

МОДЕЛИРОВАНИЕ ВЛИЯНИЯ ПАРАМЕТРОВ ВНЕШНЕЙ СРЕДЫ НА ПРОТЕКАНИЕ ЭПИДЕМИОЛОГИЧЕСКОГО ПРОЦЕССА ГРИППА/ОРВИ

Анотація. У статті пропонується підхід до дослідження впливу параметрів середовища на розвиток епідеміологічного процесу захворювання на грип / ГРВІ. Дослідження ґрунтуються на використанні реальних статистичних даних, основних положеннях теорії епідеміології та методах множинного регресійного аналізу. Як приклад наведено результати для температури повітря як параметра середовища.

Ключові слова: епідеміологічний процес, параметри середовища, регресійний аналіз.

Аннотация. В статье предлагается подход к исследованию влияния параметров среды на развитие эпидемиологического процесса заболевания гриппом/ОРВИ. Исследования основаны на использовании реальных статистических данных, основных положениях теории эпидемиологии и методах множественного регрессионного анализа. В качестве примера приведены результаты для температуры воздуха как параметра среды.

Ключевые слова: эпидемиологический процесс, параметры среды, регрессионный анализ.

Abstract. In this paper we propose an approach to the study of environmental parameters on the development of epidemic process of influenza / ARI. studies are based on the use of actual statistics, the main provisions of the theory epidemiology and methods of multiple regression analysis. As an example, the results for the temperature as a parameter of the medium.

Keywords: epidemiological process, the parameters of the medium, regression analysis

1. Вступление

Здоровье населения – показатель, влияющий на уровень экономического развития региона и характеризующий качество жизни населения. Наиболее массовыми и убыточными для народного хозяйства являются эпидемические заболевания, самое распространённое из которых грипп [1]. Поэтому моделирование эпидемиологических процессов гриппа, особенно исследование влияния на эпидемиологический процесс факторов внешней среды (например, экологических, демографических, экономических), необходимо для задач государственного управления.

Традиционные способы моделирования эпидемий гриппа [2–5] задачу по исследованию влияния параметров внешней среды, в том числе погодные условия, не включали. Преобладающее направление публикаций в настоящее время в данной области – моделирование медико-биологического процесса протекания заболевания индивидуума и учет этой модели в традиционной схеме распространения эпидемии. Безусловно, это направление наукоемкое и важное для развития эпидемиологии. Основной вывод – интенсивность и продолжительность всплеска эпидемии зависят от характеристики заболевания, и в каждом последующем городе картина эпидемии будет повторяться.

Однако оно не объясняет таких явлений, как волны эпидемий, почему один и тот же вирус гриппа в различных регионах вызывает различные по длительности и интенсивности эпидемиологические всплески. Например, на рис. 1 совмещены графики заболеваемости гриппом/ОРВИ в течение 12 недель 2009 года для всех регионов Украины. Можно видеть, что на 2-й неделе рост заболеваемости прекратился даже для тех регионов, в которых эпидемия только что началась, на 6-й – рост заболеваемости усилился почти во всех регионах, а на 7–8-й неделе наблюдается максимум.

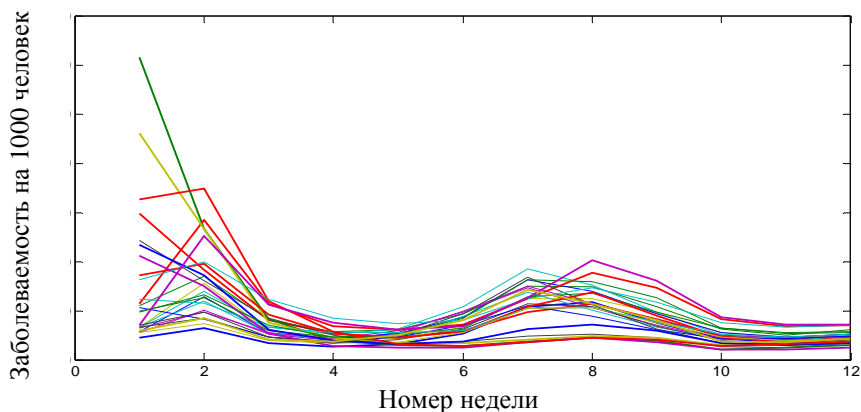


Рис. 1. Изменение заболеваемости гриппом/ОРВИ во времени для различных регионов Украины в 2009 году

