РАЗДЕЛ 1**НАУЧНЫЕ СТАТЬИ**

DOI: 10.21870/0131-3878-2020-29-3-5-13

УДК 616-006.6-02:614.876(470.3)

Радиационно-эпидемиологическая классификация онкологической заболеваемости среди российских участников ликвидации последствий аварии на Чернобыльской АЭС**Горский А.И., Максютлов М.А., Туманов К.А., Корело А.М., Кочергина Е.В.,
Лашкова О.Е., Иванов В.К.**

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

Целью работы является радиационно-эпидемиологическая классификация заболеваемости ликвидаторов злокачественными новообразованиями (ЗНО) по рубрикам и диагнозам с использованием статистических связей заболеваемости с дозой облучения. Использованы данные по онкозаболеваемости российских участников ликвидации последствий аварии на Чернобыльской АЭС (ликвидаторов, мужчин), 1986-1987 гг. въезда в зону облучения, за период наблюдения 1986-2019 гг., со средней дозой внешнего γ -облучения всего тела 0,128 Гр. Общее число случаев заболеваний – 9542. Средний возраст при диагнозе – 57,4 года. Данные накоплены в Национальном радиационно-эпидемиологическом регистре (НРЭР). Анализ статистических связей доз облучения и структуры (диагнозов) заболеваний проведён методами «Data Mining», свободными от априорных предположений о вероятностных распределениях доз и диагнозов. Для анализа использованы таблицы сопряжённости случаев заболеваний в двух дозовых категориях (группа 1 – до 0,1 Гр и группа 2 – 0,1+ Гр) и двух возрастных группах (группа 1 – до 52 лет и группа 2 – 52+ лет). Статистически значимые связи рубрик заболеваемости от дозы облучения выявлены для ЗНО пищевода (МКБ-10 C15, возрастная группа 2), желудка (C16, 1), бронхов и лёгкого (C34, 1), предстательной железы (C61, 2) и мочевого пузыря (C67, 2). Для отдельных диагнозов значимые связи выявлены для ЗНО желудка неуточнённой локализации (C16.9, 1), бронхов и лёгкого (C34.0, 1), бронхов и лёгкого неуточнённой локализации (C34.9, 1), мочевого пузыря неуточнённой локализации (C67.9, 2). Для уточнения причинно-следственных связей заболеваемости ЗНО с дозой облучения требуется дальнейшая оценка радиационных рисков с использованием более точных методов радиационно-эпидемиологического исследования стохастических эффектов.

Ключевые слова: авария на ЧАЭС, ионизирующее излучение, дозы, ликвидаторы, онкозаболеваемость, добыча данных (data mining), таблицы сопряжённости, ассоциативные связи с дозой, отношение шансов.

Введение

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). При недоучёте источников разброса наблюдений (гетерогенности данных) последнее обстоятельство может привести к ложным выводам о статистической значи-

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.; Корело А.М. – ст. научн. сотр.; Кочергина Е.В. – зав. лаб., к.м.н.; Лашкова О.Е. – научн. сотр.; Иванов В.К. – зам. директора по научн. работе, Председатель РНКРЗ, чл.-корр. РАН, д.т.н. МРНЦ им. А.Ф. Цыба – филиал ФГБУ «НМИЦ радиологии» Минздрава России.

*Контакты: 249035, Калужская обл., Обнинск, ул. Королева, 4. Тел.: (484) 399-32-60; e-mail: nrer@obninsk.com.

мости радиационных рисков для отдельных заболеваний. Поэтому представляют интерес исследования процессов заболеваемости и смертности среди ликвидаторов альтернативными современными статистическими методами.

Из современных эффективных подходов, предназначенных для анализа данных большого объема, при отсутствии априорных предположений о распределениях наблюдаемых случайных величин, можно выделить так называемые алгоритмы «Data Mining», или «интеллектуальный анализ данных», описание которого можно найти в публикациях [2-4]. Интеллектуальный анализ данных широко используется во многих сферах современной человеческой деятельности: в банковском деле, фармакологии, маркетинге, генетике, геологоразведке, криминалистике, медицине.

