

Использование метода бета-подстановки для анализа цензурированных данных об объемной активности радона в воздухе помещений

Д.В. Кононенко

Санкт-Петербургский научно-исследовательский институт радиационной гигиены имени профессора П.В. Рамзаева, Федеральная служба по надзору в сфере защиты прав потребителей и благополучия человека, Санкт-Петербург, Россия

При обработке результатов измерений объемной активности радона, проводимых в рамках радоновых обследований на различных уровнях административно-территориального деления, существует проблема учета результатов, значение которых не превышает нижней границы диапазона измерений использованного типа средств измерений, т.е. задача анализа цензурированных слева данных. В данной работе на практическом примере продемонстрирована возможность применения одного из наиболее точных, но при этом несложных с вычислительной точки зрения методов анализа цензурированных данных — метода бета-подстановки — для анализа массива результатов измерений объемной активности радона в воздухе помещений. Выполнен анализ результатов радонового обследования, проведенного в рамках региональной программы «Радон» в Красносельском районе Санкт-Петербурга в 2000 г. Показано, что грубые подходы к решению обозначенной задачи, такие как отбрасывание цензурированных значений или их замена на значение нижней границы диапазона измерений, могут приводить к существенному смещению оценок параметров или статистических характеристик получаемых распределений, которые в дальнейшем используются при расчете доз внутреннего облучения населения радоном и различных показателей риска для здоровья населения, что приводит к заметному их завышению. Результаты визуального и количественного анализа Q-Q диаграмм позволяют сделать вывод о том, что оцененные параметры распределения, полученного после применения метода бета-подстановки, характеризуются минимальным смещением. Использованный метод бета-подстановки может быть рекомендован для применения в качестве элемента в процедуре обработки результатов радоновых обследований для минимизации смещения оценок параметров распределений значений объемной активности радона, наряду с методом построения и анализа Q-Q диаграмм, применяемым для проверки соответствия полученных распределений логнормальному закону.

Ключевые слова: радон, объемная активность, анализ цензурированных данных, бета-подстановка, логнормальное распределение, параметры распределения, Q-Q диаграмма, оценка риска.

Введение

За последние несколько лет были опубликованы результаты ряда радоновых обследований, проведенных на территории нескольких субъектов РФ с различными целями (см., например, [1–6]). Вне зависимости от того, проводилась ли проверка соответствия распределений результатов измерений логнормальному закону и приведены ли в публикациях параметры полученных распределений, общим моментом для этих работ (а также целого ряда более ранних публикаций) является отсутствие информации о том, какая доля результатов измерений не превысила нижней границы диапазона измерений (НГДИ) использованного типа средств измерений (СИ) и каким образом подобные результаты учитывались при расчете параметров распределений (например, медианы,

совпадающей для логнормального распределения с геометрическим средним) или статистических характеристик (арифметического среднего и квантилей). Следует заметить, что грубые подходы к решению этой задачи, такие как отбрасывание подобных результатов или их замена на значение НГДИ, могут приводить к существенному смещению оценок [7].

Между тем для расчета параметров распределений физических величин, результаты измерений которых часто не превышают НГДИ, существует ряд методов, объединенных общим термином «анализ цензурированных данных» (см., например, [8, 9]). Методы применяются для физических величин, распределение которых носит, как правило, логнормальный характер, а данные при этом цензурированы слева (т.е. присутствуют результаты, зна-

Кононенко Дмитрий Викторович

Санкт-Петербургский научно-исследовательский институт радиационной гигиены имени профессора П.В. Рамзаева.

Адрес для переписки: 197101, Россия, Санкт-Петербург, ул. Мира, д. 8; E-mail: d.kononenko@niirg.ru

чение которых составляет менее нижней границы диапазона измерений). Некоторые методы с успехом применяются даже в том случае, когда распределение является так называемым смешанным [10], т.е. составленным из распределений результатов измерений, полученных с использованием разных методик или типов СИ с отличающимися НГДИ. В этом случае в массиве данных могут присутствовать результаты измерений, лежащие между несколькими НГДИ.

Как известно, распределение значений объемной активности (ОА) радона в воздухе помещений носит именно логнормальный характер [11, 12]. Этот факт был также подтвержден путем анализа массива данных, содержащего более 800 тыс. результатов измерений содержания радона в воздухе помещений в 78 субъектах РФ [13]. Поскольку характеристики и параметры распределений значений ОА радона (например, медиана), получаемых после обработки результатов радоновых обследований, в дальнейшем используются при расчете доз внутреннего облучения населения радоном и различных показателей риска для здоровья населения, представляется весьма важным, чтобы смещение оценок параметров распределений было минимизировано. И одним из способов решения этой задачи является использование в процессе обработки результатов измерений методов анализа цензурированных данных. В данном случае речь идет исключительно о результатах, полученных в рамках спланированных радоновых обследований, когда точно известна методика измерения, тип применявшегося СИ и его НГДИ. Это кардинально отличает подобные результаты от тех, что содержатся в банках данных (БД), задействованных в Единой государственной системе контроля и учета индивидуальных доз облучения граждан (ЕСКИД): БД лабораторий радиационного контроля «ФФ-4» [14], региональных банков данных доз облучения граждан Российской Федерации за счет естественного и техногенно измененного радиационного фона (РБДОПИ) и Федеральном банке данных доз облучения граждан Российской Федерации за счет естественного и техногенно измененного радиационного фона (ФБДОПИ). Подавляющая часть содержащихся в них данных представляет собой результаты измерений содержания радона в воздухе помещений, проводимых в рамках обследований зданий перед сдачей в эксплуатацию или в рамках контрольно-надзорных мероприятий. Однако ключевым недостатком программного обеспечения для ведения этих БД, из-за которого применение методов анализа цензурированных данных становится невозможным, является отсутствие возможности указать, что результат измерения не превышает НГДИ конкретного типа СИ, который также невозможно указать в качестве дополнительной характеристики результата измерения [13].

В отечественных публикациях использование методов анализа цензурированных данных применительно к результатам измерений содержания радона в воздухе или воде ранее описано не было, зарубежные публикации на эту тему немногочисленны [15, 16].

Цель исследования – продемонстрировать на практическом примере возможность применения одного из наиболее точных, но при этом несложных с вычислительной точки зрения методов анализа цензурированных данных – метода бета-подстановки – для анализа массива результатов радонового обследования.