Теперь сравним данные, представленные на рис. 1, с данными о температуре воздуха за этот же период для регионов Украины (рис. 2).

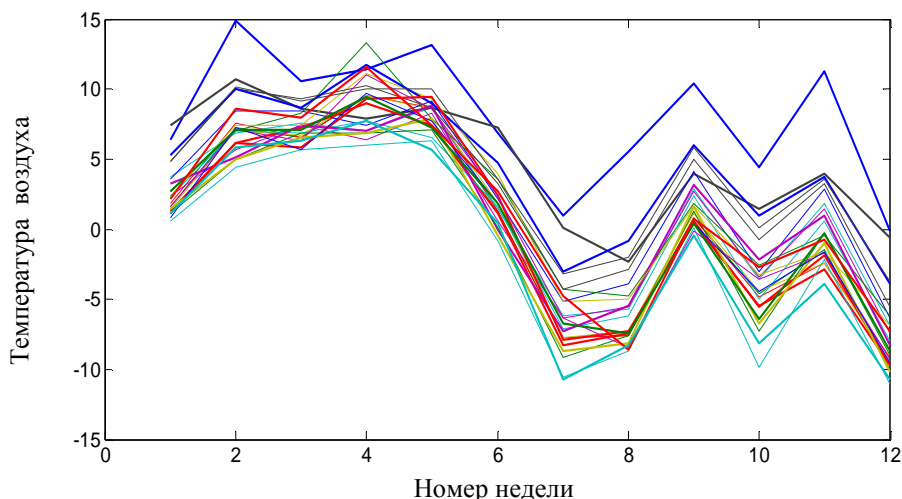


Рис. 2. Температура воздуха в $^{\circ}\text{C}$ для регионов Украины за период, соответствующий данным эпидемии

Объяснение волнообразного распространения эпидемии и несовпадение эпидемических всплесков в различных городах влиянием температуры воздуха кажется очевидным. Подтверждает эту зависимость также тот факт, что эпидемии гриппа происходят только в холодные периоды года: зимой или осенью.

Отсутствие модели влияния температуры воздуха на заболеваемость, вероятно, объясняется тем, что публикуемые статистические данные не являются полными. Даже в тех странах, в которых отчетность публикуется регулярно, отражается суммарная заболеваемость за неделю. Если учесть, что эпидемиологический всплеск заболеваемости обычно длится 2-3 недели, то очевидно, что необходим ежедневный мониторинг.

2. Исходные данные

В 2009 году в связи со вспышкой эпидемии в Ивано-Франковской области данные о заболеваемости (т.е. число обратившихся к врачу) публиковались в интернете ежедневно по областям и группам населения [6]. Температурный график за этот же период времени можно найти в [7]. Эти данные нуждаются в предварительной обработке как стандартной

(исключение грубых ошибок, пропусков и т.п.), так и специальной. Имеются в виду колебания процесса заболеваемости в зависимости от дня недели, обусловленной тем, что в некоторые дни недели люди обращаются к врачу чаще, чем в иные. По воскресным дням и праздникам данных о заболеваемости почти нет. Эти колебания присутствовали как в данных [3], так и [6].

Компенсировать колебания можно с помощью фильтрации, используя функцию, зависящую от дня недели. Эти вопросы рассматривались в работах [3] и [8] и в настоящей статье описываться не будут.

В дальнейшем будем использовать исходные данные: временные ряды соответственно $\{y_i\}$ – ряд заболеваемости, а $\{T_i\}$ – температурный ряд для каждого из 27 регионов Украины.

