

Министерство здравоохранения Кыргызской Республики

Научно-производственное объединение
“Профилактическая медицина”

Байызбекова Д. А., Розыева Р. С.,
Ерембаева А. А., Касымова Р. О.

**ПРИМЕНЕНИЕ МЕТОДОВ ОЦЕНКИ
МНОГОЛЕТНЕЙ ДИНАМИКИ В
НЕИНФЕКЦИОННОЙ ЭПИДЕМИОЛОГИИ**

(методические рекомендации)

Бишкек - 2017

УДК 616

ББК 51.9

П 76

Издается по решению научно-технического совета и редакционного совета научно-производственного объединения «Профилактическая медицина» МЗ КР

Рецензенты:

А. А. Шаршенова, доктор медицинских наук, профессор,

З. Ш. Нурматов, кандидат медицинских наук.

П 76 Применение методов многолетней оценки динамики в неинфекционной эпидемиологии : методические рекомендации / Д. А. Байызбекова, Р. С. Розыева, А. А. Ерембаева, Р. О. Касымова – Б., 2017. - 30 с.

ISBN 978-9967-603-75-2

Методические рекомендации предназначены для руководителей и специалистов организаций общественного здравоохранения, научных и практических работников, студентов медицинских ВУЗов.

П 4106000000-17

ISBN 978-9967-603-75-2

УДК 616

ББК 51.9

СОДЕРЖАНИЕ

	Стр.
Введение в неинфекционную эпидемиологию	4
Методика анализа многолетней динамики заболеваемости	5
Выравнивание динамического ряда показателей заболеваемости методом скользящей средней	7
Выравнивание динамического ряда показателей заболеваемости методом укрупнения (удлинения) периодов.	8
Анализ многолетней тенденции (тренда) заболеваемости	8
Прямолинейное выравнивание динамического ряда показателей заболеваемости методом наименьших квадратов	11
Криволинейное выравнивание динамического ряда показателей заболеваемости методом наименьших квадратов	15
Модель авторегрессии и проинтегрированного скользящего среднего	17
Пример использования модели авторегрессии и проинтегрированного скользящего среднего	20
Формулировка гипотез о причинах заболеваемости	29
Литература	30

ВВЕДЕНИЕ В НЕИНФЕКЦИОННУЮ ЭПИДЕМИОЛОГИЮ

По данным ВОЗ основным бременем для населения в последние десятилетия стали неинфекционные заболевания (НИЗ), для которых характерны высокая заболеваемость с хронизацией процесса, инвалидизацией и смертностью. Поэтому в последние десятилетия вопросы профилактики остро стали при многих неинфекционных заболеваниях. А следовательно популяционный, т.е. эпидемиологический взгляд на заболеваемость распространился на многие неинфекционные (сердечно-сосудистые и органов дыхания, эндокринные и онкологические, нервные и психические, аллергические и др.) болезни. И для изучения распространения их начали использовать эпидемиологические методы. Возникла эпидемиология неинфекционных заболеваний, т.е. эпидемиология стала общемедицинской наукой, как ее назвали В.Д. Беляков, Р.Х. Яфаев (1989).

Многие эпидемиологи были противниками введения термина «неинфекционная эпидемиология». Однако, как указывает В.Д. Беляков (1981), «сама жизнь решила этот спор», «медицина изнутри созрела для теоретического и методического обоснования эпидемиологии как общемедицинской науки: сформулированы общие принципы и методические основы изучения (в науке) и решения (в практике) любых медицинских проблем на популяционном уровне».

В.Д. Беляков (1989) дал следующее определение эпидемиологии как общемедицинской науки — это «... наука изучает причины, условия и механизмы формирования заболеваемости населения путем анализа ее распространения по территории, среди различных групп населения и во времени и использует эти данные для разработки способов профилактики заболеваний». Эпидемиология стала наукой об различных массовых заболеваниях. Она изучает эпидемиологию как инфекционных, так и неинфекционных болезней и меры борьбы с ними.

Итак, в настоящее время существуют:

- эпидемиология инфекционных болезней, изучающая эпидемический процесс;
- эпидемиология неинфекционных заболеваний, изучающая путем анализа причины, условия и механизмы возникновения различных болезней, по территории, среди различных групп и во времени и использующая эти данные для снижения и профилактики заболеваемости.

Эпиднадзор за НИЗ в Кыргызстане проводится в отношении впервые выявленных случаев эндемического зоба и железодефицитной анемии. В стране НИЗ являются основной причиной нетрудоспособности, заболеваемости и преждевременной смертности населения. По данным Республиканского медико-информационного центра, в структуре причин смертности населения Кыргызской Республики ведущее место занимают сердечно-сосудистые заболевания (52,8%), на втором месте находятся травмы и отравления (10,1%), на третьем — новообразования (9,4%) и болезни органов дыхания (7,2%).

МЕТОДИКА АНАЛИЗА МНОГОЛЕТНЕЙ ДИНАМИКИ ЗАБОЛЕВАЕМОСТИ

Многолетняя динамика заболеваемости определяется наличием и активностью комплекса разнообразных причин и условий, отличающихся по закономерности проявлений своей активностью:

первая группа - причины и условия, активность которых на протяжении многих лет не меняется или изменяется в одном направлении. Они определяют многолетнюю тенденцию (тренд);

вторая группа - причины и условия, активность которых изменяется периодически, ритмично. Эта группа обуславливает периодические (циклические) колебания в многолетней динамике;

третья группа - причины и условия, активность которых беспорядочно меняется, что вызывает нерегулярные изменения многолетней динамики.

Таким образом, задачей анализа является определение закономерных и нерегулярных проявлений многолетней динамики.

Для детального анализа следует рассчитать показатели: заболеваемости, смертности, летальности. Сопоставление абсолютных величин применяется в редких случаях, когда заболеваемость регистрируется единичными случаями.

Основным показателем является показатель заболеваемости (инцидентности). Он характеризует относительную частоту случаев заболевания среди населения в определенный период. Выражается числом заболевших на 100, 1000, 10000, 100000 человек в год или месяц.

$$(1) \quad I = \frac{A \cdot 100000}{N},$$

где: I - показатель заболеваемости,

A - абсолютное число заболевших,

N - численность населения.

Показатель смертности характеризует вероятность смерти от определенной болезни за определенный период.

$$(2) \quad D = \frac{d \cdot 100000}{N},$$

где: D - показатель смертности,

d - абсолютное число умерших,

N - численность населения.

Показатель летальности характеризует вероятность смерти для заболевшего определенной болезнью. Этот показатель используется для оценки тяжести течения заболевания, а также для оценки качества диагностики и лечения.

$$(3) \quad L = \frac{d \cdot 100}{A},$$

где: L - показатель летальности,
 d - абсолютное число умерших от данного заболевания,
 A - абсолютное число заболевших данным заболеванием.

Для характеристики структуры неинфекционной заболеваемости по нозологическим формам, по возрасту, по полу, по территориям и т.д., используется экстенсивный показатель, отражающий долю (удельный вес) какой-то части изучаемого явления в общей совокупности в процентном соотношении.

$$(4) \quad P = \frac{I' \cdot 100}{I},$$

где: P - показатель доли (удельный вес),
 I' - показатель оцениваемой заболеваемости,
 I - суммарный уровень общей неинфекционной заболеваемости.

Анализ многолетней динамики предполагает выявление многолетней тенденции, определение наличия или отсутствия периодичности (цикличности), прогноз заболеваемости на будущий год, оценку показателя заболеваемости в анализируемом году. Графически многолетнюю динамику заболеваемости изображают в виде линейной диаграммы.

При построении линейных диаграмм для лучшей наглядности и уменьшения субъективизма обычно пользуются, так называемым, правилом "золотого сечения", согласно которому ординатная часть графика должна относиться к части оси абсцисс как 1:1,5 (это касается длины осей и количества масштабных отрезков на осях). Но следует предостеречь от увлечения абсолютно строгого следования правилу "золотого сечения" - поскольку получаемая цена деления иногда затрудняет построение и анализ графиков.

После построения графика производится визуальная оценка полученного изображения. Визуальная оценка - это первый и простейший этап анализа многолетней динамики заболеваемости и необходима она для ориентировочного определения направленности многолетней тенденции. Результаты визуальной оценки позволяют выбрать продолжительность периода для выравнивания динамического ряда методом скользящей средней или методом укрупнения периодов; позволяет определить, прямолинейна или криволинейна тенденция многолетней динамики заболеваемости, а также наличие или отсутствие периодических (циклических) колебаний в многолетней динамике.

