

Научный семинар «Современная демография»

**Российская база данных краткосрочных колебаний смертности
(РосБКС)**

Щур Алексей

*Международная лаборатория исследований населения и
здоровья*

aschur@hse.ru

Москва, 24.10.2023

Предпосылки создания РосБКС:

- **Пандемия COVID-19:** возникшая необходимость перехода от годового цикла разработки и публикации данных о смертности к мониторингу, наиболее приближенному к реальному времени (месяц, неделя);
- Появление базы данных **STMF** и некоторых региональных коллекций данных о внутригодовых колебаниях смертности (Eurostat 2023; Excess and Exiguous Deaths Dashboard in Japan 2023);
- Россия – один из «антилидеров» падения ожидаемой продолжительности жизни в пандемию; наблюдавшаяся еще до ковида существенная географическая дифференциация смертности в стране вкупе с пространственным характером пандемии предопределили большое внимание к данной тематике;
- Влияние внутригодовых колебаний смертности, связанных с краткосрочными факторами риска (например, эпидемии гриппа, жара, или экономический кризис), увеличивается, в частности, в силу старения населения, общего снижения уровня смертности, изменения климата;
- Наличие д а н н ы х .

Команда РосБКС:

STMF, концептуальная рамка
и начало работы над РосБКС



Дмитрий Жданов



Владимир М.
Школьников



Сергей Тимонин

Доработка и публикация базы на сайте
МЛИНЗ в рамках проекта
**«Мониторинг и прогноз ожидаемой
продолжительности жизни на
основе данных оперативной
статистики»**



Алексей Щур



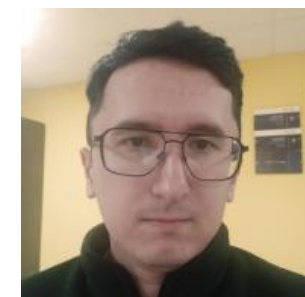
Елена Чурилова



Егор Сергеев



Ольга Родина



Буллат
Шамсутдинов



Вера Соколова

Исходные данные (числа умерших и население под риском):

Источник (числа умерших): **деперсонифицированные индивидуальные данные об умерших** (все выданные ЗАГСами свидетельства о смерти) за 2000-2021 гг. (депозитарий – Росстат)

Data Editor (Browse) - [D2020_modify]

File Edit View Data Tools

terr_zags[1] 4.02990000e+14

	terr_zags	terr_full	terr_resid	year	year_regist	month_regist	year_death	month_death	day_deat
1	4.02990e+14	4.03420e+14	4.03420e+14	2020	2020	4	2020	4	2
2	4.02990e+14	4.03420e+14	4.03420e+14	2020	2020	4	2020	4	
3	4.02990e+14	4.03420e+14	4.03420e+14	2020	2020	4	2020	4	1
4	4.02990e+14	4.03420e+14	4.03420e+14	2020	2020	4	2020	4	2
5	4.02990e+14	4.03420e+14	4.03420e+14	2020	2020	4	2020	4	1
6	4.02990e+14	4.03420e+14	4.03420e+14	2020	2020	4	2020	4	2
7	4.02990e+14	4.03420e+14	4.03420e+14	2020	2020	4	2020	4	2
8	4.02990e+14	4.03420e+14	4.03420e+14	2020	2020	4	2020	4	2
9	4.02990e+14	4.03420e+14	4.03420e+14	2020	2020	4	2020	4	2
10	4.02990e+14	4.03420e+14	4.03420e+14	2020	2020	4	2020	4	
11	4.02990e+14	4.03450e+14	4.03450e+14	2020	2020	4	2020	4	2
12	4.02990e+14	4.03840e+14	4.03840e+14	2020	2020	4	2020	4	2
13	4.02990e+14	4.09010e+14	4.09010e+14	2020	2020	4	2020	4	2
14	4.02990e+14	4.03420e+14	4.03420e+14	2020	2020	4	2020	4	2
15	4.02990e+14	4.03420e+14	4.03420e+14	2020	2020	4	2020	4	2
16	4.02990e+14	4.03420e+14	4.03420e+14	2020	2020	4	2020	4	2
17	4.02990e+14	4.03420e+14	4.03420e+14	2020	2020	4	2020	4	2
18	4.02990e+14	4.03470e+14	4.03470e+14	2020	2020	4	2020	4	2
19	4.02990e+14	4.03420e+14	4.03420e+14	2020	2020	4	2020	4	2
20	4.02990e+14	4.03810e+14	4.03810e+14	2020	2020	4	2020	4	2
21	4.02990e+14	4.03450e+14	4.03450e+14	2020	2020	4	2020	4	2
22	4.02990e+14	4.03260e+14	4.03260e+14	2020	2020	4	2020	4	2
23	4.02990e+14	4.16364e+14	4.02990e+14	2020	2020	4	2020	4	2
24	4.02990e+14	4.03450e+14	4.03450e+14	2020	2020	4	2020	4	2
25	4.02990e+14	4.16484e+14	4.02990e+14	2020	2020	4	2020	4	2
26	4.02990e+14	4.03420e+14	4.03420e+14	2020	2020	4	2020	4	2
27	4.02990e+14	2.85100e+14	4.02990e+14	2020	2020	4	2020	4	2
28	4.02990e+14	1.16484e+14	4.02990e+14	2020	2020	4	2020	4	2