Необходимо также отметить, что взаимная корреляция рядов $\{y_i\}$ и $\{T_i\}$ не дает желаемого результата, так как ряд $\{y_i\}$ очевидным образом не является стационарным случайным процессом. Для построения искомой зависимости рассмотрим основные предположения, которые используются при построении эпидемиологических моделей.

3. Исходные положения для эпидемиологического моделирования и обоснование подхода к построению модели

3.1. Разбиение населения региона на непересекающиеся группы людей

Наиболее популярные в англоязычной литературе compartment модели [4, 5] обозначаются буквами, которые используются для идентификации этих групп, например, SIR (чувствительные к заболеванию S, больные-заразные I, выздоровевшие-имунные R), или SEIR – добавляется группа E – зараженные и находящиеся в латентном периоде заболевания. Это характерно и для модели [3].

Для сравнения регионов с различным количеством населения численность этих групп нормализуется, т.е. делится на численность населения региона. Следовательно, $S + I + R = 1$ или $S + E + I + R = 1$ в зависимости от принятой модели.

Поскольку исходные данные представляют собой временные ряды, то будем рассматривать все другие вычисляемые характеристики эпидемиологического процесса (S,E,I,R) как дискретные во времени. Учитывая также предыдущие замечания, получаем

$$S_i + I_i + R_i = 1. \quad (1)$$

3.2. Предположения о процессе заражения

Распространение заболеваемости прежде всего зависит от числа контактов между инфицированными субъектами – I и субъектами, чувствительными к инфекционному заболеванию – S. Чаще всего предполагают [2], что число контактов равно произведению $I \cdot S$, а число заболевших пропорционально этому произведению, т.е.

$$y_i = \beta \cdot I_i \cdot S_i, \quad (2)$$

где β – коэффициент пропорциональности, i – отсчет времени (день). Некоторые исследователи в [5] предполагают следующую зависимость:

$$y_i = \beta \cdot I_i^{\alpha_1} \cdot S_i^{\alpha_2}, \quad (3)$$

где α_1 и α_2 – параметры «перемешивания» населения. Причем, если население «перемешано» до однородности, эти параметры равны 1.

В некоторых работах предполагается считать коэффициент β зависящим от времени, чтобы «приблизить» абстрактную теоретическую модель к реальным данным [5], т.е.

$$y_i = \beta_i \cdot I_i \cdot S_i. \quad (4)$$

Иногда используют приближение [5]:

$$S_i \cong 1, \quad (5)$$

т.е. чувствительным к заболеванию является все население региона.

Для решения поставленной задачи (идентификации зависимости процесса заболеваемости от температуры воздуха) приведенная выше формализация процесса заражения очень важна, поскольку здесь появляется коэффициент β_i , который можно рассматривать как функцию от параметров среды.

В общем случае зависимость, характеризующую заболеваемость, можно представить:

$$y_i = F(I_i, S_i, \beta_i(T_i)), \quad (6)$$

где I_i – число больных, S_i – чувствительных, β_i – значение функции.

Используя приближение (5)

$$y_i = F(I_i, \beta(T_i))$$

или

$$y_i = \Psi(I_i, T_i). \quad (7)$$

3.3. Предположения о процессе выздоровления

Число людей, выздоровевших в i -й день, представляет собой разность $R_{i+1} - R_i$ общего числа выздоровевших – иммунных и зависит прежде всего от характеристик заболевания и эпидемиологической обстановки.

В [2, 4, 5] предполагают, что количество выздоровевших пропорционально количеству больных:

$$R_{i+1} - R_i = \nu \cdot I_i, \quad (8)$$

где ν – коэффициент, зависящий от характеристик заболевания.

