

УДК 57.042; 519.246.2  
DOI: 10.21668/health.risk/2021.2.02

Читать  
онлайн



Научная статья

## ГРУППОВЫЕ ПОКАЗАТЕЛИ РИСКА ЗДОРОВЬЮ В НЕОДНОРОДНОЙ КОГОРТЕ. КОСВЕННАЯ ОЦЕНКА ПО ДИНАМИКЕ СОБЫТИЙ

**В.Ф. Обеснюк**

Южно-Уральский институт биофизики Федерального медико-биологического агентства России, Россия, 456780, г. Озерск, Озерское шоссе, 19

*Объектом внимания в данном исследовании является описание методики оценки интенсивных и кумулятивных показателей специфического риска при наблюдении когорт, подвергавшихся сочетанному воздействию нескольких внешних или внутренних факторов. Показано, каким образом хорошо известные эвристико-дескриптивные показатели, принятые в эпидемиологии отдаленных последствий, могут быть использованы для анализа динамики счетных событий в когорте на вполне строгой статистико-вероятностной основе, опирающейся на байесовский подход в трактовке условных вероятностей их наступления. Несмотря на то что в работе не использовано ни одного нового или ранее неизвестного эпидемиологического понятия или показателя, статья не является обзором литературы. Сравнительно новой является сама предложенная методика, сочетающая приемы обработки традиционной эпидемиологической информации и корректный метрологический подход на основе описания процессов.*

*Основным результатом работы является подведение читателя к пониманию того, что все основные дескриптивные эпидемиологические показатели в рамках когортного описания оказываются количественно взаимосвязанными, если рассматривать их как условные групповые процессы. Это позволяет одновременно производить взаимосогласованную оценку как годовых показателей риска, так и кумулятивных показателей Каплана – Майера (Флеминга – Харрингтона) и Нельсона – Аалена, а также других условных показателей риска или их аналогов. Показано, что при выборе основной описательной характеристики кумулятивных показателей в качестве меры измеримого пролонгированного внешнего воздействия естественно применять понятие дозы данного фактора риска, которая носит суррогатный характер. Работоспособность метода оценки подтверждается примером. Указано на отличие предложенной методики от ее прототипа, который ранее приводил к существенному смещению оценок до ~100 % по ряду параметров даже в случае штатного режима работы. Применение требует создания специализированного, но доступного программного обеспечения для ЭВМ.*

**Ключевые слова:** риск, показатель, эпидемиология, фактор риска, конкуренция, косвенная оценка, смертность, процесс, когорта, страта, модель.

Будем опираться на устоявшееся мнение о том, что риск здоровью представляет собой объективную вероятность наступления того или иного нежелательного события [1–3] при некоторых сложившихся условиях/факторах, включая период наблюдения. В перечень изучаемых событий обычно включают смерть, болезнь или, реже, субклинические надежно диагностируемые необратимые изменения состояния здоровья. Соответственно, риск может быть количественно выражен некоторым числовым или функциональным показателем. В практике прогнозирования он может характеризовать даже еще не наступившее событие, однако достичь этапа управления рисками весьма сложно, если не обобщен опыт наблюдения подобных событий в схожих условиях, то есть не выполнена апостериорная оценка риска по ранее собранной

информации. В этом смысле задача рискометрии похожа на метрологическую процедуру определения некой неизвестной (но объективно существующей) величины, или на установление связи между показателем риска и условиями его потенциальной реализации. В этой логической схеме содержится кажущееся внутреннее противоречие, связанное с тем, что понятие вероятности предполагает действие категории «случайность», противоречащей детерминированной категории «связь». Да, случайность присутствует и оказывает влияние, но противоречия нет. Применение категории «риск» имеет смысл только в условиях возможной альтернативы развития событий. Однако факторы риска могут присутствовать здесь в виде той или иной детерминированной комбинации, порождая вполне реальную связь «факторы – риск», которая

© Обеснюк В.Ф., 2021

**Обеснюк Валерий Федорович** – кандидат физико-математических наук, доцент, старший научный сотрудник (e-mail: v-f-o@subi.su; тел.: 8 (35130) 7-52-36; ORCID: <https://orcid.org/0000-0002-2446-4390>).

и должна являться предметом изучения еще до наступления этапа управления рисками.

Отметим ряд сугубо математических свойств понятия «риск», определяющих сложность процедуры оценки. Связь или потенциальная связь с факторами риска указывает на то, что мы имеем дело с условной вероятностью. Дело может усугубляться еще и тем, что при анализе специфического риска здоровью в роли одного из факторов риска обычно выделяют время или возраст, из чего можно сделать вывод, что риск здоровью – не только показатель (число), это еще и динамический случайный процесс. Наконец, если ставится задача элиминации влияния на метрологическую процедуру случайных или неконтролируемых факторов, этот риск никак не может быть оценен индивидуально, но только для однородной группы индивидуумов в качестве их некоторой общебиологической характеристики.

Типичным и широко употребляемым показателем, используемым в дескриптивной эпидемиологии профессиональных заболеваний, клинической эпидемиологии, медико-экологических и демографических исследованиях является интенсивный показатель риска<sup>1</sup> общей или специфической смертности или заболеваемости, известный также под названиями «интенсивность риска», «сила смертности/заболеваемости» (force of mortality, hazard rate), «моментный показатель риска» [4, 5]. При описании отдаленных последствий с его помощью эту величину обычно относят к году возрастной или календарной шкалы как наиболее часто употребляемой единице времени. Использование в качестве меры ряда из показателей риска обусловлено довольно неплохой воспроизводимостью его динамики при описании отдаленных последствий широкого списка эффектов, к примеру неинфекционных заболеваний, для большого количества изолированных субкогорт или субпопуляций, находящихся в сходных социально-экологических условиях. Это позволяет рассматривать интенсивный показатель во всей его динамике практически в качестве видоспецифической характеристики. Данное обстоятельство является, например, одной из причин проведения регулярного регионального скрининга всех видов онкологической заболеваемости и смертности именно по упомянутому показателю [4]. Как правило, этот показатель и есть та самая величина, избыточные значения которой на уровне 0,001–1 % в год в качестве предельно допустимых рисков<sup>2</sup> принимаются контролирующими органами как уровни принятия решений. Интенсивный групповой показатель риска известен также под не совсем точным названием «индивидуальный риск», которое противоречит его групповой природе.

#### Пример 5-летнего когортного исследования

Параметр	Экспонировано	Не экспонировано
Умерло по причине	30	10
Не умерло по причине	70	90
Всего	100	100

Понятие о названном показателе и его непосредственной связи с величиной риска и другими объективными показателями нетрудно получить из опубликованного простого примера [5], воспроизведенного в таблице.

Пусть за некоторый сравнительно небольшой период времени  $T = 5$  лет происходило наблюдение за двумя практически одинаковыми по составу группами населения (стратами), отличающимися только тем, что одна из них подвергалась, а другая не подвергалась воздействию некоторого фактора риска, влияние которого мы пытаемся оценить. Если риск – это вероятность гибели по исследуемой причине, то в экспонированной группе его оценка составит очевидные 30 из 100 случаев, или  $R_e = 0,30$ ; аналогично в неэкспонированной  $R_n = 10/100 = 0,10$ . Тогда избыточный риск смерти по изучаемой причине составит  $0,3 - 0,1 = 0,2$ ; его естественно связать с действием экспозиции фактора риска. Относительный риск для действия фактора  $RR = R_e/R_n = 3,0$ .

Приведенные величины – это кумулятивные смертности за рассматриваемый пятилетний период. По ним нетрудно оценить и интенсивные показатели, исходя из соотношений  $M_e = N_e \cdot (1 - \exp(-h_e T))$  и  $M_n = N_n \cdot (1 - \exp(-h_n T))$ , а также из предположения о постоянстве интенсивных показателей в экспонированной и неэкспонированной группах. Здесь  $M_e, M_n$  – числа «случаев» в экспонированной и неэкспонированной стратах;  $N_e, N_n$  – исходная заселенность страт;  $h_e, h_n$  – hazards – годовые показатели риска. Наличие экспонент в указанных формулах отражает тот факт, что величины  $h_e, h_n$  являются скользящими, то есть связанными с условием дожития до текущего возраста в пределах интервала наблюдения, в то время как кумулятивные показатели  $R_e, R_n$  относятся ко всему периоду наблюдения в целом по сравнению с состоянием субкогорт в начале. Благодаря этому показатели  $h_e, h_n$  по своим математическим свойствам аналогичны непрерывной норме дисконтирования в экономической теории, порождая связь с экспоненциальными соотношениями. Расчет по ним дает  $h_e \approx 71$  % в год

<sup>1</sup> Эпидемиологический словарь / под ред. Д.М. Ласта для Международной эпидемиологической ассоциации. – 4-е изд. – 316 с.

<sup>2</sup> Р 2.1.10.1920-04. Руководство по оценке риска для здоровья населения при воздействии химических веществ, загрязняющих окружающую среду. – М.: Федеральный центр госсанэпиднадзора Минздрава России, 2004. – 143 с.

и  $h_e \approx 21\%$  в год соответственно. Отношение показателей риска (hazard ratio)  $HR = h_e/h_n = 3,38 \neq RR = 3,0$  под действием фактора.