ВЫРАВНИВАНИЕ ДИНАМИЧЕСКОГО РЯДА ПОКАЗАТЕЛЕЙ ЗАБОЛЕВАЕМОСТИ МЕТОДОМ СКОльзяЩЕЙ СРЕДНЕЙ

Скользящие средние получаются путем суммирования показателей фактической заболеваемости двух, трех или более смежных лет и деление полученной суммы на число суммированных лет.

$$(5) \quad I_{теор.1} = \frac{(I_1 + I_2 + I_3)}{3}; \quad I_{теор.2} = \frac{(I_2 + I_3 + I_4)}{3} \text{ и т. д.},$$

где: $I_{теор.1,2}$ - теоретические показатели,
 $I_{1,2,3,4}$ - показатели фактической заболеваемости 1-го, 2-го и т.д. лет анализируемого периода.

В полученных средних величинах взаимно погашаются случайные отклонения. Чем шире шаг (интервал) скользящего, тем более плавной получается линия тенденции (тренд), но одновременно сокращается количество лет наблюдений - это одно из отрицательных свойств данного метода.

В нашем примере (см. табл.1) скользящая средняя рассчитывается следующим образом:

$$I_{теор.(1977-1979)} = \frac{(10,5 + 27,0 + 158,1)}{3} = 65,2;$$

$$I_{теор.(1978-1980)} = \frac{(27,0 + 158,1 + 18,2)}{3} = 67,8 \text{ и т.д.}$$

Таблица 1. Матрица для расчета скользящей средней и укрупнения периодов

Годы	$I_{факт.}$	Скользящая средняя $I_{теор.}$	Укрупнение периодов $I_{теор.}$
1977	10,5	-	
1978	27,0	65,2	
1979	158,1	67,8	67,8
1980	18,2	78,6	
1981	59,5	87,2	
1982	183,8	84,8	84,8
1983	11,1	69,8	
1984	14,5	61,9	
1985	160,0	66,1	66,1
1986	23,8	-	

ВЫРАВНИВАНИЕ ДИНАМИЧЕСКОГО РЯДА ПОКАЗАТЕЛЕЙ ЗАБОЛЕВАЕМОСТИ МЕТОДОМ УКРУПНЕНИЯ (УДЛИНЕНИЯ) ПЕРИОДОВ

При использовании этого метода рассчитываются средние хронологические величины. В нашем примере (табл.1) средние хронологические величины выбраны для трехлетнего периода, начиная с 1978 г.

Первая средняя величина равна:

$$I_{теор.}(1978-1980) = (27,0+158,1+18,2):3 = 67,8;$$

$$\text{вторая} - I_{теор.}(1981-1983) = (59,5+183,8+11,1):3 = 84,8;$$

$$\text{третья} - I_{теор.}(1984-1986) = (14,5+160,0+23,8):3 = 66,1.$$

АНАЛИЗ МНОГОЛЕТНЕЙ ТЕНДЕНЦИИ (ТРЕНДА) ЗАБОЛЕВАЕМОСТИ

По результатам оценки многолетней тенденции определяется значимость заболеваемости, выдвигаются и обосновываются гипотезы о факторах, объясняющих тенденцию, оценивается эффективность осуществляемых профилактических мероприятий и, в целом, результативность деятельности медицинской службы.

Приступая к статистической обработке исходных данных, необходимо исследовать динамический ряд с целью выявления резко выделяющихся низких и высоких (выскакивающих) показателей заболеваемости. Они формируются, как правило, под влиянием случайных факторов и не характерны для основной части совокупности показателей. Поскольку они ("выскакивающие" показатели) могут привести к существенным искажениям тенденции многолетней заболеваемости, целесообразно заменять такие показатели расчетными величинами. Последние вычисляются по формуле (6):

$$(6) \quad I_{теор.} = I_{ср.} + 2\sigma,$$

где:

$I_{теор.}$ - расчетный показатель, замещающий "выскакивающие" величины,

$I_{ср.}$ - средний арифметический многолетний показатель,

σ (сигма)- среднее квадратическое отклонение динамического ряда.

$$(7) \quad I_{ср.} = \frac{\sum I_{факт.}}{n},$$

где: \sum - знак суммы,

$I_{факт.}$ - показатели фактической заболеваемости за каждый год изучаемого периода,
 n - число лет изучаемого периода.

$$(8) \quad \sigma = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}},$$

где: $\sum x^2$ - сумма квадратов отдельных вариант динамического ряда,

$(\sum x)^2$ - квадрат суммы отдельных вариант,

n - число вариант.

Для выявления "выскакивающих" показателей требуется предварительно ранжировать вариационный ряд, причем, не имеет значения порядок ранжирования, можно от большего числа к меньшему, а можно и наоборот (табл. 2). В ранжированном ряду резко выделяющиеся минимальные и максимальные величины занимают крайние места. Для доказательства того, что они обладают свойствами резко выделяющихся величин в исследуемой совокупности показателей, в математической статистике существует ряд способов и критериев. Наиболее целесообразно применять критерий Шовене.

Для исследования "подозрительных" величин применяются формулы:

$$\text{а) для крайних минимальных показателей: } \frac{I_{ср.} - I_1}{\sigma},$$

где: I_1 - показатель первого ранга;

$$\text{б) для крайних максимальных показателей: } \frac{I_n - I_{ср.}}{\sigma},$$

где: I_n - показатель последнего ранга.

Полученные величины сравниваются с табличными критериями Шовене (табл. 3) для данного числа наблюдений. Если полученная величина будет больше табличного коэффициента, то исследуемый показатель считается резко выделяющимся.

В нашем примере (см. табл. 2) резко выделяющимися величинами являются: показатель 1977 года - $10,5 \text{ ‰}$ и 1982 года - $183,8 \text{ ‰}$; они и подлежат анализу.

Таблица 2. Матрица для расчета среднего многолетнего показателя и среднего квадратического отклонения

Годы	$I_{\text{факт.}}(x)$	x^2	Ранги	$I_{\text{ср.}}; \sigma_x$
1977	10,5	110,25	1	$I_{\text{ср.}} = 666,5 \div 10 = 66,6$ $\sigma_x = 71,15$
1978	27,0	729,0	6	
1979	158,1	24995,61	8	
1980	18,2	331,24	4	
1981	59,5	3540,25	7	
1982	183,8	33782,44	10	
1983	11,1	123,21	2	
1984	14,5	210,25	3	
1985	160,0	25600,0	9	
1986	23,8	566,44	5	
n=10	$\Sigma = 666,5$	$\Sigma = 89988,69$		

Таблица 3. Критерий Шовене для исключения резко отличающихся величин в анализируемом ряду (n - число наблюдений). По Д.Сепетлиеву, 1968.

N	критерий	n	критерий	n	критерий	n	критерий	n	критерий
5	1,68	9	1,92	16	2,16	24	2,31	50	2,58
6	1,73	10	1,95	18	2,2	26	2,36	100	2,80
7	1,79	12	2,03	20	2,24	30	2,39	200	3,02
8	1,86	14	2,10	22	2,28	40	2,50	500	3,29

Анализ показателя 10,5: $(66,6 - 10,5) : 71,15 = 0,79$. Полученная величина сравнивается с табличным коэффициентом - критерием Шовене, который для числа наблюдений 10 равен 1,95. В нашем примере $0,79 < 1,95$ и, следовательно, анализируемый показатель (10,5) не считается резко выделяющимся.

Анализ показателя 183,8: $(183,8 - 66,6) : 71,15 = 1,65$. Поскольку $1,65 < 1,95$, следовательно, и второй анализируемый показатель тоже нельзя считать "выскакивающим".

Таким образом, динамический ряд показателей заболеваемости за 10 лет не содержит резко выделяющихся величин, это указывает, что за анализируемый период резких спадов и подъемов заболеваемости, связанных с действием случайных (нерегулярных) причин, не было.

Динамический ряд может содержать несколько "выскакивающих" величин. В таких случаях, после замены первого резко отличающегося показателя ($I_{\text{ср.}} + 2\sigma$ к максимальным числам; $I_{\text{ср.}} - 2\sigma$ к минимальным числам), снова определяется средний многолетний показатель, среднее квадратическое отклонение и повторяется весь цикл вычислений.

Теоретические показатели, заменяющие резко отличающиеся, используются в ходе анализа только на этапах определения прямолинейной или криволинейной многолетней тенденции.

ПРЯМОЛИНЕЙНОЕ ВЫРАВНИВАНИЕ ДИНАМИЧЕСКОГО РЯДА ПОКАЗАТЕЛЕЙ ЗАБОЛЕВАЕМОСТИ МЕТОДОМ НАИМЕНЬШИХ КВАДРАТОВ

По сравнению другими методами выравнивания динамического ряда (скользящей средней, укрупнения периодов) метод наименьших квадратов наиболее объективный способ выявления многолетней тенденции (тренда).

