

МАТЕМАТИЧЕСКИЕ ТЕХНОЛОГИИ В ПРОГНОЗИРОВАНИИ ЭПИДЕМИОЛОГИЧЕСКИХ ПОКАЗАТЕЛЕЙ ПО ТУБЕРКУЛЕЗУ У ДЕТЕЙ

О.В. Моисеева, А.В. Соклакова

О.В. Моисеева, доктор медицинских наук, ГБОУ ВПО «Ижевская государственная медицинская академия» Минздрава РФ; ФГБОУ ВПО

А.В. Соклакова, кандидат экономических наук, «Ижевский государственный технический университет» им. М.Т. Калашникова

Автор, ответственный за переписку: Моисеева Ольга Валерьевна, Ижевск, Воровского, 156-104, 426063, ovm@e-izhevsk.ru

Резюме

Для глубокого изучения закономерностей туберкулезного эпидемического процесса необходимо применение системного анализа, одним из инструментов которого является математическое моделирование. В целях системного анализа значительный интерес представляют собой исследования параметров системы методом главных компонент. Моделирование тенденции временного ряда показателя заболеваемости туберкулезом детей при наличии структурных изменений возможно оценивать с помощью теста Грегори Чоу. Также для прогнозирования основных эпидемиологических показателей по туберкулезу среди детей и подростков Удмуртской Республики и определения влияния факторов риска используется многофакторный корреляционно-регрессионный анализ. Априорную вероятность заболевания можно считать в соответствии с формулой неинформативного распределения.

Ключевые слова: прогнозирование, эпидемиологические показатели, математическое моделирование, туберкулез, дети.

MATHEMATICAL TECHNOLOGIES IN FORECASTING EPIDEMIOLOGICAL INDEXES ON A TUBERCULOSIS AT CHILDREN

O.V. Moiseeva, A.V. Soclakova

The summary

For penetrating study of legitimacies of tubercular epidemic process the application of systems analysis is necessary, one of which instruments is the mathematical model operation. With the purposes of systems analysis the considerable interest is represented by examinations of parameters of system by a method a main builder. The model operation of the tendency of time series of an index of a case rate by tuberculosis of children at presence of structural changes is possible for estimating with the help of the paste Gregory Chou. Also for forecasting the basic epidemiological indexes on tuberculosis among children both teenagers of the Udmurt Republic and definition of influence of risk factors will be utilized multifactor correlation regression the analysis. It is possible to consider a prior probability of disease according to the formula uninformative of allocation.

Key words: forecasting, epidemiological indexes, mathematical model operation, tuberculosis, children.

Туберкулез к концу XX века вновь стал одной из актуальных проблем мирового здравоохранения. В 1993 г. в связи с ухудшением эпидемической ситуации в мире Всемирная Организация Здравоохранения объявила туберкулез глобальной опасностью [1].

В России подъем эндемии туберкулеза начался с середины 80-х годов. Основной причиной обострения эпидемической ситуации в России явилось ухудшение социально-экономических условий жизни населения. При этом задача контроля эпидемического процесса туберкулеза не может быть решена без оперативного прогнозирования эпидемической ситуации. Прогнозирование эпидемического процесса основывается на выявлении и изучении закономерностей его течения. Для глубокого изучения закономерностей процесса необходимо применение системного анализа, одним из инструментов которого является математическое моделирование [1, 2].

Применение методов математического моделирования в целях прогнозирования эпидемического процесса туберкулеза проводили некоторые авторы ранее. Разработанные ими компартментные математические модели, дающие долгосрочные прогнозы и удобные для исследовательских целей, имеют существенную ошибку при краткосрочном прогнозировании [1, 3, 4, 5].

Для краткосрочных прогнозов эпидемиологических показателей по туберкулезу применялись методы регрессионного анализа. Авторы давали прогнозы в виде сценариев развития событий, то есть предполагался определенный уровень одних параметров, а по ним с использованием математических моделей вычислялись другие. При этом не выявлялись с помощью кросскорреляционного анализа и не использовались для прогнозирования параметры, изменения в которых происходят раньше во времени [1, 6].

В целях системного анализа значительный интерес представляют собой исследования параметров системы методом главных компонент. Этот метод позволяет исследовать структуру взаимосвязей и вклады независимых переменных в главные компоненты, что важно для выявления показателей, вносящих максимальную дисперсию в изучаемый процесс [1, 7, 8, 9, 10].

Статистическая и аналитическая обработка материала, графическая визуализация результатов как правило проводится с помощью современной компьютерной техники Pentium в системе Microsoft Windows 2000 и XP с использованием стандартных пакетов прикладных компьютерных программ Microsoft Office 2003-2007, XP; Microsoft Word, Microsoft Excel [11, 12].