Если просто не обращать внимания на двусмысленность толкования относительного риска и возвратится к названию статьи, уместно задать вопрос: в чем же проблема оценки показателей в неоднородной когорте? Ведь, судя по данным таблицы, все предельно просто. На самом же деле здесь целый ворох методологических проблем:

1) показатель  $h = h(t)$  на самом деле не число, а функция возраста или времени, то есть процесс, в то время как в таблице оценки приведены к обычным скалярным (числовым) показателям;

2) в таблице указан только один фактор, оказывающий влияние на риск, в то время как в ситуации реальной выборки или когорты почти всегда приходится иметь дело с многофакторным исследованием. Здесь нужна особая технология статистического оценивания, подразумевающая стратификацию неоднородной когорты более чем на две страты с учетом всех интересующих комбинаций факторов риска;

3) как можно видеть, ни интенсивность  $h$ , ни ее кумулятивный аналог  $h \cdot T$  не являются непосредственно наблюдаемыми величинами. Живучести противоположных (устаревших) воззрений способствует известное приближенное свойство величины  $h$ , позволяющее вычислять ее как «отношение числа специфических случаев к числу человеко-лет наблюдения под риском»<sup>1</sup>. Ошибочно считать это приближенное свойство точным определением. В действительности первичной эмпирически наблюдаемой информацией является динамика отсчетов в таблице как в кумулятивной, так и в индивидуальной форме. Это обстоятельство порождает задачу *косвенной оценки* процесса  $h(t)$  или его кумулятивного аналога по наблюдениям процесса накопления отсчетов в каждой однородной страте когорты;

4) отсчеты в однородной страте, являющейся частью случайной выборки, тоже случайны. В оценке же нуждается показатель некой однородной и почти генеральной совокупности. По этой причине показателям может быть дана оценка с некоторой относительной неопределенностью, которая будет тем больше, чем меньше число случаев в изучаемой когорте/страте. Например, интервальные оценки кумулятивных рисков для рассмотренных экспонированной и неэкспонированной страт (на уровне 95%-ной доверительной вероятности) составляют  $R_e = 0,219 \dots 0,396$  и  $R_n = 0,056 \dots 0,175$ , причем видно, что для неэкспонированной группы ширина интервала неопределенности превышает центральную оценку. Это обстоятельство практически запрещает нам работать с малочисленными стратами, число «случаев» в которых менее 4, так как расширенная относительная неопределенность риска заведомо

будет превышать 100%. Менее чем по четырем точкам вообще невозможно надежно идентифицировать процесс, несмотря на то, что статистически значимые различия страт могут наблюдаться и при меньшем числе случаев [6]. Таким образом, требуется создание такого алгоритма оценивания, который сохранял бы преимущества стратификации пространства факторов вместе с возможностью идентификации *оптимальной модельной зависимости* показателей риска, учитывающей взаимосвязь со всеми факторами риска для всей совокупности страт одновременно.

Таким образом, актуальной и естественной является задача построения алгоритма оценки интенсивных и кумулятивных показателей специфического риска в неоднородной когорте на основе данных регистра о ее длительном наблюдении.

#### Описание метода оценки и его прототипа.

Отметим, что интерес к проблеме неоднородности реальных наблюдений существует давно как в случае когортных выборок, так и в случае популяционных исследований медико-демографических проблем [7]. Установлено множество различных причин неоднородности, наблюдаемых при жизни представителей выборки или устанавливаемых посмертно, включая ненаблюдаемые скрытые факторы [8].

Не претендуя на всеобщий охват, в дальнейшем сосредоточим свое внимание на изучении влияния только заранее известных факторов риска, полагая латентные переменные состояния отсутствующими. Здесь остается полагаться только на интуицию врача-исследователя при ведении регистра и первичном сборе эпидемиологической информации. Это позволяет еще до начала математического анализа данных выполнить группирование индивидуумов по стратам с возможностью постоянной привязки к своей заранее определенной страте в течение всего периода наблюдения. ИмPLICITно предполагается, что все индивидуумы в ней имеют одинаковые шансы заболеть или умереть от любой исследуемой причины, закодированной в сборниках МКБ-9 или МКБ-10. Особенно просто это сделать в отношении факторов риска, которые можно описать в виде бинарных признаков, например пола или принадлежности к группе курящих или некурящих. Бинарным признаком может быть даже interfering заболевание в анамнезе периода наблюдения. Такие факторы можно считать почти неизменными на протяжении довольно длительного времени. Некоторые количественные факторы воздействия на здоровье также могут оказаться пригодными для выбранной схемы стратификации, если характер интенсивности воздействия одинаков для всех членов когорты. Это очевидно для случаев острого однократного воздействия, либо хронического равномерного воздействия с постоянной индивидуальной интенсивностью. Тогда возможно введение такого (суррогатного) кумулятивного фактора, как доза,

накопленная на конец периода наблюдения. В этом случае представляется вполне естественным анализ условной зависимости в паре «кумулятивный риск – кумулятивная доза».

Описанный подход в эпидемиологии имеет свой функциональный прототип – модуль AMFIT пакета программ Epicure<sup>3</sup>. Этот программный продукт успешно зарекомендовал себя в эпидемиологических исследованиях во всем мире [9, 10]. В частности, в области радиационно-эпидемиологических исследований он провозглашен стандартом программного обеспечения: «... Epicure is the de-facto standard for modeling radiation health effects ...» [11]. На основании результатов, полученных с помощью AMFIT (алгоритм пуассоновской регрессии), базируются практически все национальные нормы радиационной безопасности в странах, имеющих радиационно опасные производства, включая Россию<sup>4</sup>. При этом считается установленным, что эквивалентная доза радиационного воздействия – общепризнанный фактор риска отдаленных радиационно-онкологических последствий.

Подход, предлагаемый в данной работе, несколько отличается от AMFIT, несмотря на то что имеет общую цель. Таких отличий три:

а) если придерживаться строго вероятностного метода, то описание наблюдений и построение функционала их совокупной оценки нельзя обеспечить применением пуассоновского распределения для отсчетов «случаев» в стратах, поскольку пуассоновская статистика пригодна только для описания редких событий в неограниченной выборке. Реальные когорты рано или поздно перестают быть таковыми по мере вымирания даже приближенно. Необходимо более корректное описание отклонений от основного тренда (процесса) на базе биномиальной статистики, пригодной для конечных выборок и не связанной с условием редкости событий;

б) во главу угла в алгоритме AMFIT положены оценки интенсивности специфических событий в ущерб оценке связанных с ними кумулятивных показателей. Между тем интенсивность не является непосредственно наблюдаемой дескриптивной характеристикой, однако вместо нее успешно может быть применен кумулятивный показатель Каплана – Майера, напрямую связанный с индивидуальными отсчетами «случаев» и случайными отклонениями биномиальной природы. Применение такого показателя позволило бы повысить устойчивость оценок одновременно с сохранением возможности вычисления интенсивности событий;

в) в основе алгоритма AMFIT лежит метод максимального правдоподобия в формулировке Рональда Фишера, который, строго говоря, не является вероятностным, а имеет сугубо эвристическую ос-

нову подобно своему прототипу – методу максимального правдоподобия Карла Гаусса [12].

Хотя любой алгоритм статистической обработки дает смещенные оценки, есть основания надеяться на заметное улучшение измерительной точности за счет исправления недостатков уже неплохо зарекомендовавшего себя алгоритма AMFIT, если использовать последний в качестве прототипа.

**Взаимосвязь интенсивных и кумулятивных условных статистических показателей в эпидемиологии отдаленных последствий.** Стремление к получению объективных оценок эпидемиологических показателей требует применения несколько более углубленного математического аппарата, чем тот, который был продемонстрирован при обработке данных, приведенных в таблице. Эта математика, однако, никого не должна вводить в заблуждение относительно намерений перейти в область аналитической эпидемиологии. Демонстрируемая методика по-прежнему соответствует обычной дескриптивной статистике, не связанной с природой причинно-следственных связей при описании состояния здоровья, если речь идет о событиях в условно однородной страте в составе неоднородной когорты. Дескриптивный подход подразумевает очевидный формализм: в тех случаях, когда потоки событий со специфической причиной смерти или заболевания в однородной группе можно описать простой марковской схемой с тремя состояниями и наличием двух конкурирующих причин выбывания из-под наблюдения, скорость изменения заселенности компартмента, отвечающего за исходное состояние, пропорциональна численности лиц, его заселяющих (рис. 1, формула (1)). Независимо от того, насколько близка такая модель к действительности, коэффициент пропорциональности между скоростью выбывания и численностью лиц в основном состоянии всегда может быть вычислен и представлен в виде суммы двух интенсивностей событий – изучаемой и конкурирующей.

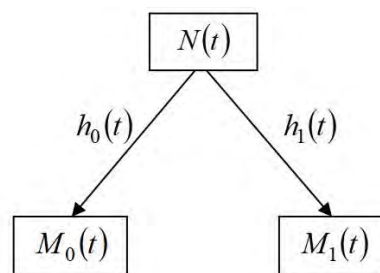


Рис. 1. Марковская структурная схема выбытия из-под наблюдения для двух конкурирующих причин смерти. Указаны компартмент исходного состояния и компартменты для двух разновидностей регистрируемых причин смерти

<sup>3</sup> EPICURE User's Guide / D. Preston, J. Lubin, D. Pierce, M. McConney. – Seattle: Hirosoft, 1988–1993. – 334 p.

<sup>4</sup> СанПиН 2.6.1.2523-09. Нормы радиационной безопасности (НРБ-99/2009). – М.: Федеральный центр гигиены и эпидемиологии Роспотребнадзора, 2009. – 100 с.

Схема рис. 1 может быть приближенно описана системой обыкновенных дифференциальных уравнений

$$\begin{aligned} \frac{dN}{dt} &\approx -(h_0(t) + h_1(t)) \cdot N, \\ \frac{dM_1}{dt} &\approx h_1(t) \cdot N, \end{aligned} \quad (1)$$

где  $t$  – время (возраст);  $h_0(t), h_1(t)$  – интенсивности событий конкурирующей и изучаемой причин смерти;  $N(t)$  – заселенность исходного состояния;  $M_1(t)$  – кумулятивное (накопленное) число случаев смерти по изучаемой причине. Приближенность системы уравнений (1) связана с тем, что не существует производных от дискретнозначных функций, подверженных к тому же случайным флуктуациям. Однако это не мешает перейти к ожидаемым условным долям, исчисляемым от исходной численности однородной подгруппы в некоторой точке начала наблюдения  $t_0$ , а значит, к условным вероятностям или показателям распространенности в страте. Тогда соотношения станут непрерывными и точными (2):

$$\begin{aligned} \frac{dS}{dt} &= -(h_0(t) + h_1(t)) \cdot S, \\ \frac{dR}{dt} &= h_1(t) \cdot S, \end{aligned} \quad (2)$$

если условно принять в точке начала наблюдения  $S(t_0) = 1$ . В силу линейности, система (2) имеет простое аналитическое решение:

$$S(t) = S(t|t_0) = S(t_0) \cdot P_0(t|t_0) \cdot P_1(t|t_0), \quad (3)$$

$$\Delta R(t|t_0) = \int_{t_0}^t h_1(\tau) \cdot S(\tau) d\tau, \quad (4)$$

где  $P_0(t|t_0) = \exp(-H_0(t|t_0))$ ;  $P_1(t|t_0) = \exp(-H_1(t|t_0))$ ;  $H(t|t_0) = \int_{t_0}^t h(\tau) d\tau$ ;  $\Delta R$  – прирост риска смерти от изучаемой причины за период  $[t_0, t]$ . По смыслу решений (3), (4) и по опыту широкого круга эпидемиологов [13–23] в этих выражениях помимо интенсивности изучаемых специфических событий  $h_1(t)$  можно узнать точную функцию дожития  $S(t|t_0)$ , добавочный условный пожизненный риск смерти от изучаемой причины  $\Delta R(t|t_0)$ , кумулятивную вероятность не умереть от изучаемой причины при условии дожития до возраста  $t_0$  и условного отсутствия конкурирующих причин смерти –  $P_1(t|t_0)$ , а кроме того, аналог показателя Нельсона – Аалена  $H_1(t|t_0)$ , известного также как кумулятив-

ная интенсивность специфической смертности при условии отсутствия иных причин смерти (cumulative hazard) [19, 20]. Может показаться, что связанные друг с другом показатели  $P_1(t|t_0)$  и  $H_1(t|t_0)$  не наблюдаемы. Однако это не так. В эпидемиологических приложениях теории мартингалов [19] установлено, что показателю  $P_1(t|t_0)$  соответствует условная, но вполне измеримая функция дожития Каплана – Майера [22], а величине  $H_1(t|t_0)$  – измеримый и уже упоминавшийся мартингальный показатель Нельсона – Аалена. Напомним, что величина  $H_1(t|t_0)$ , по сути являющаяся площадью под кривой процесса  $h_1(t)$ , успешно и эффективно более двух десятилетий контролируется в рамках онкологического мониторинга [4] в Российской Федерации. Этот показатель удобен не только для измерений в когорте, но и в популяции.

