}}\right) + b_1 \sin\left(\frac{pt}{\mathbf{l}}\right) \qquad (9)$$

В нашем случае, будучи периодической и заданной в виде таблицы, сезонная составляющая не являлась синусоидой, но была близка к ней. Изменив масштаб по оси t в $\frac{\mathbf{l}}{p}$ раз ($\mathbf{l}$ – число полупериодов сезонной составляющей) получили:

$$s(t) = a_o + a_1 \cos(z) + b_1 \sin(z) \qquad (10)$$

Сжатие графика по оси t не изменит табличные значения s(t) и коэффициентов a_o, a_1, b_1. Коэффициенты a_o, a_1, b_1 определялись из уравнения (10) методом наименьших квадратов. Была получена система из трёх уравнений, после решения которой, получены формулы для расчёта коэффициентов:

$$b_1 = \frac{\sum_{1}^{N} s_i \cdot \sin(z_i)}{\sum_{1}^{N} \sin^2(z_i)}$$

$$a_1 = \frac{N \cdot \sum_1^N s_i \cdot \cos(z_i) - \sum_1^N \cos(z_i) \cdot \sum_1^N s_i}{N \cdot \sum_1^N \cos^2(z_i) - \left(\sum_1^N \cos(z_i)\right)^2}$$

$$a_o = \frac{\sum_1^N s_i - a_1 \cdot \sum_1^N \cos(z_i)}{N}$$

$$z_i = \frac{p\,(2 \cdot i - N)}{N} \qquad (11)$$

Функция s(t) определена на отрезке [1, N] не симметрично точке t=0. Поэтому, после вычисления коэффициентов по уравнениям (11), следует сдвинуть функцию s(t), заменив аргумент $\left(\frac{pt}{\mathbf{l}}\right)$ в формуле (9) на выражение:

$$\frac{p}{\mathbf{l}}\left(t - \frac{N+1}{2}\right) \qquad (12)$$

Аппроксимация тригонометрическим многочленом менее точная, чем например степенная иди другая, но к ней прибегают если функция строго периодическая. Для более точного расчёта a_o, a_1, b_1 необходимо точно определить период сезонной составляющей вводимого временного ряда.

Условие $d_n \approx e$ - степень полинома оптимальна, выполнялось для уравнения тренда с учётом сезонной составляющей.

Весь алгоритм расчётов по формулам 2 – 12 можно автоматизировать. Нами была составлена компьютерная программа, которая позволяет провести полиномиальную аппроксимацию высокого порядка (8 порядка и выше) значений функций заданных таблично. Используя полученное уравнение тренда, проводили прогнозирование заболеваемости. Программы составлены для ПК, работающие в среде Microsoft Windows на программном языке Visual Basik.

Для составления данной программы использована база данных «Tumklech» [8]. Фактическая заболеваемость КЭ была переведена в показатели заболеваемости на 100 тыс. населения. Анализ тренда проводили в два этапа: определяли наличие тренда и выделяли тренд. Для определения наличия тренда использован критерий Стьюдента, позволяющий выявить различие выборочных средних двух половинок временного ряда. Если различие значимо, то гипотеза о наличии тренда не отвергалась. Для выделения тренда использовали модель простого статистического ряда и аддитивную модель временного ряда (временной ряд с сезонной составляющей).

В динамике заболеваемости КЭ характерна трехлетняя цикличность и в течение сезона года (с 1.04 по 30.09) также наблюдались колебания в показателях заболеваемости. Усреднив значения ряда в соответствии с трёхлетней цикличностью, получим сезонные колебания заболеваемости для каждого года трехлетнего цикла [7]. Однако для долгосрочного прогноза, предложенный способ прогноза будет недостаточно точным.

Известно, что в природе существуют циклические колебания погоды, связанные с цикличностью излучения Солнца. Большой цикл Солнца длится 11 лет. Он связан с цикличностью образования солнечных пятен. 1979-1982 годы максимальной активности; 1986-1987 годы минимальной активности. В настоящее время Солнечная активность минимальная [6].

Методом дисперсионного анализа нами также установлено наличие многолетнего цикла в динамике заболеваемости КЭ. В 1999г. наблюдалась максимальная

заболеваемость КЭ во всех климатических подзонах юга Тюменской области.

Применение многолетней цикличности к нашим расчётам, означает усложнение модели временного ряда ещё одной составляющей, зависящей от многолетней цикличности. В этом случае прогноз модели будет точнее. Условно назовём эту модель - аддитивная модель с двумя составляющими. Уравнение этой модели:

$$y(t) = f(t) + \varepsilon_1(t) + \varepsilon_2(t), \qquad (13)$$

где $\varepsilon_1(t)$ – сезонная составляющая с трёхлетним периодом, $\varepsilon_2(t)$ –сезонная составляющая с многолетним периодом. Данная модель может быть использована для долгосрочного прогнозирования заболеваемости КЭ.

СПИСОК ЛИТЕРАТУРЫ:

1. Болотин Е.И., Цициашвили Г.Ш., Голычева И.В. // Паразитология. - 2002. – Т.36.- Вып. 2. – С.89.

2. Большев Л.Н., Гольдфарб Л.Г. // Вопросы эпидемиологии и профилактики клещевого энцефалита. – М., 1970. – С. 154.

3. Большев Л.Н., Гольдфарб Л.Г., Круопис Ю.И. // Вопросы эпидемиологии и профилактики клещевого энцефалита. – М., 1970. – С. 171.

4. Венецкий И.Г., Кильдишев Г.С. Теория вероятности и математическая статистика. – М., «Статистика». – 1975. – 264 с.

5. Ивашев-Мусатов О.С. Теория вероятности и математическая статистика. – М., «Наука». – 1979. – 254 с.

6. Ишков В.Н. «Вселенная и мы» www.Астронет.ru / Солнце в текущем 23 цикле солнечной активности.

7. Козлов Л.Б., Кашуба Э.А., Цокова Т.Н. и др. Способ прогноза заболеваемости клещевыми инфекциями // Патент RU 2294697 C2. – Бюл. № 7 от 10.03.2007г.

8. Козлов Л.Б., Цокова Т.Н., Огурцов А.А. и др. / Заболеваемость клещевым энцефалитом в Тюменской области «Tumklech» // База данных. – Свидетельство №2007620363 от 18.10.2007. – Правообладатель: ФГУЗ «Центр гигиены и эпидемиологии в Тюменской области».

DEVELOPMENT OF MATHEMATICAL MODEL OF FORECASTING OF DESEASE THE VIRUS OF TICK-BORN ENCEPHALITIS

Tsokova T.N., Kozlov L.B.
Tyumen state medical academy, Tyumen

The mathematical model of forecasting of infectious desease on model of tick-born infections which activator is the virus of tick-born encephalitis is developed. The mathematical model is presented in the form of additive time of some, including a trend, casual components and seasonal components having different periodicity: less than year, 3 years and long-term.