Расчет производится по формулам:

$$(9) \quad I_{\text{теор.}} = I_{\text{ср.}} + bx,$$

$$(10) \quad I_{\text{ср.}} = \frac{\Sigma I_{\text{факт.}}}{n},$$

$$(11) \quad b = \frac{\Sigma (I_{\text{факт.}} \cdot x)}{\Sigma x^2},$$

где: $I_{\text{теор.}}$ - теоретические (расчетные) уровни заболеваемости после выравнивания,

$I_{\text{ср.}}$ - средний арифметический уровень заболеваемости за изучаемый период,

$I_{\text{факт.}}$ - фактические показатели заболеваемости за каждый год изучаемого периода,

b - коэффициент, показывающий разницу между $I_{\text{теор.}}$ за смежные годы,

x - натуральные числа, проставляемые от середины вариационного ряда в оба его конца; причем, при наличии нечетного числа лет "x" принимает значения: -3, -2, -1, 0, +1, +2 и т.д.;

при четном числе лет "x" выразится цифрами: -5, -3, -1, +1, +3 и т.д., т.е. в этом случае для двух центральных величин "x" принимает значения -1 и +1, а остальные значения "x" берутся через единицу,

n - число лет изучаемого периода.

Для построения прямой линии тенденции достаточно две точки, поэтому можно определить $I_{теор.}$ только для 1977 и 1986 гг. По полученным значениям $I_{теор.}$ для 1977 и 1986 гг. на графике провести линию многолетней тенденции (тренд).

Таблица 4. Матрица для расчета тенденции многолетней заболеваемости А населения г."О"

Годы	x	$I_{факт.}$	$I_{факт.} \cdot x$	x^2
1977	-9	10,5	-94,5	81
1978	-7	27,0	-189,0	49
1979	-5	158,1	-790,5	25
1980	-3	18,2	-54,6	9
1981	-1	59,5	-59,5	1
1982	1	183,8	183,8	1
1983	3	11,1	33,3	9
1984	5	14,5	72,5	25
1985	7	160,0	1120,0	49
1986	9	23,8	214,2	81
	n = 10	$\Sigma = 666,5$	$\Sigma = 435,7$	$\Sigma = 330$

Проведем расчеты для нашего примера:

$$I_{ср.} = 666,5 : 10 = 66,6$$

$$b = 435,7 : 330 = 1,32$$

$$I_{теор.(1977)} = 66,6 + 1,32 \cdot (-9) = 54,7 \pm 2,29$$

$$I_{теор.(1986)} = 66,6 + 1,32 \cdot 9 = 78,5 \pm 2,62$$

Для обобщения количественной оценки тенденции многолетней заболеваемости используется показатель темпа прироста (снижения), выражаемый в процентах ($T_{пр. (сн.)}$), который рассчитывается по формуле:

$$(12) \quad T_{пр. (сн.)} = \frac{b \cdot 100}{I_{ср.}}$$

где: b и $I_{ср.}$ - величины, входящие в формулу (9).

В приведенном примере: $T_{пр. (сн.)} = 1,32 \cdot 100 : 66,6 = 1,98\%$.

Если коэффициент "b" со знаком "плюс", это значит, имеется тенденция к росту заболеваемости, при "b" со знаком "минус" - тенденция к снижению.

Градации темпа прироста (снижения) (В.Д. Беляков с соавт., 1981):

при $T_{пр. (сн.)}$ от 0 до 1% - тенденция отсутствует (стабильна),

при $T_{пр. (сн.)}$ от 1,1% до 5% - умеренная тенденция,

при $T_{пр. (сн.)}$ более 5% - выраженная тенденция.

В приведенном примере наблюдается умеренная тенденция к росту заболеваемости корью ($T_{пр. (сн.)} = +1,98\%$).

Для оценки тенденции развития недостаточно определить темпы прироста или снижения, необходимо определить произошло ли за исследуемое время существенное (выраженное, достоверное) изменение заболеваемости. Для этого определяется существенность (достоверность) различий теоретических значений показателей заболеваемости в первый и последний год линии тенденции. Для оценки достоверности (существенности) различий двух показателей используется расчет и сравнение доверительных границ сопоставляемых показателей и расчет критерия "t" (Стьюдента).

В нашем примере необходимо оценить существенность различий $I_{теор.}$

для 1977 = $54,7 \text{ ‰}$ и $I_{теор.}$ для 1986 = $78,5 \text{ ‰}$. Доверительные границы указанных показателей при уровне доверия 95% составляют величину $I \pm 2m$ (показатель плюс-минус удвоенная ошибка этого показателя).

Ошибка показателя ($\pm m$) рассчитывается по формуле:

$$(13) \quad m = \pm \sqrt{\frac{I \cdot q}{n}}$$

где: m - средняя ошибка показателя,

I - показатель заболеваемости,

q - величина равная:

(100000 - I) - если I рассчитывался на 100000 населения;

(1000 - I) - если I выражена в промиле;

(100 - I) - если I рассчитывался на 100 человек определенного контингента (процент);

n - численность населения (контингента), лежащая в основе показателя (если n < 30, то в знаменателе n - 1).

В наше примере средняя ошибка для $I_{теор.(1977)}$ будет следующей:

$$m_{теор.(1977)} = \pm \sqrt{\frac{54,7 \cdot (100000 - 54,7)}{1041100}} = 2,29 \text{ ‰}$$

$$59,3 \div 50,1 \text{ ‰} \quad (54,7 \pm 2,29 \cdot 2).$$

Для $I_{теор.(1986)}$ средняя ошибка - $m = \pm 2,62 \text{ ‰}$, а доверительные границы $83,7 \div 73,3$. (Рис. 1).

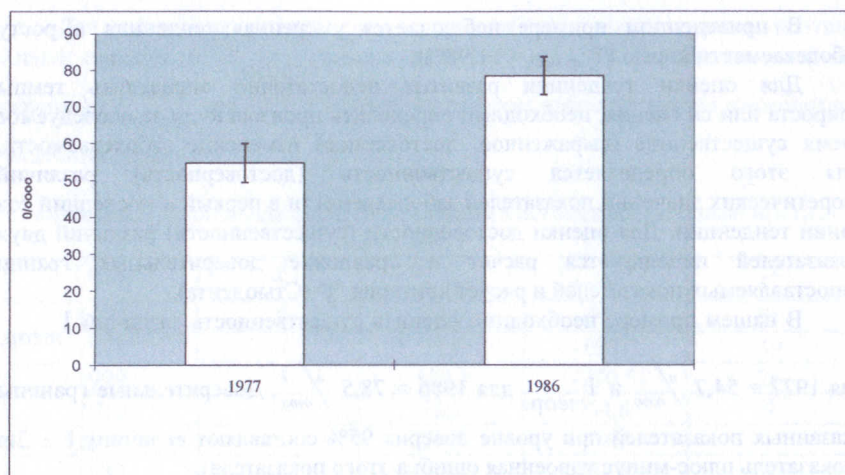


Рис. 1. Графическая оценка достоверности различий между показателями заболеваемости корью в 1977 и 1986 гг. (‰)

При сопоставлении доверительных границ показателей могут встретиться следующие варианты:

1) доверительные границы одного показателя не попадают в зону доверительных границ другого показателя. В этом случае даже без расчета критерия t можно сделать вывод о наличии существенных статистических различий между сравниваемыми показателями;

2) верхняя доверительная граница меньшего показателя на $1/3$ перекрывает нижнюю доверительную границу большего. В этом случае, чаще всего, различия между показателями могут оказаться существенными, но этот вывод необходимо проверить расчетом критерия t и, если он окажется более 2, то различия будут считаться существенными (достоверными);

3) верхняя доверительная граница меньшего показателя значительно (более чем на $1/3$) перекрывает нижнюю границу большего показателя. В этом случае различия будут статистически несущественны (рассчитывать критерий t нет необходимости) и можно будет сказать, что показатели статистически не отличаются друг от друга.

Критерий t рассчитывается по формуле:

$$(14) \quad t = \frac{I_1 - I_2}{\sqrt{m_1^2 + m_2^2}}$$

где: $I_{1,2}$ - сопоставляемые показатели,
 $m_{1,2}$ - средние ошибки сопоставляемых показателей.

