

УДК 614.39

СТАТИСТИЧЕСКИЙ ПРОГНОЗ ЗДОРОВЬЯ НАСЕЛЕНИЯ РЕГИОНА НА ОСНОВЕ МАТЕМАТИЧЕСКОГО И КОМПЬЮТЕРНОГО МОДЕЛИРОВАНИЯ

М.С.Токмачев

Институт электронных и информационных систем НовГУ, tokm@mail.natm.ru

Представлена методика оценки исчерпанной заболеваемости населения. На основе персонифицированной базы данных осуществлен вероятностный прогноз заболеваемости и смертности населения региона.

Ключевые слова: *цепь Маркова, прогнозирование, регрессионная модель, заболеваемость по обращаемости, исчерпанная заболеваемость*

The method of estimate of the population real morbidity is presented. On the basis of personified database the probability forecast of morbidity and death rate of the region population is embodied.

Keywords: *Markov chain, forecasting, regression model, morbidity of negotiability, real morbidity*

Введение

Одной из глобальных проблем современного российского общества является многолетняя устойчивая тенденция противоестественной убыли населения, объясняемая не столько низкой рождаемо-

стью, сколько запредельно высоким уровнем смертности, сопровождаемая стабильным ростом скрытой (не диагностированной) заболеваемости. Отсутствие информации о своем здоровье является серьезным фактором риска для большей части населения. Вместе с тем накопленная статистика заболеваемости и

смертности населения позволяет подойти к проблеме изучения общественного здоровья, основываясь на математических методах, используя имеющиеся массивы информации с максимальной эффективностью.

В представленном исследовании использованы методы статистического анализа и компьютерного моделирования. Теоретической базой служит авторская методика [1-3]. В основе практической численной реализации лежит единая Персоналифицированная база данных (ПБД) здоровья населения Новгородской области [4], сформированная в Новгородском научном центре СЗО РАМН и охватывающая все застрахованное население области (692 890 человек, 2000 — 2005 гг.), а также Персоналифицированная база данных выборочного обследования здоровья населения Новгородской области, 2005 г. [4-6]. Для обработки массивов ПБД создано программное обеспечение [4,7].

Отметим, что «заболеваемость» в различных базах данных формируется как «заболеваемость по обращаемости», в частности таковой является вышеуказанная единая ПБД. В то же время исследование выборки (здоровье населения Новгородской области, 2005 г.) показывает, что уровни реальной, истинной заболеваемости, называемой «исчерпанной заболеваемостью», для большинства классов болезней в отдельных половозрастных группах населения в разы отличаются от уровней официальной статистики «заболеваемости по обращаемости». Исследование смертности по причинам также выявляет слабую информативность показателя «заболеваемость по обращаемости»: в целом случаи «внезапной» и «насильственной» смерти составляют 23,4% среди всех умерших [8]. Это несоответствие официальных и реальных значений показателей связано с различными причинами: степенью развития профилактической медицины, менталитетом населения, качеством диагностики, уровнем доступности медицинских услуг и т.д. Следовательно, решение проблемы заболеваемости и убыли населения лежит вне плоскости математического исследования. Вместе с тем, опираясь на данные истинной заболеваемости по половозрастным группам и классам болезней, можно более точно оценить проблемы здоровья населения и здравоохранения. В частности, уровни исчерпанной заболеваемости используются при построении и прогнозировании интегральных показателей здоровья населения [9,10]. Масштабы и соотношения классов истинной заболеваемости в возрастно-половом аспекте должны учитываться органами управления здравоохранением.

Целью представленного в статье исследования является реализация алгоритма нахождения оценки исчерпанной заболеваемости в динамике на ближайшую и отдаленную (на протяжении поколения) перспективы. На основе именно исчерпанной заболеваемости производится пересчет значений показателей здоровья, находимых по единой ПБД с помощью разработанного программного комплекса [4]. Осуществ-

ляется уточнение тенденций и прогнозов, связанных со здоровьем и смертностью населения.

Построение модели исследования

В [1-3] представлена методика и алгоритмизация исследования здоровья и смертности населения в динамике с отображением существующих тенденций на последующий (прогнозный) период времени.

