

УДК 004.942

МОДЕЛИРОВАНИЕ И ПРОГНОЗ ЗАБОЛЕВАЕМОСТИ ВИЧ-ИНФЕКЦИЕЙ В ПЕРМСКОМ КРАЕ

Давыд Юлия Константиновна

Студент;

Березниковский филиал Пермского национального исследовательского политехнического университета;

618400, Пермский край, г. Березники, ул. Тельмана, 7;

e-mail: julia.davyd@gmail.com.

Произведено моделирование заболеваемости ВИЧ-инфекцией в Пермском крае с использованием линейной многофакторной модели, модели в пространстве состояний и авторегрессионной модели 3 порядка, а также оценка возможностей прогнозирования этих моделей для дальнейшей поддержки принятия решений.

Ключевые слова: ВИЧ, факторная модель, бесфакторная модель, прогнозирование, принятие решений.

MODELING AND PROGNOSIS OF THE INCIDENCE OF HIV INFECTION IN THE PERM KRAI

Davyd Yuliya

Student;

Berezniki branch of the Perm National Research Polytechnic University;

The Address of the Organization: 618400, Berezniki, Perm reg., Telmana str., 7;

e-mail: julia.davyd@gmail.com.

Modeling of the incidence of HIV infection in the Perm Krai was carried out. For this, a linear multi-factor model, a model in the state space and an autoregressive model of the third order are used. The possibility of predicting these models for further support of decision making was also evaluated.

Keywords: HIV, factor model, factorless model, forecasting, decision making.

Введение

ВИЧ-инфекция – заболевание, вызванное вирусом иммунодефицита человека. Вирус поражает клетки иммунной системы, в результате чего организм больного теряет возможность защищаться от инфекций и опухолей [1].

Проблема ВИЧ-инфекции уже более 30 лет продолжает оставаться актуальной для мирового сообщества и неоднократно рассматривалась на самом высоком мировом уровне, включая Генеральную ассамблею ООН, саммиты «Группы восьми», СНГ и ЕвроАзЭС. Масштабы ее распространения приобрели глобальный характер и представляют реальную угрозу социально-экономическому развитию большинства стран мира.

Согласно сведениям экспертов ПЕПФАР, в 2017 году Россия заняла лидирующую позицию по темпам распространения ВИЧ-инфекции [2].

Как указано в проекте федерального закона о бюджете на 2018 год, в этом году финансовая поддержка профилактики ВИЧ-инфекции и вирусных гепатитов в России составит 16,9 млрд. рублей, что подчеркивает важность данной проблемы.

Если рассматривать уровень заболеваемости ВИЧ-инфекцией в субъектах РФ, то одним из субъектов, где данная проблема стоит особо остро, является Пермский край. На сегодняшний день он занимает 5 место по уровню заболеваемости ВИЧ-инфекцией в России.

Одним из способов исследования развития ВИЧ-инфекции является моделирование. Так в [3] рассматриваются различные способы математического моделирования ВИЧ-инфекции, которые позволяют исследовать иммунологическую динамику в рамках одного организма, а также исследо-

вать и спрогнозировать ситуацию развития ВИЧ на популяционном уровне. Однако представленные в данной работе модели сложные и требуют большого количества данных, к которым не всегда имеется доступ (численность индивидов восприимчивых к ВИЧ, численность инфицированных в исследуемой группе, численность больных СПИДом, доля индивидов, приобретающая инфекционность группы после инфицирования и пр.). Модели, представленные в данной работе, основываются на меньшем объеме данных, предоставленных в общем доступе.

Целью настоящей работы является моделирование заболеваемости ВИЧ-инфекцией в Пермском крае для поддержки принятия решений по развитию ситуации.

1. Исходные данные

Для построения моделей используются годовые ряды данных о количестве заболевших ВИЧ-инфекцией в Пермском крае, взятые с официального сайта «Пермьстат» за последние 12 лет.

В качестве факторов, которые потенциально могут оказывать влияние на заболеваемость ВИЧ, выбраны следующие:

- x_1 – количество безработных (чел);
- x_2 – количество людей с инфекционными заболеваниями (чел);
- x_3 – численность населения с денежными доходами ниже величины прожиточного минимума (чел);
- x_4 – число преступлений, связанных с изнасилованием (шт);
- x_5 – число преступлений, связанных с наркотиками (шт);
- x_6 – число разводов (шт);
- x_7 – среднедушевой денежный доход населения (руб).