Таким образом, в рамках простой динамической схемы, приведенной на рис. 1, оказываются связанными шесть дескриптивных эпидемиологических показателей – четыре кумулятивных и два интенсивных. Важно отметить, что величина показателей, являющихся относительными, привязана к разной базе. Например, условный пожизненный риск исчисляется по отношению к точке начала наблюдения, а интенсивные показатели и их кумулятивные характеристики  $P_1(t|t_0)$  и  $H_1(t|t_0)$  являются скользящими, т.е. исчисляемыми относительно достигнутой доли выживших, поскольку связаны с условием дожития до момента наблюдения. Кроме того, истинная функция дожития оказывается сильнее связанной с конкурирующими причинами смерти и неконтролируемой историей страты до начала наблюдения, по сравнению с показателями, отвечающими за изучаемую причину смерти. По этой причине показатели  $H_1(t|t_0)$  и  $P_1(t|t_0)$  представляют самостоятельный интерес. В силу связи  $P_1(t|t_0) = \exp(-H_1(t|t_0))$  показатель  $P_1(t|t_0)$  можно трактовать как некий условный аналог «функции дожития», своеобразие которого состоит в том, что его предельное значение при  $t \rightarrow \infty$  может не достигать нуля, в отличие от истинной функции дожития. Такое поведение возможно, если изучаемая причина смертности не является ведущей, в отличие от совокупности конкурирующих причин смерти. Оно будет наблюдаться также и в тех случаях, когда фатальный потенциал изучаемой причины смерти конечен в силу конечности доли лиц, потенциально подверженных развитию изучаемого заболевания. Эти свойства позволяют интерпретировать показатели  $1 - \exp(-H_1(t|t_0))$  и  $H_1(t|t_0)$  подобно величине специфического риска и численно почти отождествлять их, если только предельное значение показателя Нельсона – Аалена не превышает примерно

$\sim 0,1$ . Этому условию практически всегда удовлетворяет популяционная кумулятивная смертность, связанная с конкретной локализацией онкологического заболевания [4]. Однако, например, смертность от всех болезней системы кровообращения в популяции обычно перекрывает это ограничение, и тогда  $H_1(t|t_0)$  не является оценкой риска. Сравнительно больших значений  $H_1(t|t_0)$  может достигать также среди контингента специализированных клиник или отделений в силу специфики отбора их пациентов. Например, в работе [24] приводятся оценки группового риска смерти от рака предстательной железы до  $1 - P_1 \approx 24\%$ , что соответствует предельному значению  $H_1 \approx 0,27$  за период более 3000 суток. Для сравнения, кумулятивный риск развития рака предстательной железы в популяции мужчин РФ за 75 лет жизни не превышает  $5,7\%$  [4]. За показателем риска  $1 - \exp(-H_1(t|t_0))$  в радиационной эпидемиологии закрепились аббревиатура RADS [15, 16].

**Принципы конструирования и параметризации модели. Байесовское интервальное статистическое оценивание.** Как было показано, оценка кумулятивных показателей для отдельных страт не представляет больших технических трудностей, однако любая подробная стратификация наблюдений в пространстве факторов приводит к уменьшению числа «случаев» в каждой страте и, соответственно, к росту неопределенности оценки рисков. Поднять точность оценок и одновременно сохранить подробность описания можно только путем применения единой аппроксимирующей математической модели для всех страт одновременно. Тем самым в расчетах будут участвовать все наблюдаемые «случаи», что при соответствующей оптимизации позволит добиться большей устойчивости оценок риска. В этом состоит роль единой модели для всех страт. Она должна быть динамической, то есть пригодной для всех наблюдений, разнесенных по времени (возрасту). Строго говоря, модель должна соответствовать природе связи факторов с показателями риска, однако обычно это и является предметом исследования, то есть до завершения анализа такой модели обычно нет. В этом случае можно опереться на ожидаемое подобие динамики реализации риска во времени для различных страт  $h = h(t|z, \beta, Data)$ , взяв за основу уже изученное поведение показателей в некоторой группе сравнения. Здесь  $z$  – вектор факторов риска;  $\beta$  – соответствующий вектор подгоночных параметров модели. При изучении онкологических эффектов разумно использовать популяционные показатели [4] и на основе метода возмущений ввести в описание связь с факторами риска и соответствующую параметризацию. Например, можно использовать тот

факт, что большинство интенсивных показателей риска смерти от аналогичных онкозаболеваний в возрастной динамике имеет вид унимодальных функций, характеризующихся приблизительно степенным ростом в диапазоне возрастов до 60–65 лет и резким спадом в возрасте старше 75 лет.

Если непрерывная модель  $h(t|z, \beta, Data)$  в некотором смысле адекватна изучаемому множеству дискретных эмпирических отсчетов  $Data$ , это будет означать попытку двойственного описания одних и тех же событий либо через отсчеты непосредственно, либо в пространстве параметров  $\beta$ . Естественному рассеянию в пространстве наблюдений будет соответствовать некоторое непрерывное условное распределение в пространстве параметров. Оно будет прямо представлять интерес для интервального оценивания множества подходящих параметрических гипотез, если процедуре оценки придать вероятностную форму.

Подходящим инструментом для этого является теорема Байеса, позволяющая связать два упомянутых типа условных распределений:

$$\psi(\beta | Data) = \frac{L(Data | \beta) \cdot prior(\beta)}{\int L(Data | \beta) \cdot prior(\beta) d\beta} \quad (5)$$

$$\text{или } \psi(\beta | Data) \propto L(Data | \beta) \cdot prior(\beta). \quad (6)$$

Несмотря на то что байесовский подход считается прямым статистическим конкурентом известному методу максимального правдоподобия, оба подхода органически связаны друг с другом. Здесь  $L(Data | \beta)$  – плотность рассеяния наблюдений для фиксированной параметрической модели;  $\psi(\beta | Data)$  – плотность распределения параметров модели для собранных наблюдений;  $prior(\beta)$  – априорное распределение параметров предположительно адекватной модели hazard. По смыслу условных распределений  $\psi$  – правдоподобие Байеса;  $L$  – правдоподобие Фишера, а ожидаемая область наиболее вероятных параметров лежит вблизи точки максимального правдоподобия функции  $L$ , по крайней мере в тех исследованиях, результат которых до анализа опытных данных неизвестен. В данном случае соотношения (5), (6) вполне строгие, если бы не неизвестность априорного распределения  $prior(\beta)$ , благодаря которому справедливо мнение<sup>5</sup> о том, что понятие параметрически зависимого правдоподобия не тождественно понятию условной плотности вероятности [25]. Однако в защиту байесовского подхода отметим, что всякое новое и в особенности однократное исследование

<sup>5</sup> Справочник по прикладной статистике / ред. Э. Ллойд, У. Ледерман. – М.: Финансы и статистика, 1989. – 510 с.

характеризуется почти полным отсутствием до-опытного знания, благодаря которому функция  $prior(\beta)$  заведомо характеризуется существенно большей шириной по сравнению с  $L(Data | \beta)$  как функцией параметров  $\beta$  в некоторой значимой области. Поэтому не будет большой ошибкой принять предположение о каком-либо неинформативном априорном распределении или даже о предположении  $prior(\beta) \propto 1$  в значимой области. Тогда формально  $\psi(\beta | Data) \approx L(Data | \beta)$ , чем фактически и пользовался Рональд Фишер, что не мешало ему в публикациях одновременно полностью отвергать байесовский подход [25]. Близость понятий байесовского и фишеровского правдоподобий оправдывает рассмотрение постоянных параметров функции правдоподобия  $L(Data | \beta)$  в качестве подгоночных переменных модели. То, что считается постоянным вектором в фриквентистской/частотной концепции Фишера и Пирсона, при более строгом рассмотрении является непрерывной случайной переменной по Лапласу / Байесу.

Отметим главное: правдоподобие для совокупности страт в силу их независимости просто равно произведению правдоподобий для каждой однородной страты. Поэтому построим правдоподобие для отдельной однородной страты. Для этого в ее пределах введем свою сетку моментов времени  $t_i$ , каждый узел которой привязан к одному из специфических событий на числовой оси возраста. Точка  $t_0$  соответствует началу наблюдения. Очень редко в одном узле может происходить до 2–3 таких событий одновременно, поэтому каждый полуоткрытый интервал  $(t_{i-1}, t_i]$  между двумя соседними узлами будет связан со своим количеством накопленных специфических случаев  $m_i$ . Обычно  $m_i = 1$ . Общее количество накопленных случаев к рассматриваемому моменту времени  $t_i$  составит  $M_i = \sum_{j=1}^i m_j$ .

