

Министерство науки и высшего образования Российской Федерации
Челябинский государственный университет

ОБРАЗОВАНИЕ МАГИСТРОВ: ПРОБЛЕМЫ И ПЕРСПЕКТИВЫ РАЗВИТИЯ

Тезисы докладов VII Всероссийской конференции
«Образование магистров: проблемы и перспективы развития»
18-19 декабря 2023 г., Челябинск

Челябинск
Издательство Челябинского государственного университета
2024

УДК 378.1
ББК Ч48
О-23

Организаторы конференции:

ФГБОУ ВО «Челябинский государственный университет»

Ответственный за выпуск С.А. Белоусова

- О-23 **Образование магистров: проблемы и перспективы развития** : тезисы докладов VII Всероссийской конференции «Образование магистров: проблемы и перспективы развития» (18 -19 декабря 2023 г., Челябинск) / отв. за вып. С.А. Белоусова. Челябинск : Изд-во Челяб. гос. ун-та, 2024. 248 с.

Сборник включает тезисы докладов участников VII Всероссийской конференции «Образование магистров: проблемы и перспективы решения». Рассматриваются вопросы совершенствования механизмов продвижения и «упаковки» образовательных программ магистерской подготовки; развития интердисциплинарности - рождение и реализация программ на стыке наук; анализируются магистратуры глазами обучающихся, ожидания от обучения и управления собственной образовательной траекторией; продвижения идей баланса компетенций Soft-Hard-Digital в контексте внедрения новых образовательных и социальных технологий; выявления лучших практик и разбор острых и проблемных моментов в реализации образовательных программ.

УДК 378.184
ББК Ч48я43

© Авторы

© Челябинский государственный университет

Выводы:

Дозы облучения мягких тканей плотвы формируются в значительной степени ^{137}Cs депонированным в донных отложениях. Величины суммарных доз на головную часть почки у исследованных особей были выше 1 Гр год^{-1} , но не превышали 3 Гр год^{-1} . Дозы облучения костной ткани ребер плотвы сопоставимы с таковыми для головной части почки, однако дозы на позвонки почти в 2 раза превышают дозовую нагрузку на мягкие ткани и могут достигать $\sim 7 \text{ Гр год}^{-1}$.

Список литературы:

1. Тюхай М. В. Влияние хронического радиационного воздействия на состояние кроветворения в головной почке у плотвы (*Rutilus rutilus*) / М. В. Тюхай, Е. А. Пряхин, А. В. Аклеев. // Адаптация биологических систем к естественным и экстремальным факторам среды: материалы VII Международной научно-практической конференции / под ред. Д.З. Шибковой, П.А. Байгужина (г. Челябинск, 11–13 октября 2018 г.) – Челябинск: Издательство ЮУрГГПУ, 2018. – с. 120 – 121. – Текст : непосредственный.
2. Тюхай М.В. Оценка площади мелано-макрофагальных центров в головной части почек у плотвы (*rutilus rutilus*), обитающей в радиоактивно-загрязненных водоемах ТКВ ПО «МАЯК» // Радиобиология. От клетки до биосферы: сборник научных тезисов всероссийской научно-практической конференции молодых ученых, 24 октября 2019 г., г. Челябинск / [редкол.: А.В. Аклеев (отв. ред.) и др.]. – Челябинск. 2019. – с. 53-54. – Текст : непосредственный.
3. Тюхай М.В. Оценка функционального состояния клеток костной и хрящевой ткани у рыб, обитающих в радиоактивно-загрязненных водных экосистемах // Радиобиология. От клетки до биосферы: сборник научных тезисов всероссийской научно-практической конференции молодых ученых, 24 октября 2019 г., г. Челябинск / [редкол.: А.В. Аклеев (отв. ред.) и др.]. – Челябинск. 2019. – с. 68-69. Текст : непосредственный.
4. Иванов А. А. Физиология рыб : учебное пособие для вузов / А. А. Иванов. – под ред. С. Н. Шестах. - Москва : Мир, 2003. – 284 с. - ISBN 5-03-003564-8. – Текст : непосредственный.

УДК 519.25, 378.4

**ПРОБЛЕМА АНАЛИЗА ДАННЫХ С ЦЕНЗУРИРОВАННЫМИ
НАБЛЮДЕНИЯМИ ТИПА «МЕНЕЕ ЧЕМ» В БИОЛОГИИ И ПРОГРАММНОЕ
ОБЕСПЕЧЕНИЕ ДЛЯ ЕЁ РЕШЕНИЯ**

Нохрин Д. Ю.

