

Стохастические модели предсказания динамики вирусной популяции

Неверов А.Д.

ФКН 2024

Необходимые для понимания моделей основы теории вероятностей

СЛУЧАЙНЫЕ ВЕЛИЧИНЫ И ИХ РАСПРЕДЕЛЕНИЯ

Распределение Бернулли и производные распределения

Распределение Бернулли:

$$X=\{0,1\}, P(X=1) = p, P(X=0) = 1-p, E[X] = p, Var[X] = p(1-p)$$

Биномиальное распределение: $Y \sim Bin(p, n)$

$$Y = \sum_{i=1}^n X_i \quad P(Y = k \mid n, n \geq k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$E[Y] = np \quad Var[Y] = np(1-p) \quad Var[Y] < E[Y]$$

Геометрическое распределение: $Y \sim Geom(p)$

Y - время **до** первой "1"

$$\underbrace{0,0,\dots,0}_k, 1 \quad P(Y = k) = p(1-p)^k, k = 0,1,\dots$$

$$E[Y] = \frac{1-p}{p} \quad Var[Y] = \frac{1-p}{p^2}$$

Отрицательное биномиальное распределение $Y \sim NegBin(p, r)$

Y – число нулей к появлению r -й “1”

$$X_i \sim Geom(p), i = 1..r \quad Y = \sum_{i=1}^r X_i$$

$\underbrace{0, 1, 0, 0, \dots, 1, 1}_{\text{"0" } k\text{-times, "1" } r\text{-times}}$

NegBin - распределение с
овердисперсией

$$E[Y] = \frac{r(1-p)}{p} \quad Var[Y] = \frac{r(1-p)}{p^2} \quad Var[Y] > E[Y]$$

$$P(Y = k \mid p, r) = \binom{k+r-1}{r-1} p^r (1-p)^k$$

Пример из эпидемиологии

$$Y \sim \text{NegBin}(p, r)$$

Коррекция обрезки временного ряда справа:

$O_i, i = 1..t$ - число новых случаев заболевания зарегистрированных в день i , t – текущий день от начала эпидемии

$k \sim G$ G – закон распределения задержек в днях от выявления заболевания до учета в статистике

Сколько случаев Y_{t-k} заболевания в день $t-k$ не попали в статистику?

$$Y_{t-k} \sim \text{NegBin}(p_k, O_{t-k} + 1), k = 0, \dots, t-1$$

$p_k = F_G(k)$ - функция распределения вероятности (ФРВ)

Мы рассматриваем зарегистрированные случаи заболевания случившиеся в день $t-k$, как успешные реализации бернулиевской случайной переменной из серии истинных случаев заболевания

Распределение Пуассона

$$Y \sim \text{Pois}(\lambda)$$

$$Y \sim \text{NegBin}(p, r) \quad E[Y] = \frac{r(1-p)}{p} \xrightarrow{r \rightarrow +\infty} \lambda$$

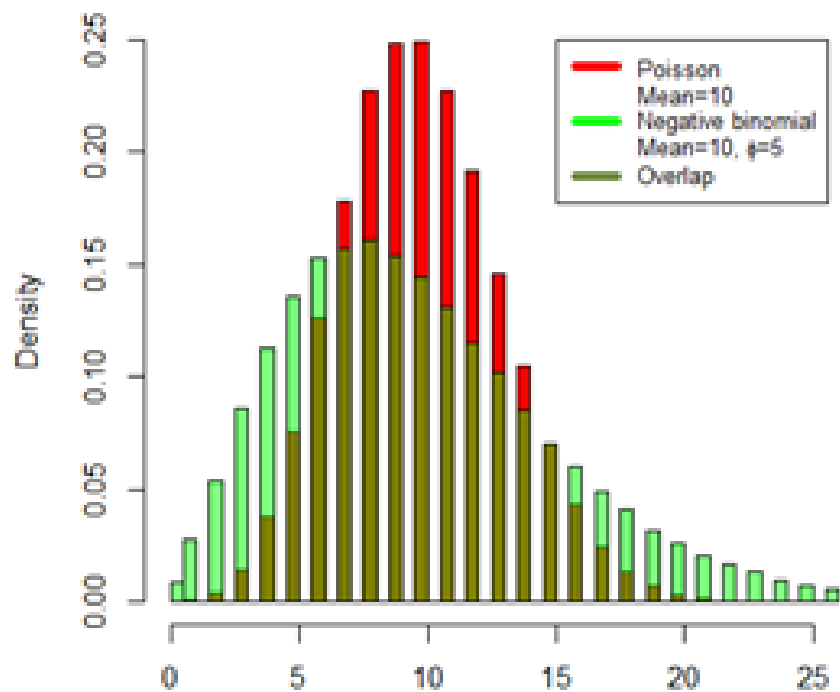
$$p = \frac{1}{1 + \frac{\lambda}{r}} \quad \text{Var}[Y] = \frac{r(1-p)}{p^2} = \lambda \left(1 + \frac{\lambda}{r}\right) \xrightarrow{r \rightarrow +\infty} \lambda$$

$$P(Y = k \mid p, r) = \frac{k+r-1}{r-1} p^r (1-p)^k = \frac{(k+r-1)!}{(r-1)!(r+\lambda)^k} \frac{1}{\left(1 + \frac{\lambda}{r}\right)^r} \frac{\lambda^k}{k!} \xrightarrow{r \rightarrow +\infty}$$

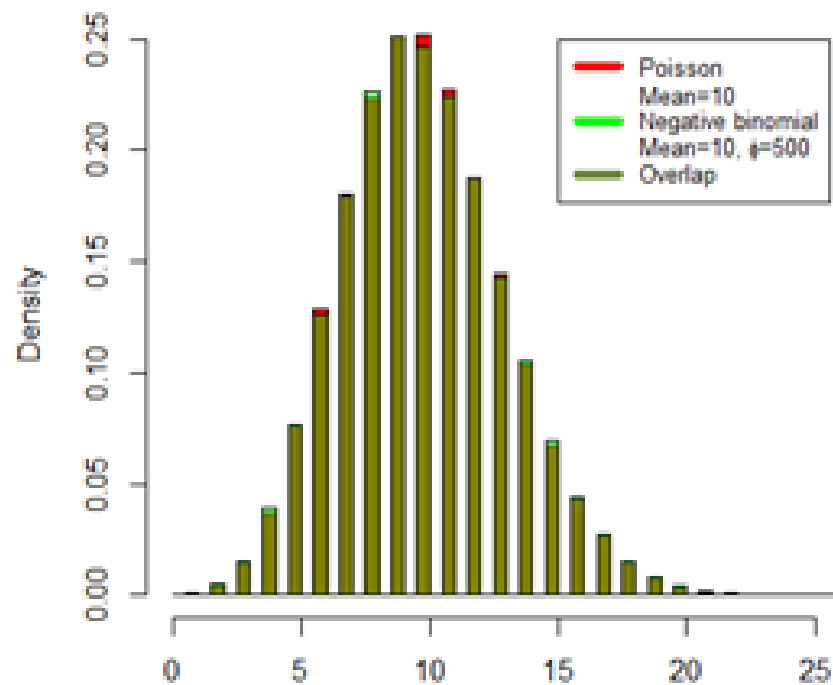
$$P(Y = k \mid \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}, k = 0, \dots, +\infty$$

$$\text{Var}[Y] = E[Y] = \lambda$$

Probability mass function



Probability mass function



Распределение Пуассона

$$Y \sim Pois(\lambda)$$

$$P(Y = k) = \frac{\lambda^k e^{-\lambda}}{k!}, k = 0, \dots, +\infty \quad Var[Y] = E[Y] = \lambda$$

$$P_{Bin}(Y = k \mid p, n) \xrightarrow{n \rightarrow +\infty, np \rightarrow \lambda} P_{Pois}(Y = k \mid \lambda)$$

Базовые предположения:

1. Y – число событий в интервале наблюдения
2. События независимы
3. λ – не зависит от появления событий, но может зависеть от времени
4. В каждый момент времени может возникнуть не более одного события