Материалы и методы

Методы анализа цензурированных данных уже не один десяток лет с успехом применяются при анализе результатов измерений содержания различных загрязнителей в воде, пищевых продуктах и других средах. Некоторые методы рекомендованы к применению авторитетными международными организациями, такими как Европейское агентство по безопасности продуктов питания [17] и Всемирная организация здравоохранения [18]. Среди наиболее часто применяющихся можно упомянуть метод оценки максимального правдоподобия (MLE) [19–22], метод лог-пробит регрессии (LPR) [22, 23], метод Каплана – Мейера (К-М) [24, 25], Байесовский метод [26, 27] и наиболее простые методы подстановки $LOD/2$ и $LOD/\sqrt{2}$ [19, 22, 28].

До 2010 г. «золотым стандартом» во многих случаях считался MLE-метод [22, 29], поскольку он давал минимальное смещение оценки и среднеквадратичную ошибку, являющуюся показателем общей точности оценки. Однако данный метод является и наиболее сложным с вычислительной точки зрения. Поиск компромиссного варианта, сочетающего в себе точность и простоту реализации, привел к разработке в 2010 г. метода бета-подстановки [29]. Ряд тестов, проведенных как разработчиками, так и группой независимых специалистов [30, 31], в ходе которых из логнормальных распределений со средним геометрическим, равным единице, и различными геометрическими стандартными отклонениями было сгенерировано от 1000 до 100 000 случайных выборок, показал, что метод бета-подстановки не только превосходит по точности методы подстановки $LOD/2$ и $LOD/\sqrt{2}$, метод Каплана – Мейера и Байесовский метод, но и ничем не уступает MLE-методу при гораздо более простой вычислительной процедуре.

Метод бета-подстановки

При анализе методов подстановки $LOD/2$ и $LOD/\sqrt{2}$, в которых каждое значение ниже НГДИ заменяется произведением НГДИ на $1/2$ и $1/\sqrt{2}$ соответственно, авторы публикации [19] пришли к выводу, что оба этих метода дают несмещенные оценки истинного геометрического среднего и геометрического стандартного отклонения (GSD) только для определенных сочетаний GSD и доли цензурированных значений в массиве данных. Это послужило авторам публикации [29] отправной точкой в поисках некоторого коэффициента β (рассчитываемого на основе нецензурированных данных в массиве), который заменит множители $1/2$ и $1/\sqrt{2}$, что в итоге даст практически нулевое смещение оценок искомых параметров распределения и минимальную среднеквадратичную ошибку.

В зависимости от того, какой параметр распределения оценивается (геометрическое или арифметическое среднее), значение коэффициента β определяется по-разному. После оценки геометрического и арифметического средних метод позволяет оценить GSD и далее квантили распределения. Процедура расчета в целом состоит из 8 шагов, ее несложно реализовать как с помощью любого табличного процессора, так и с помощью языков программирования, встроенных в распространенные программы статистической обработки данных.

Шаг 1. Из исходного массива данных объемом n значений, в котором содержится k цензурированных значений, формируется массив нецензурированных данных объемом $n-k$ значений.

Шаг 2. Рассчитываются входные и промежуточные значения:

$$\bar{y} = \frac{1}{n-k} \sum_{i=1}^{n-k} y_i, \quad (1)$$

где $y_i = \ln(x_i)$, а x_i – i -е значение из массива нецензурированных данных.

$$z = \Phi^{-1} \left[\frac{k}{n} \right], \quad (2)$$

где функция Φ^{-1} дает обратное значение стандартного нормального распределения ($\mu=0$; $\sigma=1$), т.е. z -значение, соответствующее k/n .

$$f(z) = \frac{pdf(z, 0, 1)}{1 - cdf(z, 0, 1)}, \quad (3)$$

$$\hat{s}_y = \frac{\bar{y} - \ln(L)}{f(z) - z} \quad (4)$$

$$f(\hat{s}_y, z) = \frac{1 - cdf(z - \frac{\hat{s}_y}{n}, 0, 1)}{1 - cdf(z, 0, 1)} \quad (5)$$

где $pdf(z, 0, 1)$ – функция плотности вероятности, а $cdf(z, 0, 1)$ – кумулятивная функция распределения стандартного нормального распределения;

\hat{s}_y – предварительная оценка $\ln(GSD)$;

L – значение НГДИ.

Шаг 3. Рассчитывается коэффициент β для оценки арифметического среднего (AM):

$$\beta_{AM} = \frac{n}{k} \cdot cdf(z - \hat{s}_y, 0, 1) \cdot \exp \left[-\hat{s}_y \cdot z + \frac{(\hat{s}_y)^2}{2} \right] \quad (6)$$

Шаг 4. Каждое цензурированное значение в исходном массиве данных заменяется на произведение $L \cdot \beta_{AM}$, после чего по полученному массиву рассчитывается арифметическое среднее.

Шаг 5. Рассчитывается коэффициент β для оценки геометрического среднего (GM):

$$\beta_{GM} = \exp \left[\frac{-(n-k) \cdot n}{k} \cdot \ln(f(\hat{s}_y, z)) - \hat{s}_y \cdot z - \frac{n-k}{2kn} \cdot (\hat{s}_y)^2 \right] \quad (7)$$

Шаг 6. Каждое цензурированное значение в исходном массиве данных заменяется на произведение $L \cdot \beta_{GM}$, после чего по полученному массиву рассчитывается геометрическое среднее.

Шаг 7. Пересчитывается предварительная оценка $\ln(GSD)$ и затем рассчитывается GSD :

$$s_y = \sqrt{\frac{2n}{n-1} \cdot \ln \left(\frac{AM}{GM} \right)} \quad (8)$$

$$GSD = \exp(s_y) \quad (9)$$

В случае, если отношение $AM/GM \leq 1$, что реализуется в случае выборок небольшого объема, в которых результаты близки к НГДИ, значение s_y принимается равным 0, а значение GSD равным 1.

Шаг 8. Рассчитывается 0,95-квантиль распределения:

$$\hat{X}_{0,95} = \exp \left[\ln(GM) - \frac{s_y^2}{2n} + 1,645 \cdot s_y \right] \quad (10)$$

При необходимости расчета квантиля распределения другого уровня значение 1,645 должно быть заменено на значение квантиля стандартного нормального распределения нужного уровня.

Объем выборки при применении метода бета-подстановки должен составлять не менее 5 значений, а доля цензурированных значений не должна превышать 50%.

В случае, если массив исходных данных составлен из результатов измерений, полученных с использованием разных методик или типов СИ с отличающимися НГДИ, в качестве значения L , используемого в шагах 2, 4 и 6, используется значение \bar{L} , рассчитываемое следующим образом:

$$\bar{L} = \exp \left[\frac{1}{\sum m_i} \cdot \sum (m_i \cdot \ln(L_i)) \right], \quad (11)$$

где m_i – количество результатов измерений, цензурированных по i -й НГДИ L_i .

