

О НЕКОТОРЫХ ПОДХОДАХ К ВЫЧИСЛЕНИЮ РИСКОВ ТЕМПЕРАТУРНЫХ ВОЛН ДЛЯ ЗДОРОВЬЯ

Д.А. Шапошников, Б.А. Ревич

Институт народнохозяйственного прогнозирования Российской академии наук, Россия, 117418, г. Москва, Нахимовский проспект, 47

Рассматриваются методы оценки влияния факторов окружающей среды на здоровье населения, которые за последние несколько лет стали общепринятыми в мировой практике. Основное внимание уделено современным подходам к вычислению рисков дополнительной смертности в больших группах населения во время волн холода и жары. Обсуждаются основные этапы прямого эпидемиологического исследования: определение температурных волн; формулировка статистических гипотез; спецификация моделей; чувствительность статистических критериев и статистическая достоверность полученных результатов. По результатам многолетних исследований авторов в различных городах России построены логистические кривые вероятности получения значимых результатов оценки риска на малых выборках. В качестве температурных порогов для идентификации температурных волн рекомендуется использовать процентилю многолетних распределений среднесуточных температур воздуха, поскольку такие пороги соответствуют представлению об экстремальных (для данной местности) температурах и дают сопоставимые результаты по ожидаемому количеству волн в различных климатических зонах. Показано, что Пуассоновская обобщенная линейная модель ежедневной смертности является наиболее распространенным методом вычисления рисков, обусловленных неблагоприятными факторами окружающей среды. В регрессионной модели рекомендуется учитывать явную зависимость смертности от времени и загрязнения воздуха. Учет метеорологических условий, влияющих на тепловой баланс (влажность воздуха и скорость ветра) достигается либо включением их в модель в явном виде, либо использованием биоклиматических индексов; исследования в этом направлении продолжаются. При вычислении рисков необходимо учитывать временные лаги между волнами экстремальных температур и откликом смертности. Выявлено, что минимальная численность населения типичного города, для которого еще возможно получить статистически значимые оценки риска ансамблей волн жары, приближается к 200 тысячам человек.

Ключевые слова: смертность населения, температурные волны, анализ временных рядов, оценка риска, распределение Пуассона, обобщенная линейная модель, смешивающие факторы.

Наблюдаемые изменения климата сопровождаются увеличением повторяемости, продолжительности и интенсивности волн тепла и уменьшением этих характеристик для волн холода [1]. Последствия воздействия температурных волн на здоровье населения исследуются в различных странах мира, и в основной медицинской базе данных PubMed приведено более 1000 публикаций по этому направлению исследований. Значительные нарушения здоровья во время температурных волн характеризуются не только его утратой, но и снижением числа здоровых лет жизни, что отражается и на экономических показателях – страновом и региональном валовом продукте.

В данной статье рассматриваются современные биостатистические методы, позволяющие выявить взаимосвязи между метеорологи-

ческими факторами и показателями общественного здоровья. В частности, речь пойдет о вычислении дополнительной смертности населения, обусловленной воздействием волн жары и холода. Величина суточной (ежедневной) смертности для изучаемой популяции является случайной функцией, причем эта случайность неустранима (например, она не является результатом ошибки измерения). Эпидемиологические исследования в этой сфере часто связаны с регрессионным анализом временных рядов суточной смертности. Цель работы – привести описание современных статистических моделей смертности, которые применяются для расчета рисков смертности, обусловленных воздействием температурных волн. Статья условно разделена на три раздела: определение температурных волн; моделирование ежедневной смертно-

© Шапошников Д.А., Ревич Б.А., 2018

Шапошников Дмитрий Анатольевич – кандидат физико-математических наук, старший научный сотрудник лаборатории прогнозирования качества окружающей среды и здоровья населения (e-mail: dshap@newmail.ru; тел.: 8 (926) 598-22-28).

Ревич Борис Александрович – доктор медицинских наук, профессор, заведующий лабораторией прогнозирования качества окружающей среды и здоровья населения (e-mail: revich@yandex.ru; тел.: 8 (499) 129-18-00).

сти; статистическая достоверность результатов оценки риска на малых выборках.

Определение волн жары и холода. В некоторых исследованиях ставится вопрос: что сильнее увеличивает смертность – волны жары или холода? Для корректного ответа целесообразно сформулировать «симметричные» определения волн жары и холода как в отношении их длительности, так и в отношении температурных порогов. Подразумевая такую симметрию, приведем здесь лишь определение волн жары, а волны холода определяются по аналогии.

В метеорологии под волной тепла понимается «значительное потепление, распространяющееся в определенном направлении, связанное с адвекцией теплой массы». Росгидромет дает следующее определение «аномально жаркой погоды»: «...в период с апреля по сентябрь в течение 5 дней и более значение среднесуточной температуры воздуха выше климатической нормы на 7 °С и более»¹. Это определение волн жары должно использоваться МЧС и другими государственными организациями для классификации опасных погодных явлений (ОЯ). При таком определении наибольшее число волн жары в средней полосе России приходится на апрель, а наименьшее на июль. Волны тепла в метеорологии определяются аналогично, однако снимается ограничение «с апреля по сентябрь», так что волны тепла наблюдаются во все сезоны года [2]. Однако главная проблема, связанная с этим определением, заключается в том, что вероятность наблюдения температурных волн (или их количество) будут совершенно разными в различных климатических зонах.