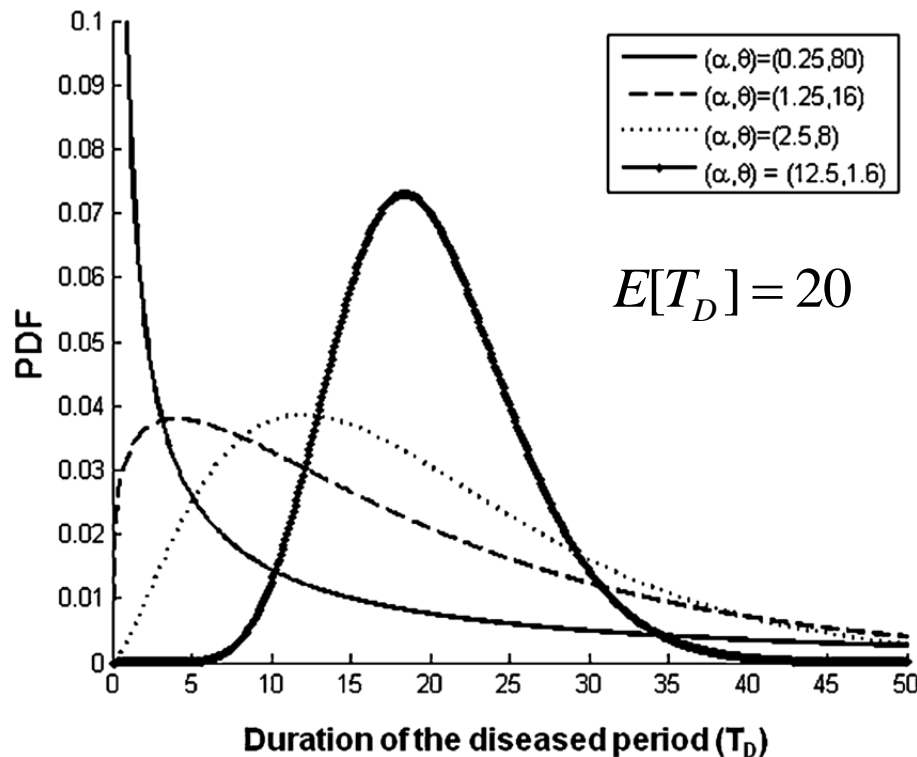
Гамма распределение

$$Y \sim \text{Gamma}(\alpha, \theta)$$

PDF: $f(x, \alpha, \theta) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-\frac{x}{\theta}}, x > 0, \alpha, \theta > 0$

α - shape
 ϑ - scale

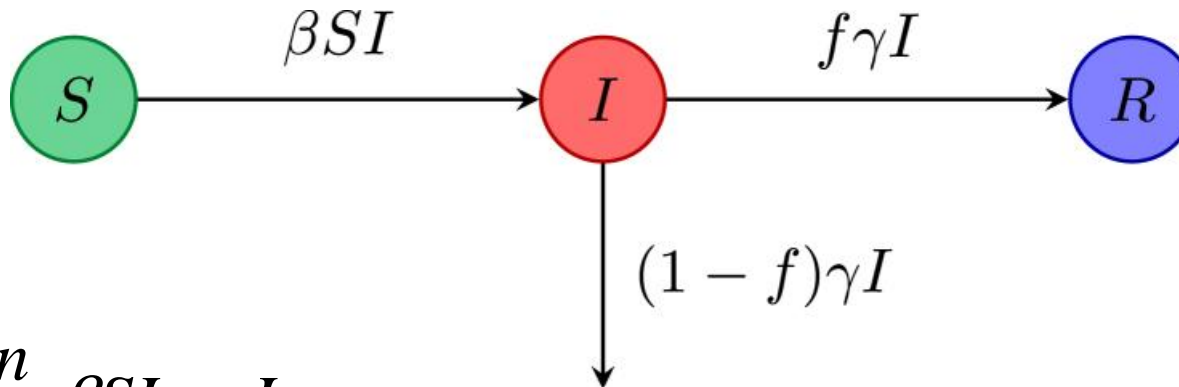
$$E[X] = \alpha\theta \quad V[X] = \alpha\theta^2$$



В качестве примера из эпидемиологии

РЕПРОДУКТИВНОЕ ЧИСЛО И РАСПРЕДЕЛЕНИЕ ГЕНЕРАТИВНЫХ ИНТЕРВАЛОВ

SIR



$$\frac{dI}{dt} = \frac{n}{N} \beta SI - \gamma I$$

$$\frac{dS}{dt} = -\frac{n}{N} \beta SI$$

$$\frac{dR}{dt} = f\gamma I$$

$$r_0 = \frac{n\beta S_0}{N\gamma}$$

n – среднее число
контактов
 N – число индивидов
 $N = S_t + I_t + R_t$

r_0 определяет показатель эксп. роста
и конечный размер эпидемии

$$I_t = I_0 \exp(\gamma(r_0 - 1)t), \quad r_0 > 1$$

$$R_\infty = N - S_0 e^{-r_0 \frac{R_\infty}{N}}, \quad f = 1$$

Базовое репродуктивное число

- r_t – репродуктивное число (РЧ) – среднее число новых инфекций, вызванных одной инфекцией
- r_0 – базовое репродуктивное число – РЧ ($r_{t=0}$) в полностью восприимчивой популяции

нормируем на размер популяции:

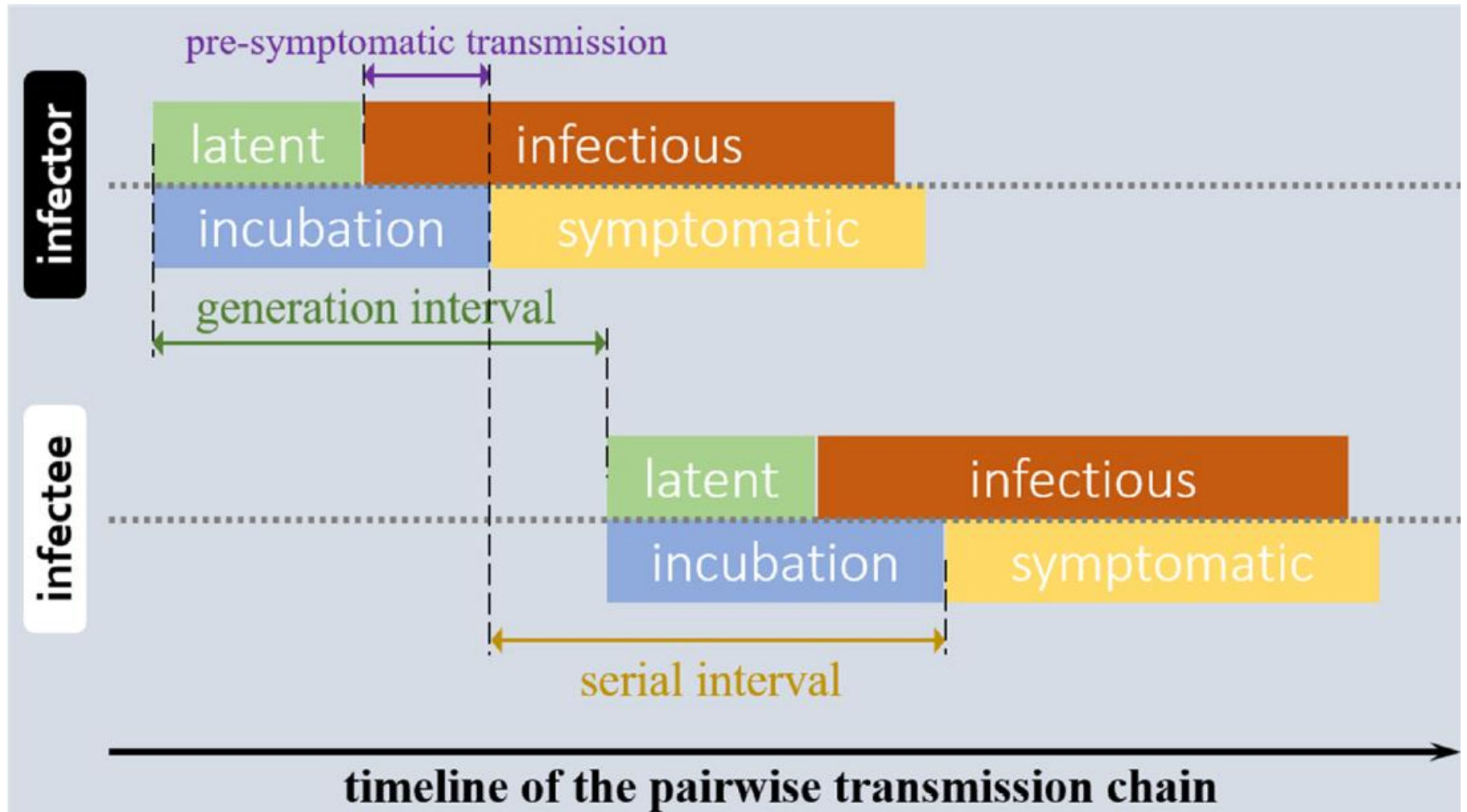
$$S(t) \equiv \frac{S(t)}{N}$$

$$r_t = r_0 S(t)$$

Стадный иммунитет (HIT, herd immunity threshold) – доля иммунизированной популяции, такой что $r_t=1$

$$S(t) = 1 - HIT \Rightarrow r_t = 1 \qquad HIT = 1 - \frac{1}{r_0}$$

Цепочка передачи инфекции



Пример из эпидемиологии

$$Y \sim \text{Pois}(\lambda)$$

Оценка динамики репродуктивного числа r_t по заболеваемости

Число новых случаев инфицирования в день i , t – текущий день от начала эпидемии

$$S \sim g$$

g – Распределение генерационных интервалов - распределение времени между последовательными инфицированиями в цепочках передачи.