Variables

Filter variables here

Name	Label
☒ terr_zags	full code of ZAGS territory (generated from TerrZAGS)
☒ terr_full	full code of the district (generated from Terr)
☒ terr_resid	full code of the Resid territory (generated from Resid)
☒ year	year of data processing
☒ year_regist	year of event registration
☒ month_regist	month of event registration
☒ year_death	year of death
☒ month_death	month of death
☒ day_death	day of death
☒ date_death	date of death (generated)
☒ dow_death	day of week of death (generated, 0=Sunday,...,6=Saturday)
☒ year_issue	year of issue of death certificate
☒ month_issue	month of issue of death certificate
☒ day_issue	day of issue of death certificate
☒ date_issue	date of issue of death certificate (generated)
☒ dow_issue	day of week of certificate issue (generated, 0=Sunday,...,6=Saturday)
☒ year_birth	year of birth

Variables Snapshots

Properties

Variables

Name	Label	Type	Format
terr_zags	full code of ZAGS territory (generated from TerrZAGS)	double	%12.0g

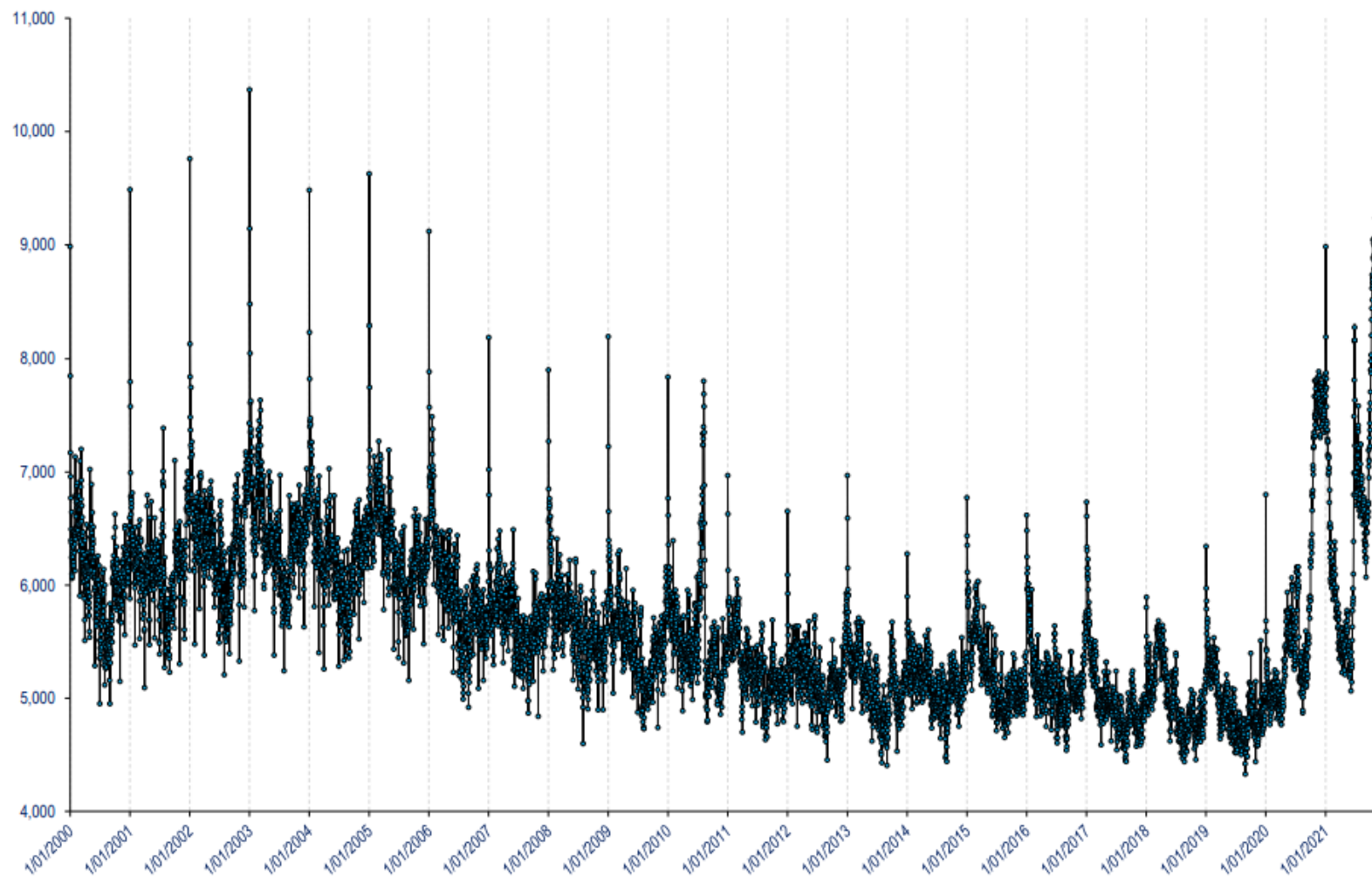
Vars: 58 Order: Dataset Obs: 223 502 Filter: Off Mode: Browse CAP NUM

О каждом умершем известны:

- ✓ Пол;
- ✓ Возраст (полных лет);
- ✓ **Дата наступления смерти (ДД.ММ.ГГ);**
- Дата регистрация смерти (ДД.ММ.ГГ);
- ✓ Регион статистической разработки;
- Причина смерти (Краткая Российская номенклатура Росстата и коды ICD-10);
- Проч. социально-демографические характеристики.