Совместное правдоподобие наблюдения всей последовательности специфических событий в  $j$ -й страте с вектором факторов  $z_j$ , как цепи последовательных переходов, составит

$$L_j = L(M_{i_{\max}}, M_{i_{\max}-1}, \dots, M_1 | z_j, \beta) = 1 \cdot \prod_{i=1}^{i_{\max}} p(M_i | M_{i-1}), \quad (7)$$

где

$$p(M_i | M_{i-1}) = \frac{N_{i-1}!}{m_i! (N_{i-1} - m_i)!} (\pi_i)^{N_{i-1} - m_i} (1 - \pi_i)^{m_i}. \quad (8)$$

Здесь также введены обозначения  $\pi_i = \exp(-H_i)$  и  $H_i = H(t_i | t_{i-1}, z, \beta)$ , где приращения кумулятив-

ной интенсивности риска являются интегралами модельной функции интенсивности

$$H_i = H_i(t_i | t_{i-1}, z, \beta) = \int_{t_{i-1}}^{t_i} h(\tau | z, \beta) d\tau. \quad (9)$$

По своей конструкции правдоподобие (7), (8) имеет дифференциальную форму, как и традиционное правдоподобие Фишера для независимых событий, однако это не совсем так. Анализируя частные правдоподобия  $L_i \sim \pi_i^{N_{i-1} - m_i} (1 - \pi_i)^{m_i}$ , нетрудно заметить, что они достигают максимального значения при  $\pi_i^{opt} = (N_{i-1} - m_i) / N_{i-1}$ , то есть локально удовлетворяют процедуре Каплана – Майера [22] на каждом  $i$ -м шаге по времени для однородной страты. Таким образом, максимизация функционала (7), (8) потенциально может привести к интерполяции кумулятивного показателя, если рассматривать оценки  $\hat{\pi}_i = \pi_i^{opt}$  в качестве параметров интерполирующей модели. Разумеется, это же свойство приближенно сохранится и при меньшем числе параметров в составе рассматриваемого вектора  $\beta$ , но с добавлением фильтрующего свойства правдоподобия как функционала оценки. Таким образом, сконструированное правдоподобие способно одновременно обеспечить и дифференциальную, и кумулятивную аппроксимацию (регрессию) одновременно. Оценка интенсивных показателей аналогична разностному численному дифференцированию меняющейся зашумленной функции. Дифференцирование эмпирических данных – плохо обусловленная (некорректная) численная операция. Напротив, производные от сглаженной кумулятивной функции будут более устойчивыми.

Технически удобнее вместо резко изменяющейся функции правдоподобия  $L(Data | \beta)$  или плотности  $\psi(\beta | Data)$  оперировать их удвоенным натуральным логарифмом, смещенным относительно точки предельного насыщения (интерполяции). Тогда для одной страты вместо поисков вблизи максимума выражения (7) следует анализировать функцию

$$\Omega(\beta | Data) = 2 \cdot \sum_i \left[ (N_{i-1} - m_i) \ln \left( \frac{N_{i-1} - m_i}{N_{i-1} \pi_i(\beta)} \right) + m_i \ln \left( \frac{m_i}{N_{i-1} (1 - \pi_i(\beta))} \right) \right] \quad (10)$$

вблизи ее минимума. Вклады (10) следует просуммировать для всей совокупности непустых страт. Окончательно в этом случае можно говорить о достигнутой девиации (функционале оценки)

$$\Omega_{\Sigma}(\beta) = \sum_j \Omega_j(\beta | Data); \quad \Omega_j(\beta | Data) \geq 0. \quad (11)$$

Согласно известным представлениям [26, 27], свойственным фишеровскому подходу, по величине отклонения  $\Omega_{\Sigma}(\beta)$  от нуля можно судить о качестве и статистической значимости выполненной аппроксимации, а по разности достигнутых оптимальных значений – следует производить селекцию моделей. Если параметрическая зависимость девиации (11) близка к квадратичной по малым отклонениям от центра (т.е.  $\psi(\beta | Data)$  – почти нормальное многомерное распределение), случайное рассеяние  $\Omega_{\Sigma}(\beta)$  вблизи минимума близко к распределению «хи-квадрат» с числом степеней свободы, равным разности между количеством группированных слагаемых в составе (11) и размерностью вектора  $\beta$ .

С учетом того, что на практике параметрическая зависимость правдоподобий  $L(Data | \beta)$  и  $\psi(\beta | Data)$  может оказаться весьма далекой от многомерного нормального распределения, более продуктивным будет завершение алгоритма оценки в логике последовательного продолжения байесовского подхода. Это означает переход к интервальному оцениванию, основанному на многомерных совместных распределениях (5), (6). Однако вместе с достаточной строгостью байесовский подход отличается высокой трудоемкостью и склонностью к накоплению вычислительных погрешностей, связанных с прямым подсчетом многомерных интегралов с участием плотности  $\psi(\beta | Data)$  в пространстве параметров  $\beta$ . По этой причине более реалистичным на практике является метод статистических испытаний (метод Монте-Карло), позволяющий по большому множеству точечных псевдонаблюдений (~1 млн или более), подчиняющихся распределению  $\psi(\beta | Data)$ , производить осреднение интересующих исследователя функций или параметров вместе с оценкой их маргинальных статистических характеристик. Алгоритм тоже трудоемкий, но, несмотря на это, – наиболее реалистичский. Теоретическая основа для одного из способов его практической реализации – алгоритма Гиббса – хорошо известна [28]. При этом оценки центра распределения по методу максимального правдоподобия в случае уни-modalной плотности  $\psi(\beta | Data)$  могут служить хорошим начальным приближением для построения стохастической последовательности псевдонаблюдений.

**Результаты оценки показателей риска в неоднородной когорте с заранее известными свойствами.** Рассмотрим искусственно созданный эпидемиологический регистр, описывающий некую «эталонную» когорту, события в которой не имеют случайной составляющей и точно подчиняются заранее известной модели. Выбор имитационных данных вместо реальных в данном случае является

предпочтительным, поскольку не существует ни одного реального регистра, результат исследования рисков для которого заведомо совпадал бы с точными и заранее известными показателями. Поскольку в искусственной когорте будет происходить только детерминированная имитация стохастического поведения ее участников, следует ожидать возможности численного достижения нулевой или почти нулевой девиации (11). Иными словами, предстоит практически ответить на вопрос, обладает ли изучаемый алгоритм оценки асимптотической сходимостью. Этот вопрос можно адресовать не только к рассматриваемому алгоритму, но и к его прототипу – алгоритму AMFIT пакета программ Erisure, несмотря на то что ранее эта проверка не проводилась никогда.

В качестве примера рассмотрим радиационно-эпидемиологическое исследование. В основу положим следующие характеристики. Создадим имитационную выборку из четко ограниченных однородных страт, различающихся по дозе гамма-облучения (от нуля до 2 Зв), полу и достигнутому возрасту. Будем считать, что кумулятивная доза радиации – фактор риска, причем вся она была получена в режиме острого однократного равномерного облучения легких в возрасте 19 лет. Основными причинами изменения показателей онкологической смертности будем считать: а) условно линейный рост кумулятивного показателя интенсивности риска смерти от рака легких с ростом дозы, известной и ограниченной по величине; б) некоторое сокращение продолжительности жизни для всех облученных членов когорты; в) пол в качестве одного из факторов неоднородности, вызывающего как несоответствие фоновых показателей риска у мужчин и женщин, так и различия в радиочувствительности. Такие причинно-следственные связи хорошо известны по серии исследований реальной когорты лиц, подвергшихся атомной бомбардировке в Хиросиме и Нагасаки [29, 30]. Фрагмент диаграммы (мужчины) с грубыми оценками годовых показателей риска по типовой схеме трехфакторной «эталонной когорты» (пол, доза, возраст), представлен на рис. 2. Всего в базе данных имелось 25 тысяч индивидуальных записей (15 тысяч мужчин; 10 тысяч женщин). Общее количество случаев смерти от специфического рака при условии полного вымирания когорты – 1118 (995 случаев среди мужчин; 123 случая среди женщин). Общая продолжительность наблюдения когорты 1 031 414 человеко-лет. При условии группирования по 14 возрастным интервалам, двум половым признакам и пяти уровням доз можно получить  $14 \cdot 2 \cdot 5 = 140$  однородных страт. Оказалось, что из них только 91 страта содержала ненулевое количество случаев специфического рака. Непустые страты соответствовали 754 106 человеко-годам наблюдения.

Представляется очевидным тот факт, что при отсутствии случайных колебаний выборочных пока-



зателей в «эталонной когорте» дозовый и возрастной тренды на рис. 2 видны невооруженным глазом. Одновременно количественные закономерности для фоновых показателей риска полностью соответствуют популяционным [4]. Исследователь риска обязан «видеть» всю заданную детализацию. Что же будут давать предлагаемый алгоритм и его ближайший прототип?

Оказывается, предлагаемый метод оценки процессов изменения кумулятивных и интенсивных показателей неплохо справляется с поставленной задачей, что можно видеть на рис. 3 и 4 как для субкогорты женщин, так и для субкогорты мужчин в рамках единой параметрической модели ( $\beta \in R^8$ ).

Вычисленная минимальная девиация  $\Omega_{\min}$  для найденного экстремального решения составила в данном случае всего 0,40 единицы при существенном увеличении составляющих «наблюдений» по Каплану – Майеру и росте числа степеней свободы по сравнению с ситуацией традиционной предварительной группировки с применением 5-летних возрастных интервалов (см. рис. 2). В расчете на один случай из 1118 это приблизительно соответствует визуально наблюдаемому среднеквадратическому относительному отклонению оцененных показателей hazard около  $\sqrt{0,40/1118} = 0,019 = 1,9\%$ . Остаточное отклонение  $\Omega_{\min}$  не достигло точного нуля, что указывает на неустранимую роль дискретной природы записей в базе данных выборки в сравнении с непрерывной природой показателей в истинной генеральной совокупности. Однако вряд ли девиацию можно снизить еще больше. Оценка коэффициента дозового тренда избыточного кумулятивного пожизненного показателя интенсивности накопления риска и его неопределенности по информационной матрице Фишера составила  $0,49 \text{ Гр}^{-1}$  (95% ДИ : 0,24 ... 0,99) для мужчин и  $0,68 \text{ Гр}^{-1}$  (95% ДИ : 0,33 ... 1,35) для женщин. Следует обратить внимание на существенную неопределенность, наблюдающуюся, несмотря на чрезвычайно малую величину достигнутой девиации. Это обстоятельство связано с преимущественной зависимостью размеров области неопределенности от кривизны функции (11) вблизи экстремума, а не от достигнутого ею минимального значения, поскольку оценка доверительных интервалов выполняется так, будто бы отсчеты в «эталонной когорте» все же подвержены случайным флуктуациям в той же мере, как если бы они были реальными. Этим биномиальная регрессия отличается от регрессии по методу наименьших квадратов.