Таблица 1. Исходные данные

	2006	2007	2008	2009	2010	2011
y	6898	7557	9171	10845	12906	15339
x_1	98500	93800	122600	146300	122600	109300
x_2	147300	143200	146000	131500	122900	113000
x_3	393249,6	375111,6	387301,2	394375	363423	379300
x_4	412	291	222	207	171	159
x_5	3897	3948	3572	3556	3348	3126
x_6	11035	12196	12295	11768	10657	11322
x_7	10982,3	13481	16119	17640,8	19426,5	20640,4

Таблица 2. Исходные данные (продолжение)

	2012	2013	2014	2015	2016	2017
y	17322	19975	22 916	26284	28585	29199
x_1	84900	86600	74800	81700	76500	82700
x_2	115900	97400	101200	93200	92600	93691
x_3	321000	300700	315500	323300	390300	392500

x_4	166	166	121	115	108	128
x_5	4416	4261	4932	4245	3849	3902
x_6	11450	12428	12569	10667	10952	11156
x_7	23328,8	26054,3	28315,5	32042,6	28399,8	28399,8

Для того чтобы значения данных попадали в сопоставимые по величине интервалы, произведем нормировку критерия и факторов системы. Нормировка данных производится по формуле:

$$x_{\text{норм}} = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}}, \tag{1}$$

где: $x_{\text{норм}}$ – нормированное значение параметра системы; x – текущее значение параметра системы; x_{min} – минимальное значение параметра системы; x_{max} – максимальное значение параметра системы.

Далее произведем анализ предварительно выбранных факторов с точки зрения их взаимной корреляции, а также корреляции с критерием. Факторы с высокой взаимной корреляцией подлежат исключению линейной многофакторной модели, так как такие факторы имеют линейную зависимость. Факторы, у которых корреляция с критерием близка к нулю, так же исключаются из моделей, так как, предположительно, не влияют на критерий [4]. Парную корреляцию рядов определим с помощью функции КОРРЕЛ (ряд1; ряд2) в MS Excel. Полученные значения представлены в табл. 2.

Таблица 3. Корреляционный анализ факторов и критерия

	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_1	1	0,6261	0,4855	0,3072	-0,7107	0,0137	-0,6136
x_2		1	0,4683	0,8428	-0,3792	0,2045	-0,9635
x_3			1	0,3909	-0,7142	-0,2914	-0,5246
x_4				1	-0,1915	0,0799	-0,8838
x_5					1	0,3785	0,4722
x_6						1	-0,1547
x_7							1
y	-0,6721	-0,9566	-0,3147	-0,8089	0,4007	-0,2491	0,9557

Корреляционный анализ показал, что факторы x_2 и x_7 имеют высокую взаимную корреляцию и необходимо исключить один из этих факторов из линейной многофакторной модели. Исключим фактор x_7 , так как он имеет меньшую корреляцию с критерием системы.

2. Построение моделей

Перейдем к построению линейной многофакторной модели. Формула данной модели имеет вид:

$$y(t_i) = \sum_{j=1}^N a_j \cdot x(t_i) + b, \tag{2}$$

где: $y(t_i)$ – значение критерия в момент времени t_i ; a_j , b – коэффициенты линейной многофакторной модели; $x(t_i)$ – значение фактора в момент времени t_i .

Найдем коэффициенты модели, воспользовавшись мастером «Поиск решения» *MS Excel*. Полученные коэффициенты представлены в табл. 3.

Таблица 4. Коэффициенты линейной многофакторной модели

a_0	a_1	a_2	a_3	a_4	a_5	a_6
0,5431	-0,1366	-0,6640	0,3328	-0,3099	0,3706	-0,0867

Критерием оценки системы является количество заболевших ВИЧ-инфекцией, следовательно, из полученных коэффициентов можно понять, что улучшению критерия (снижению заболеваемости) способствуют:

- увеличение количества безработных;
- увеличение количества людей с инфекционными заболеваниями;
- уменьшение численности населения с денежными доходами ниже величины прожиточного минимума;
- увеличение числа преступлений, связанных с изнасилованием;
- уменьшение числа преступлений, связанных с наркотиками;
- увеличение числа разводов.