Данная проблема может быть решена, если вместо критерия «7 °С» использовать процентилю «исторического» многолетнего распределения среднесуточных температур или среднеквадратичного отклонения для этого распределения – то есть статистических (вероятностных) характеристик, которые вычислены по единой выборке, включающей либо все дни года, либо все дни теплого (холодного) периодов. Всемирная метеорологическая организация предлагает использовать верхний 90-й или 95-й процентиль

распределения температур в теплый период года в качестве местного специфичного порога в следующем определении: «Волной жары называется четко очерченный двухдневный (или более) эпизод *необычно сильной жары*, измеряемой в суточных максимальных, средних, минимальных температурах, в теплый период года». Пороги предлагается установить с учетом местного климата [3].

Подобное определение используется и в эпидемиологической литературе [4]. Например, в методических рекомендациях Роспотребнадзора предложено считать волной жары пять последовательных дней и более, во время которых среднесуточная температура превышает 97-й процентиль распределения среднесуточных температур за изучаемый многолетний период². Такой порог будет в среднем за год «отсекать» $3,65 \cdot 3 \approx 11$ самых жарких дней, не все из которых войдут в волны жары. «В среднем», потому что волны жары распределены по годам неравномерно, за двумя или тремя прохладными годами обычно следует жаркое лето с несколькими волнами жары. Наибольшее число волн придется на июль – самый жаркий месяц года.

Какое из приведенных выше двух определений более строгое, то есть дает меньше волн из-за более высокого порога жары? Рассмотрим для примера г. Москву. В Москве за десятилетие 2000–2009 гг. средняя температура июня составила 16,5 °С; июля 19,9 °С; августа 17,5 °С. То есть пороги жары для трех летних месяцев, по определению Росгидромета, приблизительно (поскольку мы используем здесь среднемесячные температуры) должны составить 23,5 °С; 26,9 °С; 24,5 °С. За этот же период 97-й процентиль среднесуточных температур в Москве составил 23,5 °С. Видно, что первое определение более строгое: за 10 лет ожидается всего 14 дней с температурами выше таких порогов, то есть всего одна или (маловероятно) две волны жары. Второе определение, как показали наши исследования [5–8], в умеренном климате дает в среднем около восьми волн жары за десятилетний период. Такое количество волн, естественно, дает гораздо больше данных

¹ РД 52.27.724-2009. Наставление по краткосрочным прогнозам погоды общего назначения: руководящий документ. – Обнинск: ИГ-СОЦИН, 2009. – 66 с.

² МР 2.1.10.0057-12. Состояние здоровья населения в связи с состоянием окружающей среды и условиями проживания населения. Оценка риска и ущерба от климатических изменений, влияющих на повышение уровня заболеваемости и смертности в группах населения повышенного риска: методические рекомендации [Электронный ресурс]. – М.: Федеральная служба по надзору в сфере защиты прав потребителей и благополучия человека, 2012. – 37 с. – URL: <http://36.rosпотребнадзор.ru/documents/rekdoc1/9374> (дата обращения: 11.10.2017).

для эпидемиологических исследований и в то же время отвечает интуитивным представлениям об экстремальных температурах.

Биоклиматические индексы. В статистических моделях смертности температура, влажность, скорость ветра и другие метеофакторы могут присутствовать как независимые переменные, то есть вычисляются риски, обусловленные температурой, с поправкой на другие метеорологические переменные. Однако есть и другой подход, когда в моделях используют определенную комбинацию этих переменных, имеющую размерность температуры, называя ее, например, эффективной температурой (для жары) или ветрохолодовым индексом (для холода). Эпидемиологи продолжают искать наиболее подходящую комбинацию метеорологических переменных – такую, для которой статистический тест показал бы наибольшую силу связи с выбранным показателем здоровья [9, 10]. Соответственно, волны жары можно определить не через обычную температуру воздуха (сухого термометра), а через эффективную температуру. Одно из определений эффективной температуры приведено в методических рекомендациях Роспотребнадзора [11]:

$$AT = -2,653 + 0,994 T + 0,0153 D^2,$$

где AT – эффективная температура; T – температура воздуха; D – температура точки росы.

В одном нашем исследовании было показано, что эффективная температура теснее связана со смертностью во время волн жары, а ветрохолодовой индекс является лучшим предиктором смертности во время волн холода, чем обычная температура воздуха [5]. В данной работе используется обычная температура воздуха в качестве переменной для идентификации температурных волн (то есть волн жары и волн холода), без ограничения общности изложения.

Формулировка и тестирование статистических гипотез. Любое статистическое исследование начинается с формулировки статистических гипотез, которые соответствовали бы цели конкретного исследования. Если речь идет о влиянии температурных волн на здоровье населения, то требуется доказать наличие статистической связи между такими волнами и числом ежедневных исходов выбранного показателя здоровья. Хотя статистические методы для изучения волн холода и волн жары применяются одни и те же, есть одно важное отличие между влиянием холода и жары на смертность –

временной аспект: более отсроченное влияние холода и почти мгновенное влияние жары. Хотя объяснение этому различию надо искать в патофизиологических механизмах действия холода и жары, адаптации и краткосрочной акклиматизации организма к экстремальным температурам, целью статистического исследования является лишь выявление, а не объяснение статистических закономерностей. Поэтому биостатистик, изучающий предикторы смертности, например температуру воздуха, ограничивается корректным выводом типа «обнаружена взаимосвязь между температурой и смертностью». В этом случае температура называется фактором риска, причем этот фактор может быть причиной или всего лишь маркером эффекта.

Если речь идет об одной конкретной волне, то нулевая гипотеза (H_0) может формулироваться так: наблюдавшаяся в период волны смертность статистически не отличима от ожидавшейся при прочих равных условиях в отсутствие волны. Эта же гипотеза очевидным образом обобщается на исследование ансамбля (некоторой совокупности) волн за многолетний период исследования. При этом индивидуальные волны жары в ансамбле будут различаться как по длительности, так и по интенсивности (амплитуде). Принимая во внимание цели исследования, можно изучать зависимости конечных показателей здоровья от этих параметров, составляя соответствующие волновые ансамбли. Простейшие модели, особенно в условиях коротких временных рядов, изучают ансамбль *всех* идентифицированных волн, таким образом вычисляя лишь усредненный по ансамблю риск дополнительной смертности. По определению, риск есть отношение ожидаемой смертности во время изучаемого ансамбля волн к ожидаемой смертности в эти же календарные даты, но в отсутствие экстремальной жары.