В отличие от полученного результата, метод AMFIT в рамках своей типовой группы моделей [9, 10, 31–33] продемонстрировал значительное систематическое смещение годовых показателей риска в области старших возрастов (рис. 5; субкогорта

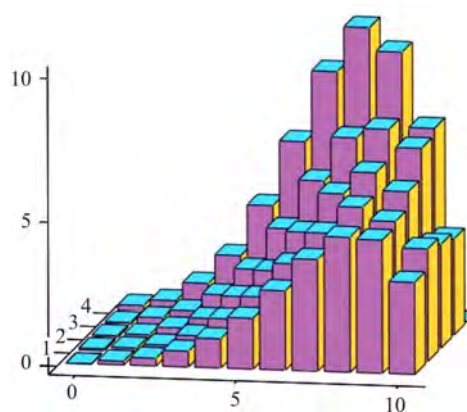


Рис. 2. Трехмерная диаграмма «возраст – доза – показатель» для мужской части «эталонной» когорты. По горизонтали отложены номера возрастных страт (14 шт.) и дозовых страт (5 шт.). По вертикали – грубые оценки показателя специфической смертности в пределах страты (% в год)

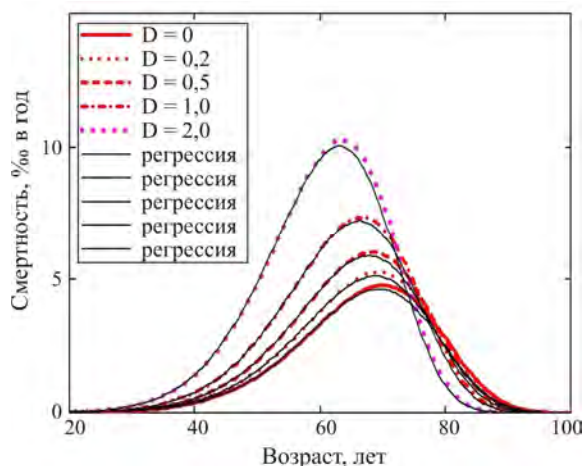


Рис. 3. Заданные и оцененные по алгоритму биномиальной регрессии динамические зависимости показателя специфической смертности среди мужчин «эталонной когорты»

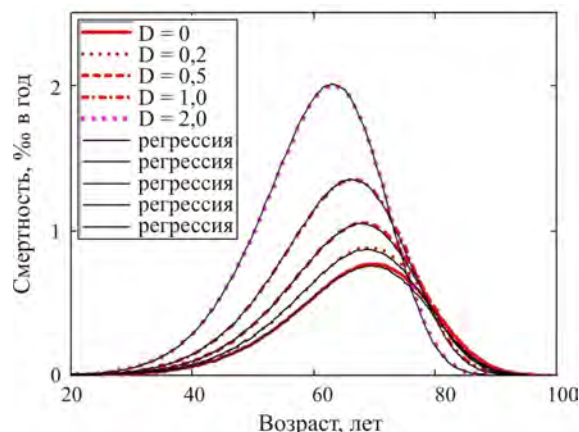


Рис. 4. Заданные и оцененные по алгоритму биномиальной регрессии динамические зависимости показателя специфической смертности среди женщин «эталонной когорты»

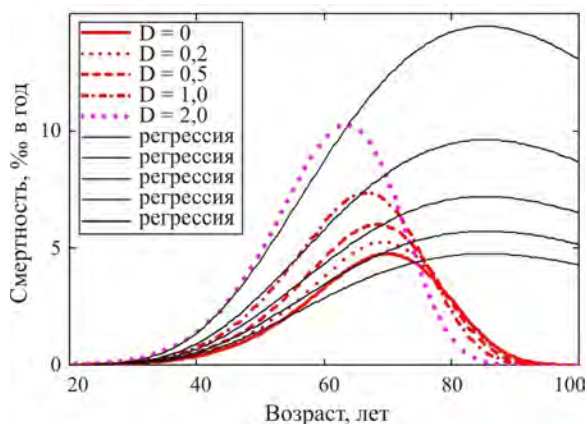


Рис. 5. Результат AMFIT-оценки дозовых и возрастных трендов показателя специфического риска в сравнении с истинным его поведением для мужской субкогорты. На рисунке видно, что типовая модель [9, 10, 31–33] поведения фонового и исследуемого рисков является источником систематического смещения всех оценок

мужчин). Неплохо аппроксимируются только восходящие участки кривых. Также модели алгоритма оказались неспособными «разглядеть» в когорте возможности взаимопересечения группы кривых hazard, связанной с некоторым сокращением продолжительности жизни облученных.

Наблюдаемая минимальная девиация (формула (11)) для найденного экстремального решения составила 24,3 единицы при числе степеней свободы 85 (число страт за вычетом числа параметров). При таком числе степеней свободы 90%-ная область ожидаемого случайного попадания минимальной девиации, казалось бы [26], распределенной по «закону хи-квадрат», должна составлять 64,7...107,5. Таким образом, наблюдаемое значение 24,3 статистически значимо меньше типичного случайного. Наблюдаемый оверфиттинг практически однозначно указывает на искусственность «эталонной когорты». Однако вызывает удивление заметное отклонение  $\Omega_{\min}$  от нуля, очевидно связанное с абсурдным систематическим смещением  $h(t|D, sex, \beta)$  в области возрастов более 75 лет и даже более 100 лет для всех дозовых воздействий. Фактически достигнутый уровень значимости  $p \sim 10^{-11}$  формально указывает на крайне малую вероятность уклонения модели от данных, что, однако, соответствует реальной картине только в области быстрого роста интенсивности риска (см. рис. 5), но не в широком диапазоне возрастов. Поскольку смещению подверглись все кривые показателей годового риска, то смещенными оказались и дозовые тренды. Например, годовой показатель  $h(t|D, sex, \beta)$  в возрасте 60 лет для мужчин смещался с темпом изменения дозы  $\approx 82\%$  на 1Зв, что почти вдвое превышает показатель дозового тренда для кумулятивной пожизненной величины, определенной предыдущим алгоритмом при тех же условиях. Несоответствие двух ти-

пов относительных трендов и, следовательно, необходимость их различения отмечалась и другими авторами [34]. Таким образом, AMFIT может иметь склонность к переоценке эффектов облучения. В ряде случаев использование годовых показателей может также приводить и к недооценке реального (кумулятивного) риска [31]. Потенциальная основа для этого хорошо видна на рис. 3 и 4.

**Обсуждение преимуществ и недостатков алгоритма оценки рисков в неоднородной когорте.** Отметим главные обстоятельства, отличающие наш рискметрический алгоритм от своих аналогов и, прежде всего, от прототипа – алгоритма AMFIT. В первую очередь, в предложенном алгоритме все отсчеты изучаемых специфических событий рассматриваются в качестве процессов биномиальной природы, а кумулятивные и интенсивные показатели определяются на вероятностной основе, а не в качестве эвристических величин. Это позволяет рассматривать оценку рисков в качестве инструментальной процедуры косвенных измерений непрерывно-распределенных оценок параметров на основе байесовского подхода. В отличие от этого, и алгоритм AMFIT, и другие, реже применяемые алгоритмы [35–38] базируются на точечной оценке всей совокупности событий и идее гладкой аппроксимации получающихся негладких эмпирических распределений в рамках частотно-дискретной статистической парадигмы Пирсона и Фишера.

Отмеченные альтернативные подходы к оценке рисков вместе с математической простотой в сравнении с оценками процессов приобретают очевидные, но игнорируемые недостатки. Например, интенсивный показатель риска в реальности не является ни константой, ни набором констант, как это предписывалось бы пуассоновской статистикой. По этой причине не ясна роль погрешности, связанной со стратификацией неоднородной когорты по возрасту. Слишком малые возрастные интервалы могут приводить к исчезновению случаев в них для когорт малого объема, при которых функционал пуассоновской регрессии может терять экстремальные свойства. Слишком большие интервалы, напротив, будут приводить к необоснованному усреднению интенсивностей накопления риска в пределах интервала. Заранее назначить оптимальную ширину интервалов при стратификации по возрасту в этом случае затруднительно. Аналогичный недостаток пуассоновской регрессии проявляется по мере вымирания когорты, когда число новых специфических случаев устремляется к нулю. На какой границе возрастного распределения следует остановиться? Более адекватен был бы биномиальный закон распределения событий, из которого закон Пуассона вытекает асимптотически как частная статистическая модель редких событий.

Отметим также, что качество аппроксимации  $h(t|D, sex, \beta)$  для описания радиационного воздействия зависит от того, насколько хорошо работает

эта функция в случае отсутствия внешнего воздействия ( $D = 0$ ). Речь идет о той самой группе сравнения, на которой держится любое сравнительное исследование. Часто на этот нюанс исследователи внимания не обращают [32, 33], полагаясь на простые модели степенного роста или модели с насыщением. Формально это означает, что фоновый пожизненный кумулятивный показатель интенсивности может достигать очень больших значений или обращаться в бесконечность, а сам специфический риск гибели от изучаемой причины достигнет единицы. Однако на самом деле таких болезней не существует, и это видно уже на этапе предварительного группирования данных (рис. 2). Учет такой информации в моделях динамики интенсивных показателей риска уже с успехом применялся в байесовских исследованиях на ограниченных выборках с неполным периодом наблюдения и потерями данных, например, по методикам *right censored spell models* [39], *cure rate models* [40], *bounded cumulative hazard models* [41]. Более того, есть основания полагать, что смещение оценок на рис. 5 преимущественно связано именно с неудачной моделью фонового риска, а не только с игнорированием природы процесса и ошибкой в выборе статистического закона. Заметим, что такой показатель, как *p-value*, определяемый в результате применения теста отношения правдоподобий, в многомерном случае не всегда может быть надежным ориентиром при оценке качества аппроксимации.

Все сказанное не означает, что вновь предлагаемый метод оценки рисков является идеальным инструментом исследователя. Напомним, что он базируется на принципе такой стратификации когорты, которая не предусматривает переход членов когорты из страты в страту в период наблюдения. Если бы такой переход происходил, страты нельзя было бы считать независимыми единицами, а значит функционал вероятностной оценки должен строиться на иных основаниях. Уязвимым местом может быть также неполнота покрытия факторного пространства имеющимися наблюдениями, свойственная любой эмпирической выборке. Стандарты CONSORT [42] распространяются на любые выборочные исследования. Если хотя бы одна пара факторов сильно коррелирует, действие одного из факторов может быть замаскировано кажущимся действием другого. По этой причине для обеспечения успешного оценивания процессов либо когорты должны быть достаточно велики и разнообразны, либо исследование связей риска с факторами должно иметь характер контролируемого эксперимента.