Применение правил ассоциаций из алгоритмов «Data Mining» сравнительно нечасто встречается в зарубежных публикациях, касающихся данных медицинских наблюдений. Ссылки на них можно найти в работе [5]. Из публикаций в отечественной литературе по данной тематике можно привести работы [5, 6].

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

Материалы и методы

В статье использованы данные наблюдений за когортой ликвидаторов с 1986 по 2019 гг. Это ликвидаторы (мужчины), имеющие следующие характеристики:

- год въезда в зону облучения – 1986-1987 гг.;
- год регистрации в когорте – до 1991 г.;
- документированные дозы внешнего облучения;
- дату и диагноз МКБ-10 [7] онкологического заболевания;
- данные о годе рождения.

Для анализа дозовой зависимости заболеваемости все случаи были разделены на две дозовые группы: с дозой внешнего облучения меньше или равной 100 мГр (группа 1) и дозой больше 100 мГр (группа 2) и две возрастные группы до 52 лет (группа 1) на момент заболевания и свыше 52 лет (группа 2).

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

Для анализа использовался программный модуль из пакета статистических программ «Статистика» [8] под названием «Sequence, Association, and Link Analysis», 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 заболевания на одного умершего члена когорты.

«Поддержка» (Support) – отношение числа транзакций (числа членов когорты), содержащих определённый набор данных (набор доз, возрастов и диагнозов заболевания) к общему

количеству транзакций (в данном случае умерших членов когорты, n). Например, «поддержка» набора данных, состоящих всего из двух элементов 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 (y члена когорты) набора A следует наличие в ней набора C : $\text{Confidence}(\text{if } A \rightarrow C) = \text{Support}(A,C)/\text{Support}(A)$. «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 $\text{lift}(\text{if } A \rightarrow C) = \text{Confidence}(\text{if } A \rightarrow C)/\text{Support}(C) = P(AC)/(P(A) \times P(C))$. 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 сопряжённости наблюдаемых O и ожидаемых E значений.

Методика расчёта компонентов двух таблиц 2×2 и необходимые комментарии к ним приведены в работе [5].

Выражение для статистики χ^2 через вероятности sup, conf и меру связи lift определяется из уравнения [5]:

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

Таблица сопряжённости для наблюдаемых частот позволяет оценить отношение шансов (OR, odds ratio), которое в случае редких заболеваний представляет оценку относительного риска. Выражение для OR, определённое через вероятности Support (sup), Confidence (conf), меру связи lift, имеет вид [5]:

$$\text{OR} = 1 + \frac{\text{conf} \cdot (1 - \text{lift})}{(\text{conf} - 1) \cdot (\text{conf} - \text{lift} \cdot \text{sup})}. \quad (3)$$

Для расчётов приближенного 95%-го доверительного интервала (ДИ) оценки **OR** используется приближение [9]:

$$\text{OR}^{(1 \pm 1.96 / \chi)}. \quad (4)$$

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 заболевания (рубрика или диагноз) в определённой возрастной категории. Правила связи между дозой облучения, возрастом и диагнозом заболевания считались статистически значимыми, если нижняя граница 95%-го ДИ оценки OR во второй дозовой группе – величина low – превышала единицу.

Результаты

Фрагмент таблицы классификации представлен в табл. 1, в которой приведены связи заболеваний с дозой облучения (дозовой группой), для которых относительный риск (OR) больше 1. Всего при заданных значениях (support=0,005) и (confidence=0,01) найдено 44 правила. В качестве причины (body) рассматривалась дозовая группа, следствия (head) заболевания и возрастная группа. Фактически в табл. 1 приведена структура риска, из которой формируется значение риска для всего класса и рубрик ЗНО. Для наличия дозовой зависимости заболевания необходимо условие, что значение OR в дозовой группе 2 было больше единицы. Однако, как следует из табл. 1, для некоторых заболеваний имеется обратная зависимость, риск в дозовой группе 1 больше, чем в дозовой группе 2. Наблюдаемый эффект наиболее вероятно обусловлен неучтёнными эффектами гетерогенности данных и он будет уменьшать эффект облучения для всего класса заболеваний.