Большинство выводов противоречит здравому смыслу и отсюда следует, что использовать эту модель для дальнейшего прогноза и поддержки принятия решений нельзя [5].

Перейдем к построению модели в пространстве состояний. Данная модель представлена двумя формулами. Формула расчета факторов:

$$\bar{x}'(t_{j+1}) = \bar{a} + B\bar{x}(t_j), \quad (3)$$

где: вектор $\bar{x}'(t_{j+1})$ – вычисляемые значения всех факторов в момент времени t_{j+1} ; вектор \bar{a} – свободные коэффициенты; матрица B – коэффициенты при факторах; вектор $\bar{x}(t_j)$ – значения всех факторов в момент времени t_j .

Формула расчета критерия:

$$y'(t_j) = c + \bar{d} \cdot \bar{x}'(t_j), \quad (4)$$

где: $y'(t_j)$ – расчетное значение критерия в момент времени t_j ; c – свободный коэффициент; вектор \bar{d} – коэффициенты при критериях; вектор $\bar{x}'(t_j)$ – значения критериев в момент времени t_j .

Результат построения модели представлен на рис. 1.

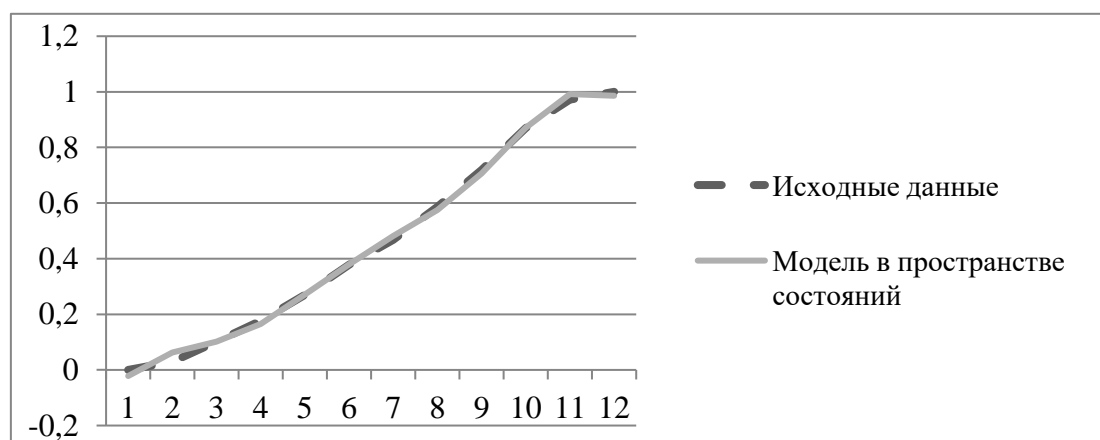


Рис. 1. График нормированного критерия (ось ординат) в зависимости от номера года (ось абсцисс), полученный с использованием модели в пространстве состояний

Так как данная модель с достаточной точностью аппроксимирует исходные данные, проверим ее способности к прогнозированию с помощью метода постпрогноза. Данный метод заключается в расчете реакции системы по модели при известных рядах факторов за последние несколько лет.

Произведем расчет постпрогноза для данной модели на 1, 2 и 3 года. Результат работы представлен на рис. 2.