При расчете критерия t в числителе, больший по величине показатель ставится на первое место, независимо от порядковых номеров показателей.

$$\text{В нашем примере: } t = \frac{78,5 - 54,7}{\sqrt{2,62^2 + 2,29^2}} = \frac{23,8}{3,48} > 2$$

Таким образом, поскольку верхняя доверительная граница I_{1977} меньше нижней доверительной границы I_{1986} , можно утверждать с достоверностью 95%, что различия показателей существенны, т.е. в период с 1977 по 1986 гг. заболеваемость A достоверно (статистически существенно) возросла и это подтверждается расчетом критерия t , который в данном случае оказался более двух ($t > 2$).

КРИВОЛИНЕЙНОЕ ВЫРАВНИВАНИЕ ДИНАМИЧЕСКОГО РЯДА ПОКАЗАТЕЛЕЙ ЗАБОЛЕВАЕМОСТИ МЕТОДОМ НАИМЕНЬШИХ КВАДРАТОВ

Выравнивание динамического ряда проводится по функции параболы 2-го порядка:

$$(15) \quad I_{\text{теор.}} = a + bx + cx^2$$

Определение неизвестных параметров "а", "b" и "с" из уравнения 15 производится в результате решения системы трех уравнений с тремя неизвестными по формулам:

$$(16) \quad \sum I_{\text{факт.}} = an + c \sum x^2$$

$$(17) \quad \sum (I_{\text{факт.}} \cdot x) = b \sum x^2$$

$$(18) \quad \sum (I_{\text{факт.}} \cdot x^2) = a \sum x^2 + c \sum x^4$$

где: "а", "b", "с" - неизвестные параметры,

$I_{\text{факт.}}$ - показатели фактической заболеваемости,

x - натуральные числа, проставленные от середины вариационного ряда в оба его конца,

n - число лет наблюдений.

Неизвестный параметр "b" из уравнения 17 определяется также как он определялся по формуле 11 - этот параметр рассчитывается одинаково как при прямолинейной, так и при криволинейной тенденции: $b = 435,7 : 330 = 1,32$

Параметры "а" и "с" уравнений 16 и 18 определяются одним из способов решения системы двух уравнений с двумя неизвестными - методом вычитания.

Таблица 5. Матрица для расчета многолетней криволинейной тенденции заболеваемости А.

Годы	$I_{\text{факт.}}$	x	x^2	x^4	$I_{\text{факт.}} \cdot x$	$I_{\text{факт.}} \cdot x^2$
1977	10,5	-9	81	6561	-94,5	850,5
1978	27,0	-7	49	2401	-189,0	1323,0
1979	158,1	-5	25	625	-790,5	3952,5
1980	18,2	-3	9	81	-54,6	163,8
1981	59,5	-1	1	1	-59,5	59,5
1982	183,8	1	1	1	183,8	183,8
1983	11,1	3	9	81	33,3	99,9
1984	14,5	5	25	625	72,5	362,5
1985	160,0	7	49	2401	1120,0	7840,0
1986	23,8	9	81	6561	214,2	1927,8
$n = 10$	$\Sigma = 666,5$	$\Sigma = 0$	$\Sigma = 330$	$\Sigma = 19338$	$\Sigma = 435,7$	$\Sigma = 16763,3$

Используя итоговые данные из табл. 5 в систему двух уравнений (16 и 18) подставляются их значения: 1) $666,5 = 10a + 330c$

2) $16763,3 = 330a + 19338c$.

Следующий этап: определяется множитель первого уравнения (16):

$Mн. = 330a : 10a = 33$.

Затем с помощью множителя уравнивается первое уравнение (16) с третьим (18)

1) $666,5 = 10a + 330c$

$$\begin{array}{r} x \quad x \quad x \\ 33 \quad 33 \quad 33 \\ \hline \end{array}$$

1a) $21994,5 = 330a + 10890c$

На следующем этапе производится вычитание сочленов третьего уравнения (18) из сочленов нового (1a) уравнения и определяется количественное значение неизвестного параметра "с":

1a) $21994,5 = 330a + 10890c$

$16763,3 = 330a + 19338c$

$5231,2 = 0 + (-8448)c$;

$c = 5231,2 : (-8448) = -0,6$

Теперь в исходное уравнение подставляется найденное значение параметра "с" и определяется количественное значение неизвестного параметра "а":

1) $666,5 = 10a + [330 \cdot (-0,6)]$

$666,5 = 10a - 198$

$666,5 + 198 = 10a$; $a = 86,45$

Таким образом, количественные значения неизвестных параметров системы уравнений составляют: "а" = 86,4; "b" = 1,32; "с" = -0,6, при помощи которых можно определить теоретические показатели для каждого анализируемого года.

Полученные результаты расчета теоретических показателей (см. табл. 6) наносятся на график.

Таблица 6. Матрица для расчета теоретических показателей криволинейной тенденции заболеваемости корью

Годы	$a + bx + cx^2 = I_{\text{теор.}}$
1977	$86,4 + 1,32 \cdot (-9) + (-0,6) \cdot 81 = 25,92$
1978	$86,4 + 1,32 \cdot (-7) + (-0,6) \cdot 49 = 47,76$
1979	$86,4 + 1,32 \cdot (-5) + (-0,6) \cdot 25 = 64,80$
1980	$86,4 + 1,32 \cdot (-3) + (-0,6) \cdot 9 = 77,04$
1981	$86,4 + 1,32 \cdot (-1) + (-0,6) \cdot 1 = 84,48$
1982	$86,4 + 1,32 \cdot 1 + (-0,6) \cdot 1 = 88,32$
1983	$86,4 + 1,32 \cdot 3 + (-0,6) \cdot 9 = 84,96$
1984	$86,4 + 1,32 \cdot 5 + (-0,6) \cdot 25 = 78,00$
1985	$86,4 + 1,32 \cdot 7 + (-0,6) \cdot 49 = 66,24$
1986	$86,4 + 1,32 \cdot 9 + (-0,6) \cdot 81 = 49,68$

МОДЕЛЬ АВТОРЕГРЕССИИ И ПРОИНТЕГРИРОВАННОГО СКОЛЬЗЯЩЕГО СРЕДНЕГО

Модель авторегрессии и проинтегрированного скользящего среднего (АРПСС) была предложена американскими учёными Боксом и Дженкинсом в 1976 г. как один из методов оценки неизвестных параметров и прогнозирования временных рядов.

Моделью авторегрессии проинтегрированного скользящего среднего называется модель, которая применяется при моделировании нестационарных временных рядов.

Нестационарный временной ряд характеризуется непостоянными математическим ожиданием, дисперсией, автоковариацией и автокорреляцией.

В основе модели авторегрессии и проинтегрированного скользящего среднего лежат два процесса:

1. процесс авторегрессии;
2. процесс скользящего среднего.

Процесс авторегрессии может быть представлен в виде:

(19) $x_t = a + \delta_1 x_{t-1} + \delta_2 x_{t-2} + \dots + \varepsilon_t$

где: а – свободный член модели, являющийся константой;

$\delta_1 \delta_2 \dots$ – параметры модели авторегрессии;

ε – случайное воздействие (ошибка модели).

Каждое наблюдение в модели авторегрессии представляет собой сумму случайной компоненты и линейной комбинации предыдущих наблюдений.

Процесс скользящего среднего может быть представлен в виде:

$$(20) \quad x_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots$$

где: μ – свободный член модели, являющийся константой;

$\theta_1, \theta_2, \dots$ – параметры модели скользящего среднего;

ε – случайное воздействие (ошибка модели).

Текущее наблюдение в модели скользящего среднего представляет собой сумму случайной компоненты в данный момент времени и линейной комбинации случайных воздействий в предыдущие моменты времени.

Следовательно, в общем виде модель авторегрессии и проинтегрированного скользящего среднего описывается формулой:

$$(21) \quad y_t = C + \sum_{i=1}^R AR_i y_{t-i} + \sum_{j=1}^M MA_j \varepsilon_{t-j} + \varepsilon_t$$

где: C – свободный член модели, являющийся константой;

ε_t – некомпенсированный моделью случайный остаток.

В обозначениях Бокса и Дженкинса модель авторегрессии и проинтегрированного скользящего среднего записывается как *АРПСС* (p, d, q) или *ARIMA* (p, d, q), где:

p – параметры процесса авторегрессии;

d – порядок разностного оператора;

q – параметры процесса скользящего среднего.

Для рядов с периодической сезонной компонентой применяется модель авторегрессии и проинтегрированного скользящего среднего с сезонностью, которая в обозначениях Бокса и Дженкинса записывается как *АРПСС* (p, d, q) (ps, ds, qs), где:

ps – сезонная авторегрессия;

ds – сезонный разностный оператор;

qs – сезонное скользящее среднее.

Моделирование нестационарных временных рядов с помощью модели авторегрессии и проинтегрированного скользящего среднего осуществляется в три этапа:

1. проверка временного ряда на стационарность;

2. идентификация порядка модели и оценивание неизвестных параметров;

3. прогноз.