Таблица 1

Структура дозовой зависимости заболеваний

Причина (Body)	=>	Следствие (Head)	Поддержка (Support)	Достоверность (Confidence)	Полезность правила (Lift)	Статистика χ^2	Отношение шансов (OR)	Нижняя граница 95% ДИ (Low)
Доз.груп=1	=>	C34.8	0,01415	0,0281	1,0431	0,4988	1,0935	0,9662
Доз.груп=1	=>	Возр.груп=2, C34.8	0,01142	0,0227	1,0021	0,0010	1,0044	1,0041
Доз.груп=1	=>	C32.9	0,00765	0,0152	1,0138	0,0279	1,0285	1,0191
Доз.груп=1	=>	Возр.груп=2, C32.9	0,00534	0,0106	1,2202	4,1184	1,5773	0,2574
Доз.груп=1	=>	C22	0,00514	0,0102	1,1185	1,2498	1,2740	0,7494
Доз.груп=1	=>	C20	0,01656	0,0329	1,0187	0,1131	1,0397	1,0134
Доз.груп=1	=>	Возр.груп=2, C20	0,01352	0,0268	1,0995	2,3963	1,2290	0,6574
Доз.груп=1	=>	C18.7	0,00901	0,0179	1,0674	0,7496	1,1484	0,9081
Доз.груп=1	=>	Возр.груп=2, C18.7	0,00744	0,0148	1,0444	0,2739	1,0951	0,9977
Доз.груп=1	=>	C16	0,01184	0,0235	1,0341	0,2620	1,0729	0,9998
Доз.груп=1	=>	Возр.груп=2, C16	0,00817	0,0162	1,0258	0,1036	1,0543	1,0197
Доз.груп=2	=>	Возр.груп=1, C64	0,00943	0,0190	1,0419	0,3068	1,0885	0,9928
Доз.груп=2	=>	Возр.груп=1, C34.9	0,00901	0,0182	1,0251	0,1065	1,0520	1,0184
Доз.груп=2	=>	Возр.груп=1, C34.1	0,00702	0,0141	1,1437	2,4335	1,3374	0,5498
Доз.груп=2	=>	Возр.груп=1, C34	0,00576	0,0116	1,0258	0,0718	1,0533	1,0249
Доз.груп=2	=>	Возр.груп=1, C16.9	0,00639	0,0129	1,0326	0,1260	1,0677	1,0201
Доз.груп=2	=>	C67.9	0,00828	0,0167	1,0267	0,1105	1,0553	1,0189
Доз.груп=2	=>	Возр.груп=2, C67.9	0,00587	0,0118	1,0349	0,1324	1,0726	1,0203
Доз.груп=2	=>	C64	0,02651	0,0534	1,0337	0,5832	1,0732	0,9655
Доз.груп=2	=>	Возр.груп=2, C64	0,01708	0,0344	1,0293	0,2789	1,0620	0,9979
Доз.груп=2	=>	C61	0,04454	0,0897	1,0048	0,0214	1,0106	1,0075

Заметим, что сумма OR для определённого диагноза и возрастной группы в двух дозовых группах равна 2. Для всего класса ЗНО относительный риск статистически значим для второй дозовой и возрастной групп и равен 1,0112 с нижним 95% доверительным пределом 1,0058.

В табл. 2 приведены статистически значимые значения относительного риска для рубрик заболеваний во второй дозовой группе ($low > 1$).

Таблица 2

Статистические связи рубрик заболеваний с дозой облучения

Причина (Body)	=>	Следствие (Head)	Поддержка (Support)	Достоверность (Confidence)	Полезность правила (Lift)	Статистика χ^2	Отношение шансов (OR)	Нижняя граница 95% ДИ (Low)
Доз.груп=2	=>	Возр.груп=1, C16	0,0170	0,0342	1,0198	0,1275	1,0415	1,0123
Доз.груп=2	=>	Возр.груп=1, C34	0,0261	0,0526	1,0133	0,0908	1,0282	1,0115
Доз.груп=2	=>	Возр.груп=1, C67	0,0050	0,0101	1,0178	0,0299	1,0363	1,0239
Доз.груп=2	=>	Возр.груп=2, C15	0,0095	0,0192	1,0017	0,0005	1,0034	1,0032
Доз.груп=2	=>	Возр.груп=2, C61	0,0429	0,0863	1,0096	0,0819	1,0212	1,0092
Доз.груп=2	=>	Возр.груп=2, C67	0,0141	0,0285	1,0223	0,1344	1,0467	1,0129