Состояния здоровья индивидуума или исследуемой группы населения на протяжении всей жизни от рождения до смерти рассматриваются в последовательные фиксированные моменты времени. Полученные действия можно интерпретировать как последовательность испытаний, в которых каждый последующий исход зависит лишь от предыдущего и в то же время не является однозначно предопределенным, а осуществляется с некоторой вероятностью. Тогда сама человеческая жизнь — случайное блуждание по состояниям здоровья, которое заканчивается поглощающим состоянием «смерть» [3]. Этот процесс блуждания по состояниям, рассматриваемый как неоднородная цепь Маркова, моделируется на компьютере.

Под системой в нашем случае понимается конкретный индивидуум или группа индивидуумов, представляющих население региона (в данном конкретном случае, население Новгородской области). Возможные исходы испытаний принято называть (в понятиях математического аппарата цепей Маркова) состояниями системы. Состояние системы в каждый из фиксированных моментов времени — состояние здоровья.

Классификация состояний здоровья индивидуума, при наличии значимых заболеваний как на начальном, так и на конечном этапе наблюдения, осуществляется по доминирующему заболеванию согласно кодировке Международной классификации болезней (МКБ-10) и степени тяжести, определяемой для всех имеющихся заболеваний (табл.1, состояния E_1, E_2, \dots, E_{19}). Для однозначного выбора состояния E_j ($j=1, \dots, 19$) при наличии у индивидуума заболеваний разных классов, но одинаковой степени тяжести, вводится иерархия классов. Иерархия классов — упорядочивание классов болезней в соответствии с реальными данными уровней смертности населения по причине болезни из соответствующего класса. Чем больше уровень смертности в классе болезней, тем выше его место в данной классификации. Отметим, что по данным государственной статистики в среднем на одного жителя области приходится около двух заболеваний, поэтому для однозначной констатации состояния является необходимым введение доминирующего заболевания и иерархии классов. Для исследования состояний здоровья всего населения в динамике кроме заболеваемости введены состояния E_0 — «относительно здоров», и E_{20} — «смерть». Таким образом, состояния E_0, E_1, \dots, E_{20} образуют полную систему событий.

Классификация состояний здоровья по кодировке МКБ-10

Состояние	Класс	
E_0	—	«Относительно здоров»
E_1	I00-I99	Болезни системы кровообращения
E_2	S00-T98	Травмы, отравления и некоторые другие последствия воздействия внешних причин
E_3	C00-D48	Новообразования
E_4	J00-J99	Болезни органов дыхания
E_5	R00-R99	Симптомы, признаки и отклонения от нормы, выявленные при клинических и лабораторных исследованиях, не классифицированные в других рубриках
E_6	K00-K93	Болезни органов пищеварения
E_7	A00-B99	Некоторые инфекционные и паразитарные болезни
E_8	G00-G99	Болезни нервной системы
E_9	N00-N99	Болезни мочеполовой системы
E_{10}	F00-F99	Психические расстройства и расстройства поведения
E_{11}	D50-D89	Болезни крови, кроветворных органов и отдельные нарушения, вовлекающие иммунный механизм
E_{12}	E00-E90	Болезни эндокринной системы, расстройства питания и нарушения обмена веществ
E_{13}	M00-M99	Болезни костно-мышечной системы и соединительной ткани
E_{14}	Q00-Q99	Врожденные аномалии (пороки развития), деформации и хромосомные нарушения
E_{15}	L00-L99	Болезни кожи и подкожной клетчатки
E_{16}	O00-O99	Беременность, роды и послеродовой период
E_{17}	P00-P96	Отдельные состояния, возникающие в перинатальном периоде
E_{18}	H60-H95	Болезни уха и сосцевидного отростка
E_{19}	H00-H59	Болезни глаза и его придаточного аппарата
E_{20}	—	«Смерть»

Практическая составляющая при проведении исследования — реальные данные состояния здоровья населения (ПБД), рассматриваемые в динамике. По ним рассчитаны конкретные значения переходных вероятностей из состояния в состояние, сформированные в виде стохастических матриц, для различных половозрастных групп за пятилетний период (шаг процесса в исследовании — 5 лет).