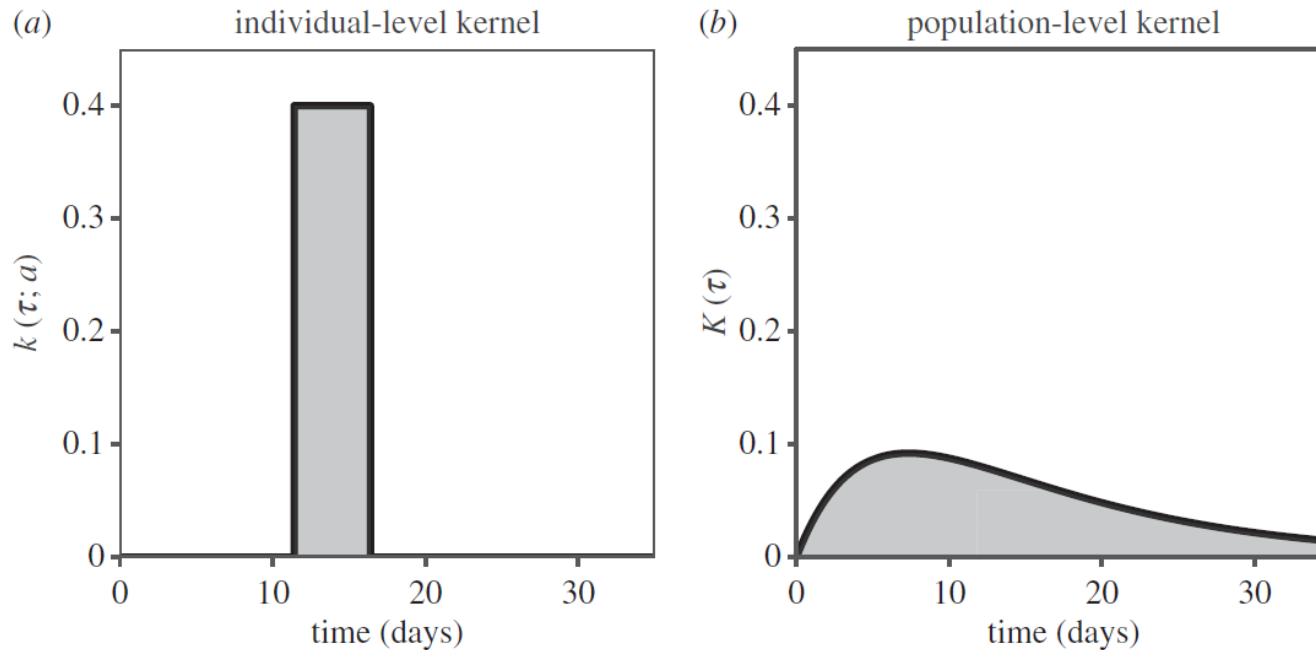
$$g_s = \int_{s-1}^s f_g(\tau) d\tau \quad f_g(s) \quad - \text{ функция плотности вероятности (ФПВ)}$$

$$\Lambda_t = \sum_{s=1}^t g_{t-s} I_{t-s} \quad \text{Инфекционный потенциал – число активных носителей}$$

$$E[I_t] = \Lambda_t r_t \quad \text{Ожидаемое число новых инфекций в момент времени } t, \text{ где } r_t - \text{ репродукционное число}$$

$$I_t \sim \text{Pois}(\Lambda_t r_t)$$

Интенсивность инфекционных контактов индивида



а) Индивидуальная реализация скоростей инфекционных контактов с латентным периодом 11.5 дней и инфекционным периодом – 5 дней.

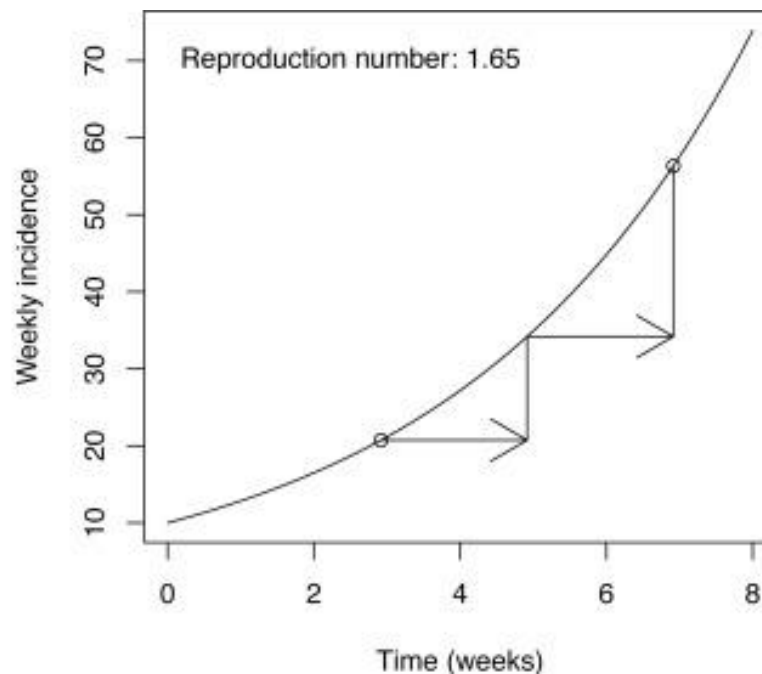
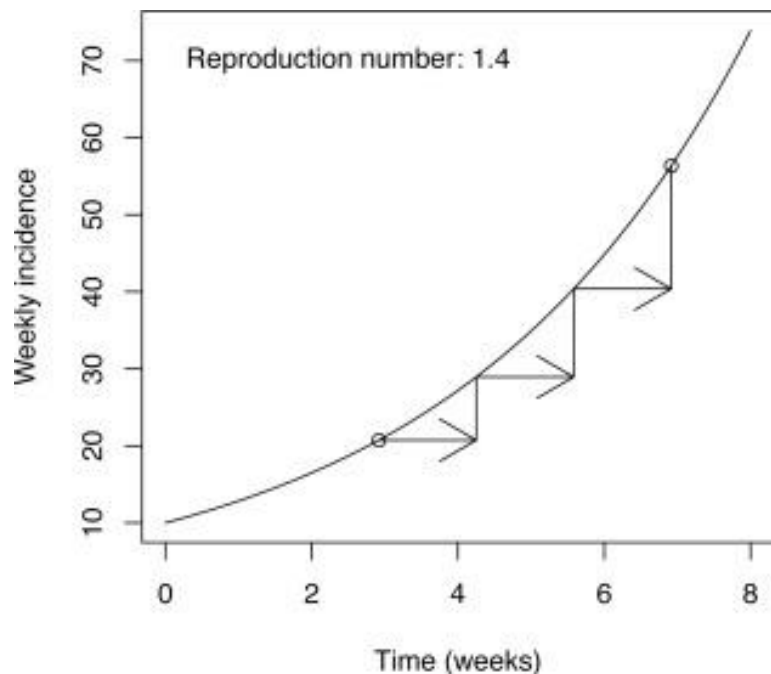
б) Популяционное распределение скоростей инфекционных контактов, где индивидуальные латентные и инфекционные периоды являются *реализациями* экспоненциальных процессов со средними 11.5 и 5 дней соответственно

$$K(\tau) = \int k(\tau, a) f(a) da$$

a - факторы, определяющие инд. особенности инфекционного процесса и контактов

Оценка базового репродуктивного числа по скорости роста заболеваемости

На начальном этапе эпидемии скорость роста числа инфицированных можно легко оценить из данных



Две гипотетические эпидемии с одинаковыми показателями эксп. скорости роста ($1/C = 0.25 \text{ нед.}^{-1}$). А) Короткий генеративный интервал (быстрая передача на инд. уровне) соответствует меньшим значениям репродуктивного числа r_t по сравнению с большим репродуктивным интервалом (В).