Исходные ланные:

*Ежедневное число умерших в России с 1.1.2000-
31.12.2021*



Преобразование исходных данных (переход от суточных чисел умерших к числам умерших по порядковым неделям, сопоставимым с классификацией ISO):

В РосБКС недели определены по стандарту ИСО 8601:2004 (**ГОСТ Р 7.0.64-2018**). Согласно этому стандарту неделя определяется как «период, состоящий из семи календарных дней, начиная с понедельника».

Номер календарной недели определяется следующим образом: «Порядковый номер, идентифицирующий календарную неделю в рамках текущего календарного года, определяется правилом, согласно которому первая календарная неделя года - это та, которая включает первый четверг текущего года, а последней календарной неделей календарного года является неделя, непосредственно предшествующая первой календарной неделе следующего календарного года».

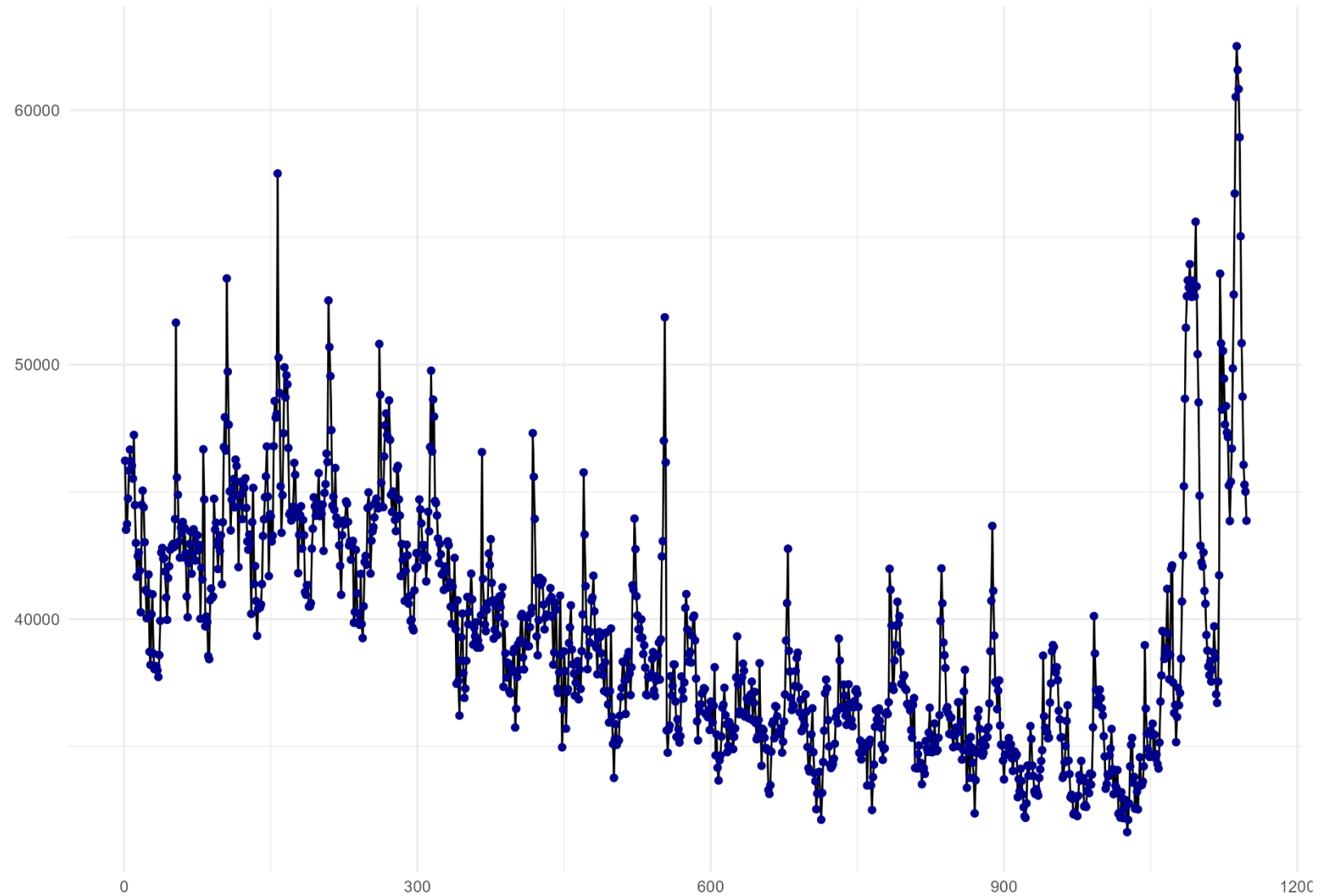
Первая неделя какого-либо календарного года может включать несколько дней (и, соответственно, смертей) из последней недели предыдущего года или не включать несколько дней текущего года (то есть начинаться со 2-го или 3-го января). Соответственно, некоторые годы включают несколько дней предыдущего года (в тех случаях, когда 1 января пришлось на вторник, среду или четверг) и наоборот (если 1 января – пятница-воскресенье).