В работе [3] определяют число выздоровевших больных в i -й день, используя временной ряд заболеваемости, т.е.:

$$R_{i+1} - R_i = \sum_j y_{i-j} \cdot g_j, \quad \sum_j g_j = 1, \quad (9)$$

где g_j – вероятность выздоровления человека, заболевшего j дней назад. Причем приводятся статистические данные для последовательности $\{g_j\}$:

$$G = (0; 0,1; 0,35; 0,25; 0,15; 0,10; 0,5).$$

В самом простом случае

$$R_{i+1} - R_i = y_{i-j}, \quad (10)$$

если заболеваемость длится ровно j дней.

3.4. Вычисление количества чувствительных индивидуумов

Предположения позволяют определить временной ряд больных (инфицированных) и чувствительных.

Количество чувствительных (S) уменьшается с каждым днем на число заболевших. Следовательно:

$$S_{i+1} = S_i - y_i. \quad (11)$$

Эта разностная схема определяет ряд:

$$S_{i+1} = S_0 - \sum_{j=1}^i y_j,$$

где S_0 – начальное значение.

Если же предположить, что $S_0 = 1$, то

$$S_{i+1} = 1 - \sum_{j=1}^i y_j. \quad (12)$$

Это решение соответствует всем перечисленным выше предположениям.

3.5. Вычисление количества больных

Каждый день число больных (инфицированных) увеличивается на количество заболевших и уменьшается на число выздоровевших за текущие сутки:

$$I_{i+1} - I_i = y_i - (R_{i+1} - R_i).$$

Далее, если воспользоваться предположением (8), то получим

$$I_{i+1} = I_i(1 - \nu) + y_i$$

или

$$I_{i+1} = I_0(1 - \nu)^{i+1} + \sum_{j=1}^i y_j(1 - \nu)^{i-j},$$

где I_0 – начальное значение. Если $I_0 = 0$, то

$$I_{i+1} = \sum_{j=1}^i y_j(1 - \nu)^{i-j}. \quad (13)$$

Решение, основанное на предположении (9), будет иметь вид

$$I_{i+1} = \sum_{j=1}^n y_{i-j} q_j, \quad (14)$$

где q_j – вероятность того, что человек, заболевший j дней тому назад, в настоящий момент будет еще больным. Последовательность q_j связана с последовательностью g_j соотношением

$$q_j = 1 - \sum_{k=1}^j g_k.$$

Статистические данные для оценки последовательности q_j равны [3]:

$$Q = \{q_i\} = (0, 1; 0, 9; 0, 55; 0, 35; 0, 25; 0, 15; 0, 1).$$

Нулевое значение первого члена последовательности соответствует определению латентного периода длиной в 1 день. На второй день после заражения все являются переносчиками инфекции, на следующий день 90% и т.д.

Третий вариант решения, соответствующий соотношению (10), имеет вид

$$I_{i+1} = \sum_{j=1}^n y_{i-j}, \quad (15)$$

где n – продолжительность заболевания.

4. Идентификация и анализ модели

Вычисление интегральных характеристик заболеваемости – количество больных и чувствительных – позволяет обойти идентификацию искомой зависимости как динамической системы. Поэтому можно объединить выборки по всем регионам (объем результирующей выборки – 1700 точек).

Стандартный подход к идентификации зависимости (7) – метод множественной регрессии [11]. Наиболее предпочтительно использование линейной регрессии – идентификации модели вида

$$y = a + bI + cT + dS + \xi, \quad (16)$$

где ξ – погрешность исходных данных. Для мультипликативных моделей можно заменить значения исследуемых величин их логарифмами [11]. Получим

$$\ln(y) = a + b \ln(I) + cT + d \ln(S) + \xi. \quad (17)$$

Идентификация обеих зависимостей дает хорошее приближение. Однако модель (17) (рис. 3) дает лучшее приближение (коэффициент детерминации для первого уравнения равен 0,8297, для второго – 0,8696).