Возможное наличие временного лага между волной температуры и откликом смертности усложняет формулировку нулевой гипотезы, поскольку требует дополнительных предположений об отсроченном характере воздействия и соответствующего математического аппарата.

Обычно объектом исследования является население одного города. Численность населения определяет математическое ожидание суточной смертности как случайной величины и дисперсию этого показателя, которые в свою очередь влияют на чувствительность применяемого для проверки H_0 статистического теста.

Анализ временных рядов. Наиболее распространенным методом исследования длинных временных рядов суточной смертности стал регрессионный анализ с помощью Гауссовых и Пуассоновских обобщенных линейных моделей [11]. Предположение о независимости исходов приводит к Пуассоновскому распределению суточной смертности. В действительности для ежедневной смертности основное свойство распределения Пуассона (математическое ожидание ежедневного числа исходов λ равно дисперсии) выполняется лишь приближенно. По мере возрастания λ появляется избыточная дисперсия, а само распределение ежедневной смертности при больших λ становится Гауссовым. Одной из причин этого явления является, по-видимому, нарушение предположения о независимости исходов.

Если зависимая переменная подчиняется распределению Пуассона, то при использовании натурального логарифма в качестве функциональной связи в регрессионной модели регрессионные остатки будут распределены по нормальному закону. Если же зависимая переменная распределена по нормальному закону, то корректно использовать линейную функциональную связь (Гауссову регрессию). Например, линейную регрессию можно использовать в Москве, где $\lambda \approx 300$. Однако большого «выигрыша» от этого не получается: при использовании Пуассоновской модели в Москве получаются сравнимые по статистической значимости результаты оценки рисков.

Спецификация модели. Рассмотрим Пуассоновскую модель ежедневной смертности M_i . Поскольку временной ряд M_i является случайной функцией номера дня i , то в регрессионном уравнении зависимой переменной является ожидаемая величина ежедневной смертности $E(M_i) \equiv \mu_i$. Удобно разделить все предикторы на явно зависящие от времени t и все остальные регрессоры суточного разрешения x_1, x_2, \dots, x_p , в том числе экологические, то есть описывающие влияние окружающей среды (метеофакторы, загрязнение), которые являются медиаторами эффекта [12]:

$$\log[E(M_i)] = \alpha + \sum_{j=1}^p g_j(x_{ij}) + f(t_i) + \beta DOW. \quad (1)$$

Достоинством обобщенной линейной регрессии является гибкость в выборе функциональных зависимостей g_j от непрерывных регрессоров и возможность включения в модель

дискретных регрессоров без ограничения общности. Примером является последнее слагаемое в модели (1) – вектор из семи категориальных переменных, маркирующих дни недели DOW , с соответствующими регрессионными коэффициентами β .

Явная зависимость ежедневной смертности от времени включает сезонность, многолетний тренд, зависимость от дня недели, праздников, периодов эпидемий гриппа и т.п. Если исследователя интересуют сезонные изменения смертности, то функция $f(t)$ может содержать периодическое слагаемое с периодом один год. В общем случае учет плавных зависимостей от времени на различных масштабах достигается включением в модель сплайнов $S(t)$ с заданным числом степеней свободы (или узлов) за весь период исследования.

Оценка риска волн жары. Если нас интересует лишь общий эффект волн жары на смертность, то регрессионная модель вообще не будет содержать температуры воздуха в явном виде – достаточно лишь бинарной переменной, маркирующей все дни, входящие в изучаемый ансамбль волн жары (с заданной временной задержкой в днях, или *лагом*). Регрессионный коэффициент при этой переменной β_{heat} (точнее, его экспонента) и будет численно характеризовать прирост смертности в среднем для данного ансамбля волн, скорректированный на все остальные факторы, влияющие на смертность и учтенные в модели (1).

Общий эффект волн жары можно разделить на два слагаемых. Первое слагаемое называется «основным эффектом» температуры, поскольку зависит от среднесуточных температур воздуха. Второе слагаемое представляет собой «волновую добавку», которая возникает лишь при продолжительном воздействии жары и является функцией номера дня в непрерывной последовательности жарких дней. Как показал А. Гаспарини, для обычных по продолжительности волн (длиной 5–10 дней) основной эффект в несколько раз превышает волновую добавку [13]. Впрочем, для аномально продолжительной волны жары в Москве летом 2010 г. мы наблюдали противоположное соотношение [14].

Исследователи давно заметили, что основной эффект температуры является не «мгновенным», а сложным образом распределен во времени, то есть влияет на смертность во все следующие за референтным дни. Были предложены различные способы исследования отсроченных зависимостей с учетом временного лага

между воздействием и эффектом [15, 16]. Статистические функции, описывающие нелинейные модели с распределенным лагом, были интегрированы в программную среду *R* и сейчас находятся в открытом доступе в CRAN (Comprehensive R Archive Network) [17].