Использование метода бета-подстановки для анализа результатов измерений содержания радона в воздухе было протестировано на массиве результатов радонового обследования, проведенного в рамках региональной программы «Радон» в жилых и общественных зданиях г. Красное Село Красносельского района Санкт-Петербурга в 2000 г. [32]. Объем массива составил 369 значений ОА радона в воздухе помещений, из них ниже НГДИ, равной 25 Бк/м³, – 106 значений (уровень цензурирования составил около 29%).

Из результатов радонового обследования было сформировано три массива значений ОА радона. Отличие массивов состояло только в том, каким образом трактовались значения ниже НГДИ. В Массиве #1 все результаты ниже НГДИ были заменены на само значение НГДИ, в Массиве #2 – просто отброшены, а в Массиве #3 – заменены на результаты расчета с помощью метода бета-подстановки. Следует упомянуть, что авторами отчета [32] был выбран первый способ трактовки цензурированных данных: «Поскольку нас мало интересовали низкие значения ОА радона в воздухе помещений, при анализе распределения этих значений диапазон от 0 до 25 Бк/м³ был просто заменен его граничным значением».

Результаты и обсуждение

В таблице приведены результаты оценки параметров распределений значений ОА радона в каждом из трех описанных массивов, а на рисунках 1–3 – гистограммы частотного распределения и Q-Q диаграммы (диаграммы квантиль-квантиль или нормальные вероятностные графики) значений ОА радона в Массивах #1, #2 и #3 соответственно. Q-Q диаграмма используется для визуальной оценки того, насколько результаты измерений физической величины отклоняются от теоретических значений, полученных из нормального распределения с теми же параметрами, что и для эмпирических данных. В ка-

честве результатов измерений для построения Q-Q диаграмм используются натуральные логарифмы ОА радона (подробнее см. [13, 33]). В случае идеального совпадения результаты измерений и теоретические значения будут находиться на одной прямой. На Q-Q диаграммах, приведенных на рисунках 1–3, точки представляют собой результаты измерений, а прямая построена по теоретическим значениям. Плато из точек на рисунках 1 и 3 представляют собой значительное количество идентичных результатов измерений. По оси ординат отложены значения натурального логарифма ОА радона, а по оси абсцисс – z-значения (т.е. обратные значения стандартного нормального распределения, имеющего среднее, равное нулю, и стандартное отклонение, равное единице).

Результаты оценки параметров логнормальных распределений значений ОА радона по трем проанализированным массивам данных

Таблица

Results of the estimation of parameters of the lognormal distributions of radon concentrations for the three analyzed arrays

[Table

Показатель [Parameter]	Массив #1 [Array #1]	Массив #2 [Array #2]	Массив #3 [Array #3]
Объем массива, значений [Array size]	369	263	369
Геометрическое среднее (95% доверительный интервал), Бк/м ³ [Geometric mean (with 95% confidence interval), Bq/m ³]	73,0 (65,4–81,6)	112,5 (99,7–126,9)	56,3 (48,9–64,8)
Геометрическое стандартное отклонение [Geometric standard deviation]	3,0	2,7	4,1
Арифметическое среднее, Бк/м ³ [Arithmetic mean, Bq/m ³]	153,5	205,3	149,9

Несмотря на то, что для всех трех массивов наблюдается в целом хорошее соответствие распределения значений ОА радона логнормальному закону, практически идеальное совпадение результатов измерений с теоретическими значениями на Q-Q диаграмме для Массива #3 говорит в пользу того, что оцененные параметры именно этого распределения характеризуются минимальным смещением, что обеспечивается использованием метода бета-подстановки для обработки результатов измерений. Результат визуального анализа Q-Q диаграмм подтверждается и расчетами суммы абсолютных значений отклонений натуральных логарифмов измеренных значений выше НГДИ от теоретических значений: для Массива #1 значение этой суммы составляет 54,54, для Массива #2 – 38,43, а для Массива #3 – 15,46.

Использование подхода к анализу цензурированных данных об ОА радона в воздухе помещений, заключаю-

щегося в замене всех результатов ниже НГДИ на значение НГДИ, допустимо только в том случае, если трактовка результатов радонового обследования включает только разделение помещений или зданий на категории соответствующих и не соответствующих санитарным нормам. В иных же случаях, когда рассчитываются средние значения ОА радона для дальнейшей оценки доз внутреннего облучения за счет радона и показателей риска для здоровья населения [34, 35], использование такого подхода, равно как и простое отбрасывание всех результатов ниже НГДИ, представляется недопустимым, т.к. приводит к заметному завышению результатов: в рассмотренном примере завышение геометрического среднего составляет 1,3 раза в случае использования первого подхода и 2 раза в случае использования второго подхода (см. табл.).

Что касается плато, образующихся на Q-Q диаграммах (и соответствующих пиков на гистограммах частотного распределения) за счет присутствия в Массивах #1 и #3 значительного количества идентичных результатов после процедур подстановки, то их устранение не имеет практического значения, поскольку в данном случае (в отличие от рассмотренного в публикации [13]) стоит задача не восстановления формы распределения, а точной оценки его параметров при известном объеме цензурированных данных и известной НГДИ. Тем не менее, причина, по которой распределения оказываются немного усеченными слева, остается неизменной – это применение методик или типов СИ с относительно высокими НГДИ. В частности, для методики измерений, основанной на использовании твердотельных трековых детекторов радона на базе нитратцеллюлозной пленки LR-115 Type II, которая использовалась в рассмотренном радоновом обследовании, значение НГДИ может быть снижено до уровня порядка 14 Бк/м³ при увеличении времени экспозиции до 90 суток и до 7 Бк/м³ – при увеличении до 180 суток.

Следует также отметить, что проверка соответствия полученных результатов измерений логнормальному закону распределения в любом случае должна быть первым шагом в процедуре анализа результатов радонового обследования перед применением (при необходимости) любого метода анализа цензурированных данных, при этом ее вполне допустимо проводить по массиву, в котором результаты ниже НГДИ отброшены (см. рис. 2).