Важно отметить, что данная статья не продвигает никакой теории в области аналитической эпидемиологии. В ней предлагается всего лишь инструмент практического анализа в рамках ментальной схемы «доза – время – риск» вместо традиционного ограниченного подхода «доза – риск», навязывающего толкование вреда от внешних воздействий ис-

ключительно в форме простых одномерных зависимостей на плоском двухкоординатном графике. Действительно, упрощающая концепция «доза – риск» способна порождать необъяснимые артефакты в виде ложных трендов. Наглядным примером является толкование радиационно-онкологического тренда, график которого нами заимствован из документа [43].

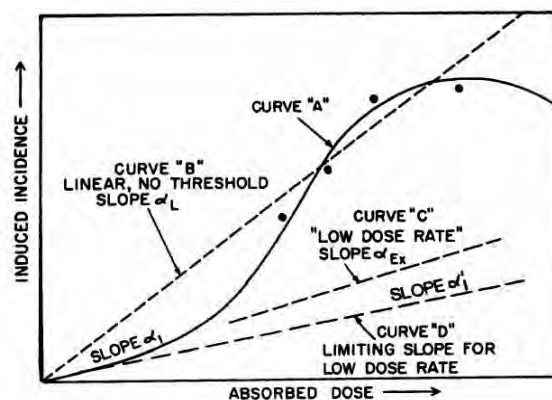


Рис. 6. Схематичное поведение избыточного показателя риска в зависимости от дозы радиационного воздействия по ожиданиям экспертов BEIR VII. Воспроизведено по отчету [43]

На нем со всей очевидностью представлен нелинейный отклик избыточной интенсивности специфических событий с явно немонотонным спадом в области больших доз. Такой тренд невозможно убедительно объяснить ни в рамках широко известной линейной беспороговой модели, ни в рамках обоснованных нелинейных моделей. Однако обратим внимание на то, что, оперируя понятием «интенсивность» вместо понятия «риск», график можно связать только с одной возрастной группой членов когорты. Тогда легко обнаружить, что в диапазоне возрастов от 60 до 65 лет на рис. 3 и 4 дозовый тренд имеет именно такой вид с характерным спадом и даже возможной сменой знака «избыточного» эффекта в еще более старших возрастных группах. При этом кумулятивный показатель риска вместе с ростом кумулятивной дозы только монотонно растет. Здесь же уместно напомнить, что известная линейно-беспороговая модель «доза – эффект» Н.В. Тимофеева-Рессовского и К.Г. Циммера [44, 45] формулировалась именно для пары кумулятивных показателей, одним из которых был аналог показателя Нельсона – Аалена задолго до изобретения последнего. Близкой точки зрения на область применения кумулятивных показателей придерживался также автор широко известной концепции «эффективной дозы» [46].

**Выводы.** Таким образом, анализ динамики наступления специфических событий в неоднородной когорте в сочетании с байесовской методологией оценки рисков имеет определенную перспективу применения, если исследователи об-

ладают детальной информацией о членах когорты в течение достаточно длительного или пожизненного наблюдения и полным исчерпывающим описанием индивидуальных факторов риска. Используемая техника вычислений лежит в рамках общепринятых эпидемиологических способов оценки вреда здоровью, комбинируя применение годовых групповых показателей риска вместе с кумулятивными. Показано, что, занимаясь прогнозированием ущерба от последствий внешнего воздействия на здоровье людей в рамках привычных ментальных схем «доза – эффект», следует отдавать предпочтение сочетанию кумулятивных доз и кумулятивных показателей риска или их дескриптивным аналогам (эффектам).

Вместе с тем нельзя не отметить повышенную трудоемкость использования описываемой параметрической версии байесовских методов. Частично преодолеть этот недостаток можно только путем создания соответствующего программного обеспечения<sup>6</sup>, способного автоматизировать процедуры группирования данных, выбора моделей, поиска экстремальных решений, моделирования статистической неопределенности байесовских оценок.

**Финансирование.** Создание методологии являлось частью исследований, финансируемых Федеральным медико-биологическим агентством (гриф «Последствия-2020»).

**Конфликт интересов.** Автор данной статьи заявляет об отсутствии конфликта интересов.

### Список литературы

1. Анализ риска здоровью в стратегии государственного социально-экономического развития: монография / Г.Г. Онищенко, Н.В. Зайцева, И.В. Май [и др.] / под ред. Г.Г.Онищенко, Н.В.Зайцевой. – М.; Пермь: Издательство Пермского национального исследовательского университета, 2014. – 738 с.
2. Commonwealth of Australia, 2012. Environmental Health Risk Assessment. Guideline for assessing human health risk from environmental hazards: Glossary. – Commonwealth of Australia, 2012. – 244 p.
3. Публикация 103 Международной Комиссии по радиационной защите (МКРЗ) / под ред. М.Ф. Киселева, Н.К. Шандалы. – М.: ООО ПКФ «Алана», 2009. – 344 с.
4. Злокачественные новообразования в России в 2018 году (заболеваемость и смертность) / под ред. А.Д. Каприна, В.В. Старинского, Г.В. Петровой. – М.: МНИОИ им. П.А. Герцена, 2019. – 250 с.
5. Ahrens W., Pigeot I. Handbook of epidemiology. – Switzerland: Springer, 2005. – 1617 p. DOI: 10.1007/978-3-540-26577-1
6. Newman J.R. Mathematics of a lady tasting tea by Sir Ronald Fisher // The World of mathematics. – Vol. III, Part VIII. – New-York: Simon and Schuster, 1956. – P. 1514–1521.
7. Vaupel J.W., Manton K.G., Stallard E. The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality // Demography. – 1979. – Vol. 16, № 3. – P. 439. DOI: 10.2307/2061224
8. Михальский А.И., Петровский А.М., Яшин А.И. Теория оценивания неоднородных популяций. – М.: Наука, 1989. – 128 с.
9. Technical Report No. 1-86. Life span study report 10. Part 1. Cancer mortality among A-bomb survivors in Hiroshima and Nagasaki, 1950-82 / D.L. Preston, H. Kato, K.J. Kopeccky, S. Fujita // RERF. – 1987. – № 111. – P. 151–178.
10. Solid cancer incidence in atomic bomb survivors exposed in utero or as young children / D.L. Preston, H. Cullings, A. Suyama, S. Funamoto, N. Nishi, M. Soda, K. Mabuchi, K. Kodama [et al.] // Journal of the National Cancer Institute. – 2008. – Vol. 100, № 6. – P. 428–436. DOI: 10.1093/jnci/djn045
11. Epicure. The premiere software for risk regression and person-year tabulation [Электронный ресурс] // EPICURE Risk Sciences International. – URL: <https://risksciences.com/epicure/> (дата обращения: 21.04.2021).
12. Гаусс К.Ф. Избранные геодезические сочинения / под ред. Г.В. Багратуни, С.Г. Судакова. – М: ИГЛ, 1957. – Т. 1. – 153 с.
13. Vaeth M., Pearce D. Calculating excess lifetime risk in relative risk models // Environmental Health Perspectives. – 1990. – Vol. 87. – P. 83–94. DOI: 10.1289/ehp.908783
14. Definition and estimation of lifetime detriment from radiation exposures: principles and methods / D. Thomas, S. Darby, F. Fagnani, P. Hubert, M. Vaeth, K. Weiss // Health Physics. – 1992. – Vol. 63, № 3. – P. 259–272. DOI: 10.1097/00004032-199209000-00001
15. Lifetime radiation risk of stochastic effects – prospective evaluation for space flight or medicine / A. Ulanowski, J.C. Kaiser, U. Schneider, L. Walsh // Ann. ICRP. – 2020. – Vol. 49, № 1. – P. 200–212. DOI: 10.1177/0146645320956517
16. On prognostic estimates of radiation risk in medicine and radiation protection / A. Ulanowski, J.C. Kaiser, U. Schneider, L. Walsh // Radiat. Environ. Biophys. – 2019. – Vol. 58, № 3. – P. 305–319. DOI: 10.1007/s00411-019-00794-1
17. Esteve J., Benhamou E., Raymond L. Statistical methods in cancer research. Descriptive epidemiology // IARC Scientific Publication. – 1994. – Vol. IV, № 128. – P. 313.
18. What is the lifetime risk of developing cancer?: the effect of adjusting for multiple primaries / P.D. Sasiени, J. Shelton, N. Ormiston-Smith, C.S. Thomson, P.B. Silcocks // Br. J. Cancer. – 2011. – Vol. 105, № 3. – P. 460–465. DOI: 10.1038/bjc.2011.250
19. History of application of martingales in survival analysis / O. Aalen, P.K. Andersen, Ø. Borgan, R.D. Gill, N. Keiding // Electronic Journal of History of Probability and Statistic. – 2009. – Vol. 5, № 1. – P. 1–28.

<sup>6</sup>Обеснюк В.Ф. Программа, управляющая многофакторной оценкой групповых рисков здоровью по индивидуальным записям регистра длительных наблюдений / зарегистрировано в БД РосПатента 23.12.2020; свид-во 2020667423.