Введенные состояния E_j ($j = 0, 1, \dots, 20$) и наборы стохастических матриц позволяют по начальным состояниям вычислять безусловные вероятности состояний [11], т.е. вероятности $p_j(t)$, с которыми система будет находиться в каждом из состояний E_j , включая смерть, через момент времени, кратный длине временного интервала (в нашем случае через 5, 10, 15, ... лет). В частности для мужского поколения 2000 г. рождения (поколения 0 лет с начальным состоянием E_0) совокупности вероятностей будущих состояний здоровья для середины каждого последующего возрастного интервала записаны в виде векторов (строки), образующих матрицу (табл.2).

Соответствующие матрицы безусловных вероятностей можно получить [4, 11] и для иных начальных возрастов и распределений по состояниям E_0, E_1, \dots, E_{20} .

Оценка исчерпанной заболеваемости и прогноз

Данные заболеваемости по обращаемости по доминирующему заболеванию в разрезе 2005 г. как результат авторского исследования [11] представлены для каждого класса болезней во всех возрастных интервалах. В том же виде сгруппированы данные выборочного исследования здоровья населения региона (репрезентативная выборка объемом 10 тыс. чел., 2005 г.), которые ввиду всестороннего углубленного обследования здоровья соответствуют исчерпанной заболеваемости. Разумеется, с учетом погрешности выборки [6] можно оценить истинные значения для всей генеральной совокупности в фиксированный момент времени (население Новгородской области, 2005 г.). Однако более интересной и сложной задачей является нахождение связи между показателями «заболеваемость по обращаемости» и «исчерпанная заболеваемость». Имеющиеся данные позволяют установить регрессионные зависимости исчерпанной заболеваемости (y) от заболеваемости по обращаемости (x) и возраста (t) для всех состояний E_1, E_2, \dots, E_{19} , а также состояний E_0 и E_{20} . Таким образом, для мужского и женского населения региона разработаны комплексы регрессионных моделей связи показателей заболеваемости по обращаемости и исчерпанной. Приведем соотношения для некоторых классов болезней.

Таблица 2

Матрица безусловных вероятностей для «поколения 0 лет», мужское население
(заболеваемость по обращаемости)

Возраст	E_0	E_1	E_2	E_3	E_4	E_5	E_6	E_7	...	E_{19}	E_{20}
0-1	1	0	0	0	0	0	0	0	...	0	0
2-4	0,353	0,013	0,006	0,024	0,147	0,002	0,009	0,002	...	0,026	0,009
5-9	0,090	0,021	0,021	0,016	0,144	0,090	0,359	0,074	...	0,009	0,021
10-14	0,046	0,035	0,038	0,014	0,133	0,053	0,500	0,044	...	0,011	0,025
15-19	0,037	0,052	0,066	0,014	0,142	0,038	0,440	0,031	...	0,014	0,034
20-24	0,104	0,032	0,101	0,013	0,166	0,018	0,282	0,030	...	0,023	0,056
25-29	0,150	0,022	0,148	0,015	0,173	0,009	0,174	0,035	...	0,021	0,086
30-34	0,149	0,025	0,151	0,014	0,161	0,010	0,161	0,024	...	0,022	0,130
35-39	0,143	0,028	0,122	0,013	0,137	0,010	0,156	0,023	...	0,023	0,186
40-44	0,128	0,039	0,107	0,013	0,113	0,009	0,140	0,018	...	0,025	0,255
45-49	0,100	0,037	0,087	0,017	0,093	0,008	0,124	0,015	...	0,029	0,346
50-54	0,081	0,040	0,064	0,014	0,068	0,007	0,105	0,013	...	0,029	0,453
55-59	0,053	0,042	0,043	0,014	0,048	0,007	0,089	0,009	...	0,024	0,568
60-64	0,041	0,035	0,026	0,013	0,031	0,004	0,081	0,007	...	0,017	0,677
65-69	0,031	0,030	0,012	0,010	0,015	0,003	0,071	0,006	...	0,012	0,766
70-74	0,023	0,019	0,009	0,007	0,010	0,002	0,051	0,004	...	0,009	0,839
75-79	0,014	0,014	0,003	0,005	0,005	0,001	0,030	0,002	...	0,007	0,901
80-84	0,008	0,009	0,001	0,002	0,003	0,001	0,016	0,001	...	0,004	0,944
85-	0,005	0,004	0,000	0,001	0,001	0,001	0,007	0,001	...	0,002	0,974
—	0	0	0	0	0	0	0	0	...	0	1