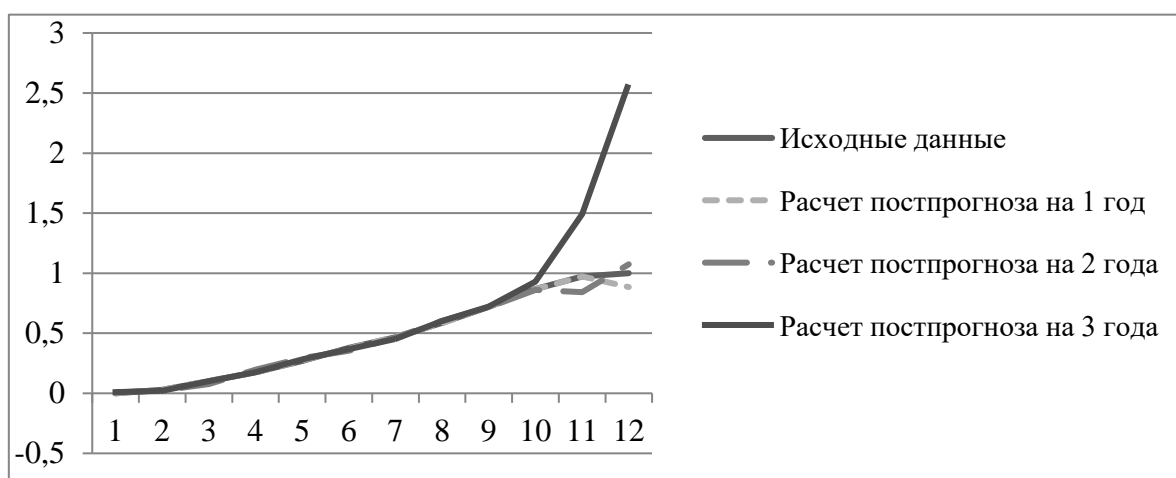


Рис. 2. Расчет постпрогноза нормированного критерия с помощью модели в пространстве состояний

Из рисунка видно, что модель в пространстве состояний в данном случае прогнозирует плохо и использовать ее для дальнейшего прогноза и поддержки принятия решений нецелесообразно.

Учитывая, что ни одна из моделей представленных выше нам не подошла, можно сделать вывод, что факторные модели в данном случае использовать не получится. Следовательно, можно попробовать использовать бесфакторную модель. К таким моделям относятся тренд и авторегрессионные модели 1, 2 и 3 порядка. Для выбора модели вычислим квадратичную погрешность аппроксимации для каждой из них:

$$S = \sum_{i=1}^N (y(t_i) - y_{расч}(t_i))^2, \tag{5}$$

где: S – квадратичная погрешность аппроксимации; $y(t_i)$ – значение критерия в момент времени t_i ; $y_{расч}(t_i)$ – расчетное значение критерия для момента времени t_i .

Таблица 5. Квадратичные погрешности аппроксимации бесфакторных моделей

$S_{тренд}$	0,022
$S_{АвРМ1}$	0,0158
$S_{АвРМ2}$	0,0042
$S_{АвРМ3}$	0,0036

Так как наименьшую квадратичную погрешность аппроксимации имеет авторегрессионная модель 3 порядка, для дальнейшей работы будем использовать ее. Данная модель представлена формулой:

$$y(t_{i+1}) = a_0 y(t_i) + a_1 y(t_{i-1}) + a_2 y(t_{i-2}) + b, \tag{6}$$

где: $y(t_{i+1})$, $y(t_i)$, $y(t_{i-1})$, $y(t_{i-2})$ – значения критерия системы в моменты времени t_{i+1} , t_i , t_{i-1} и t_{i-2} соответственно; a_0 , a_1 , a_2 , b – коэффициенты авторегрессионной модели 3 порядка.

Результат построения модели изображен на рис. 3.

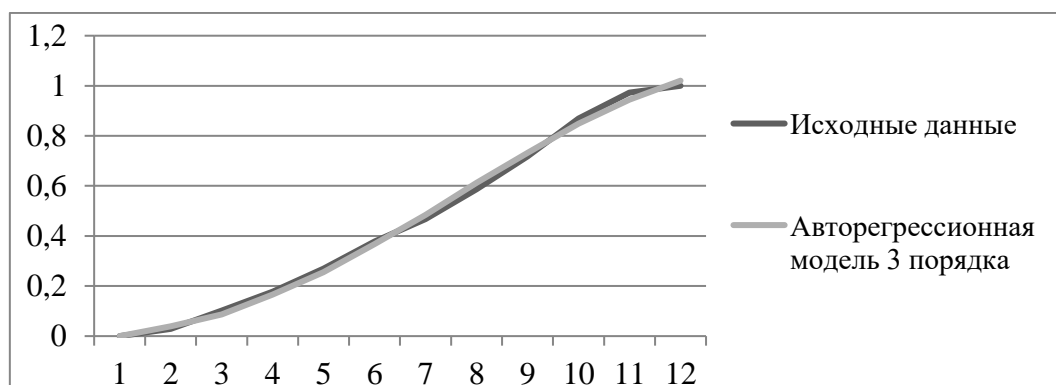


Рис. 3. График нормированного критерия (ось ординат) в зависимости от номера года (ось абсцисс), полученный с использованием авторегрессионной модели 3 порядка

Так как данная модель с достаточной точностью аппроксимирует исходные данные, произведем прогноз значения критерия системы с помощью этой модели на следующие 5 лет. Так как мы выбрали бесфакторную модель, последующие состояния системы будут зависеть только от предыдущих. Полученный прогноз представлен на рис. 4.