Применение модели АРПСС предполагает обязательную стационарность исследуемого ряда, поэтому на первом этапе данное предположение проверяется с помощью автокорреляционной и частной автокорреляционной

функций ряда остатков. Остатки представляют собой разности наблюдаемого временного ряда и значений, вычисленных с помощью модели.

Устранить нестационарность временного ряда можно с помощью метода разностных операторов.

Разностным оператором первого порядка называется замена исходного уровня временного ряда разностями первого порядка:

$$(22) \quad \Delta y_t = y_{t+1} - y_t$$

Разностные операторы первого порядка позволяют исключить линейные тренды.

Разностные операторы второго порядка позволяют исключить параболические тренды.

Сезонные разностные операторы предназначены для исключения 12-ти или 4-х периодичной сезонности:

$$(23) \quad \Delta_{12} y_t = y_{t+12} - y_t, \text{ или } \Delta_4 y_t = y_{t+4} - y_t$$

Если модель содержит и трендовую, и сезонную компоненты, то необходимо применять оба оператора.

На втором этапе необходимо решить, сколько параметров авторегрессии и скользящего среднего должно войти в модель.

В процессе оценивания порядка модели авторегрессии и проинтегрированного скользящего среднего применяется квазиньютоновский алгоритм максимизации правдоподобия наблюдения значений ряда по значениям параметров. При этом минимизируется (условная) сумма квадратов остатков модели. Для оценки значимости параметров используется t Стьюдента. Если критерий t не значим, соответствующие параметры в большинстве случаев удаляются из модели без ущерба подгонки.

Полученные оценки параметров используются на последнем этапе для того, чтобы вычислить новые значения ряда и построить доверительный интервал для прогноза.

Оценкой точности прогноза, сделанного на основе модели авторегрессии и проинтегрированного скользящего среднего является среднеквадратическая ошибка (mean square), вычисляемая по формуле (22). Чем меньше данный показатель, тем точнее прогноз.

Модель авторегрессии и проинтегрированного скользящего среднего считается адекватной исходным данным, если остатки модели являются некоррелированными нормально распределёнными случайными величинами.

Исследования Дж. Бокса и Г. Дженкинса показали, что во многих случаях от нестационарности можно избавиться, заменяя исходный ряд на ряд разностей:

$$(24) \quad Y_t = X_t - X_{t-1}$$

Если от нестационарности избавиться не удалось можно взять разность повторно:

$$(25) \quad Z_t = Y_t - Y_{t-1}$$

После проведенных преобразований к исходному ряду можно применить модель АРСС.

Полученную на основе таких рассуждений модель называют моделью Бокса-Дженкинса или моделью авторегрессии и проинтегрированного скользящего среднего АРПСС (p, d, q). В этой модели: p – порядок авторегрессии, d – порядок разности, q – порядок скользящего среднего.

Таким образом, модель АРПСС является обобщением всех рассмотренных выше моделей. Основными показателями качества модели авторегрессии и проинтегрированного скользящего среднего являются: средний относительный модуль ошибки (MAPE); максимальный относительный модуль ошибки (MaxAPE); нормализованный байесовский информационный критерий (BIC); критерий Льюнга-Бокса.

В настоящее время для оценки параметров моделей АРПСС разработаны специальные алгоритмы и программы. Провести анализ временных рядов можно используя программу "Statistica" (фирмы StatSoft) или SPSS (Statistical Package for the Social Sciences).

ПРИМЕР ИСПОЛЬЗОВАНИЯ МОДЕЛИ АВТОРЕГРЕССИИ И ПРОИНТЕГРИРОВАННОГО СКОльзяЩЕГО СРЕДНЕГО

Анализ проведен в программе SPSS (Statistical Package for the Social Sciences), использовался лаг в один месяц с января 2005 по декабрь 2014 гг.

Первым этапом нашего исследования была оценка влияния профилактических программ на показатели заболеваемости и смертности при сердечно-сосудистых заболеваний. Речь шла о внедрении патронажных медсестер по ПМСП. Период вмешательства январь 2011 по декабрь 2014 гг.

До начала анализа проверялись качество и полнота данных: характер ряда стабильный, ряд содержал достаточное число наблюдений (ежемесячные данные по заболеваемости и смертности от сердечно-сосудистых заболеваний), пропуски строк и пропуски данных отсутствуют, нет резких изменений поведения ряда.

Сравнивались графики заболеваемости (заболеваемость 1 без дифференцирования, заболеваемость 2 дифференцирование 1 порядка) (Рис. 2 и 3), для уточнения стационарности вариационного ряда, чтобы убедиться, что нет тренда, есть постоянная дисперсия, постоянное значение автокорреляции для всего ряда, если исходный ряд стационарен, достаточно дифференцирования первого порядка.

Графики последовательности (Рис. 2; 3) показывают отсутствие тренда, доказывающий, что модель имеет постоянный средний уровень и постоянную дисперсию. Данное условие отвечает требованиям стационарности вариационного ряда.

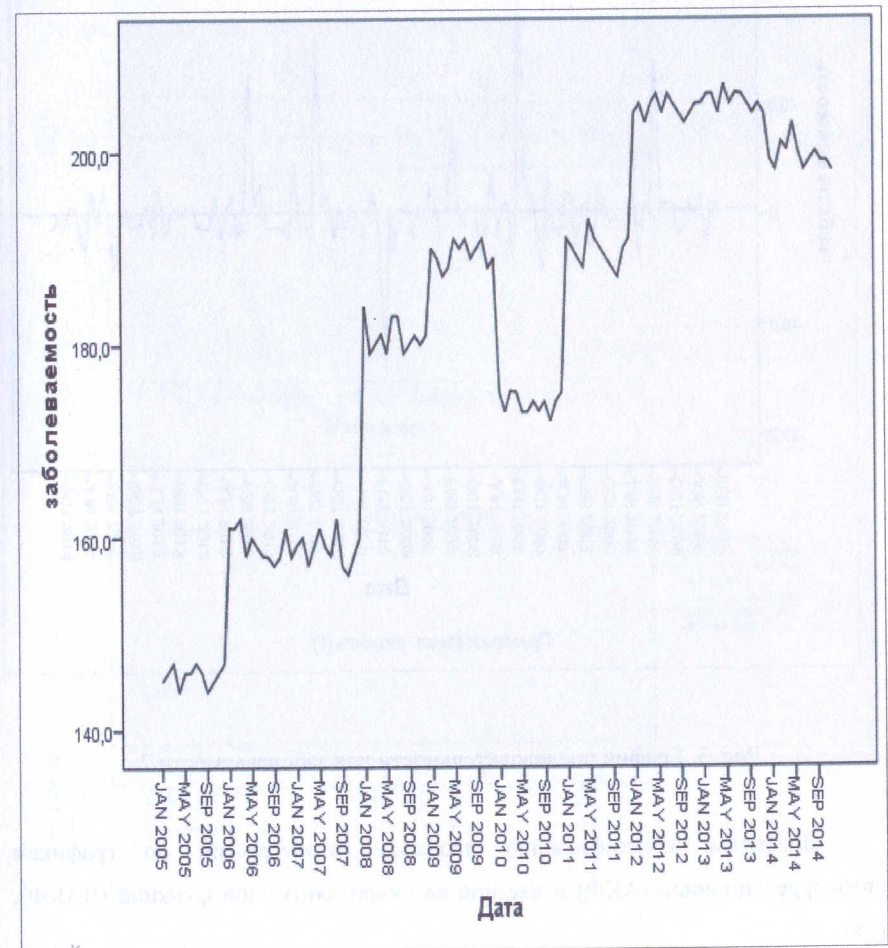


Рис. 2. График последовательности для заболеваемости 1.