Таблица 3

Точечные оценки вероятностей в «поколении 0 лет», мужское население,
по возрастам и состояниям здоровья (исчерпанная заболеваемость)

Возраст	E_0	E_1	E_2	E_3	E_4	E_5	E_6	...	E_{19}	E_{20}
0-1	1	0	0	0	0	0	0	...	0	0
2-4	0,0831	0,0102	0,1257	0,0180	0,0467	0,0738	0,4966	...	0,0129	0,0043
5-9	0,0424	0,0161	0,0713	0,0128	0,2291	0,1296	0,3592	...	0,0069	0,0164
10-14	0,0266	0,0360	0,0450	0,0104	0,1748	0,0677	0,5387	...	0,0084	0,0209
15-19	0,0216	0,0849	0,0330	0,0114	0,1907	0,0496	0,4612	...	0,0133	0,0302
20-24	0,0224	0,0722	0,0344	0,0133	0,2731	0,0333	0,3021	...	0,0282	0,0514
25-29	0,0202	0,0705	0,0592	0,0152	0,3104	0,0193	0,2076	...	0,0338	0,0823
30-34	0,0156	0,0934	0,0639	0,0145	0,2885	0,0149	0,1936	...	0,0374	0,1261
35-39	0,0113	0,1185	0,0413	0,0130	0,2465	0,0107	0,1819	...	0,0395	0,1827
40-44	0,0074	0,1597	0,0324	0,0122	0,2001	0,0065	0,1568	...	0,0397	0,2514
45-49	0,0038	0,1643	0,0229	0,0125	0,1650	0,0037	0,1304	...	0,0407	0,3427
50-54	0,0013	0,1746	0,0138	0,0098	0,1203	0,0020	0,1000	...	0,0333	0,4506
55-59	0,0002	0,1681	0,0078	0,0078	0,0815	0,0021	0,0725	...	0,0196	0,5662
60-64	0,0001	0,1371	0,0049	0,0060	0,0512	0,0002	0,0591	...	0,0092	0,6750
65-69	0,0000	0,1108	0,0045	0,0037	0,0258	0,0005	0,0505	...	0,0040	0,7653
70-74	0,0000	0,0801	0,0027	0,0025	0,0180	0,0005	0,0348	...	0,0019	0,8385
75-79	0,0000	0,0560	0,0025	0,0013	0,0096	0,0003	0,0162	...	0,0008	0,9003
80-84	0,0000	0,0348	0,0015	0,0004	0,0061	0,0005	0,0063	...	0,0002	0,9436
85-	0,0000	0,0184	0,0008	0,0002	0,0026	0,0007	0,0012	...	0,0000	0,9733
—	0	0	0	0	0	0	0	...	0	1

Мужчины

$$E_1: y = -0,0903 + 3,1452x + 0,0017t^{1,25};$$

$$R^2 = 0,9562; \text{ ст. ошибка} = 0,0768;$$

$$E_4: y = -0,2611 + 1,2786x^{0,5} + 0,0796t^{-1};$$

$$R^2 = 0,9850; \text{ ст. ошибка} = 0,0147.$$

Женщины

$$E_5: y = 0,1012 + 1,4049x - 0,0386t^{0,25};$$

$$R^2 = 0,9350; \text{ ст. ошибка} = 0,0126;$$

$$E_{14}: y = -0,0060 + 0,2822x^{1,25} + 0,0170t^{-0,25};$$

$$R^2 = 0,9387; \text{ ст. ошибка} = 0,0009.$$

Все найденные регрессионные модели корректны, статистически значимы и характеризуются высокими значениями коэффициента детерминации R^2 и приемлемыми значениями стандартной ошибки регрессии, которая представляет точность прогнозов. В соответствии с авторской методикой прогнозирования заболеваемости (по доминирующему заболеванию по обращаемости) и смертности населения региона комплекс разработанных регрессионных моделей позволяет осуществить аналогичные прогнозы и исчерпанной заболеваемости. Например, для «поколения 0 лет» точечные оценки прогноза, рассчитанные по соответствующим регрессионным соотношениям, представлены в табл.3. Заметим, что в соответствии с нормирующим условием в каждом возрастном интервале проведена корректировка значений.

