DOI: 10.21870/0131-3878-2018-27-1-22-32

УДК 312.2:614.876(470.3)

Статистические связи смертности ликвидаторов с дозой облучения

Горский А.И., Максютлов М.А., Туманов К.А., Кочергина Е.В., Корело А.М.

МРНЦ им. А.Ф. Цыба – филиал ФГБУ «НМИЦ радиологии» Минздрава России, Обнинск

Целью работы является радиационно-эпидемиологическая классификация причин смерти ликвидаторов по классам, рубрикам и диагнозам с использованием статистических связей. Использованы данные о причинах смерти российских участников ликвидации последствий аварии на Чернобыльской АЭС (ликвидаторов) 1986-1987 гг. въезда в зону облучения за период наблюдения 1986-2014 гг. Средняя накопленная доза внешнего γ -облучения всего тела у ликвидаторов этой группы – 0,110 Гр, число случаев смерти – 27840. Анализ статистических связей доз облучения и структуры причин смерти проведён методами «сбора данных» («Data Mining»), свободными от априорных предположений о вероятностных распределениях доз и диагнозов. Для анализа использованы таблицы сопряжённости случаев смерти, дозы облучения и возраста в категориях (группа 0 – до 0,1 Гр и группа 1 – 0,1+ Гр) и двух возрастных группах (группа 0 – до 52 лет и группа 1 – 52+ года). Статистически значимый положительный тренд (отношение шансов) по дозе облучения в одинаковых возрастных категориях наблюдается для классов «Злокачественные новообразования» (класс С МКБ-10 для возрастной группы 1) и «Болезни системы кровообращения» (класс I для возрастной группы 1). Среди рубрик МКБ-10 статистически значимы положительные тренды для рубрик «Злокачественное новообразование желудка» (C16, возрастная группа 1), «Злокачественное новообразование бронхов и лёгких» (C34, возрастная группа 1), «Хроническая ишемическая болезнь сердца» (I25, возрастная группа 1). Для отдельных диагнозов тренд статистически значим для диагноза «Хроническая ишемическая болезнь сердца не уточнённая» (I25.9, возрастная группа 1). Положительный тренд по дозе облучения выявлен в основном для возрастной группы старше 52 лет, что даёт основания предполагать, что модель радиационного риска скорее мультипликативная вследствие положительной корреляции радиогенных раков с возрастом, а не аддитивная. Для уточнения причинно-следственных связей между причиной смерти и дозой облучения необходимо оценить радиационные риски с использованием методов радиационно-эпидемиологического исследования стохастических эффектов.

Ключевые слова: авария на ЧАЭС, ионизирующее излучение, дозы, ликвидаторы, смертность, сбор данных («Data Mining»), МКБ-10, таблицы сопряжённости, ассоциативные связи, отношение шансов, дозовая зависимость.

Даже при однократном воздействии на человека ионизирующего излучения в дозах менее 1 Гр возможно развитие отдалённых стохастических эффектов – увеличение вероятности развития радиационно-индуцированных заболеваний у людей в течение всей жизни после облучения [1].

Дозовая зависимость частоты заболеваемости и смерти наблюдается не только в случае злокачественных новообразований, но и в случае некоторых неонкологических заболеваний [1], к ним относятся болезни системы кровообращения (БСК), так как именно БСК являются наиболее частой причиной смерти человека.

Участники ликвидации аварии на Чернобыльской АЭС были облучены в дозах, в основном, менее 1 Гр. С эпидемиологической точки зрения когорты российских ликвидаторов является выборкой из населения, проживающего по всей территории России, и характеризуется значительной гетерогенностью в отношении фоновых (отсутствие облучения) показателей заболеваемости [2].

Горский А.И.* – вед. научн. сотр., к.т.н.; Максютлов М.А. – зав. отд., к.т.н.; Туманов К.А. – зав. лаб., к.б.н.; Кочергина Е.В. – зав. лаб., к.м.н.; Корело А.М. – ст. научн. сотр. МРНЦ им. А.Ф. Цыба – филиал ФГБУ «НМИЦ радиологии» Минздрава России.

*Контакты: 249036, Калужская обл., Обнинск, ул. Королева, 4. Тел.: (484) 399-32-60; e-mail: nrer@obninsk.com.