Смешивающие факторы. Согласно «вентиляционной гипотезе», загрязнение атмосферного воздуха по-разному влияет на здоровье в жаркие дни и в дни с умеренной температурой. В жаркие дни люди постоянно открывают окна для вентиляции и больше времени проводят на улице, поэтому подвергаются большим дозам воздействия загрязнителей воздуха [18]. Если эта гипотеза справедлива, то определенная часть риска высоких температур будет на самом деле обусловлена содержанием загрязняющих веществ в воздухе (прежде всего PM_{10}) и повышенной экспозицией к таким поллютантам. Что касается приземного озона, то здесь тоже есть функциональная связь с температурой, поскольку скорости многих фотохимических реакций зависят от температуры воздуха. Поэтому загрязнение воздуха можно считать медиатором эффекта жары, а учет среднесуточных уровней PM_{10} , NO_2 , 8 ч озона в модели (1) является хорошей практикой (при наличии достаточных данных о загрязнении).

Вариации атмосферного давления тоже влияют на ежедневную смертность [19]. Если загрязнение обычно включается в регрессионную модель в виде линейных членов (согласно гипотезе о беспороговом действии), то для правильного описания влияния давления требуется более гибкое представление. Как высокое, так и низкое давление, а также быстрое падение давления статистически увеличивают смертность. Немонотонные зависимости удобно описывать кубическими сплайнами с «естественными» граничными условиями (равенство нулю второй производной на границах диапазона сплайнирования).

Влияние численности населения на достоверность результатов. Трудности изучения малых выборок. Результаты наших исследований влияния температурных волн на показатели смертности в девяти городах России с различной численностью населения можно использовать для оценки вероятности получения статистически значимых результатов оценки риска температурных волн при различных значениях λ (средней суточной смертности). Описанная ниже модель должна помочь другим исследователям при планировании исследований на малых выборках, когда возникает во-

прос: при какой минимальной численности населения можно получить значимые результаты оценки риска температурных волн? В данном случае мы исследуем бинарный признак значимости на 95%-ном уровне регрессионного коэффициента β_{heat} в модели (1), поэтому естественно предположить прирост (а не снижение) дополнительной смертности во время действия стресс-фактора и использовать односторонний *z*-тест. Моделирование вероятности получения или неполучения значимого результата в зависимости от численности населения аналогично построению кривой спроса в эконометрике (согласие или несогласие купить товар в зависимости от цены).

Исходные данные взяты из собственных работ, выполненных по сопоставимым методикам в девяти городах и на сопоставимых периодах исследования: в среднем около 10 лет. В этих исследованиях вычислялись риски смертности во время ансамблей волн (отдельно для жары и холода) длительностью от 5 дней и более, которые были идентифицированы за период исследования в данном городе. Данные о ежедневной смертности были получены из базы данных Росстата, а данные о температуре воздуха и других метеорологических переменных с сайта Росгидромета (<http://cliwave.meteo.ru/meteo/>). Риски смертности от климатозависимых причин должны быть более значимы, чем для общей смертности. Как показывает опыт, важнейшими климатозависимыми причинами являются инфаркты (коды I20–I25 по международной классификации болезней МКБ-10) и инсульты (I60–I69). Например, во время необычно сильной и продолжительной волны жары в Москве в период с 6 июня по 18 августа 2010 г. дополнительная смертность от всех причин по нашим оценкам составила 11 040 случаев, из них 5045 смертей от инфарктов (46 %) и 3712 от инсультов (34 %) [14]. Значит, сердечно-сосудистыми причинами объясняется до 80 % всей дополнительной смертности в Москве в этот период.

Для получения достаточного количества исходных данных для моделирования вероятности получения значимых результатов оценки рисков волн жары и холода в зависимости от λ вычислялись риски отдельно для «средней» и «старшей» возрастной группы: случаи смерти в возрасте 30–64 лет и случаи смерти в возрасте старше 65 лет. Такое деление на две возрастные группы является общепринятым в мировой практике, поскольку во многих странах 65 лет является пенсионным возрастом, и соответст-

венно возрастная группа от 30 до 64 лет представляет работоспособное население. Количество смертей в возрасте до 30 лет пренебрежимо мало. Многие медико-статистические показатели выявляются для этих двух групп с целью обнаружения возрастных различий. Таким образом, в каждом городе исследовались четыре показателя смертности: раздельно случаи смерти от инфарктов и инсультов в двух возрастных группах. Предполагалось, что вероятность получения статистически значимых результатов оценки риска при заданном значении параметра λ приблизительно одинакова для данных четырех показателей. В противном случае результа-

ты оценки риска для этих четырех показателей нельзя было бы объединить в одну выборку. Это предположение выполняется в том случае, если сами риски для изучаемых показателей смертности сопоставимы по величине. Для проверки справедливости сделанных предположений в таблице приводится не только бинарная переменная (0 = риск не значим; 1 = риск значим), но и сами оценки относительных приростов смертности во время волн жары или холода. При этом целью данной таблицы вовсе не является сравнение абсолютных величин полученных ранее оценок риска между холодом и жарой или между севером и югом.