Заключение

Представленные в статье результаты тестового практического применения одного из наиболее точных, но при этом несложных с вычислительной точки зрения методов анализа цензурированных данных – метода бета-подстановки – показывают, что он подходит для использования в качестве элемента в процедуре обработки результатов радоновых обследований для минимизации смещения оценок параметров распределений значений ОА радона и может быть рекомендован для использования вместе с методом построения и анализа Q-Q диаграмм, применяемым для проверки соответствия полученных распределений логнормальному закону. В рассмотренном примере такие подходы к решению задачи учета цензурированных значений, как их замена на значение НГДИ или простое их отбрасывание, приводили к завышению значения геометрического среднего

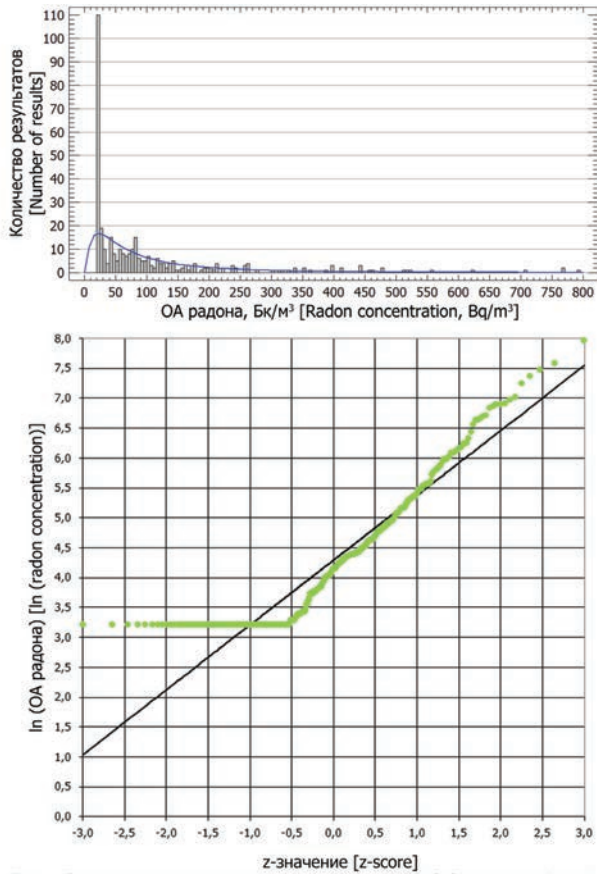


Рис. 1. Гистограмма частотного распределения (вверху) и Q-Q диаграмма (внизу) результатов измерений ОА радона в Массиве #1

[Fig. 1. Frequency distribution histogram (on the top) and Q-Q plot (in the bottom) of radon concentrations in Array #1]

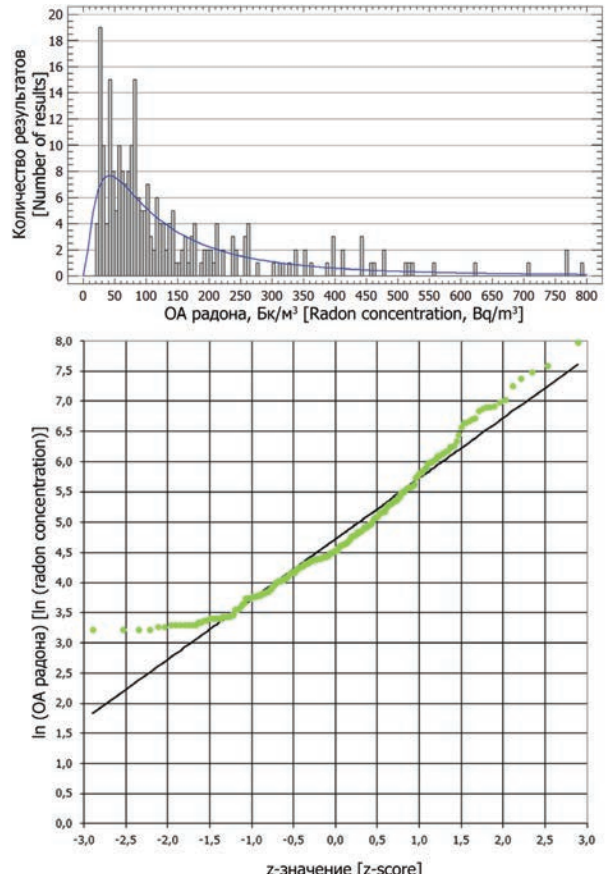


Рис. 2. Гистограмма частотного распределения (вверху) и Q-Q диаграмма (внизу) результатов измерений ОА радона в Массиве #2

[Fig. 2. Frequency distribution histogram (on the top) and Q-Q plot (in the bottom) of radon concentrations in Array #2]

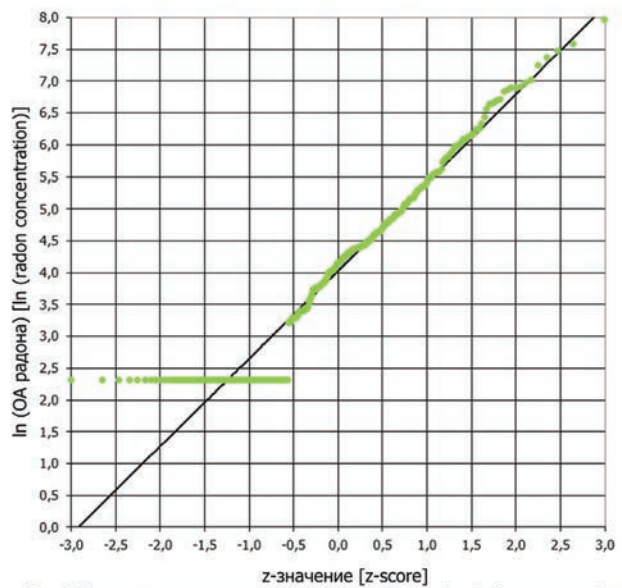
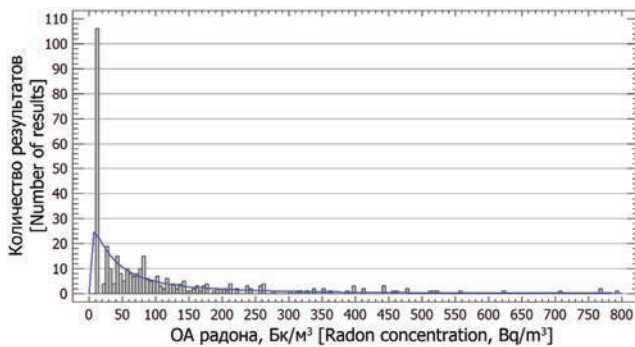


Рис. 3. Гистограмма частотного распределения (слева) и Q-Q диаграмма (справа) результатов измерений ОА радона в Массиве #3
[Fig. 3. Frequency distribution histogram (on the left) and Q-Q plot (in the right) of radon concentrations in Array #3]

распределения в 1,3 и 2 раза соответственно, по сравнению со значением, получаемым после применения метода бета-подстановки. Результаты визуального и количественного анализа Q-Q диаграмм позволяют сделать вывод о том, что оцененные параметры распределения, полученного после применения метода бета-подстановки, характеризуются минимальным смещением.