Источники вариации наблюдаемых показателей заболеваемости в когорте российских ликвидаторов весьма разнообразны, и традиционные модели радиационных рисков [1] учитывают, возможно, далеко не все из них. С другой стороны, статистическая значимость оценок радиационных рисков в традиционных моделях растёт с увеличением объёма данных (числа человеко-лет под риском в когорте) [1]. При недоучёте источников разброса наблюдений последнее обстоятельство может привести к ложным выводам о статистической значимости радиационных рисков. Поэтому представляет интерес исследование процессов заболеваемости и смертности среди ликвидаторов альтернативными современными статистическими методами.

Из современных эффективных подходов, предназначенных для анализа данных большого объёма, при отсутствии априорных предположений о распределениях наблюдаемых случайных величин, можно выделить так называемые алгоритмы «Data Mining», или «интеллектуальный анализ данных» [3-5], а именно раздел «Data Mining», относящийся к ассоциативным правилам «Sequence, Association, and Link Analysis». Интеллектуальный анализ данных широко используется во многих сферах современной человеческой деятельности: в банковском деле, фармакологии, маркетинге, генетике, геологоразведке, криминалистике, медицине.

Применение правил ассоциаций сравнительно нечасто встречается в публикациях, касающихся данных медицинских наблюдений. В ряде публикаций [6-9] подход «Sequence, Association, and Link Analysis» был использован для анализа больших объёмов данных с выводом правил, имеющих отношение к тематике данного исследования.

В работе [6] этот подход был использован для получения правил связей, отражающих соотношения между процедурами, проведёнными пациентам, и документированными диагнозами. Целью работы [7] было выведение ассоциативных правил для предсказания отсутствия или наличия заболеваний сердца. В работе [8] были получены последовательности симптомов и лабораторных анализов, которые свидетельствуют о переходе из состояния, предшествующего тромбозу, к активному тромбозу. Работа [9] посвящена определению правил для предсказания сердечных приступов.

В данном исследовании алгоритмы определения правил ассоциаций применены в области радиационной эпидемиологии: анализе исследований дозовой зависимости смертности среди ликвидаторов последствий аварии на ЧАЭС.

Материалы и методы

Для анализа использованы данные мониторинга состояния здоровья в когорте российских ликвидаторов последствий аварии на Чернобыльской АЭС (мужчин), накопленные в Национальном радиационно-эпидемиологическом регистре (НРЭР) [2]. Средняя по когорте накопленная за период работ доза внешнего γ -облучения всего тела для умерших ликвидаторов составляла 0,110 Гр, а максимальные дозы достигали 1 Гр. Более детальную информацию о НРЭР и когорте ликвидаторов можно найти в монографии [2].

В анализе использованы данные о смертности для когорты ликвидаторов с документированной дозой облучения (27840 случаев смерти) от всех причин, зарегистрированных с 1986 по 2014 гг. Структура смертности по классам МКБ-10 приведена в табл. 1, где представлены классы, численность которых превышала 150 случаев. Наибольшее число смертей наблюдается от болезней системы кровообращения (I), злокачественных новообразований (C), травм и отравлений (T, S), болезней органов дыхания и пищеварения (J, K). Для анализа дозовой зависимо-

сти смертности все случаи смерти были разделены на 2 дозовые группы: с дозой меньше или равной 100 мГр и дозой больше 100 мГр, а так же 2 возрастные группы: до 52 лет (группа 0) на момент смерти и свыше 52 лет (группа 1) с примерно равной численностью.

Таблица 1

Структура смертности среди ликвидаторов

Класс МКБ-10	Численность	%
Некоторые инфекционные и паразитарные болезни (A)	564	2,0
Новообразования (C)	4468	16,0
Психические расстройства и расстройства поведения (F)	401	1,4
Болезни нервной системы (G)	365	1,3
Болезни системы кровообращения (I)	11276	40,5
Болезни органов дыхания (J)	1484	5,3
Болезни органов пищеварения (K)	1985	7,1
Болезни мочеполовой системы (N)	151	0,5
Симптомы, признаки и отклонения от нормы, выявленные при клинических и лабораторных исследованиях, не классифицированные в других рубриках (R)	575	2,1
Отравления (S)	1708	6,1
Травмы (T)	4449	16,0
Другие	414	1,5
Все	27840	100

Элементами данных для каждого члена когорты были условный номер ликвидатора, коды (диагноз) причины смерти по классификации МКБ-10 [10] и групповые переменные – дозовая группа и возрастная группа ликвидатора.