Как следует из табл. 2, значимые связи заболеваний с дозой выявлены для рубрик МКБ-10: C15 – ЗНО пищевода, C16 – желудка, C34 – бронхов и лёгкого, C61 – предстательной железы, C67 – мочевого пузыря.

В табл. 3 представлены статистические связи диагнозов заболеваний с дозой облучения.

Таблица 3

Статистические связи диагнозов заболеваний с дозой облучения

Причина (Body)	=>	Следствие (Head)	Поддержка (Support)	Достоверность (Confidence)	Полезность правила (Lift)	Статистика χ^2	Отношение шансов (OR)	Нижняя граница 95% ДИ (Low)
Доз.груп=2	=>	Возр.груп=1, C16.9	0,0064	0,0129	1,0326	0,1260	1,0677	1,0201
Доз.груп=2	=>	Возр.груп=1, C34	0,0058	0,0116	1,0258	0,0718	1,0533	1,0249
Доз.груп=2	=>	Возр.груп=1, C34.9	0,0090	0,0182	1,0251	0,1065	1,0520	1,0184
Доз.груп=2	=>	Возр.груп=2, C16.8	0,0050	0,0101	1,0072	0,0049	1,0145	1,0125
Доз.груп=2	=>	Возр.груп=2, C34.9	0,0121	0,0243	1,0116	0,0310	1,0238	1,0155
Доз.груп=2	=>	Возр.груп=2, C61	0,0427	0,0859	1,0097	0,0816	1,0212	1,0093
Доз.груп=2	=>	Возр.груп=2, C67.9	0,0059	0,0118	1,0349	0,1324	1,0726	1,0203

Значимые связи выявлены для диагнозов: C16.9 – ЗНО желудка неуточнённой локализации, C34.0 – бронхов и лёгкого, C34.9 – бронхов и лёгкого неуточнённой локализации, C67.9 – мочевого пузыря неуточнённой локализации.

Обсуждение результатов

Данная работа является продолжением поисковых исследований для определения рубрик и диагнозов заболеваний, связанных с фактором облучения с использованием алгоритмов «Data mining».

К достоинствам данного метода исследования следует отнести отсутствие априорных предположений о форме дозовой зависимости, законах распределения случайных величин. Применение методов «Data mining», разработанных для анализа больших объёмов данных «Big data» даёт возможность провести анализ для отдельных диагнозов, что затруднительно при стандартных методах регрессионного и непараметрического анализа. Использование в иссле-

довании групповых доз вместо индивидуальных уменьшает возможное смещение оценки радиационного риска, обусловленного погрешностью индивидуальных доз.

К недостаткам метода можно отнести малое число категорий по дозам и возрасту, 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 данных изучаемой когорты.

Результаты данного исследования для всего класса ЗНО, рубрик и групп отдельных диагнозов заболеваемости ЗНО в общем согласуются с результатами исследований [10, 11]. В работе [10] тренды рисков по дозе для всех ЗНО, для ЗНО системы пищеварения (рубрики С15-С16) и органов дыхания (С33-С34) положительны, но риски статистически не значимы из малого периода наблюдения до 2004 г., в работе [11] (период наблюдения до 2017 г.) риски для всего класса ЗНО и указанных групп диагнозов статистически значимы.

Выводы

Анализ статистических связей доз облучения и структуры (диагнозов) заболеваний проведён методами «Data Mining», свободными от априорных предположений о вероятностных распределениях доз и диагнозов. Для анализа использованы таблицы сопряжённости случаев заболеваний в двух дозовых категориях (группа 1 – до 0,1 Гр и группа 2 – 0,1+ Гр) и двух возрастных группах (группа 1 – до 52 лет и группа 2 – 52+ лет).