Приведем примеры использования математических технологий в прогнозировании эпидемиологических показателей по туберкулезу у детей Удмуртии.

Для определения влияния химиопрофилактики на основные эпидемиологические показатели: заболеваемость туберкулезом, РПИ, инфицированность микобактериями туберкулеза детей и подростков нами

использован метод многомерной статистики – теорема гипотез (формула Байеса), что позволило прогнозировать некоторые эпидемиологические показатели у детей и подростков при различных уровнях одного или нескольких изучаемых факторов. Расчет проводился по формуле В.Е. Гмурман (2003):

$$PA_i(x_1 \cdot x_2 \cdot \dots \cdot x_n) = \frac{PA_i(P_{x_1|A_i} \cdot P_{x_2|A_i} \cdot \dots \cdot P_{x_n|A_i})}{PA_i(P_{x_1|A_i} \cdot P_{x_2|A_i} \cdot \dots \cdot P_{x_n|A_i}) + PA_k(P_{x_1|A_k} \cdot P_{x_2|A_k} \cdot \dots \cdot P_{x_n|A_k})}, \quad (1)$$

где, $PA_i(x_1 x_2 \dots x_n)$ – уровень эпидемиологических показателей в долях единицы в зависимости от совокупного влияния ряда факторов; $PA_i(P_{x_1|A_i})$ – фактическая величина (в долях единицы) распространенности изучаемого явления; PA_k – фактическая величина состояния противоположного изучаемому и дополняющая его до единицы; P_{x/A_i} и P_{x/A_k} – число лиц (в долях единиц) с наличием того или иного признака в основной и контрольной группах.

Моделирование тенденции временного ряда показателя заболеваемости туберкулезом детей России и Удмуртии при наличии структурных изменений мы оценивали с помощью теста Грегори Чоу. Использование теста позволило оценить значимость влияния структурных изменений проводимых профилактических противотуберкулезных мероприятий, внедренных в практическое здравоохранение в середине XX века, изменения социально-экономического курса на характер тенденции ряда заболеваемости туберкулезом. Для этого были построены линейно-кусочные модели регрессии, начиная с 1970 г. Таким образом, исходная совокупность была разделена на две подсовкупности с определением точек перегиба. Сделаны выводы о структурной стабильности временного ряда с отклонением или принятием нулевой гипотезы. Соблюдены предпосылки о нормальном распределении остатков в уравнениях и независимость их распределений [13].

Чаще всего тест Чоу используется в экономике для анализа наличия структурных изменений временных рядов. В частности, с помощью этого теста определяют стабильность временных рядов курсов акций и валюты, или стабильность прироста акционерных компаний в России [14, 15]. Кроме того, тест Чоу пригоден для установления необходимости оценивать одну объединенную регрессию или отдельные регрессии для каждой подвыборки, если выборка наблюдений состоит из двух и более подвыборок [16, 17]. В других областях науки и практической деятельности тест Чоу используется, в частности, для выделения сегментов обработки речи на основе анализа корреляционных зависимостей в речевом сигнале [18].

Априорную вероятность заболевания можно считать в соответствии с формулой неинформативного распределения Э.Т. Джейнса (1968):

$$p(A)=p(B)=p(C)=\frac{a}{n} \quad (2)$$

где $p(A)=p(B)=p(C)$ - вероятности событий, a - число единиц наблюдения, у которых имеется данная степень риска заболевания, n - общее число единиц наблюдения.

Исследование связей в условиях массового наблюдения и действия случайных факторов осуществлялось с помощью статистических моделей, представляющих логическое или математическое описание компонентов и функций, отображающих существенные свойства моделируемого объекта или процесса. Это дало возможность установить основные закономерности изменения оригинала. Выражение и модели в виде функциональных уравнений использованы нами для расчёта средних значений моделируемого показателя по набору заданных величин и для выявления степени влияния на него отдельных факторов. В нашей работе применена многофакторная модель и пакет программы Statistica 6.0.

Важнейшим этапом построения регрессионной модели (уравнения регрессии) являлось установление в процессе анализа исходной информации математической функции. Были использованы уравнения прямолинейной

связи. Адекватность построенных статистических моделей проверялась, т.к. была исследована не вся генеральная совокупность. Применен корреляционный анализ, являющийся одним из методов статистического анализа взаимосвязи нескольких признаков. Изучалась множественная корреляция – зависимость двух или более факторных признаков, включенных в исследование. Теснота связи количественно выражается величиной коэффициентов корреляции. Коррелограмма (автокоррелограмма) показывает численно и графически коэффициенты автокорреляции (и их стандартные ошибки) для последовательности лагов из определенного диапазона. В состав Microsoft Excel входил набор средств анализа данных, который позволил представить результаты работы в графическом виде.