Статистический метод анализа. Для анализа использовался программный модуль из пакета статистических программ «Статистика» [11] под названием «Sequence, Association, and Link Analysis» [3-5], который осуществляет один из методов интеллектуального анализа данных: «Анализ последовательности, ассоциации и связи». Модуль определяет ассоциативные связи переменных, их временные последовательности и кластеры. Ниже приведены основные термины, используемые в этих методах анализа.

«Транзакция» (Transaction) в контексте данной задачи – дозовая группа, возрастная группа и диагноз смерти на одного умершего члена когорты.

«Поддержка» (Support) – отношение числа транзакций (числа членов когорты), содержащих определённый набор данных (набор доз, возрастов и диагнозов смерти) к общему количеству транзакций (в данном случае умерших членов когорты). Например, «поддержка» набора данных, состоящих всего из двух элементов данных (A и C), является отношением числа транзакций, содержащих элементы A и C, к общему числу транзакций, т.е. совместной вероятностью событий вхождения элементов A и C, $P(A,C)$, в набор данных.

«Достоверность» (Confidence) показывает вероятность осуществления правила, что из наличия в транзакции (у члена когорты) набора (A) следует наличие в ней набора (C): $\text{Confidence (if } A \rightarrow C) = \text{Support}(A,C) / \text{Support}(A)$. «Достоверность» в данном контексте представляет собой условную вероятность (C) при условии (A), $P(C|A) = P(A,C) / P(A)$. Здесь «достоверность» правила не является достаточной характеристикой для использования правила на практике.

«Полезность» правила (lift) определяется как $\text{lift (if } A \rightarrow C) = \text{Confidence (if } A \rightarrow C) / \text{Support}(C) = P(AC) / (P(A) \times P(C))$. Из этого определения следует, что если $\text{lift} = 1$, то события A и C независимы, и статистическая связь A и C отсутствует. Чем больше величина lift, тем больше мера статистической связи A и C, и тем более полезно правило связи. Критерием полезности правила считается условие $\text{lift} > 1$.

Для определения статистической значимости правила используем вероятности Support (sup), Confidence (conf), меру связи lift, таблицы сопряжённости 2×2 и статистику χ^2 :

$$\chi^2 = \sum_{0 \leq i, j \leq 1} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}}, \quad (1)$$

где $O_{i,j}$ и $E_{i,j}$ – наблюдаемые и ожидаемые частоты, соответственно, в ячейке (i, j) таблицы сопряжённости 2×2. Для расчёта статистики χ^2 для пары переменных необходимо создать две таблицы (табл. 2, 3) сопряжённости наблюдаемых (O) и ожидаемых значений (E).

Таблица 2

Таблица сопряжённости для наблюдаемых частот

	C	$\bar{C}$
A	$n \cdot P(A \cap C)$	$n \cdot P(A \cap \bar{C})$
$\bar{A}$	$n \cdot P(\bar{A} \cap C)$	$n \cdot P(\bar{A} \cap \bar{C})$

Таблица 3

Таблица сопряжённости для ожидаемых частот

	C	$\bar{C}$
A	$n \cdot P(A) \cdot P(C)$	$n \cdot P(A) \cdot (1 - P(C))$
$\bar{A}$	$n \cdot (1 - P(A)) \cdot P(C)$	$n \cdot (1 - P(A)) \cdot (1 - P(C))$

Табл. 2 представляет наблюдаемые частоты. Таблица имеет четыре ячейки, соответствующих четырём возможным комбинациям событий A и C . Черта над переменной означает отсутствие события, или событие, дополнительное по отношению ко всему множеству рассматриваемых событий. Например, если групповая переменная дозы, как в данном исследовании, может принимать два значения, 0 и 1. Если событие A будет обозначать наличие у ликвидатора дозы в группе 1, то дополнительное к нему событие $\bar{A}$ – дозы в группе 0. Если событие C будет обозначать определённый диагноз смерти у ликвидатора, то дополнительное к нему событие $\bar{C}$ – отсутствие данного диагноза.

В табл. 3 приведены ожидаемые (E) частоты для статистически независимых событий A и C .