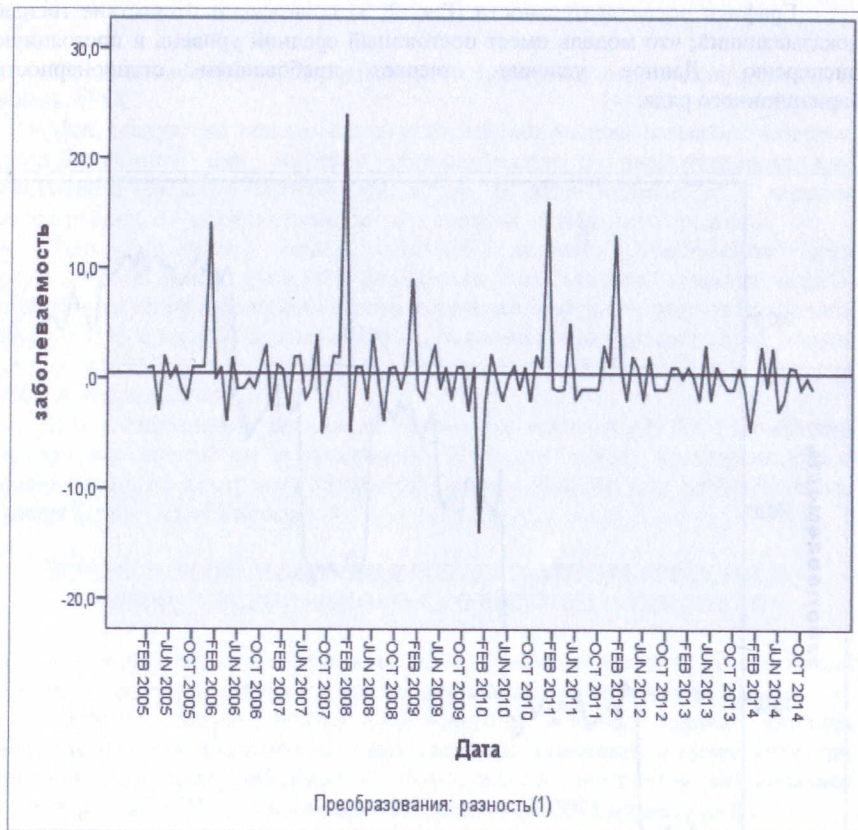


Рис. 3. График последовательности для заболеваемости 2.

Проводим идентификацию процесса авторегрессии по графикам автокорреляционной (АКФ) и частной автокорреляционной функций (ЧАКФ). (Рис. 4).

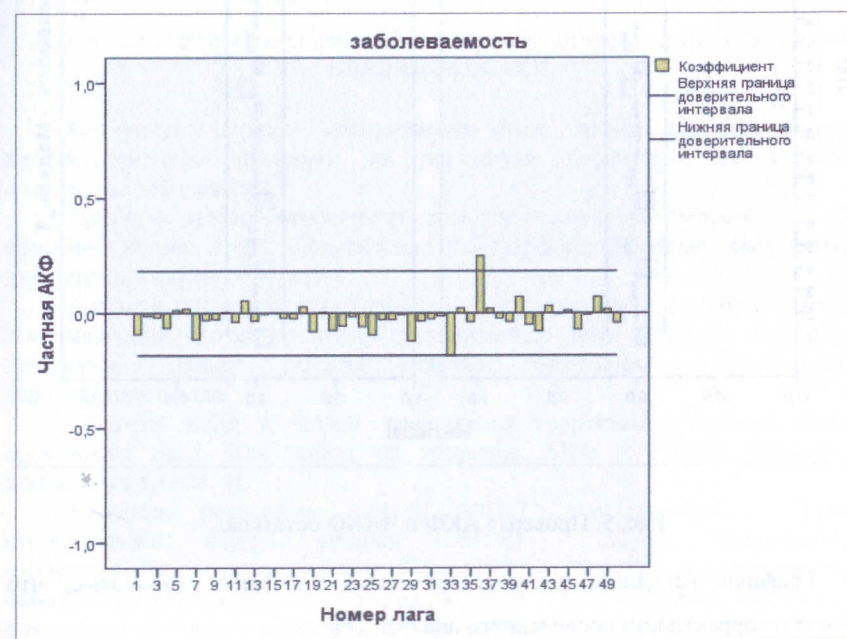
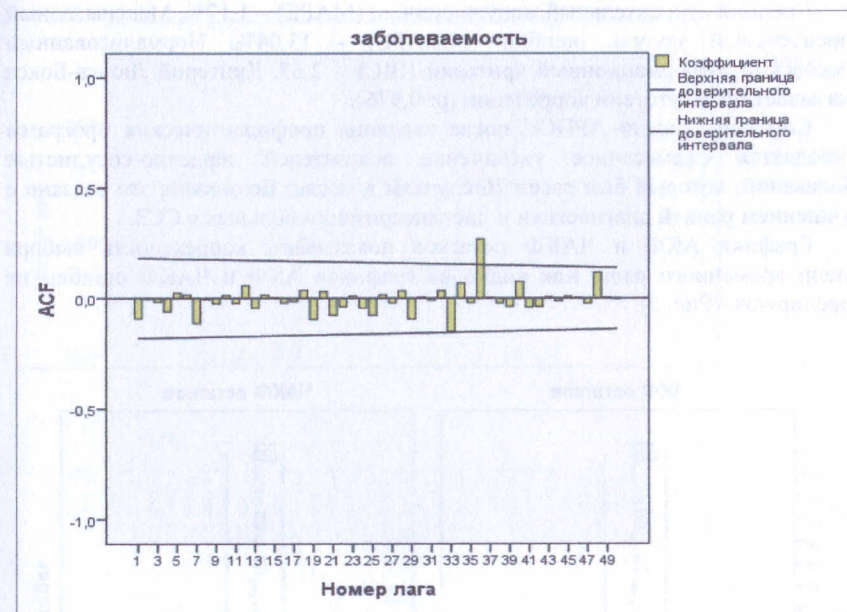


Рис. 4. Автокорреляции и идентификация модели (построение АКФ и ЧАКФ).

Средний относительный модуль ошибки (MAPE) – 1,17%, Максимальный относительный модуль ошибки (MaxAPE) - 13,04%. Нормализованный байесовский информационный критерий (BIC) – 2,67. Критерий Лjung-Бокса показывает об отсутствии корреляции ($p=0,976$).

Согласно модели АРСС после введения профилактических программ наблюдается ежемесячное увеличение показателей сердечно-сосудистых заболеваний, который был равен 16 случаям в месяц. Возможно, это связано с улучшением ранней диагностики и диспансеризации больных с ССЗ.

Графики АКФ и ЧАКФ остатков показывают корректность выбора модели временного ряда. Как видно из графиков АКФ и ЧАКФ ошибки не коррелируют. (Рис. 5).

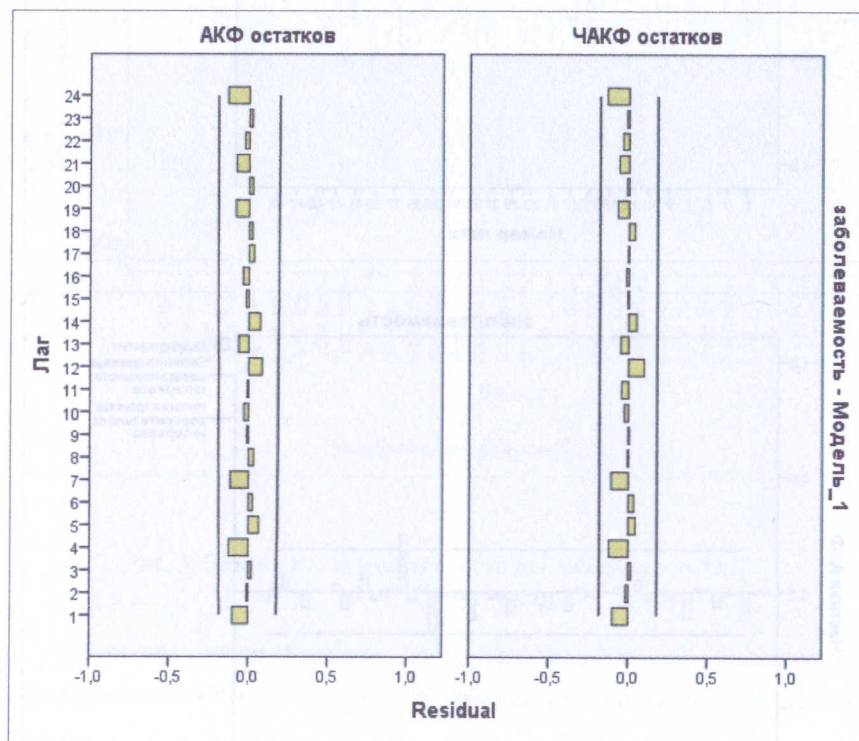


Рис. 5. Проверка АКФ и ЧАКФ остатков.

Графики исходных и смоделированных значений одинаковые, что говорит о корректности проведенного анализа. (Рис. 6).