Статистически значимые связи рубрик заболеваемости от дозы облучения выявлены для ЗНО пищевода (МКБ-10 С15, возрастная группа 2), желудка (С16, 1), бронхов и лёгкого (С34, 1), предстательной железы (С61, 2) и мочевого пузыря (С67, 2).

Для отдельных диагнозов значимые связи выявлены для ЗНО желудка, неутончённой локализации (С16.9, 1), бронхов и лёгкого (С34.0, 1), бронхов и лёгкого неутончённой локализации (С34.9, 1), мочевого пузыря неутончённой локализации (С67.9, 2).

Окончательный вывод о зависимости заболеваемости от дозы облучения для конкретных диагнозов заболеваемости может дать только классический, параметрический и непараметрический радиационно-эпидемиологический анализ, хотя и для него влияние гетерогенности данных также не исключено.

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. **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.
3. **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.
4. **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 22.06.2020).
5. **Горский А.И., Максютлов М.А., Туманов К.А., Кочергина Е.В., Корело А.М.** Статистические связи смертности ликвидаторов с дозой облучения //Радиация и риск. 2018. Т. 27, № 1. С. 22-32.
6. **Горский А.И., Максютлов М.А., Туманов К.А., Власов О.К., Кочергина Е.В., Зеленская Н.С., Чекин С.Ю., Иванов С.А., Каприн А.Д., Иванов В.К.** Анализ статистических связей смертности от злокачественных новообразований с дозой облучения радионуклидами населения регионов, загрязнённых вследствие аварии на ЧАЭС //Медицинская радиология и радиационная безопасность. 2019. Т. 64, № 6. С. 5-11.
7. Международная статистическая классификация болезней и проблем, связанных со здоровьем, 10-й пересмотр (МКБ-10). Т. 1 (часть 1). Женева: ВОЗ, 1995. 698 с.
8. Пакет статистических программ «Статистика». [Электронный ресурс]. URL: <http://www.statsoft.ru> (дата обращения 18.06.2020).
9. **Miettinen O.S.** Confounding and effect modification //Am. J. Epidemiol. 1974. V. 100. P. 350-353.
10. **Ivanov V.K., Gorski A.I., Tsyb A.F., Ivanov S.I., Naumenko R.N., Ivanova L.V.** Solid cancer incidence among the Chernobyl emergency workers residing in Russia: estimation of radiation risks //Radiat. Environ. Biophys. 2004. V. 43, N 1. P. 35-42.
11. **Иванов В.К., Карпенко С.В., Кащеев В.В., Чекин С.Ю., Максютлов М.А., Туманов К.А., Щукина Н.В., Кочергина Е.В., Зеленская Н.С., Лашкова О.Е.** Радиационные риски российских участников ликвидации последствий аварии на Чернобыльской АЭС за период 1992-2017 гг. Часть I: заболеваемость солидными раками //Радиация и риск. 2019. Т. 28, № 4. С. 16-30.

Radiation epidemiology classification of malignant neoplasms in Chernobyl clean-up workers

Gorski A.I., Maksoutov M.A., Tumanov K.A., Korelo A.M., Kochergina E.V.,
Lashkova O.E., Ivanov V.K.