Результаты оценки рисков температурных волн в девяти городах

Город *	Причина смерти и возраст на этот момент	λ	RR_{heat}	Значимость (RR_{heat})	RR_{cold}	Значимость (RR_{cold})
Архангельск (350 тыс.)	Инфаркт 30–64	1,4	0,94	0	1,18	1
	Инфаркт ≥ 65	2,1	0,93	0	1,22	1
	Инсульт 30–64	0,62	1,01	0	1,13	0
	Инсульт ≥ 65	2,5	1,30	1	1,19	1
Мурманск (325 тыс.)	Инфаркт 30–64	1,7	1,03	0	1,18	1
	Инфаркт ≥ 65	1,7	0,76	0	1,09	0
	Инсульт 30–64	0,65	0,88	0	1,07	0
	Инсульт ≥ 65	1,7	1,25	1	1,14	0
Якутск (236 тыс.)	Инфаркт 30–64	0,42	1,15	0	1,38	1
	Инфаркт ≥ 65	0,55	0,90	0	1,41	1
	Инсульт 30–64	0,23	0,91	0	0,8	0
	Инсульт ≥ 65	0,32	1,61	1	1,69	1
Магадан (100 тыс.)	Инфаркт 30–64	0,31	1,44	0	1,01	0
	Инфаркт ≥ 65	0,31	1,23	0	1,39	1
	Инсульт 30–64	0,19	1,57	0	1,37	0
	Инсульт ≥ 65	0,25	1,23	0	1,66	1
Волгоград (989 тыс.)	Инфаркт 30–64	3,5	1,25	1	1,12	1
	Инфаркт ≥ 65	7,6	1,39	1	1,10	1
	Инсульт 30–64	1,5	1,36	1	1,29	1
	Инсульт ≥ 65	10,7	1,55	1	1,08	1
Ростов (1053 тыс.)	Инфаркт 30–64	2,1	1,20	1	1,16	1
	Инфаркт ≥ 65	7,0	1,39	1	1,12	1
	Инсульт 30–64	1,4	1,51	1	1,23	1
	Инсульт ≥ 65	9,7	1,75	1	1,15	1
Астрахань (500 тыс.)	Инфаркт 30–64	1,7	1,42	1	1,23	1
	Инфаркт ≥ 65	4,2	1,58	1	1,14	1
	Инсульт 30–64	0,72	1,40	1	1,04	0
	Инсульт ≥ 65	3,4	1,57	1	1,28	1
Краснодар (710 тыс.)	Инфаркт 30–64	1,4	1,24	1	1,14	0
	Инфаркт ≥ 65	5,6	1,37	1	1,17	1
	Инсульт 30–64	0,93	1,50	1	1,33	1
	Инсульт ≥ 65	6,4	1,76	1	1,08	1
Красноярск (932 тыс.)	Инфаркт 30–64	2,2	1,10	0	1,17	1
	Инфаркт ≥ 65	5,3	1,14	1	1,04	0
	Инсульт 30–64	0,96	1,19	0	1,24	1
	Инсульт ≥ 65	4,6	1,44	1	1,11	1

Примечание: * – численность населения приведена в тысячах на середину периода исследования; λ – средняя смертность за период исследования, число случаев в день; RR_{cold} и RR_{heat} – относительные риски смертности во время волн холода и жары; значимость: 1 = риск значим на 95%-ном уровне; 0 = риск не значим на 95%-ном уровне.

В источниках данных по городам (Архангельск, Мурманск, Якутск, Магадан [8]; Волгоград, Ростов, Астрахань, Краснодар [6, 7]; Красноярск [5]) подробно описаны Пуассоновские модели, использованные авторами для оценки рисков.

Условие максимизации вероятности получения достоверной оценки риска при данном значении λ записывается следующим образом: требуется найти такую функцию $\pi(\lambda) \in \{0; 1\}$, что

$$\mu = E(y|\lambda) = P(y = 1|\lambda) \equiv \pi(\lambda), \quad (2)$$

где λ – предиктор; E – ожидаемая величина; P – вероятность; $y \in \{0; 1\}$ – бинарный признак или отклик. Предполагая разумные граничные условия для такой задачи, вероятность наступления события $y_i = 1$ в зависимости от λ_i принято аппроксимировать бинарной логистической регрессией:

$$\pi(\lambda) = \frac{\exp(\beta_0 + \beta_1\lambda)}{1 + \exp(\beta_0 + \beta_1\lambda)}, \quad (3)$$

где β_0 и β_1 – оценочные значения регрессионных коэффициентов, которые вычисляются линейной регрессией обратного преобразования:

$$\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1\lambda + \varepsilon. \quad (4)$$

Регрессионный коэффициент β_0 характеризует вероятность получения значимого результата оценки риска при нулевой численности населения, а β_1 – предельный эффект (marginal effect) при различных значениях λ .

На рис. 1 и 2 треугольниками показаны исходные данные для регрессии, взятые из таблицы, а также изображены логистические кривые, аппроксимирующие вероятность $\pi(\lambda)$ в диапазоне значений $0,1 < \lambda < 5$. Пунктиром показана стандартная ошибка регрессии $\mu \pm \sigma$. Регрессия была выполнена с помощью команды logistic в программе Stata 14.0.

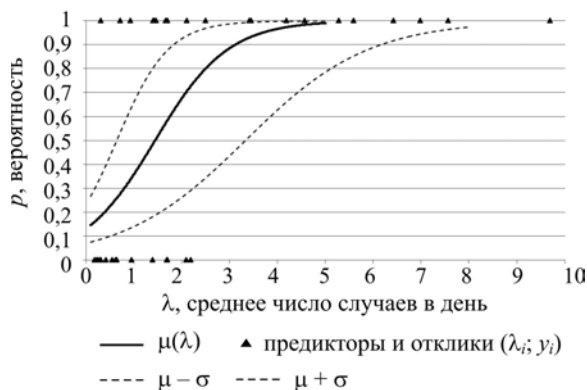


Рис. 1. Вероятность получения значимого результата оценки риска для волн жары

Для волн жары (рис. 1) оба регрессионных коэффициента β_0 и β_1 статистически значимы: $\beta_0 = -1,91 \pm 0,82$, $p = 0,020$; $\beta_1 = 1,32 \pm 0,54$, $p = 0,014$. Поэтому логистическая кривая позволяет предсказать, при каких λ можно ожидать получения достоверной оценки риска. Условие $\mu > 1/2$ (получение значимого результата более вероятно, чем незначимого) выполняется при $\lambda > 1,5$ (этот результат справедлив для периодов исследования около 10 лет). Какой же численности населения соответствует $\lambda \approx 1,5$?