Соответствующие интервальные оценки на 10 тыс. чел. в начальном возрастном интервале, сгруп-

пированные по состояниям здоровья, с надежностью $\gamma = 0,95$ представлены в табл.4,5. С учетом смертности спрогнозированы значения количества человек в поколениях по возрастным интервалам. Произведен пересчет на количество оставшегося населения в каждой категории (возраст — состояние здоровья), указаны нижняя и верхняя границы оценок.

Отметим, что разработанная методика позволяет осуществлять вероятностный прогноз и для групп населения, находящихся в начальный момент времени и в других возрастных интервалах, и с иным состоянием здоровья E_j , $j = 0, 1, \dots, 19$.

Заключение

Выявить уровни истинной заболеваемости можно лишь при сплошном обследовании населения. Для страны или отдельного региона эта задача, требующая колоссальных ресурсов (материальных, организационных, технических и др.), неосуществима. Выборочное обследование позволяет с заданной точностью оценить масштаб заболеваемости лишь в фиксированный момент времени. Компьютерная методика, представленная в статье и реализованная на данных ПБД здоровья населения Новгородской области, позволяет получить оценки в динамике не только заболеваемости по обращаемости, но и исчерпанной заболеваемости и смертности населения. Соответствующий прогноз для перспектив реального населения может быть использован органами управления здравоохранением для расчета структуры и параметров обеспечения населения медицинской и лекарственной помощью.

Таблица 4

Прогноз исчерпанной заболеваемости и смертности в «поколении 0 лет», мужское население (кол-во чел.)

Возраст	Чел.	E_0	E_1	E_2	E_3	E_4	E_5	...	E_{19}
0-1	10000	10000						...	
2-4	9957	381-1148	0-1826	646-1666	71-261	75-784	285-1073	...	44-193
5-9	9836	243-813	0-1903	398-1381	65-255	2539-3175	1137-2096	...	10-163
10-14	9791	131-712	0-2237	231-1196	69-259	2454-3085	668-1476	...	58-209
15-19	9698	52-633	0-2975	52-994	87-274	2706-3335	402-1168	...	136-285
20-24	9486	23-574	0-2575	0-920	86-269	3318-3948	77-809	...	304-447
25-29	9177	0-517	0-2456	268-1209	101-277	3559-4178	0-598	...	352-490
30-34	8739	0-445	0-2654	336-1247	96-263	3281-3867	0-523	...	398-529
35-39	8173	0-375	57-2840	92-917	80-237	2743-3282	0-447	...	422-544
40-44	7486	0-306	691-3209	20-770	78-221	2198-2685	0-370	...	429-540
45-49	6573	0-236	885-3091	0-603	89-214	1784-2209	0-300	...	443-540
50-54	5494	0-175	1202-3040	0-437	67-171	1284-1637	0-237	...	364-445
55-59	4338	0-129	1396-2857	0-310	58-141	892-1171	0-194	...	215-280
60-64	3250	0-95	1235-2336	0-222	46-110	582-772	0-129	...	96-145
65-69	2347	0-68	1109-1919	0-176	28-74	275-429	0-98	...	37-72
70-74	1615	0-47	779-1332	0-115	16-49	184-291	0-69	...	13-37
75-79	997	0-29	572-922	0-82	7-28	95-161	0-43	...	3-19
80-84	564	0-16	361-563	0-49	0-11	62-99	0-28	...	0-8
85-	267	0-8	188-267	0-24	0-5	24-42	0-20	...	0-3

Таблица 5

Прогноз исчерпанной заболеваемости и смертности в «поколении 0 лет», женское население (кол-во чел.)