Согласно определениям sup, conf и lift:

$$P(A \cap C) = \text{sup},$$

$$P(A) = \frac{\text{sup}}{\text{conf}}, \quad P(C) = \frac{\text{conf}}{\text{lift}},$$

$$P(A \cap \bar{C}) = P(A) - P(A \cap C) = \frac{\text{sup}}{\text{conf}} - \text{sup}, \quad (2)$$

$$P(\bar{A} \cap C) = P(C) - P(A \cap C) = \frac{\text{conf}}{\text{lift}} - \text{sup},$$

$$P(\bar{A} \cap \bar{C}) = 1 - P(A \cap C) - P(A \cap \bar{C}) - P(\bar{A} \cap C) = 1 - \frac{\text{sup}}{\text{conf}} - \frac{\text{conf}}{\text{lift}} + \text{sup}.$$

Подставив значения вероятностей в табл. 2, 3 и ф-лу (1), получим выражение для статистики χ^2 через вероятности sup, conf и меру связи lift.

$$\chi^2 = n \cdot (\text{lift} - 1)^2 \cdot \frac{\text{sup} \cdot \text{conf}}{(\text{conf} - \text{sup}) \cdot (\text{lift} - \text{conf})}. \quad (3)$$

Табл. 2 сопряжённости для наблюдаемых частот позволяет оценить отношение шансов (OR, odds ratio), которое в случае редких заболеваний представляет оценку относительного риска. Получим выражение для OR через вероятности Support (sup), Confidence (conf), меру связи lift, табл. 2 и соотношения (3).

Согласно табл. 2, отношение шансов события С при условии события А к шансам события С равно:

$$OR = \frac{P(A \cap C) \cdot P(\bar{A} \cap \bar{C})}{P(A \cap \bar{C}) \cdot P(\bar{A} \cap C)}. \quad (4)$$

После подстановки выражений для вероятностей (2) и упрощения выражения получим:

$$OR = 1 + \frac{\text{conf} \cdot (1 - \text{lift})}{(\text{conf} - 1) \cdot (\text{conf} - \text{lift} \cdot \text{sup})}. \quad (5)$$

Для расчётов приближённого 95%-го доверительного интервала (ДИ) оценки OR используем приближение [12]:

$$OR^{(1 \pm 1.96 / \chi)}. \quad (6)$$

Задание граничных значений для «поддержки» и «достоверности» имеет практическое значение. Задание больших значений для нижней границы «поддержки» приведёт к выявлению очевидных связей переменных, имеющих большие частоты, задание малых значений позволит выявить скрытые, неочевидные связи.

В данном исследовании поддержка задана в пределах 0,5-100%, достоверность – в пределах 1-100%.

В анализе ассоциаций рассматриваются правила, в которых из наличия одного набора элементов (body), который трактуется как причина (или условие, в терминах вероятностного анализа), следует наличие другого набора элементов (head), который трактуется как вероятностное следствие. Строго говоря, в анализе рассматриваются не причинно-следственные связи, а статистические.

В данном исследовании в качестве причины выбирали дозовые группы, а в качестве следствия – диагноз смерти (класс, рубрика или диагноз) в определённой возрастной категории. Правила связи между дозой облучения, возрастом и диагнозом смерти считали статистически значимыми, если нижняя граница 95%-го ДИ оценки OR (величина low), превышала единицу.

Результаты и обсуждение

Результаты анализа связей смертности с дозой для классов причин смерти приведены в табл. 4. Статистически значимый положительный тренд (отношение шансов) по дозе облучения в одинаковых возрастных категориях наблюдается для классов «Злокачественные новообразования» (класс С МКБ-10 для возрастной группы 1) и «Болезни системы кровообращения» (класс I для возрастной группы 1). Эти классы и возрастные группы выделены в табл. 4. Положительные, но статистически незначимые тренды выявлены для классов «Заболевания органов дыхания» (класс J, возрастная группа 1), «Симптомы, признаки и отклонения от нормы, выявленные

при клинических и лабораторных исследованиях» (класс R, возрастная группа 0), «Отравления» (класс S, возрастная группа 1). Определение групп с положительными, но не значимыми трендами полезно для проведения последующих уточняющих радиационно-эпидемиологических исследований традиционными методами анализа стохастических эффектов облучения.

Таблица 4

Результаты анализа связей смертности с дозой для классов причин смерти

Причина смерти, МКБ-10	Возрастная группа	Численность	Поддержка (support)	Достоверность (confidence)	Полезность (lift)	Статистика Хи-квадрат	Отношение шансов (OR) по дозе	Нижняя граница 95% ДИ, (OR _{low})
C	0	718	0,026	0,051	0,995	0,040	0,989	0,889
C	1	1623	0,058	0,115	1,048	8,056	1,115	1,034
I	0	2220	0,080	0,157	0,968	5,735	0,925	0,868
I	1	3618	0,130	0,256	1,053	25,564	1,152	1,091
J	0	353	0,013	0,025	0,958	1,355	0,916	0,790
J	1	405	0,015	0,029	1,050	2,010	1,110	0,961
R*	0	159	0,006	0,011	1,016	0,080	1,033	0,825
S	0	662	0,024	0,047	1,002	0,006	1,004	0,898
S	1	214	0,008	0,015	1,032	0,439	1,068	0,879

* Для возрастной группы 1 данные отсутствуют согласно ограничениям по support и confidence.