Челябинский государственный университет, Челябинск, Россия

nokhrin8@mail.ru

В своей практике биологи-исследователи нередко сталкиваются с ситуацией, когда определённые аналитическими методами концентрации интересующих соединений оказываются ниже пределов чувствительности этих методов. В лабораторных журналах и протоколах исследований такие случаи маркируются обычно как «н.ч.м.» (ниже чувствительности метода), «ND» (Not detected) или с указанием конкретной границы, например: «менее $0,2 \text{ пг/мл}$ », « $<0.2 \text{ pg/ml}$ », а при заполнении массивов данных для

последующего анализа часто вносятся как нулевые значения. В медицинской микробиологии концентрации менее 10^3 КОЕ/мл не учитываются как этиологически значимые, а потому точные данные для меньших концентраций просто не получают, хотя технически это возможно. Гистограмма распределения показателя в подобных случаях имеет характерный вид с избытком значений в области нуля (Рис.).

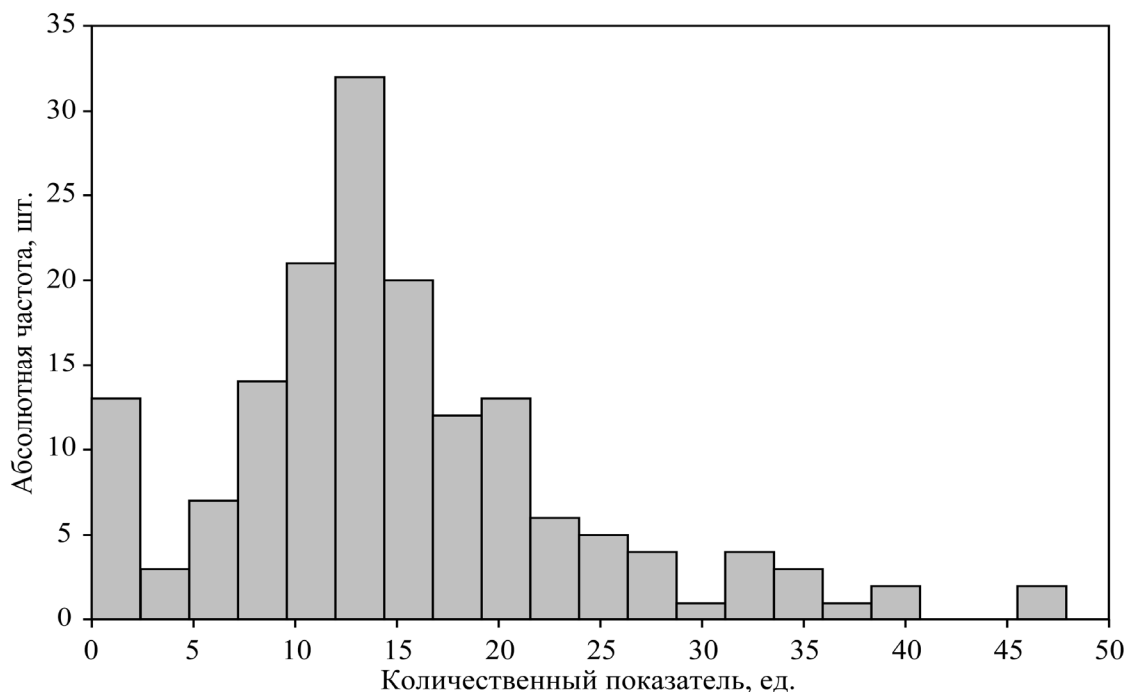


Рис. – Гистограмма типичного распределения биологического показателя в случае наличия цензурированных слева наблюдений

В математической статистике такие неполные данные называются цензурированными наблюдениями (ЦН, censored observations), в данном случае – цензурированными слева, поскольку граница «менее чем» находится в левой части распределения. Анализ публикаций по широкому спектру биомедицинских направлений показывает, что врачи и биологи (как отечественные, так и зарубежные) в своей массе не умеют работать с такими данными и даже не всегда осознают саму проблему. Вместе с тем ясно, что учёт пусть малого, но ненулевого значения в качестве нуля: 1) занижает среднее значение, 2) ухудшает статистические свойства выборки и ограничивает способы их статистической обработки. В некоторых научных коллективах практикуется замена цензурированных наблюдений типа «менее чем» (nondetects) не 0, а $\frac{1}{2}$ порога чувствительности метода. Однако это не решает проблемы (1), поскольку проблема занижения среднего в этом случае может смениться на проблему его завышения; также это не решает проблемы (2), поскольку распределение всё равно остаётся бимодальным, с неизвестными перспективами возможности и качества преобразования данных к более симметричному виду и/или использования ряда распространённых техник анализа данных. Поэтому ясно, что работа с ЦН требует преимущественно особого подхода.