Год содержит 52 или 53 недели В период с 2000 по 2021 год было 4 года, содержащих 53 недели: **2004, 2009, 2015, 2020**.

В объединенной базе деперсонифицированных данных об умерших за 2000-2021 гг. смерти с неизвестной датой наступления (неизвестен год, месяц или день) составили около 0,3% от всех смертей и были исключены из расчета понедельных возрастных и стандартизованных коэффициентов смертности.

Исходные данные:

*Недельные числа умерших в России, 1-1148 недели
(3.1.2000-31.12.2021/2.1.2022)*



Исходные данные (расчет населения под риском):

Тип данных: среднегодовая численность населения по полу и возрасту в России и ее регионах, 2000-2021;

Источник: Российская база данных рождаемости и смертности (РосБРИС), первоисточник – формы Росстата;

Расчет прожитых человеко-лет жизни (E) на неделе i в году T :

$$E^i = E^T / 365 * 7$$

E^T - средняя численность населения в году T

Расчет (возрастных, общих, стандартизованных) недельных коэффициентов:

$$CDR^i = \frac{D^i}{E^i}, \text{ где}$$

D^i – число смертей на неделе i ,

E^i – число человеко-лет жизни на неделе i .

$$SDR^i = \sum p_x^S m_x^i, \text{ где}$$

p_x^S – веса возрастных группы в стандартном населении (Европейский стандарт населения 2013 года),

m_x^i – возрастные коэффициенты смертности на неделе i в возрасте от x до $x + n$.

Для расчета недельных стандартизованных коэффициентов смертности нами были предварительно посчитаны недельные возрастные коэффициенты смертности по пятилетним возрастным интервалам (0, 1-4, 5-9, ... 95+) как отношение чисел умерших на неделе i в половозрастной группе x к средней численности населения на неделе i в половозрастной группе x :

$$m_x^i = \frac{D_x^i}{E_x^i}, \text{ где}$$

D_x^i – число смертей на неделе i в возрасте от x до $x + n$,

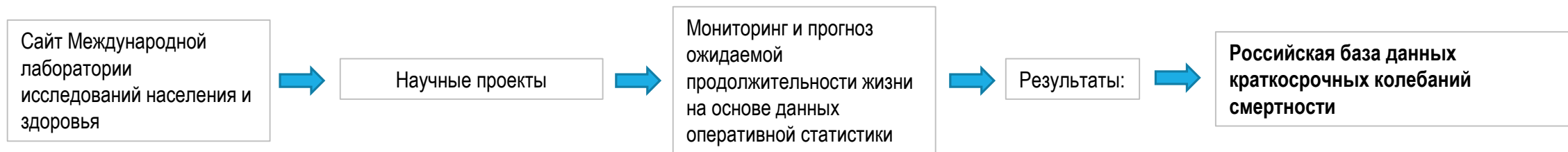
E_x^i – число человеко-лет жизни на неделе i в возрасте от x до $x + n$.

Формат данных и доступ:

Российская база данных краткосрочных колебаний смертности (РосБКС) содержит ряды недельных общих и стандартизованных коэффициентов смертности для мужчин, женщин и обоих полов по России в целом и ее регионам за период с 2000 до 2021 гг. в едином текстовом файле в формате .csv. Данный формат позволяет легко загрузить файл с данными в любую систему статистического анализа.

Данные доступны бесплатно для всех желающих на сайте Международной лаборатории исследований населения и здоровья НИУ ВШЭ - <https://demogr.hse.ru/russtmf/> (Russian Short-Term Mortality Fluctuations Data Series 2023).

В разделе «визуализация» на страничке базы содержится ссылка на набор инструментов визуализации Российской базы данных краткосрочных колебаний смертности - <https://mlinz.shinyapps.io/RosBKS/>.



Цитирование: Российская база данных краткосрочных колебаний смертности. Международная лаборатория исследований населения и здоровья НИУ ВШЭ (Москва, Россия). База данных доступна по адресу <https://demogr.hse.ru/russtmf/> Данные скачаны (дата)

Структура базы данных и описание переменных:

Каждая запись (строка) в файле данных содержит показатели смертности для заданных календарного года, недели, территории, и пола. Если для некоторой комбинации года, пола и территории показатель не может быть рассчитан, то соответствующие показатели в файле данных заменены на "NA". Описание переменных в базе данных приведено в Таблице 1.