Результаты анализа связей смертности с дозой для рубрик причин смерти приведены в табл. 5. Среди рубрик МКБ-10 статистически значимы положительные тренды для рубрик «Злокачественное новообразование желудка» (C16, возрастная группа 1), «Злокачественное новообразование бронхов и лёгких» (C34, возрастная группа 1), «Хроническая ишемическая болезнь сердца» (I25, возрастная группа 1). Положительные, но статистически незначимые тренды выявлены для рубрик C34, возрастная группа 0, «Острый инфаркт миокарда» (I21, возрастная группа 1), «Внутричерепное кровоизлияние» (I61, возрастная группа 1), «Инфаркт мозга» (I63, возрастная группа 1), «Пневмония без уточнения возбудителя» (J18, возрастная группа 1), «Фиброз и цирроз печени» (K74, возрастная группа 1).

Таблица 5

Результаты анализа связей смертности с дозой для рубрик причин смерти

Причина смерти, МКБ-10	Возрастная группа	Численность	Поддержка (support)	Достоверность (confidence)	Полезность (lift)	Статистика Хи-квадрат	Отношение шансов (OR) по дозе	Нижняя граница 95% ДИ, (OR _{low})
C34	0	208	0,007	0,015	1,026	0,277	1,055	0,865
C34	1	575	0,021	0,041	1,079	6,949	1,181	1,044
C16*	1	199	0,007	0,014	1,122	5,432	1,288	1,041
I21*	1	223	0,008	0,016	1,023	0,234	1,048	0,866
I24*	0	155	0,006	0,011	1,010	0,031	1,021	0,813
I25	0	934	0,034	0,066	0,952	4,999	0,900	0,820
I25	1	1652	0,059	0,117	1,054	10,323	1,131	1,049
I61*	1	177	0,006	0,013	1,072	1,742	1,159	0,931
I63*	1	234	0,008	0,017	1,071	2,260	1,158	0,956
J18	0	172	0,006	0,012	0,937	1,477	0,879	0,714
J18	1	187	0,0067	0,013	1,054	1,077	1,119	0,905
K74	0	180	0,006	0,013	0,976	0,225	0,951	0,773
K74	1	250	0,009	0,018	1,042	0,883	1,091	0,909
T75*	0	154	0,006	0,011	1,013	0,056	1,028	0,818

* Данные для другой возрастной группы отсутствуют согласно ограничениям по support и confidence.

Результаты анализа связей смертности с дозой для отдельных диагнозов причин смерти приведены в табл. 6. Для отдельных диагнозов тренд статистически значим для диагноза «Хроническая ишемическая болезнь сердца не уточнённая» (I25.9, возрастная группа 1). Положительные, но статистически незначимые тренды выявлены для диагнозов «Атеросклеротическая болезнь сердца» (I25.1, возрастная группа 1), «Перенесённый в прошлом инфаркт миокарда» (I25.2, возрастная группа 1), «Другой и не уточнённый цирроз печени» (K74.6, возрастная группа 1).

Таблица 6

Результаты анализа связей смертности с дозой для отдельных диагнозов причин смерти

Причина смерти, МКБ-10	Возрастная группа	Численность	Поддержка (support)	Достоверность (confidence)	Полезность (lift)	Статистика Хи-квадрат	Отношение шансов (OR) по дозе	Нижняя граница 95% ДИ, (OR _{low})
I25.1	0	437	0,016	0,031	0,921	6,294	0,846	0,742
I25.1	1	898	0,032	0,063	1,041	3,080	1,092	0,990
I25.2*	1	260	0,009	0,018	1,025	0,334	1,054	0,883
I25.9	0	290	0,010	0,020	0,929	3,240	0,863	0,736
I25.9	1	363	0,013	0,026	1,109	8,107	1,257	1,074
K74.6	0	161	0,006	0,011	0,943	1,149	0,889	0,717
K74.6	1	227	0,008	0,016	1,027	0,328	1,057	0,874

* Данные для другой возрастной группы отсутствуют согласно ограничениям по support и confidence.