Мы говорим «преимущественно», т.к. бывают ситуации, когда при работе с такими данными возможность применения самых распространённых статистических техник сохраняется. Так, например, если при описании данных используют медианы и квартили, при сравнении выборок – ранговые критерии типа Манна – Уитни или Краскела –

Уоллиса, а при анализе связей – ранговые корреляции Спирмена или Кендалла, то замена ЦН нулём или другим самым малым числом никак не скажется на результате, поскольку оно всё равно получит самый низкий ранг (или самый низкий средний ранг в случае нескольких ЦН). В случае же необходимости получения среднего значения с 95%-ным доверительным интервалом (95% ДИ) или его верхней доверительной границы (upper confidence limit, UCL), моделирования значения показателя регрессионными техниками, использования многомерных техник редукции данных с обобщением и т.п., способ замены ЦН числом будет влиять на результаты и искажать их. Вместе с тем к настоящему времени разработан целый арсенал методов работы с ЦН, хотя используют его главным образом геохимики и гидрогеологи. На примере расчёта среднего значения с 95% ДИ рассмотрим кратко основные принципы таких методов и укажем статистические пакеты с их реализацией.

1. Метод максимального правдоподобия (Maximum Likelihood Estimation, MLE).

Данная группа параметрических методов основана на использовании статистических свойств нецензурированной части данных, скорректированных с учётом теоретических эффектов, привносимых долей цензурированных наблюдений. Теоретически такой подход – самый точный, поскольку задействует в расчётах параметры известного а priori распределения (на практике – нормального, логнормального или гамма-распределения). Однако оценки, полученные с его помощью, являются смещёнными и требуют поправки на смещение, особенно в случае малых выборок (в данном случае с $n < 20$). Различные варианты таких поправок – Коэна (Cohen, 1959, 1960), Перссона и Рутцена (Persson and Rootzen, 1977), Шнейдера (Schneider, 1986) и др. – увеличивают число возможных вариантов MLE-техники и затрудняют сопоставление результатов, полученных в разных пакетах анализа данных. Более существенным ограничением является практическое, поскольку истинное распределение показателя в генеральной совокупности обычно неизвестно, а наличие в выборке даже 1-2 выбросов способно сильно исказить оценку даже в случае $n=50$ [3, 6].

2. Регрессия на порядковой статистике (Regression on Order Statistics, ROS).

Полупараметрический (semi-parametric) метод, основанный на ранжировании всего набора наблюдений и пересчёте его полной части в квантили известного распределения: обычно нормального, логнормального или гамма-распределения. В координатах квантиль-квантильного графика (Q-Q plot) эта часть данных спрямляется и по ней строится линейная регрессия, которая продлевается в область квантилей цензурированной части данных с расчётом для них теоретических значений исходного распределения. Данная «восстановленная» ROS-методом серия убывающих виртуальных значений используется далее в расчётах среднего, дисперсии и 95% ДИ. Поскольку метод задействует как порядковую статистику, так и параметры известного распределения, метод относится к полупараметрическим. В отличие от MLE, данная процедура считается робастной и иногда так и называется: Robust ROS. Различия в программной реализации касаются способа расчёта нормальных меток (normal scores) для соотнесения наблюдаемых значений с квантилями, который может быть проведён по Ван дер Вардену (Van der Waerden, 1952) или по Блому (Blom, 1958), а также использования нормального, логнормального или гамма-распределения. [3, 6]. Робастный метод Хелсела (Helsel's robust method), предложенный в 1990 г., опирался на метод Блома и логнормальное распределение (технически – на нормальное для предварительно логарифмированных по

основанию 10 данных) и более 15 лет считался наиболее точным и предпочтительным при работе с гидрогеохимическими данными.