Реализация Указа Президента РФ от 13.10.2018 г. № 585 «Об утверждении Основ государственной политики в области обеспечения ядерной и радиационной безопасности Российской Федерации на период до 2025 года и дальнейшую перспективу», в котором в качестве одного из основных направлений фигурирует «совершенствование государственного контроля (надзора) за воздействием на здоровье человека природных источников ионизирующего излучения, в том числе радона и продуктов его распада, в жилых домах, детских учреждениях, общественных и производственных зданиях», а также перспектива разработки региональных, а возможно, и национальной радоновых программ и планов действий по радону в рамках постепенного внедрения в практику современных рекомендаций МАГАТЭ, изложенных в Требовании 50 Международных основных норм безопасности [36], неизбежно приведут в ближайшем будущем к росту числа радоновых обследований в России на различных уровнях административно-территориального деления с использованием «контрольно-измерительных приборов для измерения содержания радона и продуктов его распада в жилых помещениях с применением интегральных методов», о чем прямо сказано в Указе. С учетом принятого курса на внедрение в практику санитарно-эпидемиологического надзора методологии оценки риска^{1,2,3,4}, а не простого разделения помещений и зданий на категории соответствующих и не соответствующих санитарным нормам, представляется весьма важным, чтобы при обработке результатов радоновых обследований, являющихся базой

для дальнейшей оценки рисков, применялись методы, позволяющие минимизировать смещение оценок параметров и статистических характеристик распределений значений ОА радона в воздухе помещений. Простота и эффективность описанного в статье метода анализа цензурированных данных должны способствовать его широкому внедрению в практику и, как следствие, повышению достоверности представляемых результатов радоновых обследований, в том числе точности оцениваемых доз внутреннего облучения населения радоном и различных показателей риска для здоровья населения.

Автор выражает признательность Т.А. Балабиной (ФБУН НИИРГ им. П.В. Рамзаева) за предоставленные архивные данные.

Литература

1. Маренный, А.М. Обследование зданий города Лесной на содержание радона в помещениях / А.М. Маренный, М.Е. Киселева, Н.А. Нефедов, Г.И. Сутягина, Л.Э. Карл // Радиационная гигиена. – 2018. – Т. 11, № 3. – С. 92–106.
2. Онищенко, А.Д. Влияние строительных характеристик зданий на накопление радона в детских дошкольных учреждениях Свердловской области / А.Д. Онищенко, А.В. Васильев, Г.П. Малиновский, М.В. Жуковский // Радиационная гигиена. – 2018. – Т. 11, № 2. – С. 28–36.
3. Романович, И.К. Результаты выборочного обследования содержания радона в помещениях детских дошкольных и школьных организаций Ленинградской области / И.К. Романович, И.П. Стамат, Т.А. Кормановская, Т.А. Балабина, Н.А. Королева, О.А. Историк, Л.А. Еремина // Здоровье населения и среда обитания. – 2017. – № 10 (295). – С. 46–49.
4. Поцелуев, Н.Ю. Гигиеническая оценка индивидуальных годовых доз природного облучения населения модельных территорий Алтайского края / Н.Ю. Поцелуев, И.П. Салдан, Б.А. Баландович, О.В. Околелова, С.П. Филиппова // Радиационная гигиена. – 2016. – Т. 9, № 3. – С. 28–35.

¹ Постановление Главного Государственного санитарного врача РФ от 10 ноября 1997 г. № 25 и Главного Государственного инспектора по охране природы от 10 ноября 1997 г. № 03-19/24-3483 «Об использовании методологии оценки риска для управления качеством окружающей среды и здоровья населения в Российской Федерации».

[Resolution of the Chief State Sanitary Inspector of the Russian Federation No. 25 of November 10, 1997 and the Chief State Inspector of Nature Protection No. 03-19/24-3483 of November 10, 1997 “On the use of risk assessment methodology for environmental quality management and public health management in the Russian Federation”. (In Russian)]

² Решение коллегии Федеральной службы по надзору в сфере защиты прав потребителей и благополучия человека от 5 февраля 2010 г. «О внедрении методологии оценки риска для здоровья населения и задачи по ее совершенствованию».

[Decision of the Board of the Federal Service for Surveillance on Consumer Rights Protection and Human Well-Being of February 5, 2010 “On the implementation of a methodology for assessing the risk to public health and the task to improve it”. (In Russian)]

³ Приказ Федеральной службы по надзору в сфере защиты прав потребителей и благополучия человека от 18 февраля 2010 г. № 57 «О реализации решения коллегии Роспотребнадзора от 5 февраля 2010 г. «О внедрении методологии оценки риска для здоровья населения и задачи по ее совершенствованию».

[Order of the Federal Service for Surveillance on Consumer Rights Protection and Human Well-Being No. 57 of February 18, 2010 “On the implementation of the decision of the Board of Rospotrebnadzor of February 5, 2010 “On the implementation of a methodology for assessing the risk to public health and the task to improve it”. (In Russian)]

⁴ Приказ Федеральной службы по надзору в сфере защиты прав потребителей и благополучия человека от 10 марта 2010 г. № 86 «О создании межведомственной рабочей группы по гармонизации гигиенических нормативов».

[Order of the Federal Service for Surveillance on Consumer Rights Protection and Human Well-Being No. 86 of March 10, 2010 “On the creation of an interdepartmental working group on the harmonization of hygienic standards”. (In Russian)]