В данной работе для анализа дозовой зависимости смертности ликвидаторов использованы современные статистические подходы, основанные на правилах ассоциаций. Подходы используют только «сырые» данные большого объёма, без каких-либо гипотез о законах связей и распределений переменных, и основаны исключительно на частотных характеристиках данных, полученных в результате сканирования информационной базы. Стандартные метрики ассоциативных правил были переведены в формат, традиционный в клинических и эпидемиологических исследованиях, а именно, к отношению шансов или относительному риску. Очевидным преимуществом данного подхода является возможность оперативно использовать для анализа связей с дозой отдельные диагнозы причин смерти, что затруднительно при проведении обычного радиационно-эпидемиологического анализа.

Использование в исследовании групповых доз вместо индивидуальных уменьшает возможное смещение оценки радиационного риска, обусловленного погрешностью индивидуальных доз.

Выявленные в исследовании статистические связи отдельных причин смерти с дозой облучения в основном не противоречат данным других исследований смертности и заболеваемости среди ликвидаторов [13-15]. Несколько неожиданным результатом проведённого исследования является тот факт, что для большинства причин смерти статистически значимая зависимость смертности от дозы выявляется в старшей возрастной группе (старше 52 лет), что даёт основания предполагать, что модель радиационного риска скорее мультипликативная вследствие положительной корреляции радиогенных раков с возрастом, а не аддитивная. Полученный результат находится в противоречии с моделями риска, полученными в японской когорте LSS, где относительные радиационные риски уменьшаются при увеличении достигнутого возраста, что требует дальнейшего изучения.

В заключение следует отметить, что выявленные правила имеют сравнительно невысокую поддержку и достоверность, хотя некоторые статистически значимы. Выявленные правила могут быть случайны, особенно для групп с небольшим числом случаев, вследствие возможной гетерогенности изучаемой когорты, поэтому окончательный вывод о зависимости смертности от дозы облучения для конкретных причин смерти может дать только классический, параметрический и непараметрический радиационно-эпидемиологический анализ.

Выводы

1. Статистически значимый положительный тренд (отношение шансов) по дозе облучения в одинаковых возрастных категориях наблюдается для классов «Злокачественные новообразования», «Болезни системы кровообращения».

2. Среди рубрик МКБ-10 статистически значимы положительные тренды для рубрик «Злокачественное новообразование желудка», «Злокачественное новообразование бронхов и лёгких», «Хроническая ишемическая болезнь сердца».

3. Для отдельных диагнозов тренд статистически значим для диагноза «Хроническая ишемическая болезнь сердца не уточнённая».

4. Выявленные статистически значимые связи отдельных причин смерти относятся к возрастной группе старше 52 лет, что даёт основания предполагать, что модель радиационного риска скорее мультипликативная, вследствие положительной корреляции радиогенных раков с возрастом, а не аддитивная. Полученный результат находится в противоречии с моделями риска, полученными в японской когорте LSS, где относительные радиационные риски уменьшаются при увеличении достигнутого возраста, и требует дальнейшего изучения.

5. Для уточнения выявленных связей причин смерти с дозой облучения требуется дальнейшая оценка радиационных рисков с использованием традиционных методов радиационно-эпидемиологического исследования стохастических эффектов.