20. Aalen O., Borgan Ø., Gjessing H. Survival and Event history analysis: A process point of view. – New-York: Springer Science + Business Media B.V., 2008. – P. 539.
21. Actual and actuarial probabilities of competing risks: apples and lemons / G.L. Grunkemeier, R. Jin, M.J.C. Eijkemans, J.J.M. Takkenberg // The Annals of Thoracic Surgery. – 2007. – Vol. 83, № 5. – P. 1586–1592. DOI: 10.1016/j.athoracsur.2006.11.044
22. Kaplan E.L., Meier P. Nonparametric estimation from incomplete observations // Journal of the American Statistical Association. – 1958. – Vol. 53, № 282. – P. 457–481. DOI: 10.1007/978-1-4612-4380-9\_25
23. Nelson W. Theory and applications of hazard plotting for censored failure data // Technometrics. – 1972. – Vol. 14, № 4. – P. 945–966. DOI: 10.1080/00401706.2000.10485975
24. Анализ выживаемости по медицинской базе данных больных раком предстательной железы / В.М. Буре, Е.М. Парилина, А.И. Рубша, Л.В. Свиркина // Вестник СПбГУ. Серия 10. – 2014. – Т. 10, № 2. – С. 27–35.
25. Fisher R.A. On the mathematical foundations of theoretical statistics // Phil. Trans. of the Royal Soc. of London. Series A. – 1922. – Vol. 222. – P. 309–368. DOI: 10.1098/rsta.1922.0009
26. Wilks S.S. The large-sample distribution of the likelihood ratio for testing composite hypotheses // The Annals of Mathematical Statistics. – 1938. – Vol. 9, № 1. – P. 60–62. DOI: 10.1214/aoms/1177732360
27. Fan J., Hung H., Wong W. Geometric understanding of likelihood ratio statistics // JASA. – 2000. – Vol. 95, № 451. – P. 836–841.
28. Gelfand A.E., Smith A. Sampling-based approaches to calculating marginal densities // Journal of the American Statistical Association. – 1990. – Vol. 85, № 410. – P. 398–409. DOI: 10.1080/01621459.1990.10476213
29. Sources and effects of ionizing radiation. UNSCEAR 1994 report to General Assembly. – New-York: United Nations Scientific Committee on the Effects of Atomic Radiation, 1994. – 272 p.
30. Effect on ionizing radiation. UNSCEAR 2006. Report to General Assembly. – New-York: United Nations Scientific Committee on the Effects of Atomic Radiation, 2008. – Vol. 1A. – 16 p.
31. Оценка радиогенного риска заболеваемости раком предстательной железы от внешнего  $\gamma$ -излучения в когорте работников ПО «МАЯК», подвергшихся профессиональному пролонгированному облучению / Л.В. Финашов, И.С. Кузнецова, М.Э. Сокольников, С.Г. Скуковский // Вопросы радиационной безопасности. – 2020. – № 2. – С. 37–48.
32. Риск развития радиационной катаракты у работников атомной промышленности – участников ликвидации последствий аварии на ЧАЭС / А.Р. Туков, И.Л. Шафранский, О.Н. Прохорова, М.Н. Зиядинов // Радиация и риск. – 2019. – Т. 28, № 1. – С. 37–46.
33. Lung cancer mortality among nuclear workers of the Mayak facilities in the former Soviet Union / M. Kreishermer, M.E. Sokolnikov, N.A. Koshurnikova, V.F. Khokhryakov, S. A. Romanow, N.S. Shilnikova, P.V. Okatenko, E.A. Nekolla, A.M. Kellerer // Radiat. Environ. Biophys. – 2003. – Vol. 42, № 2. – P. 129–135. DOI: 10.1007/s00411-003-0198-3
34. Zöllner S., Sokolnikov M.E., Eidemüller M. Beyond two-stage models for lung carcinogenesis in the Mayak workers: implications for plutonium risk // PLoS ONE. – 2015. – Vol. 10, № 5. – P. e0126238. DOI: 10.1371/journal.pone.0126238
35. Демин В.Ф., Иванов С.И., Новиков С.М. Общая методика оценки риска воздействия на здоровье человека разных источников опасности // Медицинская радиология и радиационная безопасность. – 2009. – Т. 54, № 1. – С. 5–15.
36. Latent period in induction of radiogenic solid tumors in the cohort of emergency workers / V.K. Ivanov, A.I. Gorsky, V.V. Kashcheev, M.A. Maksoutov, K.A. Tumanov // Radiation and Environmental Biophysics. – 2009. – Vol. 48, № 3. DOI: 10.1007/s00411-009-0223-2
37. Оптимизация радиационной защиты: «дозовая матрица» / В.К. Иванов, А.Ф. Цыб, А.П. Панфилов, А.М. Агапов. – М.: Медицина, 2006. – 304 с.
38. Lung cancer risk of Mayak workers: modelling of carcinogenesis and bystander effect / P. Jacob, R. Meckbach, M. Sokolnikov, V.V. Khokhryakov, E. Vasilenko // Radiat. Environ. Biophys. – 2007. – Vol. 46, № 4. – P. 383–394. DOI: 10.1007/s00411-007-0117-0
39. Chen M., Ibrahim J., Sinha D. A new bayesian model for survival data with a surviving fraction // Journal of the American Statistical Association. – 1999. – Vol. 94, № 447. – P. 909–919. DOI: 10.1080/01621459.1999.10474196
40. A unified view on lifetime distributions arising from selection mechanisms / J. Rodrigues, N. Balakrishnan, G. Cordeiro, M. de Castro // Computational Statistics and Data Analysis. – 2011. – Vol. 55, № 12. – P. 3311–3319. DOI: 10.1016/j.csda.2011.06.018
41. Tsodikov A.D., Ibrahim J.G., Yakovlev A.Y. Estimating cure rates from survival data: an alternative to two-component mixture models // Journal of the American Statistical Association. – 2003. – Vol. 98, № 464. – P. 1063–1067. DOI: 10.1198/01622145030000001007
42. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomized trials / D. Moher, S. Hopewell, K.F. Schulz, V. Montori, P.C. Gøtzsche, P.J. Devereaux, D. Elbourne, M.E. Douglas, G. Altman // International Journal of Surgery. – 2012. – Vol. 10, № 1. – P. 28–55. DOI: 10.1016/j.ijsu.2011.10.001
43. Health risks from exposure to low levels of ionizing radiation. BEIR VII, phase 2. – Washington D.C.: Committee to Assess Health Risks from Exposure to Low Levels of Ionizing Radiation, 2006. – 406 p.
44. Тимофеев-Рессовский Н.В., Циммер К.Г. Теория мишени радиобиологического действия (в изложении) // Биосфера. – 2010. – Т. 2, № 3. – С. 432–450.
45. Zimmer K.G. Ergebnisse und Grenzen der treffertheoretischen Deutung von strahlenbiologischen Dosis-Effekt kurven // Biol. Zent. – 1941. – № 63. – P. 72–107.
46. Jacobi W. The concept of the effective dose – a proposal of the combination of the organ doses // Radiat. And Environm. Biophys. – 1975. – Vol. 12, № 2. – P. 101–109. DOI: 10.1007/BF01328971

Обеснюк В.Ф. Групповые показатели риска здоровью в неоднородной когорте. Косвенная оценка по динамике событий // Анализ риска здоровью. – 2021. – № 2. – С. 17–32. DOI: 10.21668/health.risk/2021.2.02

Research article

**GROUP HEALTH RISK PARAMETERS IN A HETEROGENEOUS COHORT.  
INDIRECT ASSESSMENT AS PER EVENTS TAKEN IN DYNAMICS****V.F. Obesnyuk**The Southern Urals Biophysics Institute of the RF Federal Medical and Biological Agency, 19 Ozerskoe drive,  
Ozersk, 456780, Russian Federation

*The present work focuses on describing a procedure for assessing intensive and cumulative parameters of specific risk when observing cohorts under combined exposure to several external or internal factors.*

*The research goal was to reveal how to use well-known heuristic-descriptive parameters accepted in remote consequences epidemiology for analyzing dynamics of countable events in a cohort; analysis should be performed on quite strict statistic-probabilistic grounds based on Bayesian approach to explaining conditional probabilities that such countable events might occur. The work doesn't contain any new or previously unknown epidemiologic concept or parameters; despite that, it is not a simple literature review. It is the suggested procedure itself that is comparatively new as it combines techniques used to process conventional epidemiologic information and a correct metrological approach based on process description.*

*The basic result is providing a reader with understanding that all basic descriptive epidemiologic parameters within cohort description framework turn out to be quantitatively interlinked in case they are considered as conditional group processes. It allows simultaneous inter-consistent assessment of annual risk parameters and Kaplan-Meier (Fleming-Harrington) and Nelson-Aalen cumulative parameters as well as other conditional risk parameters or their analogues. It is shown that when a basic descriptive characteristic of cumulative parameters is chosen as a measure for measurable long-term external exposure, it is only natural to apply such a concept as a dose of this risk factor which is surrogate in its essence. Operability of the procedure was confirmed with an example. The suggested procedure was proven to differ from its prototype that previously allowed achieving only substantially shifted estimates, up to ~100 % even in case an operation mode was normal. Application requires creating specific but quite available PC software.*

**Key words:** risk, parameter, epidemiology, risk factor, competition, indirect estimate, mortality, process, cohort, strata, model.

**References**

1. Onishchenko G.G., Zaitseva N.V., May I.V. [et al.]. Health risk analysis in the strategy of state social and economical development: monograph. In: G.G. Onishchenko, N.V. Zaitseva eds. Moscow, Perm, Publishing house of the Perm National Research Polytechnic University Publ., 2014, 738 p. (in Russian).
2. Commonwealth of Australia, 2012. Environmental Health Risk Assessment. Guideline for assessing human health risk from environmental hazards: Glossary. Commonwealth of Australia, 2012, 244 p.
3. Publikatsiya 103 Mezhdunarodnoi Komissii po radiatsionnoi zashchite (MKRZ) [Publication No. 103 issued by the International Commission on Radiological Protection (ICRP 103)]. In: M.F. Kiselev, N.K. Shandala eds. Moscow, OOO PKF «Alana» Publ., 2009, 344 p. (in Russian).
4. Zlokachestvennye novoobrazovaniya v Rossii v 2018 godu (zabolevaemost' i smertnost') [Malignant neoplasms in Russia in 2018 (morbidity and mortality)]. In: A.D. Kaprin, V.V. Starinskii, G.V. Petrova eds. Moscow, MNIOI im. P.A. Gertsena Publ., 2019, 250 p. (in Russian).
5. Handbook of epidemiology. In: W. Ahrens, I. Pigeot eds. Switzerland, Springer Publ., 2005, 1617 p. DOI: 10.1007/978-3-540-26577-1
6. Newman J.R. Mathematics of a lady tasting tea by Sir Ronald Fisher. The World of mathematics. Vol. III, Part VIII. New-York, Simon and Schuster Publ., 1956, pp. 1514–1521.
7. Vaupel J.W., Manton K.G., Stallard E. The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality. *Demography*, 1979, vol. 16, no. 3, pp. 439. DOI: 10.2307/2061224
8. Mikhal'skii A.I., Petrovskii A.M., Yashin A.I. Teoriya otsenivaniya neodnorodnykh populyatsii [Theory of heterogeneous population estimation]. Moscow, Nauka Publ., 1989, 128 p. (in Russian).
9. Preston D.L., Kato H., Kopecky K.J., Fujita S. Technical Report No. 1-86. Life span study report 10. Part 1. Cancer mortality among A-bomb survivors in Hiroshima and Nagasaki, 1950–1982. *RERF*, 1987, no. 111, pp. 151–178.