Также для прогнозирования основных эпидемиологических показателей по туберкулезу среди детей и подростков Удмуртской Республики и определения влияния факторов риска на них был проведен многофакторный корреляционно-регрессионный анализ.

Для этого проведен анализ динамического ряда показателей заболеваемости туберкулезом детей, рассчитан коэффициент автокорреляции, процент объясненной дисперсии и ошибка аппроксимации. В результате получаем уравнение тренда, например, по данным заболеваемости за исследуемые годы. Считаем прогноз заболеваемости туберкулезом детей на последующие пять лет.

На следующем этапе уточняем коэффициенты парной линейной корреляции между изучаемыми факторами и эпидемиологическими показателями. В дальнейшем мультиколлинеарность дает возможность заменить все эти факторы каким-то одним, оказывающим наибольшее влияние на результативный признак.

Для построения линейной модели влияния ряда факторов, например на заболеваемость туберкулезом детей необходимо выбрать фактор, связь с которым наиболее существенна. Получаем: множественный R (показатель множественной корреляции), который свидетельствует о тесноте связи

между включенными в модель факторами и результативным признаком, т.к. данный коэффициент принимает значения $[0; 1]$, и чем он ближе к единице, тем связь теснее; R-квадрат, который объясняет процент вариации результирующего признака.

На заключительном этапе производится оценка критерия Фишера F, который сравнивается с табличными значениями при заданном числе степеней свободы и вероятности 95,0%. Таким образом, с вероятностью 95,0% можно утверждать, что уравнение статистически значимо или получено случайным образом.

Таким образом, в современных условиях можно использовать многообразные математические технологии для проведения прогнозирования эпидемической ситуации по туберкулезу у детей.

Список литературы:

1. Чижова О.В. Исследование и прогнозирование эпидемиологических показателей туберкулезного процесса: автореф. дис. ... канд. мед. наук. М., 2004.
2. Альмитов Р.А. Прогнозирование туберкулеза легких по медико-социальным факторам риска и пути совершенствования его профилактики : автореф. дис. ... канд. мед. наук. Казань, 2005.
3. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика : основы моделирования и первичная обработка данных. М.: Финансы и статистика. 1983.
4. Герасимов А.Н. Медицинская статистика. М.: Медицинское Информационное Агентство. 2007.
5. Гмурман В.Е. Теория вероятностей и математическая статистика. М., 2003.
6. Жижин К.С. Медицинская статистика. Ростов н /Д, Феникс. 2007.
7. Мельниченко О.А. Моделирование распространения туберкулеза и анализ факторов, влияющих на эпидемический процесс : автореф. дис. ... канд. физ.- мат. наук. М., 2008.

8. Мостеллер Ф., Тьюки Дж. Анализ данных и регрессия. М.: Финансы и статистика. 1982.
9. Симанков В.С., Халафян А.А. Системный анализ и современные информационные технологии в медицинских системах поддержки принятия решений. М.: БИНОМ, 2009.
10. Суфиянов В.Г. Разработка адаптивных статистических моделей классификации и прогнозирования : автореф. дис. ... канд. физ.-мат. наук. Пермь, 2004.
11. Аффифи А., Эйзен С. Статистический анализ : подход с использованием ЭВМ. М.: Мир. 1982.
12. Бюллер А.Б. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей : пер. с нем. СПб.: ДиаСофтЮП. 2001.
13. Gregory Ch. Nests of equality between sets of coefficients in two linear regressions. *Econometrica*. 1960. 28 (3): 591-605.
14. Кретинин И. А. Моделирование числовых характеристик финансовых активов в задаче формирования портфеля ценных бумаг: автореф. дисс... к.э.н. Воронеж, 2011.
15. Дроздова Н.П., Кормилицына И.Г. Экономическая политика государства и формирование инвестиционного климата: опыт России конца XIX — начала XX вв. СПб., ВШМ 2010. № 8.
16. Каморников С. Ф., Каморников С.С. Эконометрика. М.: Интеграция, 2012.
17. Отчет о научно-исследовательской работе «Моделирование принятия решений в социальных и экономических системах», / рук. Ф.Т. Алескеров, НИУ ВШЭ, Москва, 2011.
18. Патент на изобретение №:2445718, Автор: Афанасьев Андрей Алексеевич (RU), Новиков Евгений Иванович (RU), Трубицын Владимир Геннадьевич (RU), Титов Олег Николаевич (RU), Патентообладатель: Государственное образовательное учреждение высшего профессионального

образования Академия Федеральной службы охраны Российской Федерации (Академия ФСО России) (RU), начало действия патента: 31 Августа, 2010.