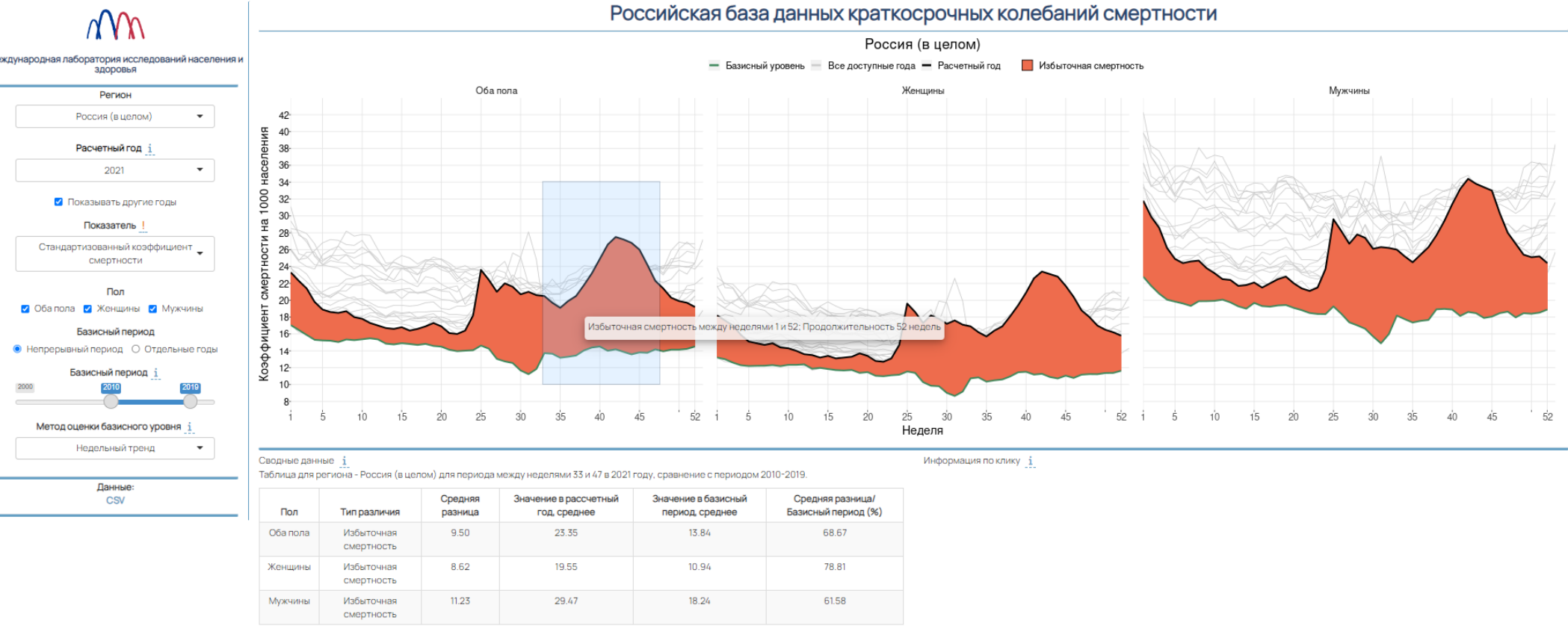
Таблица 1. Переменные в Российской базе данных краткосрочных колебаний смертности (файл «RusSTMD.csv»)

Заголовок столбца	Описание и комментарии
PopCode	Территориальный код. Коды и названия территорий приведены в соотв. файле.
Year	Год
Week	Неделя года
Sex	Пол (F – женский, M – мужской, B – оба пола)
CDR	Общий коэффициент смертности
SDR	Стандартизованный коэффициент смертности (на 1000 населения)

В сравнении с набором переменных в STMF: отказ от публикации абсолютных чисел умерших и **возрастных коэффициентов смертности** (по широким возрастным группам)

Набор инструментов визуализации РосБКС, реализованный, используя R (пакет Shiny):

Адаптация STMF visualization toolkit (Nemeth et al.2020)



При создании системы визуализации РосБКС был взят и адаптирован код, используемый в системе визуализации Short-Term Mortality Fluctuations (STMF visualization toolkit)
The code from the Short-Term Mortality Fluctuations visualization toolkit was taken and adapted when creating the RosBKS visualization system

Методы оценки избыточной смертности, используемые в системе визуализации РосБКС:

Средние за неделю (Week-specific averages)

$$y_{t,i} = \frac{1}{r} \cdot \sum_{h=t-r}^{t-1} \alpha_{h,i} + \varepsilon_{t,i}, i = 1, 2, \dots, n; h = t - r, t - r + 1, \dots, t - 1$$

Для расчетного года h и недели i базисный уровень показателя равен среднему арифметическому наблюдаемых значений за неделю i каждого года, входящего в базисный период

Средние за неделю с годовым трендом (Specific-Average with Trend)

$$y_{t,i} = \frac{1}{r} \cdot \sum_{h=t-r}^{t-1} \alpha_{h,i} + \beta h + \varepsilon_{t,i}, i = 1, 2, \dots, n; h = t - r, t - r + 1, \dots, t - 1$$

Для расчетного года h и недели i базисный уровень показателя равен среднему арифметическому наблюдаемых значений за неделю i каждого года, входящего в базисный период, с дополнительной поправкой на годовой линейный тренд

Недельный тренд (Specific-Trend)

$$y_{t,i} = \beta_i h + \varepsilon_{t,i}, i = 1, 2, \dots, n; h = t - r, t - r + 1, \dots, t - 1$$

Для расчетного года h и недели i базисный уровень показателя является ожидаемым значением линейной регрессионной модели, оцененной методом наименьших квадратов на основе наблюдаемых значений за неделю i каждого года, входящего в базисный период