3. Метод Каплана – Мейера (Kaplan–Meier method). Данная непараметрическая техника является базовой в анализе выживаемости (Survival analysis) – отдельном разделе прикладной статистики, который используется для анализа цензурированных данных в биологии, медицине и промышленности. В ходе анализа выживаемости обычно строится график кривой выживаемости – зависимости доли выживших особей (рецидивов заболевания, работающих приборов и т.п.) от времени, прошедшего с начала испытания. Такой график имеет характерную ступенчатую форму (от 100% в начале и далее – ниже) и используется обычно для графического сопровождения результатов анализа и расчёта медианы выживаемости или работы прибора до отказа. Однако по той же таблице данных можно построить другой график – кумулятивную функцию распределения (Cumulative Distribution Function, CDF). В ней каждому полному наблюдению присваивается соответствующий процентиль, а цензурированному – не присваивается, но его наличие влияет на расчёт процентиля для следующего за ним полного наблюдения. В этом случае площадь под кривой CDF представляет собой непараметрическую оценку среднего значения, рассчитанную с учётом ЦН. Классический анализ выживаемости работает с данными, цензурированными справа, т.е. когда ЦН – самые большие значения типа «более чем», однако его можно развернуть и в обратную сторону – для работы с ЦН «менее чем». На первых порах это делалось с использованием уже имевшихся в статпакетах алгоритмов после предварительного вычитания наблюдаемых данных из большой константы. В современных пакетах метод полностью адаптирован для решения рассматриваемой проблемы, включая как прямой расчёт среднего значения, так и 95% ДИ для него разными способами. Недостатком подхода является достаточно сильное смещение оценок в случаях, когда самое маленькое значение в наборе – ЦН, что как раз чаще всего и наблюдается. Сильные стороны подхода: полная независимость от параметров каких-либо распределений и возможность анализа данных с несколькими уровнями цензурирования [3, 6]. Последнее важно для случаев, когда воедино сводятся данные, полученные по методикам с разной чувствительностью. В наших собственных исследованиях для описания содержащих ЦН выборок мы использовали именно описанный непараметрический метод Каплана – Мейера для расчёта среднего значения, а 95% ДИ находили бутстрэпом (метод процентилей, $n=9999$) для этого же метода [1-2].

В заключении следует отметить, что точность или даже принципиальная возможность использования всех описанных методов работы с ЦН зависит от доли таких наблюдений в наборе данных (её желательно указывать в публикациях). Ясно, что полных наблюдений должно быть больше неполных, однако насколько – зависит от метода и пакета, поэтому в ходе анализа данных с ЦН необходимо ознакомиться с ограничениями используемых методов по прилагаемому руководству или рекомендуемой справочной литературе.

Статистические пакеты. Первым известным нам пакетом для расчёта описательной статистики с учётом ЦН был пакет UnCensor[®], разработанный в Лаборатории экологии реки Саванна университета Джорджии (округ Эйкен, Северная Каролина, США) под руководством Майкла Ньюмена. Вероятно, первые версии пакета появились в промежутке между 1989 и 1995 гг. и были написаны под операционную систему MS DOS. Последняя версия UnCensor (version 4.0) вышла в 1995 г., находилась в

открытом доступе (public domain) и работала под старые версии Windows вплоть до Windows XP. Пакет содержал 7 вариантов из всех 3 описанных выше групп методов, включая робастный метод Хелсела, а также метод винзоризации, основанный на симметричном отсечении как самых малых (в данном случае – ЦН), так и самых больших наблюдений в выборке перед расчётом описательной статистики [6].

В середине 2000-х гг. Агентство по охране окружающей среды США (U.S. Environmental Protection Agency, USEPA) спонсировало разработку пакетов для статистического анализа данных, учитывающих специфику задач организации и получаемых данных. Итогом этой работы стало появление пакета Scout 2008 (version 1.0) и серии современных версий пакета ProUCL (с 1999 г.), основанных преимущественно на многочисленных статистических наработках Аниты Сингх и Джона Носерино с коллегами, а также на написанных при их участии руководящих документах USEPA [7]. Пакет Scout 2008 работает с предварительно заполненными по требуемым правилам таблицами MS Excel и позволяет полноценно обрабатывать данные, в том числе с ЦН типа «менее чем»: рассчитывать средние значения, доверительные, прогнозные и толерантные интервалы, сравнивать выборки, проводить корреляционный и регрессионный анализы, а также анализ главных компонент, дискриминантный анализ, строить графики и др. Для расчёта среднего с 95% ДИ в зависимости от данных используется 8-12 вариантов описанных выше методов, в том числе с возможностью применения ресэмплинг-техники бутстрэпа, а в случае несоблюдения каких-либо требований в результатах появляется соответствующее предупреждение. Сейчас пакет Scout 2008 не поддерживается, однако находится в свободном доступе и может быть скачан из архива USEPA по ссылке: <https://archive.epa.gov/esd/archive-scout/web/html/> вместе с подробнейшим руководством в 3 частях. Он не требует установки (нужно только распаковать архив) и полноценно работает в ОС Windows 10 с файлами MS Excel расширения xls. Пакет ProUCL продолжает развиваться и на текущий момент имеет доступную для свободного скачивания версию 5.2: <https://www.epa.gov/land-research/proucl-software>. Он не содержит большинства методов Scout 2008, в т.ч. не рассчитывает нижние границы ДИ, однако содержит методы и подробные руководства по другим статистическим техникам, рекомендуемых и регламентируемых техническими документами USEPA, знакомство с которыми будет полезным и для отечественных экологов.