Как видно из таблицы, среди включенных в модель четырех показателей смертности преобладает инсульт в старшей возрастной группе, для которой средняя ежедневная смертность составляет 40 случаев на суммарное население девяти городов 5,195 миллиона. Пренебрегая возможной гетерогенностью соотношений между показателями смертности в различных климатических зонах, получим, что 1,5 случая инсульта в день соответствует численности населения $5195 \cdot 1,5/40 = 195$ тысяч человек. Это и есть минимальная численность населения, при которой можно ожидать достоверной оценки риска волн жары хотя бы для одного из четырех выбранных показателей смертности.

Для волн холода логистическая кривая не опускается ниже $p = 0,5$ (рис. 2), так что рассматриваемую задачу о минимальной численности населения не удастся решить предложенным способом. Причина этого, скорее всего, кроется в переоценке числа достоверных оценок рисков волн холода в области малых значений λ , связанной с перебором авторами большого числа лагов (от одного дня до трех недель) при поиске наиболее вероятного лага между холодом и откликом смертности. Перебор лагов увеличивает вероятность ошибки второго рода (т.е. обнаружение различий там, где их нет). Для волн жары процедура отбора

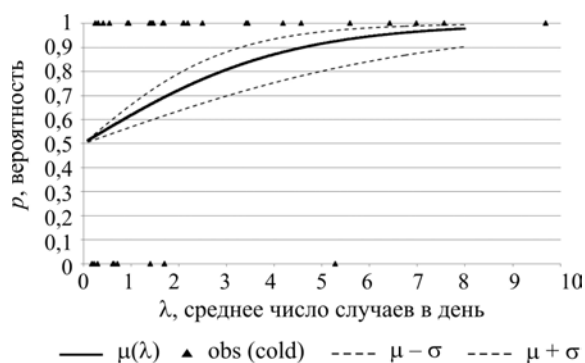


Рис. 2. Вероятность получения значимого результата оценки риска для волн холода

гораздо более строгая, поскольку допускались лишь короткие лаги до 5 дней.

Выводы. В качестве температурных порогов для идентификации температурных волн рекомендуется использовать процентилю многолетних распределений среднесуточных температур воздуха. Такие пороги соответствуют представлению об экстремальных (для данной местности) температурах и дают сопоставимые результаты по ожидаемому количеству волн в различных климатических зонах. Пуассоновская обобщенная линейная модель ежедневной смертности является наиболее распространенным методом вычисления рисков, обусловленных неблагоприятными факторами окружающей среды. В регрессионной модели рекомендуется учитывать явную зависимость смертности от времени и загрязнения воздуха.

Учет метеорологических условий, влияющих на тепловой баланс (влажность воздуха и скорость ветра), достигается либо включением их в модель в явном виде, либо использованием биоклиматических индексов; исследования в этом направлении продолжаются. При вычислении рисков необходимо учитывать временные лаги между волнами экстремальных температур и откликом смертности. При исследованиях в малых городах, в которых используются статистические данные о ежедневной смертности за периоды около 10 лет, обычно не удается получить достоверных оценок риска температурных волн при населении менее 200 тысяч человек. В данной работе этот результат получен для волн жары, однако, вероятнее всего, он должен быть справедливым и в отношении волн холода.

Список литературы

1. Второй оценочный доклад Росгидромета об изменениях климата и их последствиях на территории Российской Федерации. Раздел 1. Наблюдаемые изменения климата. – М.: Федеральная служба по гидрометеорологии и мониторингу окружающей среды, 2014. – С. 18–235.
2. Клещенко Л.К. Волны тепла и холода на территории России // Анализ изменений климата и их последствий: сборник трудов Всероссийского научно-исследовательского института гидрометеорологической информации – Мирового центра данных / под ред. В.Н. Разуваева, Б.Г. Шерстюкова. – 2010. – № 175. – С. 76–91.
3. Guidelines on the definition and monitoring of extreme weather and climate events: Draft version – first review by TT-Dewce desember 2015 [Электронный ресурс] // World Meteorological Organization. – 2016. – 62 p. [Электронный ресурс]. – URL: <https://www.wmo.int/pages/prog/wcp/ccl/opace/opace2/documents/DraftversionoftheGuidelinesontheDefinitionandMonitoringofExtremeWeatherandClimateEvents.pdf> (дата обращения: 12.10.2017).
4. Radinovic D., Curic M. Criteria for heat and cold wave duration // Theor. Appl. Climatol. – 2012. – Vol. 97, № 3–4. – P. 505–510.
5. Ревич Б.А., Шапошников Д.А. Особенности воздействия волн жары и холода на смертность в городах с резко-континентальным климатом // Сибирское медицинское обозрение. – 2017. – № 2. – С. 84–90.
6. Ревич Б.А., Шапошников Д.А. Волны холода в южных городах европейской части России и преждевременная смертность населения // Проблемы прогнозирования. – 2016. – № 2. – С. 125–131.
7. Волны жары в южных городах европейской части России как фактор риска преждевременной смертности населения / Б.А. Ревич, Д.А. Шапошников, М.А. Подольная, Т.Л. Харьковская, Е.А. Кваша // Проблемы прогнозирования. – 2015. – № 2. – С. 56–67.
8. Shaposhnikov D., Revich B. Towards meta-analysis of impacts of heat and cold waves on mortality in Russian North // Urban. Climate. – 2016. – Vol. 15. – P. 16–24. DOI: 10.1016/j.uclim.2015.11.007
9. Anderson G.B., Bell M.L., Peng R.D. Methods to calculate the heat index as an exposure metric in environmental health research // Environ. Health Perspect. – 2013. – Vol. 121, № 10. – P. 1111–1119. DOI: 10.1289/ehp.1206273
10. Report on Wind Chill Temperature and Extreme Heat Indices: Evaluation and Improvement Projects [Электронный ресурс]. – Washington, DC: Federal Coordinator for Meteorological Services and Supporting Research, 2003. – URL: <http://solberg.snr.missouri.edu/gcc/OFCMWindchillReport.pdf> (дата обращения: 08.09.2017).
11. Рябенко Е.А. Прикладной статистический анализ данных. Обобщения линейной регрессии [Электронный ресурс]. – 2016. – URL: http://www.machinelearning.ru/wiki/images/a/a6/Psad_otherreg.pdf (дата обращения 13.06.2017).
12. Richiardi L., Bellocco R., Zugna D. Mediation analysis in epidemiology: methods, interpretation and bias // Int. J. Epidemiol. – 2013. – Vol. 42, № 5. – P. 1511–1519. DOI: 10.1093/ije/dyt127
13. Gasparrini A., Armstrong B. The impact of heat waves on mortality // Epidemiology. – 2011. – Vol. 22, № 1. – P. 68–73. DOI: 10.1097/EDE.0b013e3181fdcd99
14. Mortality Related to Air Pollution with the Moscow Heat Wave and Wildfire of 2010 [Электронный ресурс] / D. Shaposhnikov, B. Revich, T. Bellander [et al.] // Epidemiology. – 2014. – Vol. 25, № 3. – P. 359–364. – URL: <https://www.ncbi.nlm.nih.gov/pubmed/24598414> (дата обращения: 08.09.2017).