Средняя за год (Yearly average week)

$$y_{t,i} = \frac{\sum_{i \in W} y_{t,i}}{|W|}$$

Для расчетного года h и недели i базисный уровень показателя равен среднему арифметическому наблюдаемых значений за все недели каждого года, входящего в базисный период

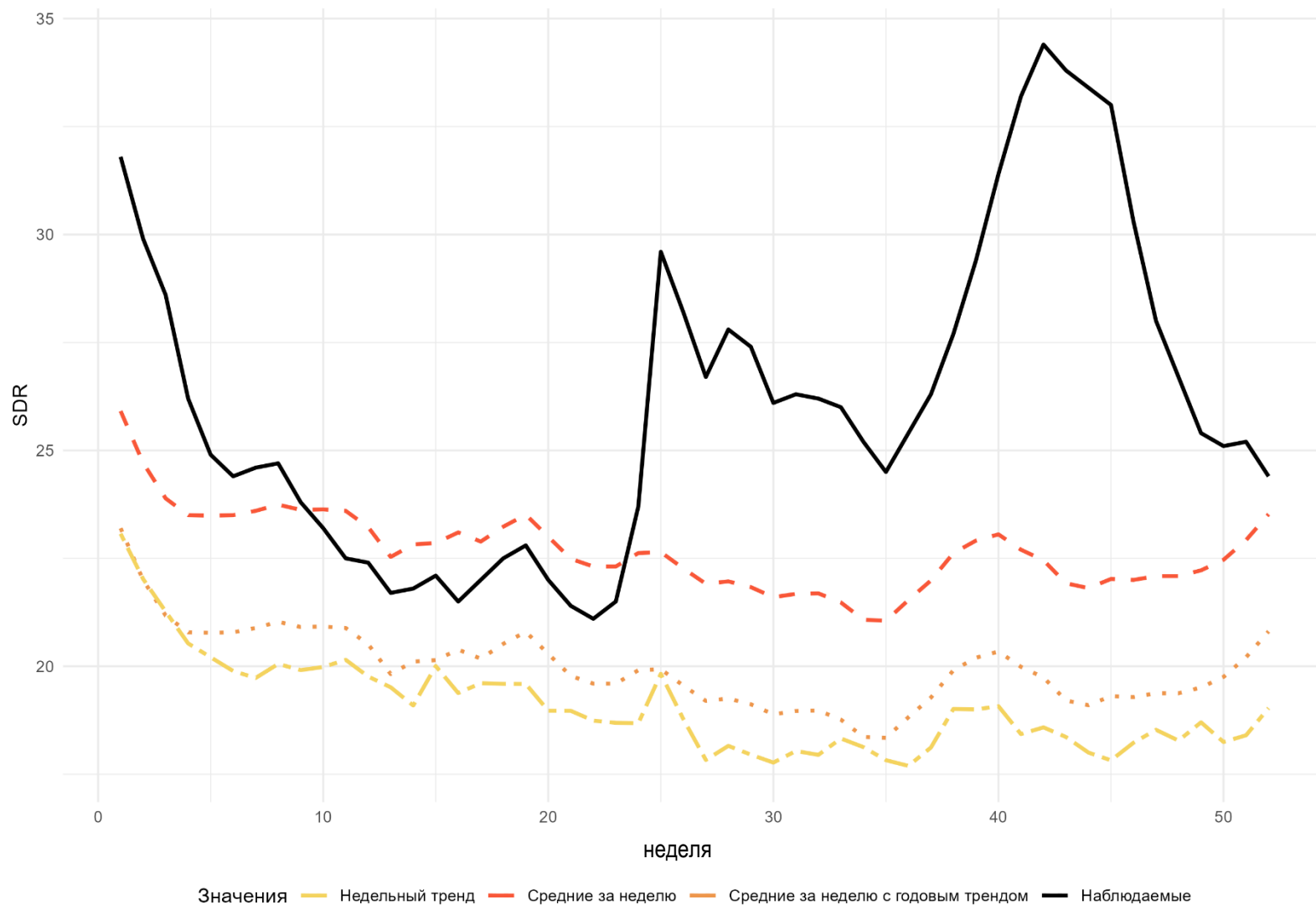
Годовой тренд с учетом сезонности (Harmonic with Trend)

$$y_{t,i} = \beta h + \delta \sin \left(2\pi \frac{i}{n} \right) + \eta \cos \left(2\pi \frac{i}{n} \right) + \theta \sin \left(2\pi \frac{i}{\left(\frac{n}{2}\right)} \right) + \nu \cos \left(2\pi \frac{i}{\left(\frac{n}{2}\right)} \right) + \varepsilon_{t,i},$$

Для расчетного года h и недели i базисный уровень показателя является ожидаемым значением линейной регрессионной модели на основе наблюдаемых значений за каждый год, входящий в базисный период, с учетом наблюдаемой сезонности внутри базисных лет