В настоящее время методы работы с ЦН «менее чем» ограниченно присутствуют в таких крупных и известных коммерческих статистических пакетах как NCSS и Minitab, однако несравненно большие возможности предоставляет открытая программно-статистическая среда R [8]. Наряду с более чем 20 тыс. бесплатными пакетами в хранилище проекта CRAN содержится 2 специализированных пакета для работы именно с цензурированными слева данными: NADA [5] и NADA2 [4]. Оба пакета основаны на работах одного из пионеров разработки и внедрения методов анализа таких данных в практику – Дэнниса Хелсела (Dennis R. Helsel), доктора наук по окружающей среде и инженерии, автора многочисленных публикаций, учебников, курсов и вебинаров, лауреата Премии за жизненные достижения (Lifetime Achievement Award) Калифорнийской ассоциации ресурсов подземных вод (2018 г.) и Премии за выдающиеся заслуги (Distinguished Service Award) Американской статистической ассоциации (2003 г.). Пакет NADA включает основные методы, описанные во 2-м издании книги «Statistics for Censored Environmental Data using Minitab and R» (2012 г.), а пакет NADA2 разрабатывался

при непосредственном участии Д. Хелсела и включает ряд дополнительных техник (новости о развитии пакета есть на авторском сайте: <https://www.practicalstats.com>). В сочетании с возможностями других пакетов анализа данных и визуализации в R проведение качественной обработки собственных данных с ЦН и презентации результатов возможны на самом современном научном уровне.

Информация о проблеме цензурированных наблюдений типа «менее чем» включена в лекционный курс «Основы биометрического анализа и планирование эксперимента», читаемом студентам 3 курса биологического факультета ЧелГУ направления подготовки 06.03.01 «Биология» программ бакалавриата направленностей: «Биоэкология», «Биофизика», «Генетика», «Микробиология», «Гистология и гистологическая техника», а также в курс «Компьютерные технологии в биологии. Математическое моделирование биологических процессов» для обучающихся 1 курса направления подготовки 06.04.01 программ магистратуры направленностей: «Микробиология и вирусология», «Биотехнология», «Медико-биологические науки», «Гистология», «Биология развития», «Радиационная биология», «Экология», «Генетика». Анализ данных с ЦН типа «менее чем» не входит в обязательный практический курс, но может быть проведён в ходе индивидуальных консультаций, в том числе с применением среды R.

Заключение. Для анализа данных, содержащих цензурированные наблюдения типа «менее чем» существует развитый аппарат статистических техник и программное обеспечение, знакомые преимущественно гидрогеохимикам, но которые могут быть успешно применены в анализе данных биологами и врачами, в том числе – бакалаврами и магистрантами при подготовке выпускных квалификационных работ.

Список литературы:

1. Нохрин, Д. Ю. Гидрохимическая характеристика и качество подземных вод Сосновского района Челябинской области для целей орошения /Д.Ю. Нохрин, Н.А. Давыдова // Аграрный вестник Урала. – 2020. Специальный выпуск «Биология и биотехнологии». С. 56–69. DOI: 10.32417/1997-4868-2021-14-56-69
2. Andreeva, S. V. Variability of the structure of winter microbial communities in Chelyabinsk lakes / S.V. Andreeva, Y.Y. Filippova, E.V. Devyatova, D.Y. Nokhrin // Biosystems Diversity. – 2021. – V. 29, № 4. – P. 311-318. DOI: 10.15421/012139
3. Helsel, D.R. More Than Obvious: Better Methods for Interpreting Nondetect Data / D.R. Helsel // Environmental Science & Technology, 2005. – 39(20). – P. 419A-423A. DOI: 10.1021/es053368a
4. Julian P. NADA2: Data Analysis for Censored Environmental Data, Version 1.1.5 / P. Julian, D. Helsel [Электронный ресурс]. – Режим доступа: <https://cran.r-project.org/web/packages/NADA2/NADA2.pdf> (дата обращения: 20.10.2023). – 66 p.
5. Lee, L. Package 'NADA'. Nondetects and Data Analysis for Environmental Data, Version 1.6-1.1 / L. Lee [Электронный ресурс]. – Режим доступа: <https://cran.r-project.org/web/packages/NADA/NADA.pdf> (дата обращения: 20.10.2023). – 64 p.
6. Newman, M. C. UnCensor[©] v4.0. / M. C. Newman, K. D. Greene, Ph. M. Dixon. – Aiken : Savannah River Ecology Laboratory, 1995. – 94 p.
7. Nocerino, J. Scout 2008 Version 1.0. User Guide Part I. / J. Nocerino, A. Singh, R. Maichle, N. Armbya, A. K. Singh. – Washington: U.S. Environmental Protection Agency, 2008. – 244 p.

8. R Core Team. R: A language and environment for statistical computing. Austria, Vienna: R Foundation for Statistical Computing, 2022 [Электронный ресурс]. Режим доступа: <https://www.R-project.org> (дата обращения: 20.10.2023).

УДК 572+611/612

СООТВЕТСТВИЕ БИОКИНЕТИЧЕСКОЙ МОДЕЛИ ^{90}Sr ПРИЖИЗНЕННЫМ ИЗМЕРЕНИЯМ В ТЕЛЕ ЧЕЛОВЕКА

Терентьева А.М.¹, Шишкина Е.А.^{1,2}

¹Челябинский государственный университет, ²Уральский научно-практический центр радиационной медицины ФМБА России, Челябинск, Россия
nastyaterentyeva@mail.ru, elena.shishkina@gmail.com

Изучение радиационной ситуации, сложившейся в Уральском регионе в результате загрязнения окружающей среды техногенными радионуклидами, включая остеотропный радионуклид ^{90}Sr , проводилось с использованием комплекса методов, среди которых было прижизненное измерение содержания ^{90}Sr во всем теле. Эти измерения проводились в УНПЦ РМ в 1974 – 1996 гг. на уникальном приборе – спектрометре излучения человека (СИЧ-9.1). За эти годы было накоплено около 52000 измерений для более 21000 человек разного возраста и пола. Для многих из этих людей измерения проводились неоднократно в указанный период времени. Это позволило создать биокинетическую модель ^{90}Sr , которая по форме является аналогом камерной модели МКРЗ, однако параметры модели отличаются для лиц разного пола и возраста и подогнаны так, чтобы наилучшим образом соответствовать большому объему экспериментальных данных. Поступление радионуклидов в организм людей, проживающих в прибрежных селах р.Теча, началось в 1950 г., именно поэтому подгоночная модель достоверно предсказывает содержание радионуклидов скелете до 46-ти лет после поступления. Однако, даже спустя 2 и более периодов полураспада ^{90}Sr ($T_{1/2} \sim 28$ лет) у ряда лиц по-прежнему достоверно детектируется содержание радионуклида в организме. Проверить правильность предсказаний модели для содержания ^{90}Sr в скелете спустя ~60 лет после поступления можно путем сопоставления расчетов и измерений содержания ^{90}Sr во всем теле, выполнявшихся в 2006 – 2007 гг. для жителей прибрежных территорий р. Течи.

Цель работы: проверка адекватности биокинетической модели ^{90}Sr УНПЦ РМ для описания содержания радионуклида в скелете через 2 периода полураспада с момента поступления в организм.

Для достижения поставленной цели было необходимо решить следующие задачи:

1. Проанализировать спектры излучения человека, измеренные в 2006 – 2007 гг. и оценить активности ^{90}Sr в организме жителей радиоактивно - загрязненных территорий р. Течи.
2. Для лиц с надежными измерениями в 2006 – 2007 гг. выбрать лиц с множеством повторных измерений, проводившихся в период 1974 – 1996 гг. по архивным данным.
3. Провести биокинетическое моделирование и сопоставить модельные предсказания и измерения, полученные на двух временных интервалах: 1974 – 1996 гг. и 2006 – 2007 гг.