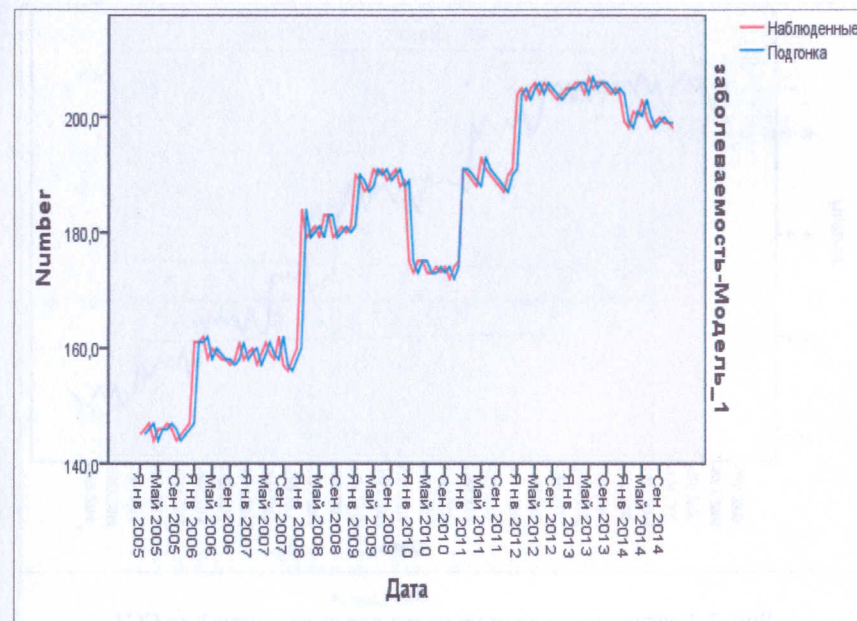


Рис. 6. Исходные (наблюдаемые) и смоделированные (подгонка) значения заболеваемости ССЗ.

Следующим этапом исследования была оценка влияния введения профилактических программ на показатели смертности от сердечно-сосудистых заболеваний.

Графики последовательности для показателя смертности от ССЗ отражены на рис. 7 и 8. Смертность 1 без дифференцирования, смертность 2 дифференцирование 1 порядка.

Графики последовательности (Рис. 7; 8) показывают отсутствие тренда, доказывающий, что модель имеет постоянный средний уровень и постоянную дисперсию. Данное условие отвечает требованиям стационарности вариационного ряда.

Графики АКФ и ЧАКФ показывают корректность выбора модели временного ряда. Как видно из графиков АКФ и ЧАКФ ошибки не коррелируют. (Рис. 9).

Результат построения модели АРСС для смертности. Средний относительный модуль ошибки (MAPE) – 5,29%, Максимальный относительный модуль ошибки (MaxAPE) - 20,61%. Нормализованный байесовский информационный критерий (BIC) – 1,41. Критерий Лjung-Бокса показывает об отсутствии корреляции ($p=0,097$).

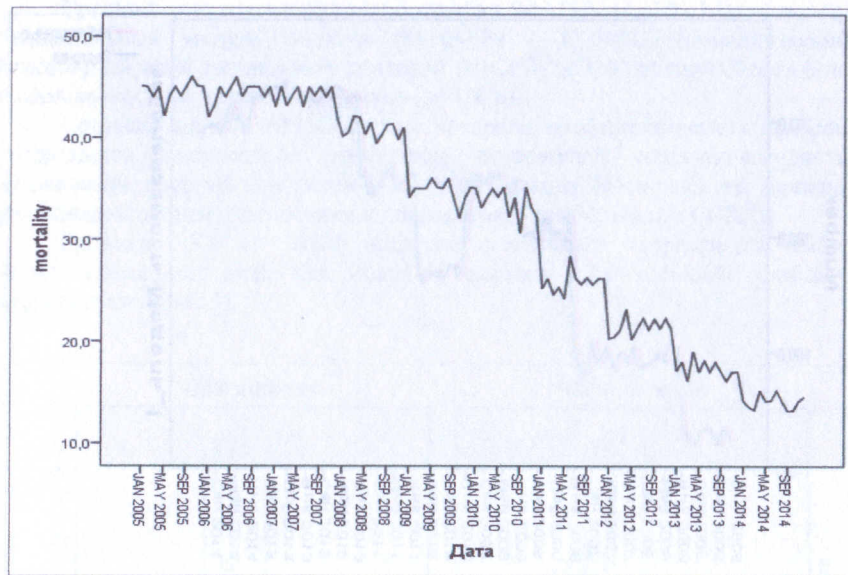


Рис. 7. График последовательности для смертности 1 от ССЗ

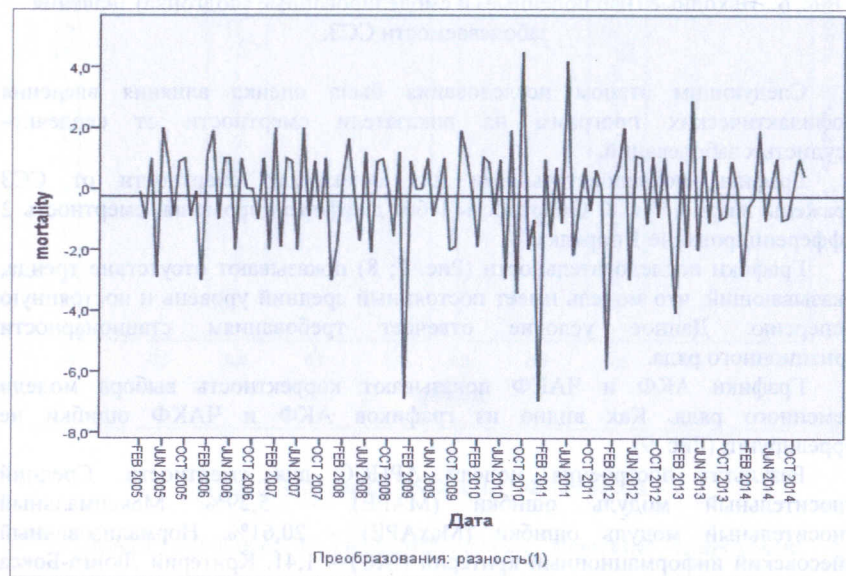


Рис. 8. График последовательности для смертности 2 от ССЗ.

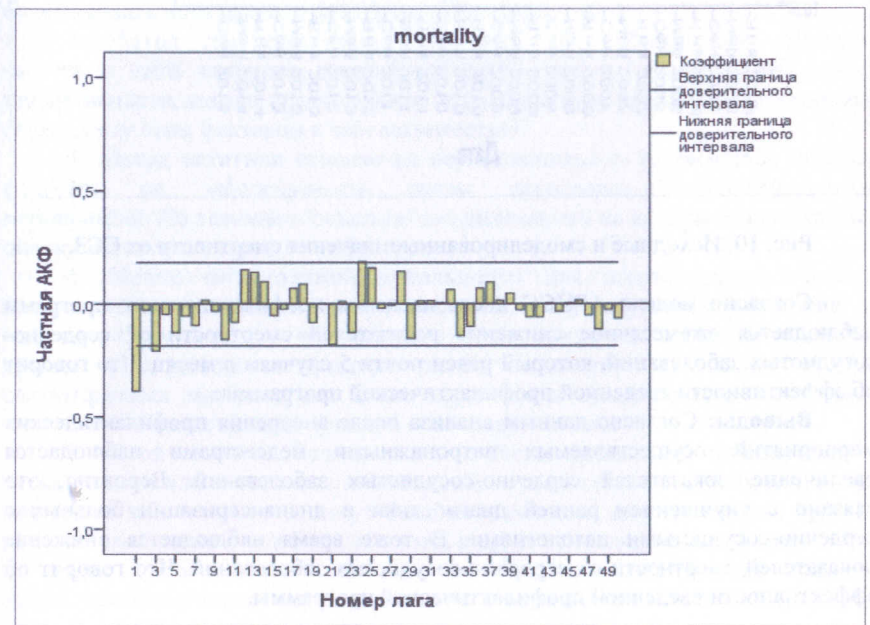
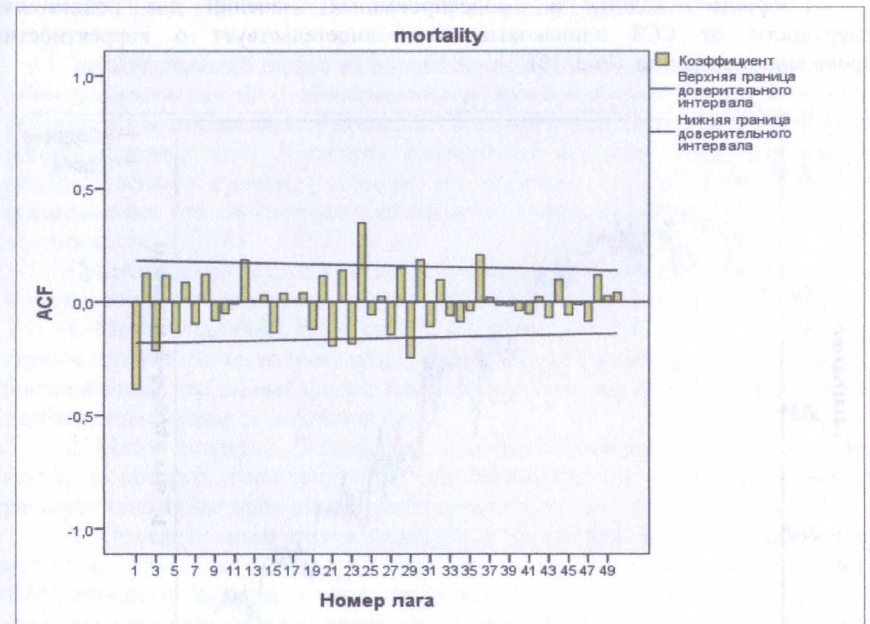


Рис. 9. Автокорреляции и идентификация модели (построение АКФ и ЧАКФ).