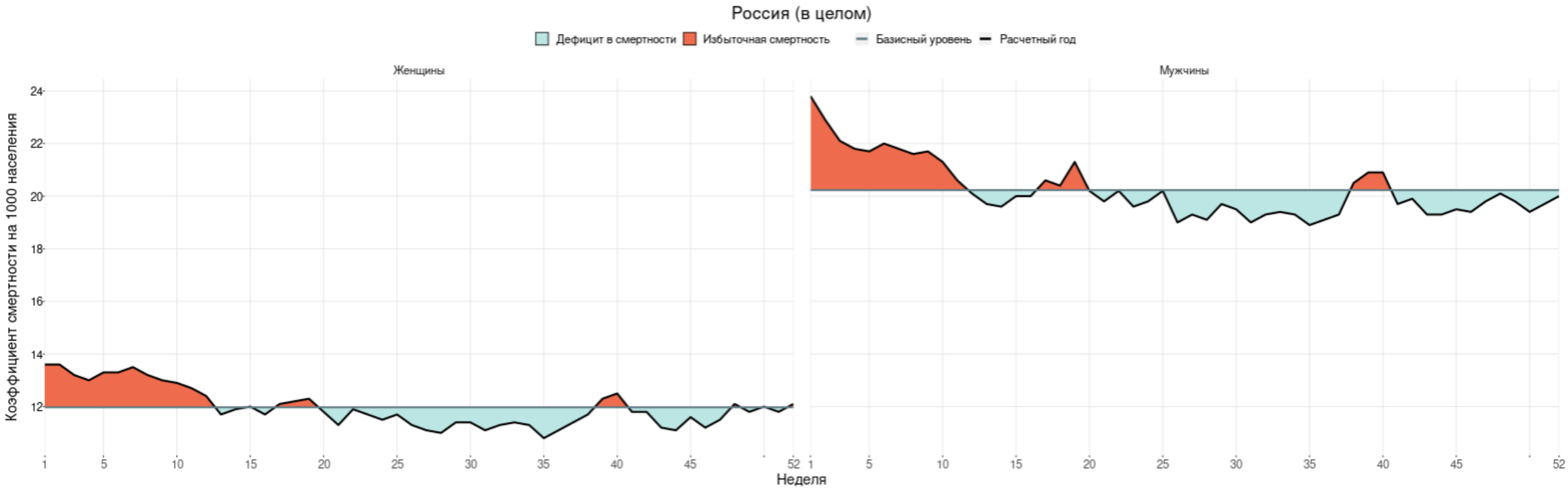
Методы оценки избыточной смертности, используемые в системе визуализации РосБКС:

Россия, мужчины, 2021 год (референтный период – 2011-2019 гг.)



Методы оценки избыточной смертности, используемые в системе визуализации РосБКС:

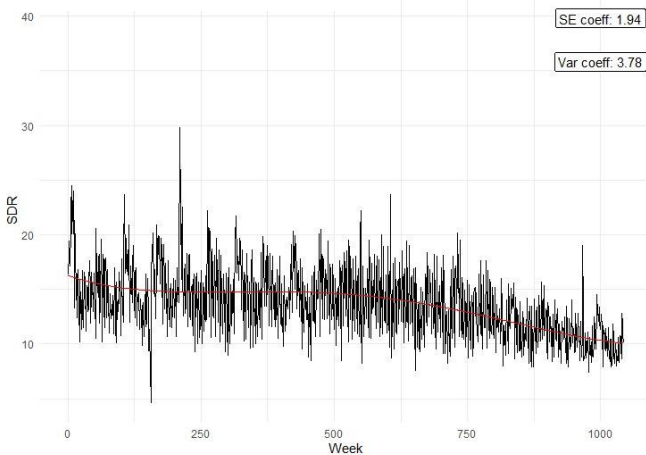
Российская база данных краткосрочных колебаний смертности



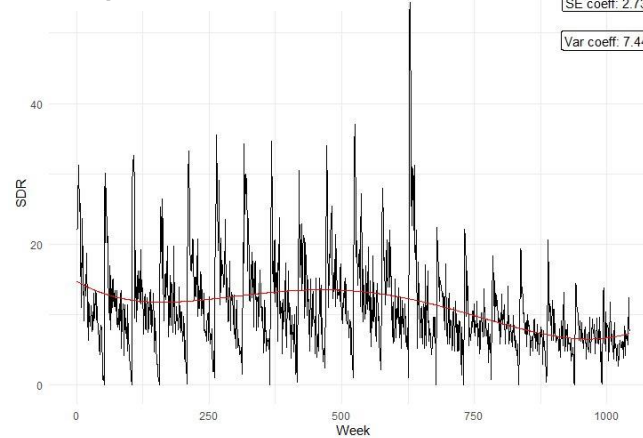
Степень устойчивости недельных коэффициентов смертности в небольших популяциях (и возможные искажения в данных):

- Малые числа умерших (случайные колебания);
- Возможные искажения в датах смерти