Литература

1. United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR). Sources and effects of ionizing radiation. UNSCEAR 2006 Report Vol. I, Annex A: Epidemiological studies of radiation and cancer. New York: United Nation, 2008.
2. **Ivanov V., Tsyb A., Ivanov S., Pokrovsky V.** Medical radiological consequences of the Chernobyl catastrophe in Russia: estimation of radiation risks. St. Petersburg: Nauka, 2004. 388 p.
3. **Piatetsky-Shapiro G.** Discovery, analysis and presentation of strong rules. Knowledge discovery in databases /Eds.: G. Piatetsky-Shapiro, W.J. Frawley. Cambridge, MA: AAAI/MIT Press, 1991. P. 229-248.
4. **Agrawal R., Imieliński T., Swami A.** Mining association rules between sets of items in large databases. Proceedings of the 1993 ACM SIGMOD International conference on Management of data (SIGMOD'93). New York, 1993. P. 207-216. doi 10.1145/170035.170072.
5. **Hahsler M.** A Probabilistic Comparison of Commonly Used Interest Measures for Association Rules, 2015. [Электронный ресурс]. URL: http://michael.hahsler.net/research/association_rules/measures.html (дата обращения 16.10.2017).
6. **Doddi S., Marathe A., Ravi S.S., Torney D.C.** Discovery of association rules in medical data //Med. Inform. Internet Med. 2001. V. 26, N 1. P. 25-33.
7. **Ordonez C., Ezquerro N., Santana C.A.** Constraining and summarizing association rules in medical data //Knowledge Information System. 2006. V. 9, N 3. P. 259-283.
8. **Jensen S.** Mining medical data for predictive and sequential patterns //Proceedings of the Fifth European Conference on Principles and Practice of Knowledge Discovery in Databases. Freiburg, Germany, September 3-5, 2001. P. 1-10.
9. **Shantakumar B.P., Kumaraswamy Y.S.** Extraction of significant patterns from heart disease warehouses for heart attack prediction //International Journal of Computer Science and Network Security (IJCSNS). 2009. V. 9, N 2. P. 228-235.
10. Международная статистическая классификация болезней и проблем, связанных со здоровьем, 10-й пересмотр (МКБ-10). Т. 1 (часть 1). Женева: ВОЗ, 1995. 698 с.
11. Пакет статистических программ «Статистика». [Электронный ресурс]. URL: <http://www.statsoft.ru> (дата обращения 16.10.2017).
12. **Miettinen O.S.** Confounding and effect modification //Am. J. Epidemiol. 1974. V. 100. P. 350-353.
13. **Горский А.И., Максюттов М.А., Туманов К.А., Щукина Н.В., Чекин С.Ю., Иванов В.К.** Непараметрический анализ радиационных рисков смертности среди ликвидаторов последствий аварии на ЧАЭС //Радиационная биология. Радиоэкология. 2016. Т. 56, № 2. С. 138-148.
14. **Иванов В.К., Цыб А.Ф., Горский А.И., Максюттов М.А., Чекин С.Ю., Петров А.В., Туманов К.А., Кашеев В.В.** Онкозаболеваемость и онкосмертность среди участников ликвидации последствий аварии на ЧАЭС: оценка радиационных рисков //Радиационная биология. Радиоэкология. 2006. Т. 46, № 2. С. 159-166.
15. **Ivanov V.K., Maksoutov M.A., Chekin S.Yu., Petrov A.V., Biryukov A.P., Kruglova Z.G., Matyash V.A., Tsyb A.F., Manton K.G., Kravchenko J.S.** The risk of radiation-induced cerebrovascular disease in Chernobyl emergency workers //Health Phys. 2006. V. 90, N 3. P. 199-207.

Association rules for discovery relationship between mortality among Chernobyl liquidators and radiation dose

Gorski A.I., Maksimov M.A., Tumanov K.A., Kochergina E.V., Korelo A.M.

A. Tsyb MRRC, Obninsk

The aim of the study is to classify causes of death of the Chernobyl emergency accident workers (liquidators) in compliance with ICD-10 using statistical links. For these purposes we used mortality data (27840 cases) on Russian liquidators entered the Chernobyl zone in 1986-1987. The data were collected for the period from 1986 over 2014, average external dose of γ -radiation to a whole body is 0.110 Gy. To analyze statistical links between radiation doses and causes of death data mining algorithms free of a priori statements on probabilistic distributions of doses and diagnoses were used. Tables of contingency of death cases in two dose groups, group 0, dose <0.1 Gy, and group 1, dose ≥ 0.1 Gy, and two age groups, group 0, age <52 years, group 1, age ≥ 52 years, were used for analysis. Statistically significant positive association of radiation dose and death causes in similar age groups for disease in the Chapter C "Malignant Neoplasms" (age group 1) and the Chapter I "Diseases of the circulatory system" (age group 1). Statistically significant associations are in the blocs of categories C16 "Malignant neoplasm of stomach", C34 "Malignant neoplasm of bronchus and lung", I25 "Chronic ischaemic heart disease", all in the age group 1. The association is statistically significant for the diagnosis I25.9 "Chronic ischaemic heart disease, unspecified" in the age group 1. The positive odds ratio was found in the age group >52 years. This finding allows us to assume that the model of radiation risk is multiplicative because of the existence of positive correlation between radiation associated cancers and age, rather than additive. To determine correlation between causes of death and radiation dose more precisely, radiation risks should be assessed with account of results of radiation epidemiological studies of stochastic effects.