Оценка базового репродуктивного числа по скорости роста заболеваемости

$$I(t) = S(t) \int_0^t K(s) I(t-s) ds$$

Базовое репродуктивное число:

$$r_0 = \int K(s) ds$$

$$r_t = r_0 S(t)$$

$$I(t) = r_t \int_0^t g(s) I(t-s) ds$$

Распределение генеративных интервалов:

$$g(s) = \frac{K(s)}{r_0}$$

В начале эпидемии популяция инфицированных растет экспоненциально:

$$I(t) = I(0) \exp(t / C)$$

При малых t :

$$\frac{1}{r_0} \approx \frac{1}{r_t} = \int_0^t g(s) \exp(-s / C) ds$$

Оценка Rt (продолжение)

Предполагая:

$$g(s) = f_{\text{gamma}}(s, k, \bar{G})$$

Альтернативная параметризация:

k – квадрат коэффициента вариации

\bar{G} – среднее

$$E(s) = \bar{G}, k = CV^2 = \frac{V(s)}{E(s)^2}$$

$$Y \sim \text{Gamma}(\alpha, \theta)$$

α - shape

θ - scale

$$\alpha = \frac{1}{k}, \theta = \bar{G}k$$

$$g(s, \frac{1}{k}, \bar{G}k) = \frac{1}{\Gamma(1/k)(\bar{G}k)^{1/k}} s^{1/k-1} e^{-\frac{s}{\bar{G}k}}$$

преобразование Лапласа

$$\frac{1}{r_0} \approx \frac{1}{r_t} = \int_0^t g(s) \exp(-s/C) ds$$

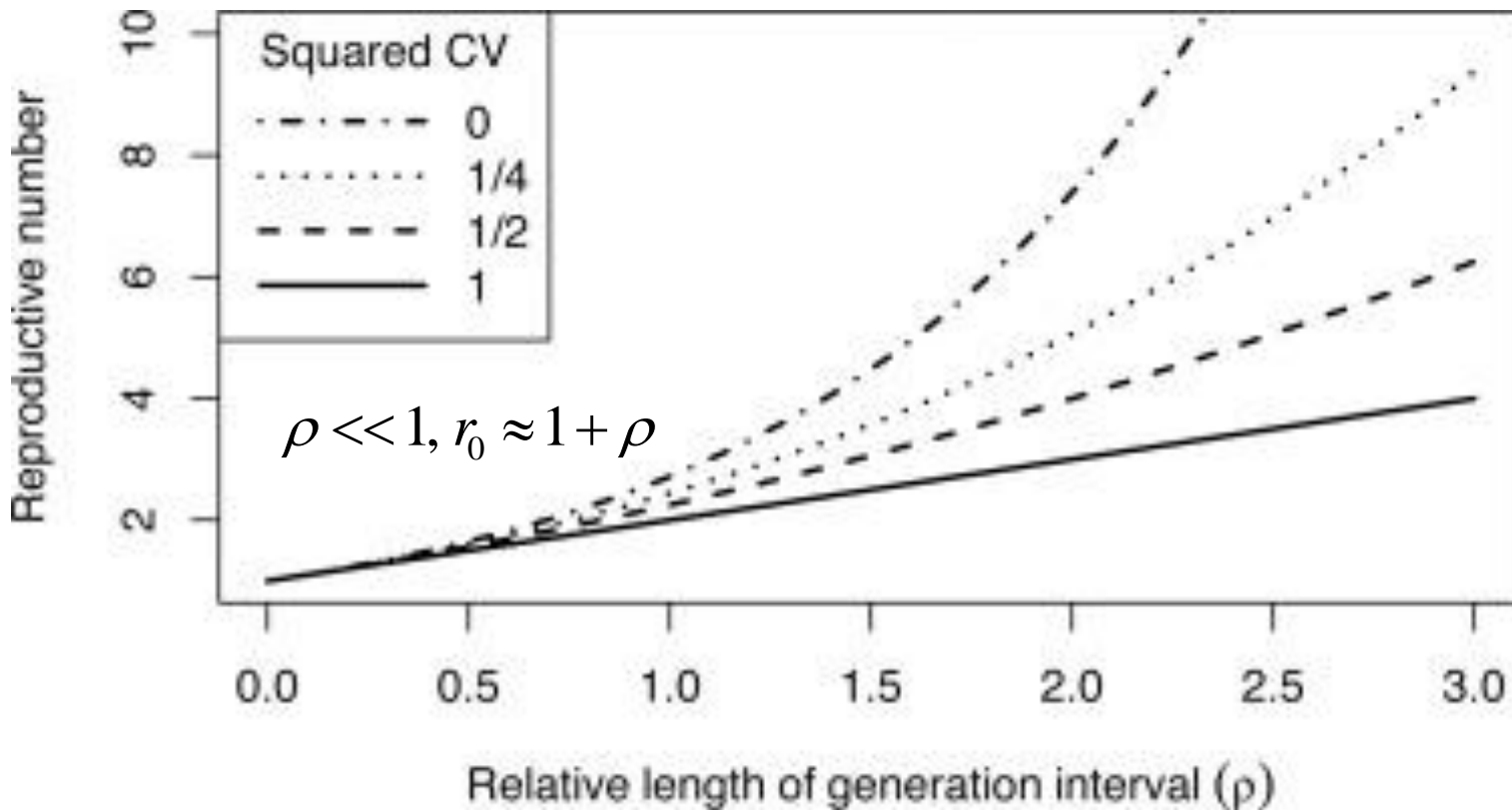
$$r_0 \approx (1 + k\rho)^{\frac{1}{k}} \quad \rho = \frac{\bar{G}}{C}$$

$$\frac{1}{r_0} \approx \frac{1}{r_t} \approx \frac{1}{r_\infty} = (1 + \frac{\bar{G}k}{C})^{-1/k}$$

$$r_0 \approx (1 + k\rho)^{\frac{1}{k}} \quad \rho = \frac{\bar{G}}{C} \quad E(s) = \bar{G}, k = CV^2 = \frac{V(s)}{E(s)^2}$$

Постоянный генеративный интервал:

$$k = 0 \Rightarrow r_t = \exp(\rho)$$

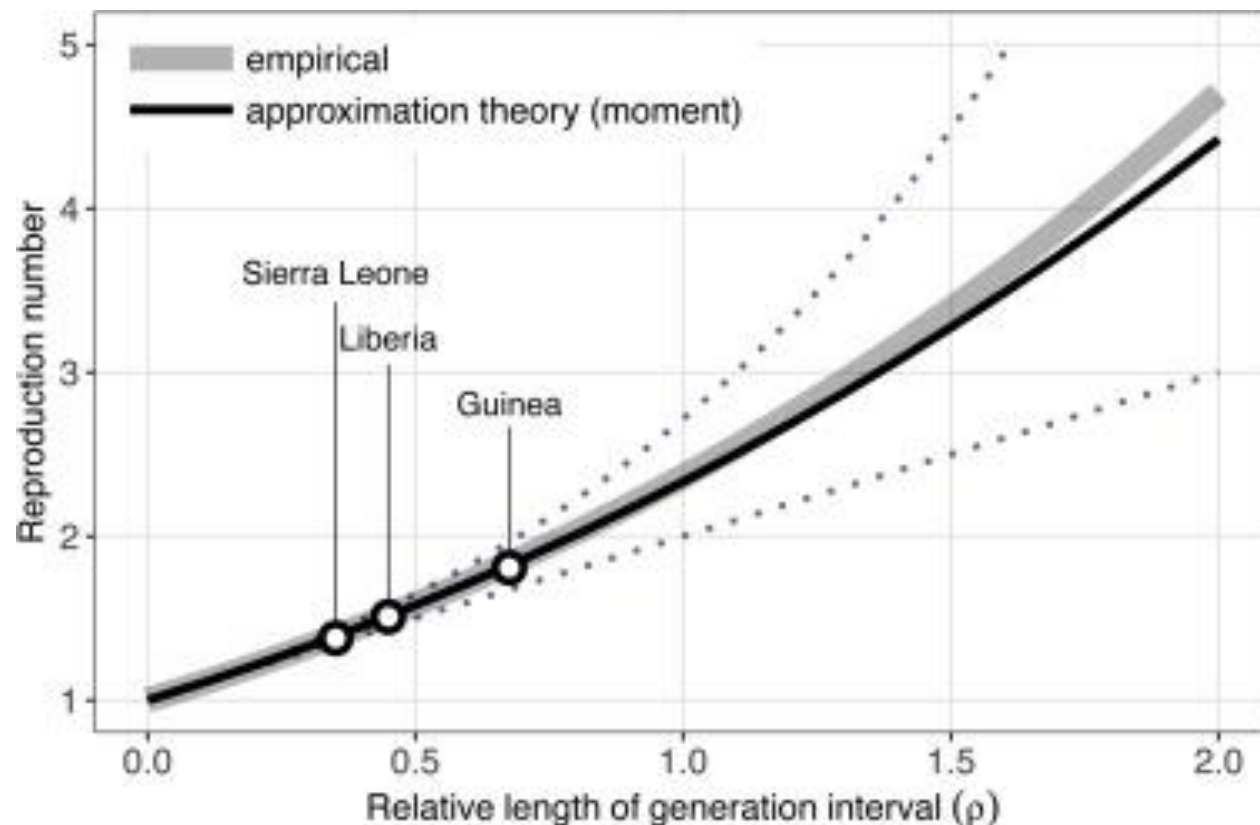


Скорость эксп. роста в SIR модели - $\frac{1}{C} = \gamma(r_0 - 1)$

Ср. $k = 1 \Rightarrow \frac{1}{C} = \frac{r_0 - 1}{\bar{G}}$

Гамма распределение хорошо аппроксимирует распределение генеративных интервалов

Ebola virus [Park 2019 DOI: 10.1016/j.epidem.2018.12.002]



Appl. To SARS-Cov-2 [Park 2020 DOI: 10.1098/rsif.2020.0144]

Рассмотренные модели

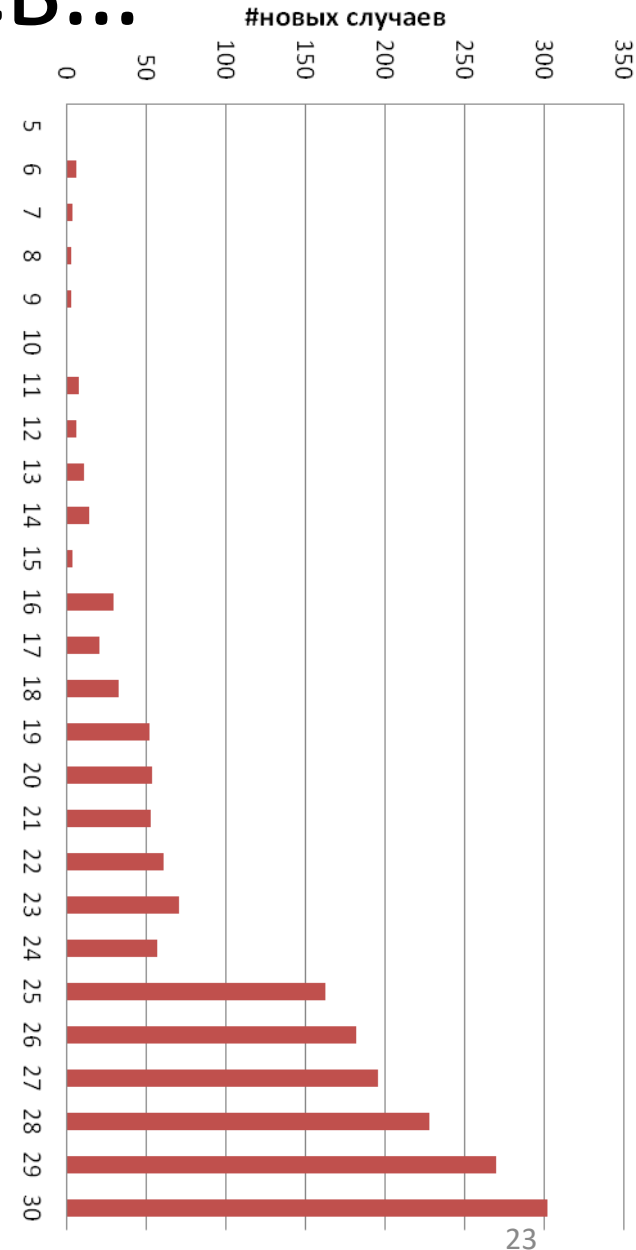
- Корректировка обеспечения справа данных о заболеваемости
- Оценка динамики репродуктивного числа методом максимального правдоподобия по заболеваемости (модель EpiForecast)
- Оценка базового репродуктивного числа по динамике заболеваемости в начальной фазе эпидемии

Результаты

НАЧАЛО ЭПИДЕМИИ SARS-COV-2 В РФ

Как все начиналось...

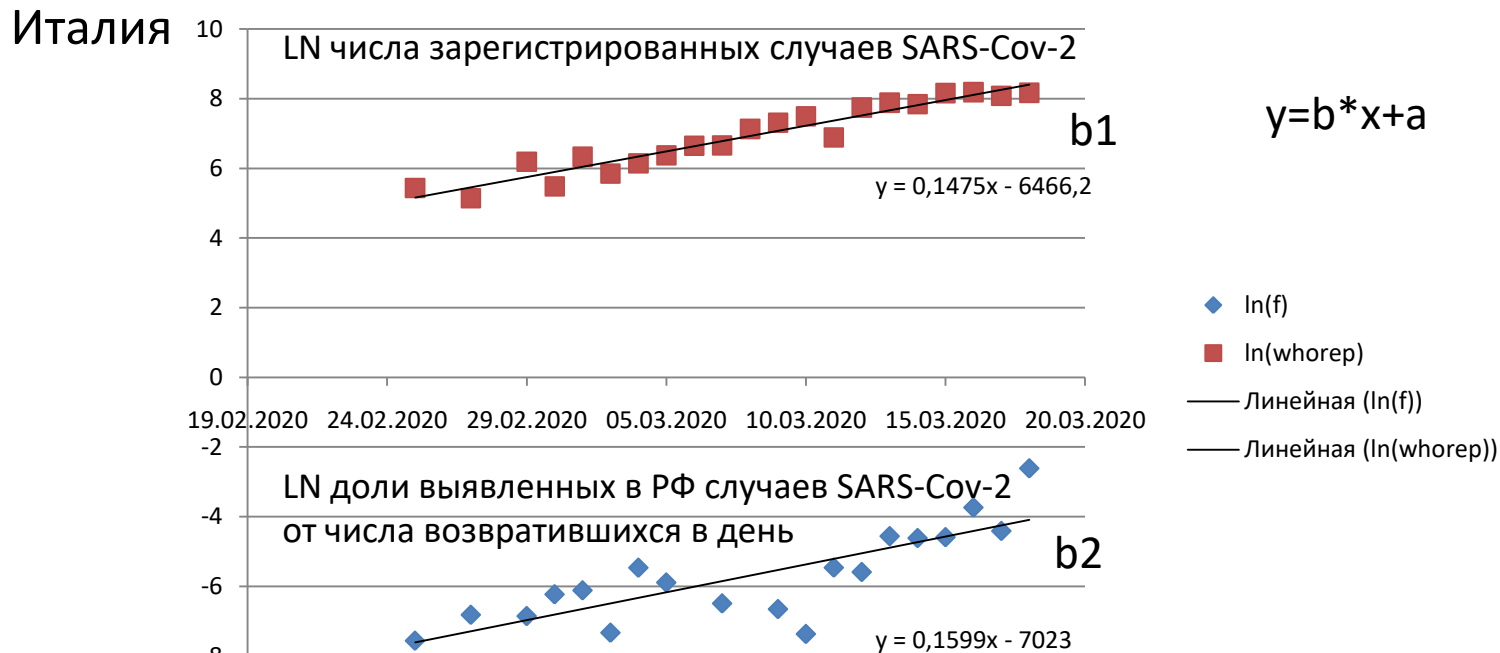
- **2 марта** у вернувшегося из Италии россиянина обнаружен коронавирус.
- **5 марта** в Москве вводится режим повышенной готовности.
- **11 марта** ВОЗ объявило о начале пандемии.
- **С 13 марта** ограничены пассажирские воздушные перевозки с Италией, Германией, Испанией, Францией.
- **15 марта** появилась информация о том, что в России трое человек заразились коронавирусом внутри страны. Эти случаи не являются завозными.
- **С 16 марта** Россия ограничивает авиасообщение со странами Евросоюза, а также Швейцарией и Норвегией.
- **С 16 марта** в Москве домашний карантин обязаны пройти все прилетающие из-за границы.
- **16 марта** в Москве запрещено проведение мероприятий с числом участников более 50 человек.
- **С 18 марта** правительство из-за коронавируса закрывает Россию для въезда всех без исключения иностранных граждан и лиц без гражданства.
- **С 21 марта** закрытие общеобразовательных и спортивных школ.
- **С 23 марта** Россия вводит временные ограничения на полеты в Турцию, Таиланд, Вьетнам, Японию.
- **С 27 марта** Россия прекращает все авиасообщение с другими странами.
- **С 30 марта** Россия ограничивает движение через все пограничные пункты пропуска. Введена нерабочая неделя