15. Gasparrini A. Distributed lag linear and non-linear models in R: the package dlrm // Journal of Statistical Software. – 2011. – Vol. 43, № 8. – P. 1–20. DOI: 10.18637/jss.v043.i08
16. Schwartz J. The distributed lag between air pollution and daily deaths // Epidemiology. – 2000. – Vol. 11. – P. 320–326.
17. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2016 [Электронный ресурс] // R. Core. Team. – 2017. – URL: <http://www.R-project.org/> (дата обращения: 18.12.2017).
18. Does Temperature Modify the Association between Air Pollution and Mortality? A Multicity Case-Crossover Analysis in Italy / M. Stafoggia, J. Schwartz, F. Forastiere, C.A. Perucci, and the SISTI Group // Am. J. Epidemiol. – 2008. – Vol. 167, № 12. – P. 1476–1485. DOI: 10.1093/aje/kwn074
19. Pope C.A. 3-rd, Hill R.W., Villegas G.M. Particulate air pollution and daily mortality on Utah's Wasatch Front // Environ Health Perspect. – 1999. – Vol. 107, № 7. – P. 567–573.

Шапошников Д.А., Ревич Б.А. О некоторых подходах к вычислению рисков температурных волн для здоровья // Анализ риска здоровью. – 2018. – № 1. – С. 22–31. DOI: 10.21668/health.risk/2018.1.03

UDC 613.1: 616 02

DOI: 10.21668/health.risk/2018.1.03.eng

ON SOME APPROACHES TO CALCULATION OF HEALTH RISKS CAUSED BY TEMPERATURE WAVES

D.A. Shaposhnikov, B.A. Revich

National Economic Prediction Institute of the Russian Academy of Sciences, 47 Nakhimovsky Prospect, Moscow, 117418, Russian Federation

The paper dwells on techniques applied for assessing impacts exerted by environmental factors on population health which have become conventional all over the world over recent years. The greatest attention is paid to up-to-date approaches to calculating risks of additional mortality which occurs in big population groups during cold and hot temperature waves. The authors consider basic stages in direct epidemiologic research: temperature waves definition; statistics hypotheses formulation; models specification; statistical criteria sensitivity, and statistical validity of the obtained results. As per long-term research performed by us in various Russian cities, we constructed logistic curves which show probability of obtaining significant risk assessment results for small samplings. We recommend to apply percentiles of long-term average daily temperature distributions as temperature thresholds when identifying temperature waves; in our opinion, such thresholds correspond to perceptions of extreme (for this or that region) temperatures and provide comparable results in terms of expected waves quantity in different climatic zones. Poisson's generalized linear model for daily mortality is shown to be the most widely spread technique for calculating risks caused by hazardous environmental factors. It is advisable to allow for an apparent correlation between mortality and time and air contamination in any regression model. We can allow for meteorological conditions which influence heat balance (air humidity and wind speed) either via including them apparently into a model or via bioclimatic indexes application; research in this sphere is going on. When calculating risks, it is advisable to allow for time lags between extreme temperatures waves and changes in mortality. We revealed that minimal population of a typical city for which it is possible to obtain statistically significant assessment of risks caused by heat waves ensembles is about 200 000 people.

Key words: population mortality, temperature waves, time rows analysis, risk assessment, Poisson's distribution, generalized linear model, mixing factors.

© Shaposhnikov D.A., Revich B.A., 2018

Dmitry A. Shaposhnikov – Candidate of Physical-Mathematical Sciences, Senior Researcher at Laboratory for Environment Quality Prediction and Population Health (e-mail: dshap@newmail.ru; tel.: +7 (926) 598-22-28).

Boris A. Revich – Doctor of Medical Sciences, Professor, Head of Laboratory for Environment Quality Prediction and Population Health (e-mail: revich@yandex.ru; tel.: +7 (499) 129-18-00).