Key words: Chernobyl accident, ionizing radiation, doses, Chernobyl liquidators, mortality, data mining, ICD-10, tables of contingency, association rules, odds ratio, dose response.

References

1. United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR). Sources and effects of ionizing radiation. UNSCEAR 2006 Report Vol. I, Annex A: Epidemiological studies of radiation and cancer. New York, United Nation, 2008.
2. Ivanov V., Tsyb A., Ivanov S., Pokrovsky V. Medical radiological consequences of the Chernobyl catastrophe in Russia: estimation of radiation risks. St. Petersburg, Nauka, 2004. 388 p.
3. Piatetsky-Shapiro G. Discovery, analysis and presentation of strong rules. Knowledge discovery in databases. Eds.: G. Piatetsky-Shapiro, W.J. Frawley. Cambridge, MA, AAAI/MIT Press, 1991. P. 229-248.
4. Agrawal R., Imieliński T., Swami A. Mining association rules between sets of items in large databases. Proceedings of the 1993 ACM SIGMOD International conference on Management of data (SIGMOD'93). New York, 1993. P. 207-216. doi 10.1145/170035.170072.
5. Hahsler M. A Probabilistic Comparison of Commonly Used Interest Measures for Association Rules, 2015. Available at: http://michael.hahsler.net/research/association_rules/measures.html (Accessed 16.10.2017).
6. Doddi S., Marathe A., Ravi S.S., Torney D.C. Discovery of association rules in medical data. *Med. Inform. Internet Med.*, 2001, vol. 26, no. 1, pp. 25-33.
7. Ordonez C., Ezquerro N., Santana C.A. Constraining and summarizing association rules in medical data. *Knowledge Information System*, 2006, vol. 9, no. 3, pp. 259-283.
8. Jensen S. Mining medical data for predictive and sequential patterns. Proceedings of the Fifth European Conference on Principles and Practice of Knowledge Discovery in Databases. Freiburg, Germany, September 3-5, 2001. P. 1-10.

Gorski A.I.* – Lead. Researcher, C. Sc., Tech.; Maksimov M.A. – Head of Dep., C. Sc., Tech.; Tumanov K.A. – Head of Lab., C. Sc., Biol.; Kochergina E.V. – Head of Lab., C. Sc., Med.; Korelo A.M. – Senior Researcher. A. Tsyb MRRC.

*Contacts: 4 Korolyov str., Obninsk, Kaluga region, Russia, 249036. Tel.: (484) 399-32-60; e-mail: nrer@obninsk.com.

9. **Shantakumar B.P., Kumaraswamy Y.S.** Extraction of significant patterns from heart disease warehouses for heart attack prediction. *International Journal of Computer Science and Network Security (IJCSNS)*, 2009, vol. 9, no. 2, pp. 228-235.
10. *International Statistical Classification of Diseases and Related Health*, 10th revision (ICD-10). Vol. 1 (Part 1). Geneva, WHO, 1995. 696 p. (In Russian).
11. Software «Statistics». Available at: <http://www.statsoft.ru> (Accessed 16.10.2017).
12. **Mietenen O.S.** Confounding and effect modification. *Am. J. Epidemiol.*, 1974, vol. 100, pp. 350-353.
13. **Gorsky A.I., Maksioutov M.A., Tumanov K.A., Shchukina N.V., Chekin S.Yu., Ivanov V.K.** Non-parametric analysis of radiation risks of mortality among Chernobyl clean-up workers. *Radiacionnaja biologija. Radiojekologija – Radiation Biology. Radioecology*, 2016, vol. 56, no. 2, pp. 138-148. (In Russian).
14. **Ivanov V.K., Tsyb A.F., Gorsky A.O., Maksioutov M.A., Chekin S.Yu., Petrov A.V., Tumanov K.A., Kashcheev V.V.** Cancer morbidity and mortality among Chernobyl emergency workers: estimation of radiation risks. *Radiacionnaja biologija. Radiojekologija – Radiation Biology. Radioecology*, 2006, vol. 46, no. 2, pp. 159-166. (In Russian).
15. **Ivanov V.K., Maksioutov M.A., Chekin S.Yu., Petrov A.V., Biryukov A.P., Kruglova Z.G., Matyash V.A., Tsyb A.F., Manton K.G., Kravchenko J.S.** The risk of radiation-induced cerebrovascular disease in Chernobyl emergency workers. *Health Phys.*, 2006, vol. 90, no. 3, pp. 199-207.