Материалы «Российской Газеты» <https://rg.ru/2020/03/03/koronavirus-hronika-rasprostraneniia.html>

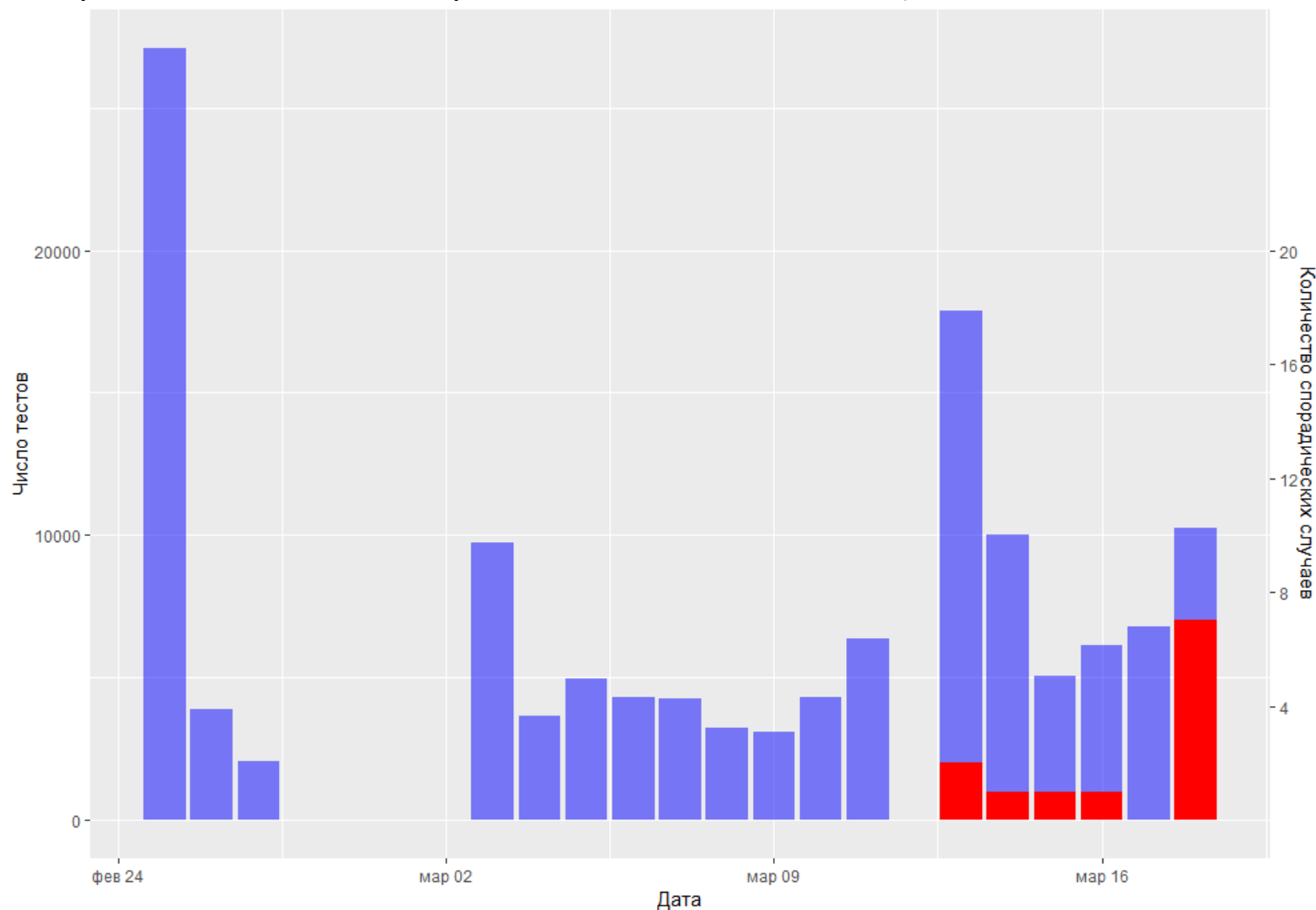
Динамика выявления COVID-19 в марте 2020 в РФ отражала динамику эпидемии в странах Европы



страна	b1	SE(b1)	b2	SE(b2)	P(b1=b2), t-test
Италия	0.147	0.01	0.16	0.028	0.681
Франция	0.233	0.017	0.178	0.046	0.284
Германия	0.224	0.026	0.123	0.047	0.076
Швейцария	0.18	0.034	0.06	0.033	0.022

Начало внутренней циркуляции вируса

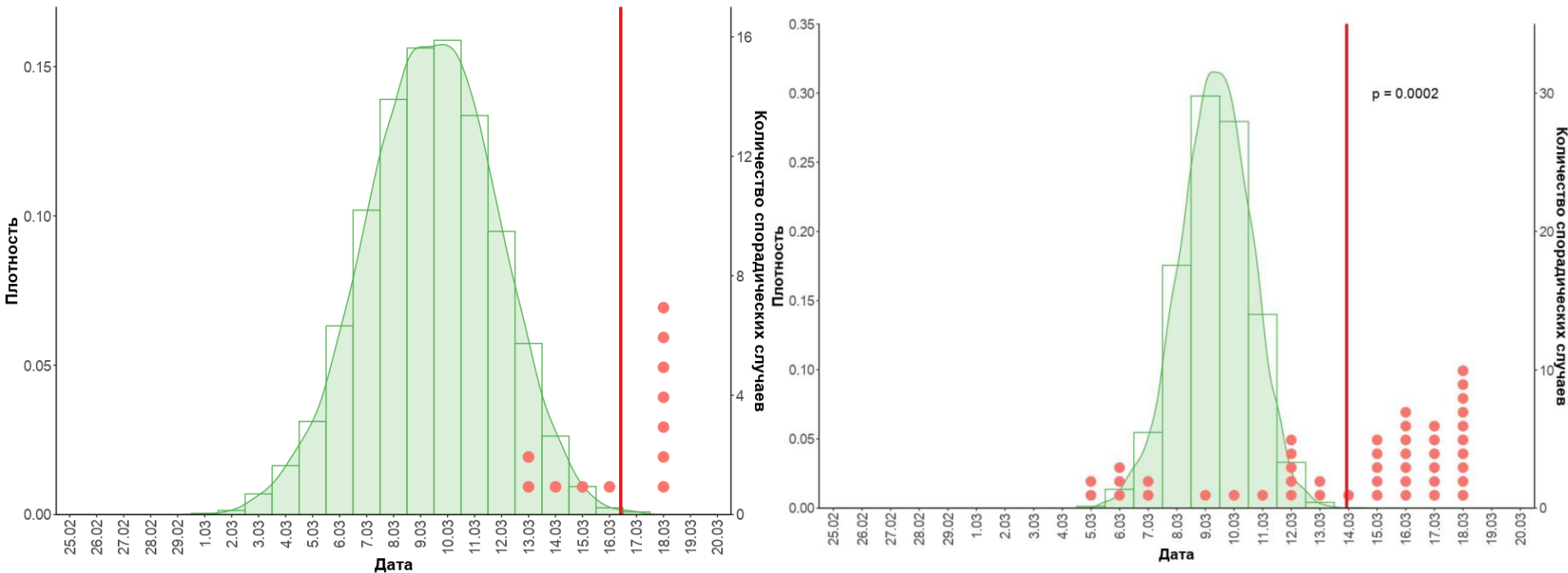
Число спорадических случаев, выявленных на территории РФ по дням, в сравнении с числом проведенных в эти дни ПЦР-тестов.