References

1. Vtoroi otsenochnyi doklad Rosgidrometa ob izmeneniyakh klimata i ikh posledstviyakh na territorii Rossiiskoi Federatsii. Razdel 1. Nablyudaemye izmeneniya klimata [The second assessment report by the Federal Service for Hydrometeorology and Environmental Monitoring on climatic changes and their consequences on the RF territory. Section 1. Observed climatic changes]. Moscow, Federal'naya sluzhba po gidrometeorologii i monitoringu okruzhayushchei sredy Publ., 2014, pp. 18–235 (in Russian).
2. Kleshchenko L.K. Volny tepla i kholoda na territorii Rossii [Cold and heat waves on the RF territory]. *Analiz izmenenii klimata i ikh posledstviy: sbornik trudov Vserossiiskogo nauchno-issledovatel'skogo instituta gidrometeorologicheskoi informatsii – Mirovogo tsentra dannykh* [Analysis of climatic changes and their consequences: collection of works issued by the Russian Scientific-research Institute for Hydrometeorological Information – World data center]. In: V.N. Razuvaev, B.G. Sherstyukov, eds. 2010, no. 175, pp. 76–91 (in Russian).
3. Guidelines on the definition and monitoring of extreme weather and climate events: Draft version – first review by TT-Dewce desember 2015. World Meteorological Organization, 2016, 62 p. Available at: <https://www.wmo.int/pages/prog/wcp/ccl/opace/opace2/documents/DraftversionoftheGuidelinesontheDefinitionandMonitoringofExtremeWeatherandClimateEvents.pdf> (12.10.2017).
4. Radinovic D., Curic M. Criteria for heat and cold wave duration. *Theor. Appl. Climatol.*, 2012, vol. 97, no. 3–4, pp. 505–510.
5. Revich B.A., Shaposhnikov D.A. Osobennosti vozdeystviya voln zhary i kholoda na smertnost' v gorodakh s rezko-kontinental'nym klimatom [Influence features of cold and heat waves to the population mortality – the city with sharply continental climate]. *Sibirskoe meditsinskoe obozrenie*, 2017, no. 2, pp. 84–90 (in Russian).
6. Revich B.A., Shaposhnikov D.A. Volny kholoda v yuzhnykh gorodakh evropeiskoi chasti Rossii i prezhdevremennaya smertnost' naseleniya [Cold waves in southern cities of European Russia and premature mortality]. *Problemy prognozirovaniya*, 2016, no. 2, pp. 125–131 (in Russian).
7. Revich B.A., Shaposhnikov D.A., Podol'naya M.A., Khar'kova T.L., Kvasha E.A. Volny zhary v yuzhnykh gorodakh evropeiskoi chasti Rossii kak faktor riska prezhdevremennoi smertnosti naseleniya [Heat waves in southern cities of European Russia as a risk factor for premature mortality]. *Problemy prognozirovaniya*, 2015, no. 2, pp. 56–67 (in Russian).
8. Shaposhnikov D., Revich B. Towards meta-analysis of impacts of heat and cold waves on mortality in Russian North. *Urban Climate*, 2016, vol. 15, pp. 16–24. DOI: 10.1016/j.uclim.2015.11.007
9. Anderson G.B., Bell M.L., Peng R.D. Methods to calculate the heat index as an exposure metric in environmental health research. *Environ. Health Perspect*, 2013, vol. 121, no. 10, pp. 1111–1119. DOI: 10.1289/ehp.1206273
10. Report on Wind Chill Temperature and Extreme Heat Indices: Evaluation and Improvement Projects. Washington, DC, Federal Coordinator for Meteorological Services and Supporting Research, Publ., 2003. Available at: <http://solberg.snr.missouri.edu/gcc/OFCMWindchillReport.pdf> (08.09.2017).
11. Ryabenko E.A. Prikladnoi statisticheskii analiz dannykh. Obobshcheniya lineinoi regressii. [Applied statistical data analysis. Linear regression generalizations]. 2016. Available at: http://www.machinelearning.ru/wiki/images/a/a6/Psad_otherreg.pdf (13.06.2017) (in Russian).
12. Richiardi L., Bellocco R., Zugna D. Mediation analysis in epidemiology: methods, interpretation and bias. *Int. J. Epidemiol.*, 2013, vol. 42, no. 5, pp. 1511–1519. DOI: 10.1093/ije/dyt127
13. Gasparrini A., Armstrong B. The impact of heat waves on mortality. *Epidemiology*, 2011, vol. 22, no. 1, pp. 68–73. DOI: 10.1097/EDE.0b013e3181fdcd99
14. Shaposhnikov D., Revich B., Bellander T. [et al.]. Mortality Related to Air Pollution with the Moscow Heat Wave and Wildfire of 2010. *Epidemiology*, 2014, vol. 25, no. 3, pp. 359–364. Available at: <https://www.ncbi.nlm.nih.gov/pubmed/24598414> (08.09.2017).
15. Gasparrini A. Distributed lag linear and non-linear models in R: the package dlnm. *Journal of Statistical Software*, 2011, vol. 43, no. 8, pp. 1–20. DOI: 10.18637/jss.v043.i08
16. Schwartz J. The distributed lag between air pollution and daily deaths. *Epidemiology*, 2000, vol. 11, pp. 320–326.
17. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2016. *R. Core. Team*. 2017. Available at: <http://www.R-project.org/> (18.12.2017).
18. Stafoggia M., Schwartz J., Forastiere F., Perucci C.A., and the SISTI Group. Does Temperature Modify the Association between Air Pollution and Mortality? A Multicity Case-Crossover Analysis in Italy. *Am. J. Epidemiol.*, 2008, vol. 167, no. 12, pp. 1476–1485. DOI: 10.1093/aje/kwn074
19. Pope C.A. 3-rd, Hill R.W., Villegas G.M. Particulate air pollution and daily mortality on Utah's Wasatch Front. *Environ Health Perspect*, 1999, vol. 107, no. 7, pp. 567–573.

Shaposhnikov D.A., Revich B.A. On some approaches to calculation of health risks caused by temperature waves. *Health Risk Analysis*, 2018, no. 1, pp. 22–31. DOI: 10.21668/health.risk/2018.1.03.eng

Получена: 07.03.2018

Принята: 15.03.2018

Опубликована: 30.03.2018