A. Tsyb MRRC, Obninsk

This article presents radiation epidemiology approach to classification of malignant neoplasms in accordance with ICD10 diagnostic categories and codes with the use of statistical links between radiation dose and cancer in clean-up workers of the Chernobyl accident. Modern statistical methods of intelligent data analysis were used for examination of statistical relationships (association) between radiation dose and the disease of a correspondent diagnostic category and code. Cancer cases registered from 1986-over 2019 in Chernobyl clean-up workers (males), who entered the Chernobyl 30-km exclusion zone in 1986 through 1987, were taken in the analysis. Average dose of external whole body γ -radiation exposure was 0.128 Gy. The total number of cancer cases was 9542; the average age at diagnosis was 57.4 years. The data are accumulated in the National Radiation and Epidemiological Register (NRER). To detect statistical relationships between radiation doses and cancer diagnoses Data Mining techniques free of apriori assumptions of probability distribution of dose and diagnosis were used. The analysis was made with the use of tables of contingency of disease cases in two dose groups (group 1: dose <0.1 Gy; group 2: dose >0.1 Gy) and in two age groups (age group 1: <52 years; age group 2: >52 years). Statistically significant relationships between categories of diagnosis and radiation dose were found for malignant neoplasm of esophagus (category C15; age group 2); malignant neoplasm of stomach (category C16; age group 1); malignant neoplasm of bronchus and lung (category C34; age group 1); malignant neoplasm of prostate (category C61, age group 2); malignant neoplasm of bladder (category C67, age group 2). Significant relationships for individual codes were found for malignant neoplasm of stomach unspecified (code C16.9, age group 1); malignant neoplasm of main bronchus (code C34.0, age group 1); malignant neoplasm of unspecified part of bronchus or lung (code C34.9, age group 1); malignant neoplasm of bladder unspecified (code C67.9, age group 2). To make radiation dose-death relationship more clear research of stochastic effects using intelligent data analysis techniques should be continued.

Key words: Chernobyl accident, ionizing radiation, doses, Chernobyl clean-up workers, cancer incidence, data mining, contingency tables, association links, odds ratio.

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. **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.
3. **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.
4. **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 22.06.2020).
5. **Gorski A.I., Maksoutov M.A., Tumanov K.A., Kochergina E.V., Korelo A.M.** Association rules for discovery relationship between mortality among Chernobyl liquidators and radiation dose. *Radiatsiya i risk – Radiation and Risk*, 2018, vol. 27, no. 1, pp. 22-32. (In Russian).

Gorski A.I.* – Lead. Researcher, C. Sc., Tech.; **Maksoutov M.A.** – Head of Dep., C. Sc., Tech.; **Tumanov K.A.** – Head of Lab., C. Sc., Biol.; **Korelo A.M.** – Senior Researcher; **Kochergina E.V.** – Head of Lab., C. Sc., Med.; **Lashkova O.E.** – Researcher; **Ivanov V.K.** – Deputy Director, Chairman of RSCR, Corresponding Member of RAS, D. Sc., Tech. A. Tsyb MRRC.

*Contacts: 4 Korozyov str., Obninsk, Kaluga region, Russia, 249035. Tel.: (484) 399-32-60; e-mail: nrer@obninsk.com.

6. **Gorski A.I., Maksiutov M.A., Tumanov K.A., Vlasov O.K., Kochergina E.V., Zelenskaya N.S., Chekin S.Yu., Ivanov S.A., Kaprin A.D., Ivanov V.K.** Analysis of statistical correlation between radiation dose and cancer mortality among the population residing in areas contaminated with radionuclides after the Chernobyl nuclear power station. *Meditinskaya radiologiya i radiatsionnaya bezopasnost' – Medical Radiology and Radiation Safety*, 2019, vol. 64, no. 6, pp. 5-11.
7. *International Statistical Classification of Diseases and Related Health*, 10th revision (ICD-10). Vol. 1 (Part 1). Geneva, WHO, 1995. 696 p. (In Russian).
8. Software «Statistics». Available at: <http://www.statsoft.ru> (Accessed 18.06.2020).
9. **Mietenen O.S.** Confounding and effect modification. *Am. J. Epidemiol.*, 1974, vol. 100, pp. 350-353.
10. **Ivanov V.K., Gorski A.I., Tsyb A.F., Ivanov S.I., Naumenko R.N., Ivanova L.V.** Solid cancer incidence among the Chernobyl emergency workers residing in Russia: estimation of radiation risks. *Radiat. Environ. Biophys.*, 2004, vol. 43, no. 1, pp. 35-42.
11. **Ivanov V.K., Karpenko S.V., Kashcheev V.V., Chekin S.Yu., Maksiutov M.A., Tumanov K.A., Shchukina N.V., Kochergina E.V., Zelenskaya N.S., Lashkova O.E.** Radiation risks of Russian liquidators of the Chernobyl accident for the period 1992-2017. Part I: Solid cancer incidenceю *Radiatsiya i risk – Radiation and Risk*, 2019, vol. 28, no. 4, pp. 16-30. (In Russian).