5. Маренный, А.М. Проведение обследований зданий различного назначения на содержание радона на территориях, обслуживаемых ФМБА России / А.М. Маренный, В.В. Романов, В.И. Астафуров, А.Т. Губин, С.М. Киселев, Н.А. Нефедов, А.В. Пенезев // Радиационная гигиена. – 2015. – Т. 8, № 1. – С. 23–29.
6. Маренный, А.М. Обследование города Краснокаменск на содержание радона в помещениях / А.М. Маренный, С.М. Киселев, А.В. Титов, С.Б. Золотухина, В.И. Астафуров, В.А. Дмитриев, Л.А. Журавлева, М.А. Маренный, Н.А. Нефедов, А.В. Пенезев, Е.А. Хохлова // Радиационная гигиена. – 2013. – Т. 6, № 3. – С. 47–52.
7. Longtin J.P. Occurrence of Radon, Radium, and Uranium in Groundwater. *JAmWaterWorksAssoc.* 1988 Jul;80(7):84–93.
8. Hecht H., Honikel K.O. Assessment of data sets containing a considerable number of values below the detection limits. *Z Lebensm Unters Forsch.* 1995 Dec;201(6):592–7.
9. Vlachonikolis I.G., Marriott F.H. Evaluation of censored contamination data. *Food Addit Contam.* 1995 Sep-Oct;12(5):637–44.
10. Дубнер, П.Н. Англо-русский статистический глоссарий: основные понятия: <http://masters.donntu.org/2002/fvti/spivak/library/book2/book2.htm> (дата обращения: 16.06.2019)
11. Sources and Effects of Ionizing Radiation. UNSCEAR 2000 Report to the General Assembly, with Scientific Annexes. Volume I: Sources. Annex B: Exposures from natural radiation sources. New York: United Nations; 2000, 76 p. ISBN 92-1-142238-8.
12. Effects of Ionizing Radiation. UNSCEAR 2006 Report to the General Assembly with Scientific Annexes. Volume II: Scientific Annexes C, D and E. Annex E: Sources-to-effects assessment for radon in homes and workplaces. New York: United Nations; 2009, 142 p. ISBN 978-92-1-142270-2.
13. Кононенко, Д.В. Анализ распределений значений объемной активности радона в воздухе помещений в субъектах Российской Федерации / Д.В. Кононенко // Радиационная гигиена. – 2019. – Т. 12, № 1. – С. 85–103.
14. Кононенко, Д.В. Типичные ошибки операторов при внесении результатов измерений содержания радона в воздухе помещений в банк данных лабораторий радиационного контроля «ФФ-4» / Д.В. Кононенко, Т.А. Кормановская // Матер. междунар. науч.-практ. конф. «Актуальные вопросы радиационной гигиены», Санкт-Петербург, 23-24 октября 2018 г. – СПб., 2018. – С. 176–178.
15. Barry T.M., Brattin W.J. Distribution of Radon-222 In Community Groundwater Systems: Analysis of Type I Left-Censored Data with Single Censoring Point. *Hum Ecol Risk Assess.* 1998;4(2):579–603.
16. Burmaster D.E., Wilson A.M. Fitting second-order finite mixture models to data with many censored values using maximum likelihood estimation. *Risk Anal.* 2000 Apr;20(2):261–71.
17. European Food Safety Authority. Management of left-censored data in dietary exposure assessment of chemical substances. Scientific report of EFSA. *EFSA Journal* 2010 Jun;8(3):1557–653.
18. GEMS/Food-EURO Second Workshop on Reliable Evaluation of Low-Level Contamination of Food. Report on a Workshop in the Frame of GEMS/Food-EURO (with Addendum, January 2013); 1995 May 26-27; Kulmbach, Federal Republic of Germany, 10 p.
19. Hornung R.W., Reed L.D. Estimation of average concentration in the presence of nondetectable values. *Appl Occup Environ Hyg.* 1990;5(1):46–51.
20. Perkins J.L., Cutter G.N., Cleveland M.S. Estimating the mean, variance, and confidence limits from censored (<Limit of Detection), lognormally distributed exposure data. *Am Ind Hyg Assoc J.* 1990;51(8):416–19.
21. Finkelstein M.M., Verma D.K. Exposure estimation in the presence of nondetectable values: another look. *AIHAJ.* 2001 Mar-Apr;62(2):195–8.
22. Hewett P., Ganser G.H. A comparison of several methods for analyzing censored data. *Ann Occup Hyg.* 2007 Oct;51(7):611–32.
23. Sinha P., Lambert M.B., Trumbull V.L. Evaluation of statistical methods for left-censored environmental data with nonuniform detection limits. *Environ Toxicol Chem.* 2006 Sep;25(9):2533–40.
24. Schmoeyer R.L., Beaucamp J.J., Brandt C.C., Hoffman F.O.Jr. Difficulties with the lognormal model in mean estimation and testing. *Environ Ecol Stat.* 1996 Mar;3(1):81–97.
25. She N. Analyzing censored water quality data using a nonparametric approach. *J Am Water Res Assoc.* 1997 Jun;33(3):615–24.
26. Busschaert P., Geeraerd A.H., Uyttendaele M., Van Impe J.F. Hierarchical Bayesian analysis of censored microbiological contamination data for use in risk assessment and mitigation. *Food Microbiol.* 2011 Jun;28(4):712–9.
27. Paulo M.J., van der Voet H., Jansen M.J., ter Braak C.J., van Klaveren J.D. Risk assessment of dietary exposure to pesticides using a Bayesian method. *Pest Manag Sci.* 2005 Aug;61(8):759–66.
28. Glass D.C., Gray C.N. Estimating mean exposures from censored data: exposure to benzene in the Australian petroleum industry. *Ann Occup Hyg.* 2001 Jun;45(4):275–82.
29. Ganser G.H., Hewett P. An accurate substitution method for analyzing censored data. *J Occup Environ Hyg.* 2010 Apr;7(4):233–44.
30. Huynh T., Ramachandran G., Banerjee S., Monteiro J., Stenzel M., Sandler D.P., Engel L.S., Kwok R.K., Blair A., Stewart P.A. Comparison of methods for analyzing left-censored occupational exposure data. *Ann Occup Hyg.* 2014 Nov;58(9):1126–42.
31. Huynh T., Quick H., Ramachandran G., Banerjee S., Stenzel M., Sandler D.P., Engel L.S., Kwok R.K., Blair A., Stewart P.A. A Comparison of the β -substitution method and a Bayesian method for analyzing left-censored data. *Ann Occup Hyg.* 2016 Jan;60(1):56–73.
32. Выполнение региональной программы «Радон» (Пушкинский и Красносельский районы): отчет заключительный по Договору № 24-02 от 06.06.2000 / СПбНИИРГ Минздрава России; рук. И.П. Стамат; исполн.: Крисюк Э.М., Терентьев Р.П., Королева Н.А., Балабина Т.А. – СПб., 2000. – 69 с.
33. National and Regional Surveys of Radon Concentration in Dwellings. Review of Methodology and Measurement Techniques. IAEA Analytical Quality in Nuclear Applications Series No. 33 (IAEA/AQ/33). Vienna: International Atomic Energy Agency; 2013. 35 p. ISSN 2074–7659.
34. Кононенко, Д.В. Дифференцированный подход к оценке риска для здоровья населения при облучении радоном / Д.В. Кононенко // Радиационная гигиена. – 2017. – Т. 10, № 1. – С. 76–83.
35. Кононенко, Д.В. Оценка риска для здоровья населения субъектов Российской Федерации при равномерном позитивном облучении радоном / Д.В. Кононенко, Т.А. Кормановская // Медико-биологические проблемы жизнедеятельности. – 2019. – № 1(21). – С. 56–61.
36. Радиационная защита и безопасность источников излучения: Международные основные нормы безопасности. Общие требования безопасности. Серия норм безопасности МАГАТЭ, № GSR Part 3. – Вена: МАГАТЭ, 2015. – 477 с. – ISBN 978-92-0-409915-7, ISSN 1020-5845.