Графики исходных и смоделированных значений для показателя смертности от ССЗ одинаковые, что свидетельствует о корректности проведенного анализа. (Рис. 10).

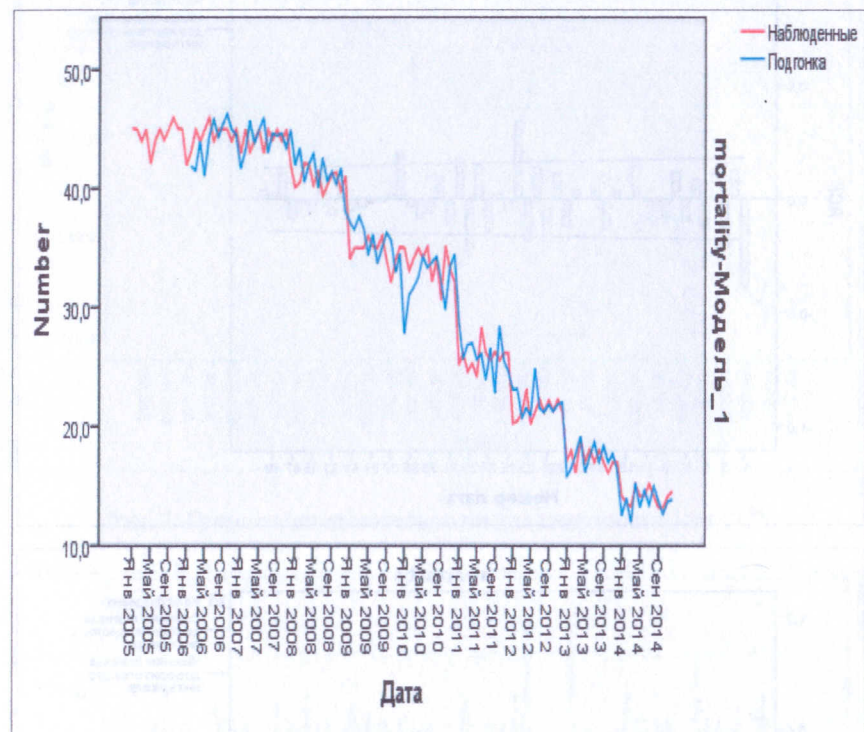


Рис. 10. Исходные и смоделированные значения смертности от ССЗ.

Согласно модели АРПСС после введения профилактических программ наблюдается ежемесячное снижение показателей смертности от сердечно-сосудистых заболеваний, который равен почти 5 случаям в месяц. Что говорит об эффективности введенной профилактической программы.

Выводы: Согласно данным анализа после внедрения профилактических мероприятий осуществляемых патронажными медсестрами наблюдается увеличение показателей сердечно-сосудистых заболеваний. Вероятно, это связано с улучшением ранней диагностики и диспансеризации больных с сердечно-сосудистыми патологиями. В тоже время наблюдается снижение показателей смертности от сердечно-сосудистых заболеваний. Что говорит об эффективности введенной профилактической программы.

ФОРМУЛИРОВКА ГИПОТЕЗ О ПРИЧИНАХ ЗАБОЛЕВАЕМОСТИ

Заключительный раздел в эпидемиологическом анализе неинфекционной заболеваемости связан с обоснованием и проверкой гипотез о причинах и условиях их возникновения и развития. При этом общие гипотезы должны быть детализированы, что позволит соизмерить влияние всех возможных предполагаемых причин, условий и выделить среди них наиболее существенные для конкретных контингентов населения в разных условиях их деятельности.

Существенную помощь в анализе заболеваемости по факторам риска может оказать использование **методов формальной логики**.

1. Метод различия. Если частота заболеваний в двух группах различна и группы отличаются по интенсивности действия какого-либо фактора, то можно предположить, что данный фактор или его отсутствие является причиной либо необходимым условием заболеваемости.

2. Метод сходства. Если фактор является общим для многих различных групп, в которых регистрируется заболеваемость, то этот фактор может рассматриваться как причина или необходимое условие заболеваемости.

3. Объединенный метод сходства и различий. Если при сравнении нескольких групп методом сходства удастся выявить предполагаемую причину заболеваемости, а затем в ряде групп эту причину устранить, т.е. создать различие групп по фактору риска, то по изменениям заболеваемости можно будет оценить роль данного фактора в заболеваемости.

4. Метод сопутствующих изменений. При обнаружении фактора, частота и сила которого изменяется параллельно с изменениями уровня заболеваемости, можно предположить, что существует причинно-следственная связь между этим фактором и заболеваемостью.

5. Метод остатков основан на последовательном исключении влияния факторов на заболеваемость путем проведения целенаправленных мероприятий. По величине "остатков" заболеваемости на каждом этапе и строят оценку значимости каждого из факторов.

6. Метод согласований используется для подтверждения гипотез построенных методами сходства, различий, сопутствующих изменений и остатков. Суть метода - чем большее число признаков согласуется друг с другом и с действием фактора риска, выявленного на основе ранее рассмотренных методов, тем выше вероятность существования причинно-следственной связи данного фактора с заболеваемостью.

7. Метод аналогий. Если выявлено сходство проявлений эпидемического процесса двух или нескольких заболеваний, то закономерности, установленные для одной заболеваемости, могут быть использованы для объяснения причин распространения при подобных условиях и других нозологических форм.

При обосновании гипотез о причинно-следственных связях заболеваемости с факторами риска используют два методических подхода:

- при первом фактор риска предшествует объясняющей его гипотезе;
- при втором - гипотеза предшествует фактору риска, подтверждающему ее.

Используя первый подход, выявленные особенности проявлений эпидемической картины пытаются объяснить на основе представлений о механизме процесса, закономерностях его развития при различных условиях и данных о возможных при изучаемой заболеваемости факторах риска.

При втором подходе гипотезу о роли конкретного фактора риска выдвигают на основе теоретических представлений об распространенности заболеваемости.

В результате анализа заболеваемости в группах населения, с использованием статистических и логических приемов, выявляются причины и условия, определяющие заболеваемость, на которые можно воздействовать потенциально эффективными мероприятиями. Окончательным критерием оценки истинности любой гипотезы является эффективность назначаемых в соответствии с данной гипотезой мероприятий.

ЛИТЕРАТУРА

1. Беляков В.Д., Яфаев Р.Х. «Эпидемиология». - Медицина, 1989.
2. Бессмертный Б.С., Ткачёва М.Н. Статистические методы в эпидемиологии. М., Медгиз, 1961.
3. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. - М. Мир, 1974. - Вып. 1, 2.
4. Дегтярёв А.А. Основы эпидемиологического анализа. Ленинград, 1982.
5. Качество и эффективность противоэпидемических мероприятий / В. Д. Беляков, А.А. Дегтярев, Ю.Г. Иванников, и др. // Медицина, Ленингр. Отделение, 1981.
6. Лакин Г.Ф. Биометрия. М.: Высшая школа, 1980.
7. Мерков А.М., Поляков Л.Е. Санитарная статистика (пособие для врачей). Л., Медицина, 1974.
8. Сепетлиев Дм. Статистические методы в научных медицинских исследованиях. М., Медицина, 1968.
9. Славин М.Б. Методы системного анализа в медицинских исследованиях. М.: Медицина, 1989.
10. Хейфец Л.В., Хазанов М.И. Методика и организация эпидемиологических методов по изучению эффективных вакцин против кишечных инфекций. М., 1959.

Подписано к печати 25.01.2017 г.
Заказ №9. Тираж 100 экз. Бумага офсетная.
Формат бумаги 60 x 90/16. Объем 2,0 п. л.
Отпечатано в ОсОО «Монополия цвета»
720005, г. Бишкек, ул. Байтик-Баатыра, 36