Рост числа проводимых в сутки ПЦР тестов не объяснял выявления спорадических случаев

А. Оперативные данные на 20 марта

Б. ОД на 23 марта



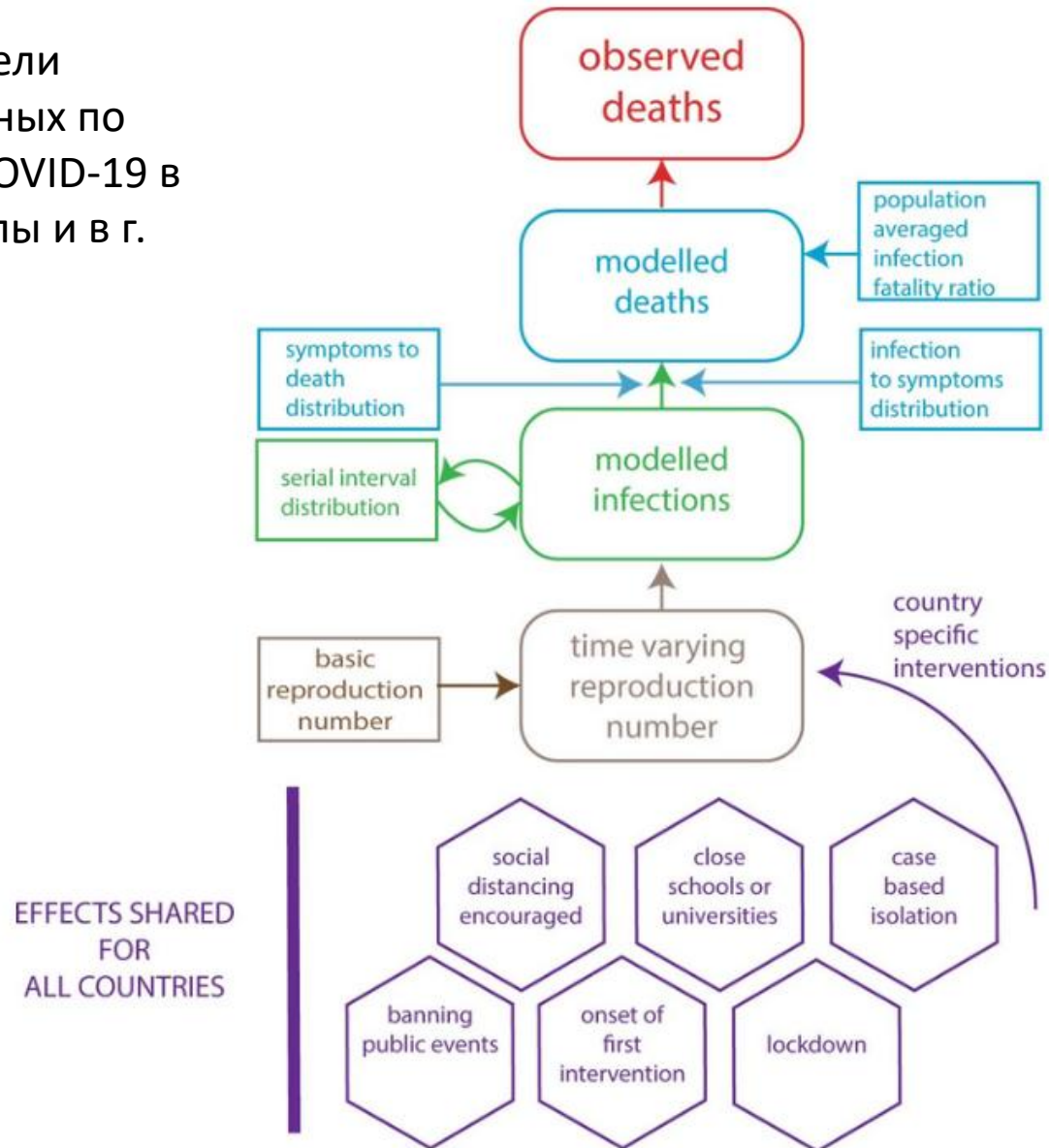
Реальные спорадические случаи (красные точки). Распределения средних дат по множествам N случайных спорадических случаев, разбросанных по оси времени пропорционально числу ПЦР-тестов, выполненных на каждую дату (зеленые столбики). $N=12$ (А) и $N=50$ (Б).

Оценка эффективности введенных в Москве мер эпидемического сдерживания

- Мы оценили эффективность на 18.04.2020 противоэпидемических мер в г. Москва на основе динамики ежедневной смертности с помощью модели Flaxman et. al. 2020, а также:
 - распределение числа новых случаев инфицирования по дням с начала эпидемии
 - r_0 и r_t
 - выявляемость COVID-19

Структура модели

Параметры модели
оценены на данных по
смертности от COVID-19 в
11 странах Европы и в г.
Москве



Предположения модели

- r_t зависит только от применяющихся противоэпидемических мер
- Аналогичные противоэпидемические меры в разных странах имеют одинаковый эффект - снижают r_t от предыдущих значений в одинаковой степени
- Приверженность населения не меняется с течением времени
- Летальность на инфекцию неизменна по времени
- Доля идентифицированных смертей от COVID-19 среди всех смертей от COVID-19 неизменна по времени
- Длительность серийного интервала имеет гамма-распределение со средним 6.5, одинаковое для всех стран

Модель ежедневной смертности

$D_{t,m} \sim \text{NegBin}(d_{t,m}, d_{t,m} + \frac{d_{t,m}^2}{\psi})$ Распределение наблюдаемой ежедневной смертности в момент t в стране m

$\psi \sim N^+(0,5)$

$d_{t,m} = \sum_{s=0}^{t-1} I_{t-s,m} \pi_{s,m}$ $I_{t,m}$ – число инфицированных, $\pi_{s,m}$ – вероятность, что инфекция приведет к смерти через s дней

$s = s_{i-o} + s_{o-d}$

$s_{i-o} \sim g_{i-o} = \text{Gamma}(5.1; 0.86)$ - инкубационный период

$s_{o-d} \sim g_{o-d} = \text{Gamma}(18.8; 0.45)$ время от появления симптомов до смерти

$f_{i-d}(\tau) = \int_0^\tau g_{o-d}(y) g_{i-o}(\tau - y) dy$ - ФПВ интервалов от инфицирования до смерти

$\text{if}r_m$ – средняя по популяции вероятность, что инфекция закончится смертью (infection to fatality rate)

$\pi_{s,m} = \text{if}r_m * P(s)$

$P(s) = F_{i-d}(s) - F_{i-d}(s-1)$

Модель ежедневной смертности (прод.)

$$D_{t,m} \sim \text{NegBin}(d_{t,m}, d_{t,m} + \frac{d_{t,m}^2}{\psi})$$

Распределение наблюдаемой ежедневной смертности в момент t в стране m

$$\psi \sim N^+(0,5)$$

$$d_{t,m} = \sum_{s=0}^{t-1} I_{t-s,m} \pi_{s,m}$$

$I_{t,m}$ – число инфицированных, $\pi_{s,m}$ – вероятность, что инфекция приведет к смерти через s дней

ifr_m – средняя по популяции вероятность, что инфекция закончится смертью (infection to fatality rate)

$$\pi_{s,m} = \text{ifr}_m * P(s)$$

$$I_{t,m} = R_{t,m} \sum_{s=1}^{t-1} I_{t-s,m} g_s$$

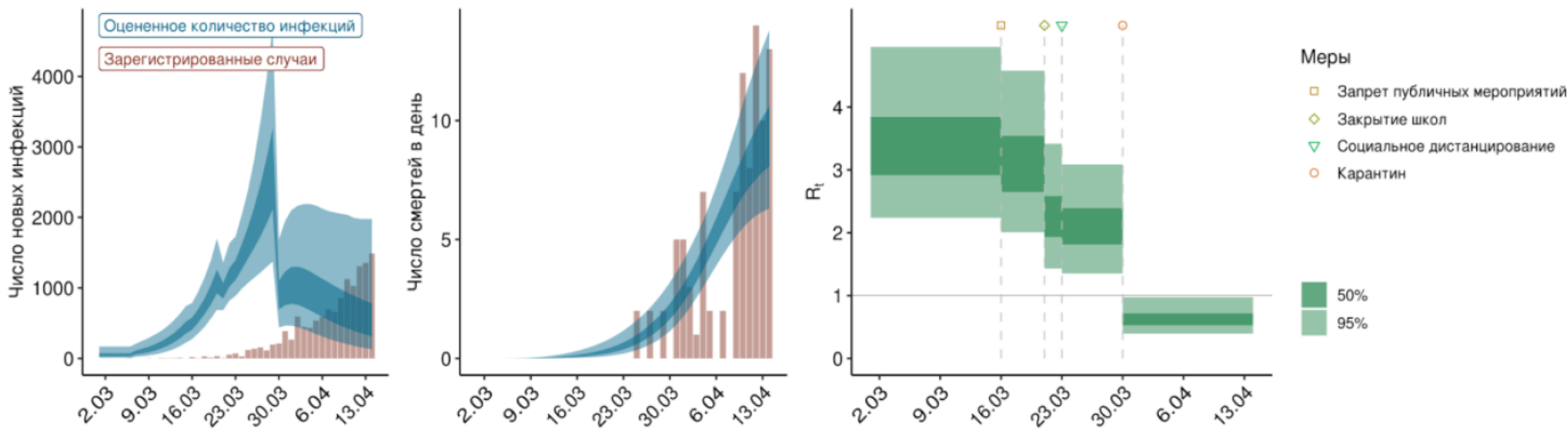
$R_{t,m}$ – репродуктивное число, g_s – распределение репродуктивных интервалов

$$g_s = \text{Gamma}(6.5, 0.62)$$

$R_{t,m}$ – моделируется как ступенчатая функция отражающая совокупность принятых противоэпид. мер.