Возраст	Чел.	E_0	E_1	E_2	E_3	E_4	E_5	...	E_{19}
0-1	10000	10000						...	
2-4	9967	754-1025	0-1356	782-1063	67-254	498-2791	249-809	...	0-247
5-9	9916	463-665	0-1423	729-976	0-137	1137-3174	1249-1936	...	67-368
10-14	9871	260-462	50-2051	742-985	27-214	1373-3446	711-1285	...	113-414
15-19	9798	147-349	230-2212	752-992	90-274	1530-3663	566-1124	...	109-405
20-24	9692	159-357	0-1519	753-991	161-340	1392-3442	163-695	...	119-411
25-29	9562	139-333	0-1448	742-976	169-346	1131-3098	48-575	...	163-454
30-34	9454	63-255	0-1702	685-913	219-392	1155-3105	0-521	...	192-481
35-39	9289	2-191	394-2291	633-858	286-454	1154-3073	0-500	...	178-460
40-44	9033	0-139	1016-2848	577-796	371-536	1133-3002	0-452	...	181-454
45-49	8729	0-92	1730-3488	541-751	433-598	1015-2810	0-408	...	151-413
50-54	8240	0-84	2442-4097	423-628	390-545	892-2580	0-424	...	130-377
55-59	7609	0-78	2997-4533	354-542	320-462	820-2379	0-352	...	120-347
60-64	6967	0-71	3213-4626	272-445	239-368	490-1916	0-305	...	108-317
65-69	6101	0-62	3046-4283	211-360	190-303	119-1387	0-248	...	84-267
70-74	5133	0-53	2979-4031	156-280	156-252	0-971	0-181	...	53-209
75-79	3898	0-40	2636-3451	70-165	73-145	0-754	0-127	...	17-140
80-84	2581	0-26	2033-2581	27-90	7-54	0-513	0-72	...	7-90
85-	1434	0-15	1212-1434	0-20	0-14	0-267	0-41	...	0-46

При наличии соответствующих персонифицированных баз данных методика воспроизводима и в других регионах.

1. Токмачев М.С. Цепи Маркова в прогнозировании медико-социальных показателей // Обозрение прикладной и промышленной математики. Т.10. Вып.2. М., 2003. С.517-518.
2. Токмачев М.С. Разработка ряда показателей общественного здоровья на основе цепей Маркова // Приложение к: Вестник НовГУ. Сер.: Техн. науки. 2004. №28. С.3-7. Препринт.
3. Токмачев М.С. Разработка новых показателей общественного здоровья на основе статистических данных // Сб. науч. тр. Новгородского науч. центра СЗО РАМН. М.: Медицина, 2005. Т.4. С.119-127.
4. Токмачев М.С., Рязанцев П.П. // Информационные технологии. 2010. №3. С.64-68.
5. Бачманов А.А., Рязанцев П.П. Некоторые вопросы формирования единой базы данных «Здоровье населения

Новгородской области» // Сб. тр. Новгородского науч. центра СЗО РАМН. М.: Медицина, 2006. Т.5. С.81-85.

6. Токмачев М.С., Прохорова А.В. Исследование репрезентативности выборочной совокупности, сформированной для изучения здоровья населения Новгородской области // Там же. С.61-67.
7. Рязанцев П.П., Токмачев М.С. Разработка программного комплекса для расчета новых показателей здоровья населения // Сб. мат. Всерос. науч.-практ. конф. и тр. Новгородского науч. центра СЗО РАМН. М.: Медицина, 2007. Т.6. С.214-219.
8. Лисицин В.И. Персонифицированная база данных умерших — основа изучения смертности населения // Там же. С.104-110.
9. Кирьянов Б.Ф. К теории построения интегральных показателей здоровья населения // Там же. С.198-203.
10. Кирьянов Б.Ф., Токмачев М.С. Математические модели в здравоохранении. В. Новгород: НовГУ. 2009. 280 с.
11. Токмачев М.С. Математическая модель процесса здоровья населения региона // Четвертая Междунар. конф. по проблемам управления: Сб. тр. М: Институт проблем управления им. В.А.Трапезникова РАН, 2009. С.893-906.