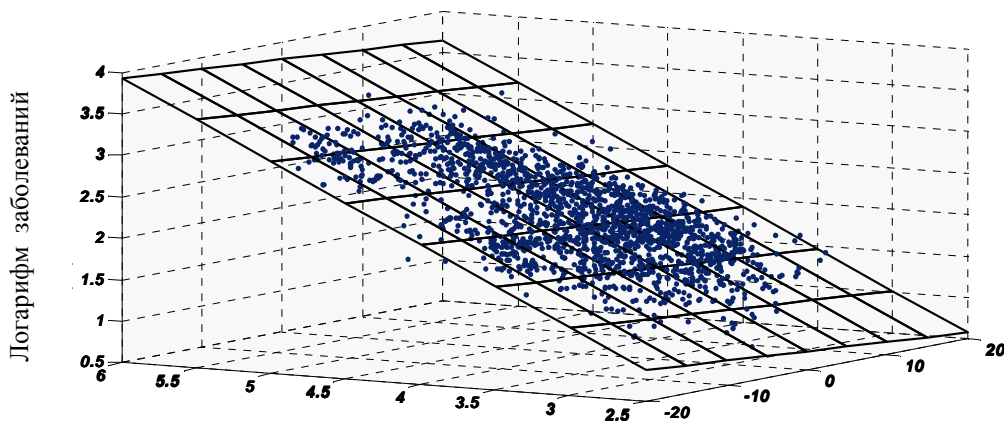


Рис. 3. Регрессионная зависимость заболеваемости от количества больных и температуры воздуха

Изменение коэффициента детерминации в зависимости от сочетания факторов в уравнении (17) представлено в табл. 1.

Таблица 1. Коэффициент детерминации в зависимости от вида уравнения регрессии

Вид уравнения	Значение коэффициента детерминации
$y = aI + b + cT + \xi$	0,8297
$\ln(y) = a + b \ln(I) + cT + \xi$	0,8696
$\ln(y) = a + b \ln(I) + cS + \xi$	0,8469
$\ln(y) = a + b \ln(I) + \xi$	0,8454

Данные, приведенные в таблице, позволяют применить критерий F^2 Фишера для оценки значимости факторов. Фактор S оказался не значимым, а факторы I и T – значимыми. Следовательно, выражение (17) можно представить как

$$\ln(y) = a + b \ln(I) + cT + \xi. \quad (18)$$

Чтобы определить степень влияния факторов на заболеваемость, авторы определили коэффициенты регрессии для стандартизованных переменных в уравнении (18), т.е. для уравнения

$$v_y = \alpha_I v_I + \alpha_T v_T + \xi,$$

где α_I и α_T – искомые коэффициенты регрессии, а

$$v_y = \frac{\ln(y) - m_y}{\sigma_y}, v_I = \frac{\ln(I) - m_I}{\sigma_I}, v_T = \frac{T - m_T}{\sigma_T}$$

– стандартизованные переменные (буквами m и σ обозначены соответствующие средние и среднеквадратичные отклонения). Результат: $\alpha_I = 0,9007$ и $\alpha_T = -0,1446$. Следовательно, заболеваемость значительно больше зависит от числа инфицированных индивидуумов в регионе, чем от температуры воздуха. Этим объясняются хорошие результаты применения ARX-моделей для краткосрочного прогнозирования по временному ряду заболеваемости.

Однако влияние температуры воздуха, если отклонение от среднего значения продержится достаточно долго, может накапливаться день ото дня и оказать существенное влияние на развитие эпидемии. Для того чтобы продемонстрировать это, вернемся к мультипликативному представлению уравнения (18):

$$y = e^a \cdot I^b \cdot e^{cT},$$

где $a = -1,462$, $b = 0,8747$ и $c = -0,0073$. Изменение коэффициента

$$k_T = e^{cT}$$

по температурным данным на период эпидемии показано на рис. 4.