© Obesnyuk V.F., 2021

Valery F. Obesnyuk – Candidate of Physical-Mathematical Sciences, Associate Professor, Senior researcher (e-mail: v-f-o@subi.su; tel.: +7 (35130) 7-52-36; ORCID: <https://orcid.org/0000-0002-2446-4390>).

10. Preston D.L., Cullings H., Suyama A., Funamoto S., Nishi N., Soda M., Mabuchi K., Kodama K. [et al.]. Solid cancer incidence in atomic bomb survivors exposed in utero or as young children. *Journal of the National Cancer Institute*, 2008, vol. 100, no. 6, pp. 428–436. DOI: 10.1093/jnci/djn045
11. Epicure. The premiere software for risk regression and person-year tabulation. «EPICURE» *Risk Sciences International*. Available at: <https://risksciences.com/epicure/> (21.04.2021).
12. Gauss K.F. Izbrannye geodezicheskie sochineniya [Selected geodesic works]. In: G.V. Bagratun, S.G. Sudakov eds. Moscow, IGL Publ., 1957, vol. 1, 153 p. (in Russian).
13. Vaeth M., Pearce D. Calculating excess lifetime risk in relative risk models. *Environmental Health Perspectives*, 1990, vol. 87, pp. 83–94. DOI: 10.1289/ehp.908783
14. Thomas D., Darby S., Fagnani F., Hubert P., Vaeth M., Weiss K. Definition and estimation of lifetime detriment from radiation exposures: principles and methods. *Health Physics*, 1992, vol. 63, no. 3, pp. 259–272. DOI: 10.1097/00004032-199209000-00001
15. Ulanowski A., Kaiser J.C., Schneider U., Walsh L. Lifetime radiation risk of stochastic effects – prospective evaluation for space flight or medicine. *Ann. ICRP*, 2020, vol. 49, no. 1, pp. 200–212. DOI: 10.1177/0146645320956517
16. Ulanowski A., Kaiser J.C., Schneider U., Walsh L. On prognostic estimates of radiation risk in medicine and radiation protection. *Radiat. Environ. Biophys*, 2019, vol. 58, no. 3, pp. 305–319. DOI: 10.1007/s00411-019-00794-1
17. Esteve J., Benhamou E., Raymond L. Statistical methods in cancer research. Descriptive epidemiology. *IARC Scientific Publication*, 1994, vol. IV, no. 128, pp. 313.
18. Sasieni P.D., Shelton J., Ormiston-Smith N., Thomson C.S., Silcocks P.B. What is the lifetime risk of developing cancer? The effect of adjusting for multiple primaries. *Br. J. Cancer*, 2011, vol. 105, no. 3, pp. 460–465. DOI: 10.1038/bjc.2011.250
19. Aalen O., Andersen P.K., Borgan Ø., Gill R.D., Keiding N. History of application of martingales in survival analysis. *Electronic Journal of History of Probability and Statistics*, 2009, vol. 5, no. 1, pp. 1–28.
20. Aalen O., Borgan Ø., Gjessing H. Survival and Event history analysis: A process point of view. New-York, Springer Science + Business Media B.V. Publ., 2008, pp. 539.
21. Grunkemeier G.L., Jin R., Eijkemans M.J.C., Takkenberg J.J.M. Actual and actuarial probabilities of competing risks: apples and lemons. *The Annals of Thoracic Surgery*, 2007, vol. 83, no. 5, pp. 1586–1592. DOI: 10.1016/j.athoracsur.2006.11.044
22. Kaplan E.L., Meier P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 1958, vol. 53, no. 282, pp. 457–481. DOI: 10.1007/978-1-4612-4380-9\_25
23. Nelson W. Theory and applications of hazard plotting for censored failure data. *Technometrics*, 1972, vol. 14, no. 4, pp. 945–966. DOI: 10.1080/00401706.2000.10485975
24. Bure V.M., Parilina E.M., Rubsha A.I., Svirkina L.V. Survival analysis of medical database of patients with prostate cancer. *Vestnik SPbGU. Seriya 10*, 2014, vol. 10, no. 2, pp. 27–35 (in Russian).
25. Fisher R.A. On the mathematical foundations of theoretical statistics. *Phil. Trans. of the Royal Soc. of London. Series A*, 1922, vol. 222, pp. 309–368. DOI: 10.1098/rsta.1922.0009
26. Wilks S.S. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 1938, vol. 9, no. 1, pp. 60–62. DOI: 10.1214/aoms/1177732360
27. Fan J., Hung H., Wong W. Geometric understanding of likelihood ratio statistics. *JASA*, 2000, vol. 95, no. 451, pp. 836–841.
28. Gelfand A.E., Smith A. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 1990, vol. 85, no. 410, pp. 398–409. DOI: 10.1080/01621459.1990.10476213
29. Sources and effects of ionizing radiation. UNSCEAR 1994 report to General Assembly. New-York, United Nations Scientific Committee on the Effects of Atomic Radiation Publ., 1994, 272 p.
30. Effect on ionizing radiation. UNSCEAR 2006. Report to General Assembly. New-York, United Nations Scientific Committee on the Effects of Atomic Radiation Publ., 2008, vol. 1A, 16 p.
31. Finashov L.V., Kuznetsova I.S., Sokol'nikov M.E., Skukovskii S.G. Radiation risk of prostate cancer incidence due to external gamma-exposure in the cohort of «Mayak» PA workers occupationally subjected to prolonged radiation exposure. *Voprosy radiatsionnoi bezopasnosti*, 2020, no. 2, pp. 37–48 (in Russian).
32. Tukov A.R., Shafranskii I.L., Prokhorova O.N., Ziyatdinov M.N. The incidence of cataracts and the radiation risk of their occurrence in liquidators of the Chernobyl accident, workers in the nuclear industry. *Radiatsiya i risk*, 2019, vol. 28, no. 1, pp. 37–46 (in Russian).
33. Kreisheimer M., Sokolnikov M.E., Koshurnikova N.A., Khokhryakov V.F., Romanov S.A., Shilnikova N.S., Okatenko P.V., Nekolla E.A., Kellerer A.M. Lung cancer mortality among nuclear workers of the Mayak facilities in the former Soviet Union. *Radiat. Environ. Biophys*, 2003, vol. 42, no. 2, pp. 129–135. DOI: 10.1007/s00411-003-0198-3
34. Zöllner S., Sokolnikov M.E., Eidemüller M. Beyond two-stage models for lung carcinogenesis in the Mayak workers: implications for plutonium risk. *PLoS ONE*, 2015, vol. 10, no. 5, pp. e0126238. DOI: 10.1371/journal.pone.0126238
35. Demin V.F., Ivanov S.I., Novikov S.M. Common methodology of health risk assessment for impact of different harm sources. *Meditsinskaya radiologiya i radiatsionnaya bezopasnost'*, 2009, vol. 54, no. 1, pp. 5–15 (in Russian).
36. Ivanov V.K., Gorsky A.I., Kashcheev V.V., Maksoutov M.A., Tumanov K.A. Latent period in induction of radiogenic solid tumors in the cohort of emergency workers. *Radiation and Environmental Biophysics*, 2009, vol. 48, no. 3. DOI: 10.1007/s00411-009-0223-2
37. Ivanov V.K., Tsyb A.F., Panfilov A.P., Agapov A.M. Optimizatsiya radiatsionnoi zashchity: «dozovaya matritsa» [How to optimize radiological protection: «dose matrix»]. Moscow, Meditsina Publ., 2006, 304 p. (in Russian).
38. Jacob P., Meckbach R., Sokolnikov M., Khokhryakov V.V., Vasilenko E. Lung cancer risk of Mayak workers: modeling of carcinogenesis and bystander effect. *Radiat. Environ. Biophys*, 2007, vol. 46, no. 4, pp. 383–394. DOI: 10.1007/s00411-007-0117-0

39. Chen M., Ibrahim J., Sinha D. A new bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, 1999, vol. 94, no. 447, pp. 909–919. DOI: 10.1080/01621459.1999.10474196

40. Rodrigues J., Balakrishnan N., Cordeiro G., de Castro M. A unified view on lifetime distributions arising from selection mechanisms. *Computational Statistics and Data Analysis*, 2011, vol. 55, no. 12, pp. 3311–3319. DOI: 10.1016/j.csda.2011.06.018

41. Tsodikov A.D., Ibrahim J.G., Yakovlev A.Y. Estimating cure rates from survival data: an alternative to two-component mixture models. *Journal of the American Statistical Association*, 2003, vol. 98, no. 464, pp. 1063–1067. DOI: 10.1198/01622145030000001007

42. Moher D., Hopewell S., Schulz K.F., Montori V., Gøtzsche P.C., Devereaux P.J., Elbourne D., Douglas M.E., Altman G. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomized trials. *International Journal of Surgery*, 2012, vol. 10, no. 1, pp. 28–55. DOI: 10.1016/j.ijsu.2011.10.001

43. Health risks from exposure to low levels of ionizing radiation. BEIR VII, phase 2. Washington D.C., Committee to Assess Health Risks from Exposure to Low Levels of Ionizing Radiation Publ., 2006, 406 p.

44. Timofeev-Resovskii N.V., Tsimmer K.G. Teoriya misheni radiobiologicheskogo deistviya (v izlozhenii) [Theory of radiological-biological effects targeting (a presentation)]. *Biosfera*, 2010, vol. 2, no. 3, pp. 432–450 (in Russian).

45. Zimmer K.G. Ergebnisse und Grenzen der treffertheoretischem Deutung von strahlenbiologischen Dosis-Effekt kurven. *Biol. Zent*, 1941, no. 63, pp. 72–107.

46. Jacobi W. The concept of the effective dose – a proposal of the combination of the organ doses. *Radiat. And Environm. Biophys*, 1975, vol. 12, no. 2, pp. 101–109. DOI: 10.1007/BF01328971

**Funding.** The procedure was created within a research program funded by the Federal Medical Biological Agency («Consequences – 2020»).

*Obesnyuk V.F. Group health risk parameters in a heterogeneous cohort. Indirect assessment as per events taken in dynamics. Health Risk Analysis, 2021, no. 2, pp. 17–32. DOI: 10.21668/health.risk/2021.2.02.eng*

Получена: 26.05.2021

Принята: 04.06.2021

Опубликована: 30.06.2021