Поступила: 25.06.2019 г.

Кононенко Дмитрий Викторович – научный сотрудник лаборатории дозиметрии природных источников Санкт-Петербургского научно-исследовательского института радиационной гигиены имени профессора П.В. Рамзаева Федеральной службы по надзору в сфере защиты прав потребителей и благополучия человека. **Адрес для переписки:** 197101, Россия, Санкт-Петербург, ул. Мира, д. 8; E-mail: d.kononenko@niirg.ru

Для цитирования: Кононенко Д.В. Использование метода бета-подстановки для анализа цензурированных данных об объемной активности радона в воздухе помещений // Радиационная гигиена. – 2019. – Т.12, № 4. – С. 47–55. DOI: 10.21514/1998-426X-2019-12-4-47-55

Application of the β -substitution method for analyzing censored data on indoor radon concentrations

Dmitry V. Kononenko

Saint-Petersburg Research Institute of Radiation Hygiene after Professor P.V. Ramzaev, Federal Service for Surveillance on Consumer Rights Protection and Human Well-Being, Saint-Petersburg, Russia

When analyzing the results of radon surveys, there could be an issue with measurements that are below the limit of detection (LOD) of the measurement technique used. This issue is usually referred to as the management of left-censored data. The paper presents the results of the practical application of one the most accurate and easy to calculate methods for analyzing censored data – the β -substitution method – to the analysis of an indoor radon concentrations dataset. The results of the radon survey conducted in the framework of the regional program “Radon” in Krasnoselskiy district of St. Petersburg in 2000 have been used as the test dataset. It is shown that inappropriate approaches to the management of the censored data, such as discarding the censored values or substituting them with the LOD value, can lead to considerably biased estimates of parameters or statistics of the resulting distributions. Further, this could result in an overestimation of doses and risk estimates which are based on the median values. The visual and quantitative analysis of Q-Q plots leads to conclusion that the estimated parameters of the distribution after the application of the β -substitution method are characterized by a minimum bias. The applied β -substitution method could be recommended for use as an element in the procedure of analyzing the results of radon surveys to minimize the bias of the estimates of the parameters of the radon concentration distributions alongside with Q-Q plots used to verify the conformity of radon concentrations with a lognormal distribution.

Key words: radon, concentration, censored data analysis, beta-substitution, β -substitution, lognormal distribution, distribution parameters, Q-Q plot, risk assessment.

References

- Marennyy A.M., Kiseleva M.E., Nefedov N.A., Sutiagina G.I., Karl L.E. Survey of buildings in the city Lesnoy on the indoor radon concentration. Radiatsionnaya Gygiena = Radiation Hygiene. 2018;11(3):92–106. (In Russian)
- Onishchenko A.D., Vasilyev A.V., Malinovsky G.P., Zhukovsky M.V. Influence of building features on radon accumulation in kindergartens of the Sverdlovsk region. Radiatsionnaya Gygiena = Radiation Hygiene. 2018;11(2):28–36. (In Russian)
- Romanovich I.K., Stamat I.P., Kormanovskaya T.A., Balabina T.A., Koroleva N.A., Istorik O.A., Eremina L.A. Results of sampling analysis of radon content in preschool and school organizations in the Leningrad region. Zdorovye nasele-niya i sreda obitaniya = Public Health and Life Environment. 2017;10(295):46–9. (In Russian)
- Potseluev N.Y., Saldan I.P., Balandovich B.A., Okolelova O.V., Filippova S.P. Annual individual hygienic assessment of natural exposure doses of the Altai territory model areas population. Radiatsionnaya Gygiena = Radiation Hygiene. 2016;9(3):28–35. (In Russian)
- Marennyy A.M., Romanov V.V., Astafurov V.I., Gubin A.T., Kiselev S.M., Nefedov N.A., Penesev A.V. Survey for indoor radon in dwellings on the territories supervised by FMBA of Russia. Radiatsionnaya Gygiena = Radiation Hygiene. 2015;8(1):23–9. (In Russian)
- Marennyi A.M., Kiselev S.M., Titov A.V., Zolotukhina S.B., Astafurov V.I., Dmitriev V.A., Zhuravleva L.A., Marennyi M.A., Nefedov N.A., Penezev A.V., Khokhlova E.A. Survey in Krasnokamensk city on the content of indoor radon. Radiatsionnaya Gygiena = Radiation Hygiene. 2013;6(3):47–52. (In Russian)

Dmitry V. Kononenko

Saint-Petersburg Research Institute of Radiation Hygiene after Professor P.V. Ramzaev.

Address for correspondence: Mira Str., 8, Saint-Petersburg, 197101, Russia; E-mail: d.kononenko@niirg.ru