Как показывают несложные вычисления, если значение коэффициента k_T минимально, то даже для максимальной заболеваемости и макси-

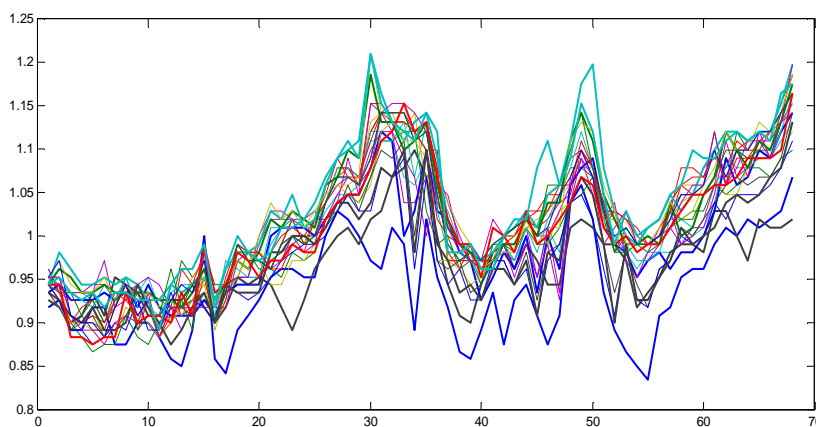


Рис. 4. Изменение коэффициента, обусловленного влиянием температуры воздуха на заболеваемость, в течение эпидемиологического периода

мального количества больных (по данным эпидемии 2009 года) заболеваемость падает до минимального уровня меньше чем за неделю. Но, поскольку заболеваемость никогда не равна нулю, то вирусная инфекция сохраняется, и при изменении погодных условий может возникнуть следующий всплеск заболеваемости. Таким образом, рассмотренная модель объясняет механизм распространения заболеваемости волнами.

5. Выводы

В статье изложены вопросы исследования влияния внешних параметров на развитие эпидемиологического процесса заболевания гриппом/ОРВИ. Предлагается подход, основанный на использовании реальных статистических данных в виде временных рядов заболеваемости, на основных положениях теории эпидемиологии и методах множественного регрессионного анализа. Исследования проведены для температуры воздуха как параметра среды в качестве примера. Подтверждается значимость данного фактора и объясняется возникновение волн заболеваемости его влиянием.

Результаты можно использовать для разработки систем моделирования и поддержки принятия решений при возникновении эпидемий на уровне регионов и страны в целом.

СПИСОК ЛИТЕРАТУРЫ

1. Смородинцев А.А. Грипп и его профилактика / Смородинцев А.А. – Л.: Медицина, 1984. – 383 с.
2. Бейли Н. Математика в медицине и биологии / Бейли Н. – М.: Мир, 1970. – 326 с.
3. Бароян О.В. Математика и эпидемиология / О.В. Бароян, Л.А. Рвачев.– М.: Знание, 1977. – 63 с.
4. Hethcote H. Qualitative analyses of communicable disease models / H. Hethcote // Math. Biosci. – 1976. – N 28. – P. 335 – 356.
5. Capasso V. Mathematical Structures of Epidemic Systems / Capasso V. // Lecture Notes in Biomath. – Heidelberg: Springer-Verlag, 1993. – N 97, XVI. – 284 p.
6. Сайт МОЗ: <http://www.moz.gov.ua/ua/portal>.
7. <http://www.gismeteo.ru/diary>.
8. Об одном подходе к прогнозированию эпидемиологической обстановки по гриппу/ОРВИ с использованием временных рядов / В.И. Вьюн, Т.К. Еременко, Г.Е. Кузьменко [и др.] // Математичні машини і системи. – 2011. – № 2. – С. 41 – 49.
9. Finkenstard B.F. Time series modeling of childhood diseases: a dynamical systems approach / B.F. Finkenstard, B.T. Grenfell // Appl. Statist. – 2000. – N 49, Part 2. – P. 187 – 205.
10. Льюинг Л. Идентификация систем. Теория пользователя / Льюинг Л. – М.: Наука, 1991. – 432 с.

Стаття надійшла до редакції 28.04.2012