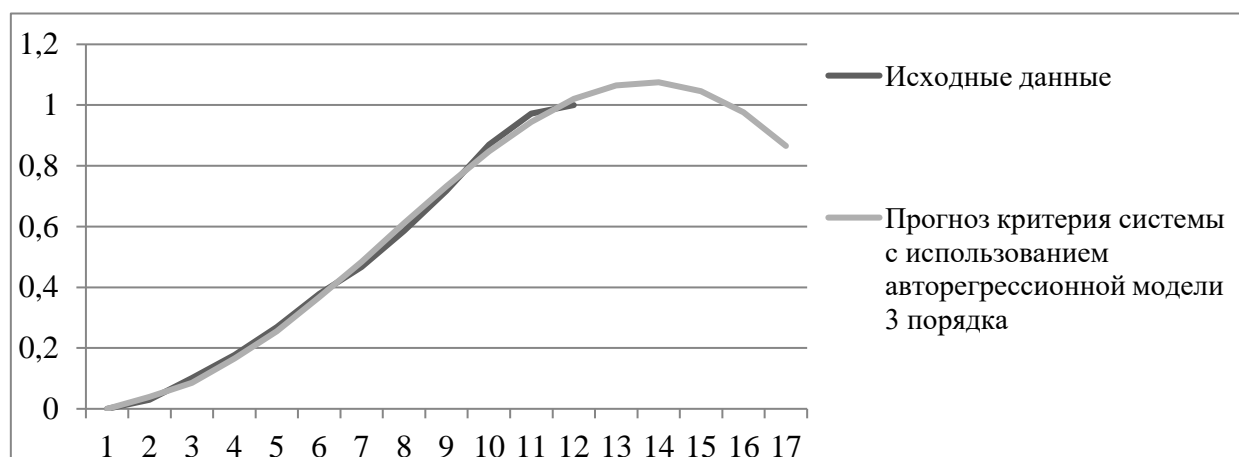


Рис. 4. Прогноз нормированного значения критерия системы на 5 лет

Заключение

Отсюда можно сделать вывод, что количество заболевших ВИЧ-инфекцией в Пермском крае к 2019 году возрастет на 7,5% и составит 31388 человек, но далее имеет тенденцию к снижению по параболе и к 2022 году снизится на 13,4% и составит 25286 человек. Судя по выводам из исследования линейной многофакторной модели и модели в пространстве состояний, с помощью изменения выбранных факторов управлять заболеваемостью ВИЧ в Пермском крае невозможно.

Список литературы

1. Вирус иммунодефицита человека [Электронный ресурс] – режим доступа: https://ru.wikipedia.org/wiki/Вирус_иммунодефицита_человека.
2. Special Briefing Deborah L. Birx, M.D., U.S. Global AIDS Coordinator and Special Representative for Global Health Diplomacy Ambassador, Washington, DC, November 30, 2017. [Электронный ресурс]. — Режим доступа: <https://www.state.gov/r/pa/prs/ps/2017/11/276123.htm>.
3. Носова Е. А. Модели контроля и распространения ВИЧ-инфекции // Матем. биология и биоинформ. — 2012. — Т. 7. — Вып. 2. — С.632-675.
4. Сиротина Н.А., Янченко Т.В., Затонский А.В. Об аппроксимации факторов дифференциальной модели социально-экономической системы // Современные исследования социальных проблем

(электронный журнал). – 2012. — № 11(19). [Электронный ресурс]. — Режим доступа: <http://sisp.nkras.ru/e-ru/issues/2012/11/sirotina.pdf>.

5. Янченко Т.В., Затонский А.В. Определение оптимальной ранжировки частных критериев оценки краевого социального ресурса // Экономика и менеджмент систем управления. — 2013. — Т. 10. — № 4. — С. 99-104.