7. Longtin J.P. Occurrence of Radon, Radium, and Uranium in Groundwater. *J Am Water Works Assoc.* 1988 Jul;80(7):84–93.
8. Hecht H., Honikel K.O. Assessment of data sets containing a considerable number of values below the detection limits. *Z Lebensm Unters Forsch.* 1995 Dec;201(6):592–7.
9. Vlachonikolis I.G., Marriott F.H. Evaluation of censored contamination data. *Food Addit Contam.* 1995 Sep-Oct;12(5):637–44.
10. Dubner P.N. English-Russian glossary on statistics: basic concepts. – Available from: <http://masters.donntu.org/2002/fvti/spivak/library/book2/book2.htm> (Accessed: 16.06.2019) (In Russian)
11. Sources and Effects of Ionizing Radiation. UNSCEAR 2000 Report to the General Assembly, with Scientific Annexes. Volume I: Sources. Annex B: Exposures from natural radiation sources. New York: United Nations; 2000, 76 p. ISBN 92-1-142238-8.
12. Effects of Ionizing Radiation. UNSCEAR 2006 Report to the General Assembly with Scientific Annexes. Volume II: Scientific Annexes C, D and E. Annex E: Sources-to-effects assessment for radon in homes and workplaces. New York: United Nations; 2009, 142 p. ISBN 978-92-1-142270-2.
13. Kononenko D.V. Analysis of distributions of indoor radon concentrations in the regions of the Russian Federation. *Radiatsionnaya Gygiena = Radiation Hygiene.* 2019;12(1):85–103. (In Russian)
14. Kononenko D.V., Kormanovskaya T.A. Typical mistakes made by users while entering indoor radon measurements results into the laboratory database “FF-4”. In: Current issues of radiation hygiene. Materials of the international scientific and practical conference; 2018 Oct 23–24; St.Petersburg, Russia. St.Petersburg; 2018, pp. 176–178. (In Russian)
15. Barry T.M., Brattin W.J. Distribution of Radon-222 In Community Groundwater Systems: Analysis of Type I Left-Censored Data with Single Censoring Point. *Hum Ecol Risk Assess.* 1998;4(2):579–603.
16. Burmaster D.E., Wilson A.M. Fitting second-order finite mixture models to data with many censored values using maximum likelihood estimation. *Risk Anal.* 2000 Apr;20(2):261–71.
17. European Food Safety Authority. Management of left-censored data in dietary exposure assessment of chemical substances. Scientific report of EFSA. *EFSA Journal* 2010 Jun;8(3):1557–653.
18. GEMS/Food-EURO Second Workshop on Reliable Evaluation of Low-Level Contamination of Food. Report on a Workshop in the Frame of GEMS/Food-EURO (with Addendum, January 2013); 1995 May 26–27; Kulmbach, Federal Republic of Germany, 10 p.
19. Hornung R.W., Reed L.D. Estimation of average concentration in the presence of nondetectable values. *Appl Occup Environ Hyg.* 1990;5(1):46–51.
20. Perkins J.L., Cutter G.N., Cleveland M.S. Estimating the mean, variance, and confidence limits from censored (<Limit of Detection), lognormally distributed exposure data. *Am Ind Hyg Assoc J.* 1990;51(8):416–19.
21. Finkelstein M.M., Verma D.K. Exposure estimation in the presence of nondetectable values: another look. *AIHAJ.* 2001 Mar-Apr;62(2):195–8.
22. Hewett P., Ganser G.H. A comparison of several methods for analyzing censored data. *Ann Occup Hyg.* 2007 Oct;51(7):611–32.
23. Sinha P., Lambert M.B., Trumbull V.L. Evaluation of statistical methods for left-censored environmental data with nonuniform detection limits. *Environ Toxicol Chem.* 2006 Sep;25(9):2533–40.
24. Schmoeyer R.L., Beaucamp J.J., Brandt C.C., Hoffman F.O.Jr. Difficulties with the lognormal model in mean estimation and testing. *Environ Ecol Stat.* 1996 Mar;3(1):81–97.
25. She N. Analyzing censored water quality data using a nonparametric approach. *J Am Water Res Assoc.* 1997 Jun;33(3):615–24.
26. Busschaert P., Geeraerd A.H., Uyttendaele M., Van Impe J.F. Hierarchical Bayesian analysis of censored microbiological contamination data for use in risk assessment and mitigation. *Food Microbiol.* 2011 Jun;28(4):712–9.
27. Paulo M.J., van der Voet H., Jansen M.J., ter Braak C.J., van Klaveren J.D. Risk assessment of dietary exposure to pesticides using a Bayesian method. *Pest Manag Sci.* 2005 Aug;61(8):759–66.
28. Glass D.C., Gray C.N. Estimating mean exposures from censored data: exposure to benzene in the Australian petroleum industry. *Ann Occup Hyg.* 2001 Jun;45(4):275–82.
29. Ganser G.H., Hewett P. An accurate substitution method for analyzing censored data. *J Occup Environ Hyg.* 2010 Apr;7(4):233–44.
30. Huynh T., Ramachandran G., Banerjee S., Monteiro J., Stenzel M., Sandler D.P., Engel L.S., Kwok R.K., Blair A., Stewart P.A. Comparison of methods for analyzing left-censored occupational exposure data. *Ann Occup Hyg.* 2014 Nov;58(9):1126–42.
31. Huynh T., Quick H., Ramachandran G., Banerjee S., Stenzel M., Sandler D.P., Engel L.S., Kwok R.K., Blair A., Stewart P.A. A Comparison of the β -substitution method and a Bayesian method for analyzing left-censored data. *Ann Occup Hyg.* 2016 Jan;60(1):56–73.
32. Stamat I.P., Krisyuk E.M., Terentiev R.P., Koroleva N.A., Balabina T.A. Implementation of the regional program “Radon” (Pushkin and Krasnoselsky districts): Final report under Contract No. 24-02 dated 06.06.2000. St.Petersburg: SPbNIIRG, Ministry of Health; 2000, 69 p. (In Russian)
33. National and Regional Surveys of Radon Concentration in Dwellings. Review of Methodology and Measurement Techniques. IAEA Analytical Quality in Nuclear Applications Series no. 33 (IAEA/AQ/33). Vienna: International Atomic Energy Agency; 2013, 35 p. ISSN 2074-7659.
34. Kononenko D.V. A differentiated approach to the risk assessment from exposure to radon. *Radiatsionnaya Gygiena = Radiation Hygiene.* 2017;10(1):76–83. (In Russian)
35. Kononenko D.V., Kormanovskaya T.A. Risk assessment for the population of the regions of the Russian Federation from constant lifelong exposure to radon. *Mediko-biologicheskie problemy zhiznedeyatel'nosti = Medical and biological problems of life activity.* 2019;1(21):56–61. (In Russian)
36. Radiation protection and safety of radiation sources: International basic safety standards. IAEA safety standards series no. GSR Part 3. Vienna: International Atomic Energy Agency; 2014, 436 p. ISSN 1020-525X.

Received: June 25, 2019

For correspondence: Dmitry V. Kononenko – Researcher, Saint-Petersburg Research Institute of Radiation Hygiene after Professor P.V. Ramzaev, Federal Service for Surveillance on Consumer Rights Protection and Human Well-Being (Mira str., 8, Saint-Petersburg, 197101, Russian Federation. E-mail: d.kononenko@niirg.ru) ORCID <https://orcid.org/0000-0002-1392-1226>

For citation: Kononenko D.V. Application of the β -substitution method for analyzing censored data on indoor radon concentrations. *Radiatsionnaya gygiena = Radiation Hygiene*, 2019, Vol.12, No 4, pp. 47–55. (In Russian) DOI: 10.21514/1998-426X-2019-12-4-47-55