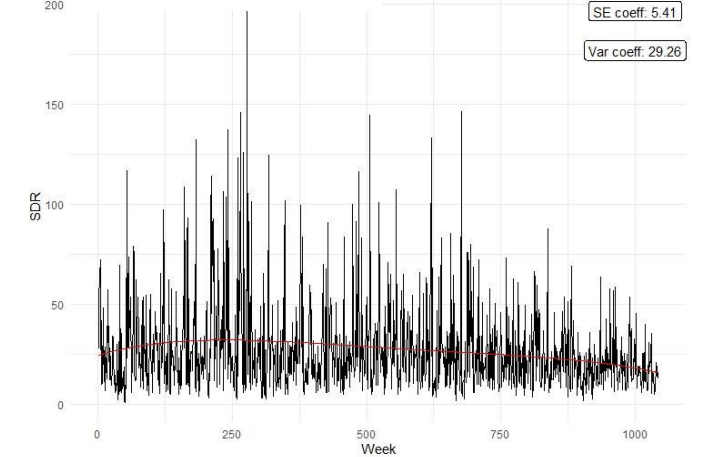
Дагестан



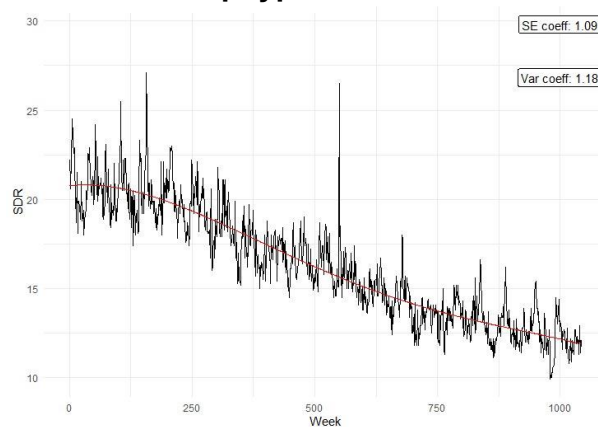
Ингушетия



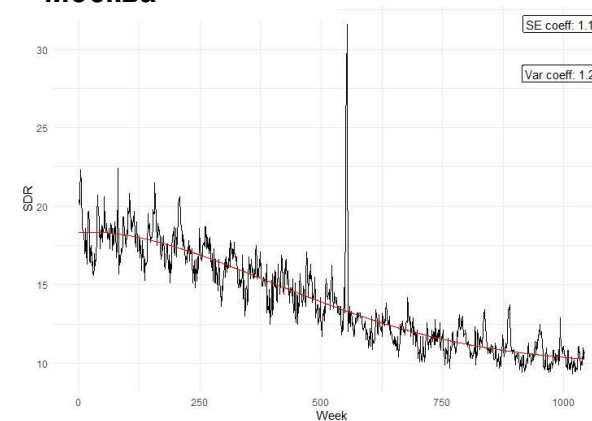
Чукотка



Санкт-Петербург



Москва



Степень устойчивости недельных коэффициентов смертности в небольших популяциях (и возможные искажения в данных):

Отношение коэффициента вариации несглаженной и сглаженной кривой недельных СКС-в, 2000-2019, по регионам (без Чукотки)



68	Республика Дагестан	1,76
69	Республика Адыгея	1,78
70	Мурманская область	1,81
71	Сахалинская область	1,82
72	Карачаево-Черкесская Республика	1,83
73	Республика Крым	1,95
74	Севастополь	2,33
75	Республика Тыва	2,49
76	Республика Калмыкия	2,68
77	Ингушская республика	2,88
78	Чеченская республика	3,02
79	Республика Алтай	3,11
80	Камчатский край	3,22
81	Еврейская автономная область	3,22
82	Магаданская область	3,39
83	Чукотский автономный округ	7,68



68	Кабардино-Балкарская Республика	1,78
69	Орловская область	1,80
70	Республика Адыгея	1,90
71	Камчатский край	1,93
72	Республика Калмыкия	1,94
73	Карачаево-Черкесская Республика	1,97
74	Республика Алтай	2,08
75	Республика Дагестан	2,28
76	Республика Тыва	2,34
77	Республика Крым	2,35
78	Чеченская республика	2,37
79	Еврейская автономная область	2,55
80	Магаданская область	2,71
81	Ингушская республика	2,95
82	Севастополь	3,52
83	Чукотский автономный округ	7,13

В визуализации не представлены следующие регионы: те, в которых в 2000-2021 гг. среднее недельное число умерших было меньше 100 и/или соотношение коэффициентов вариации сглаженных и не сглаженных СКС-ов превышает >1,5)

Кабардино-Балкарская Республи	1183
Орловская область	1154
Республика Адыгея	1179
Карачаево-Черкесская Республи	1191
Республика Калмыкия	1185
Камчатский край	1130
Республика Тыва	1193
Республика Алтай	1184
Республика Дагестан	1182
Республика Крым	1135
Еврейская автономная область	1199
Севастополь	1167
Магаданская область	1144
Чеченская республика	1196
Ингушская республика	1126
Чукотский автономный округ	1177

Идеи по дальнейшему развитию РосБКС:

- Использование данных оперативной статистики для прогнозирования ожидаемой продолжительности жизни в текущем году (разработка методологии и апробация на ретроспективных данных);
- Актуализация данных – добавление 2022 и 2023 гг.(?);
- Продвижение РосБКС и использование данных для ретроспективного анализа (space timing of COVID-19, heat waves, многолетние паттерны внутригородовой сезонности: географический аспект).



В. Поленов, диорама «Елка»