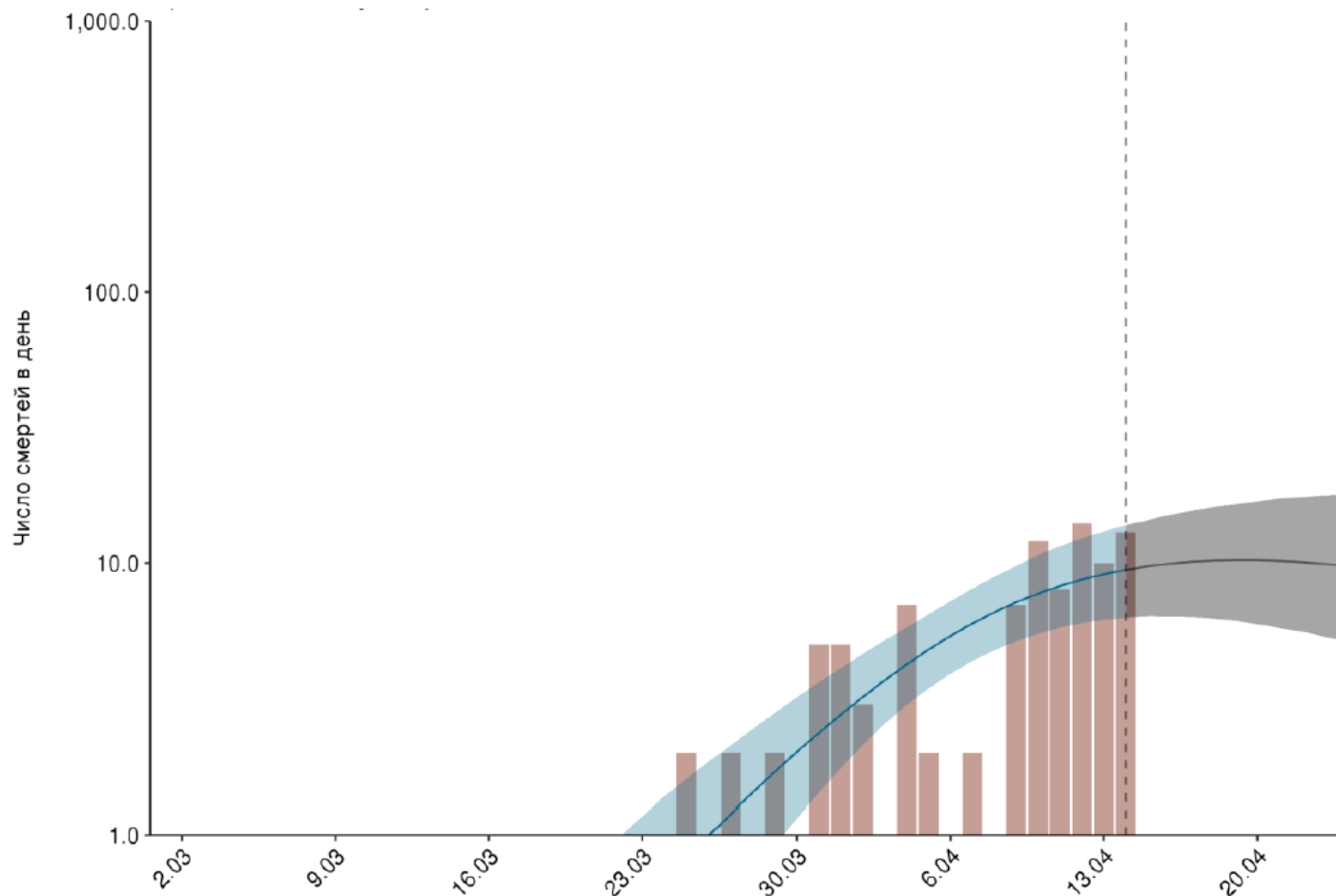
По данным моделирования введенные меры позволили взять эпидемию под контроль

- Совокупность введенных мер позволила снизить R_t с исходного значения 3.42 (95% ДИ: 2.24 - 4.96) до 16 марта до **0.65** после 30 марта.
- На 08.04.2020 суммарное оцененное число случаев инфицирования в г. Москве и области – 48017
- На 18.04.2020 суммарное число выявленных случаев в Москве и области – 3077
- Выявляемость COVID-19 $\sim 6\%$



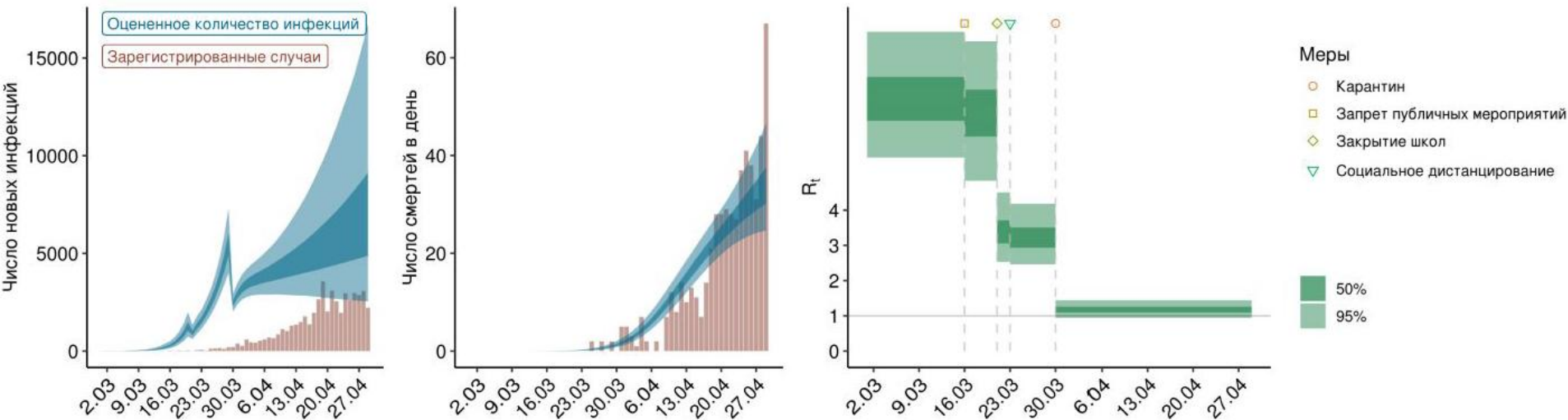
Прогноз смертности 15 - 24 апреля 2020

- Из модели ожидалась стабилизация динамики ежедневной смертности



Добавление новых данных изменило оценки модели

- На новых данных (от 02 марта - 29 апреля) $R_t = 1.18$ (95% ДИ [0.94 – 1.44]) для г. Москвы
- $R_0 = 7.18$ новых случаев на инфекцию
- Стабильное снижение ежедневной смертности в Москве началось после 30 мая



- Нарушено одно или несколько базовых предположений модели?
- Наблюдаемая смертность на 13.04 объяснялась, в основном, завозными случаями?

Модель динамики эпидемического процесса в условиях активного выявления контактных лиц



Модели:

m1: Два компартмента:

R_{spv} – для поднадзорной популяции

R_{usv} – для общей популяции

m2: Один компартмент для всей популяции

Модель m2 похожа на модель

EpiForecast (DOI:

[10.12688/wellcomeopenres.16006.1](https://doi.org/10.12688/wellcomeopenres.16006.1))

Входные данные модели m1



Оценка динамики репродуктивного числа R_t по заболеваемости

$I_i, i = 1..t$ Число новых случаев инфицирования случившихся в день i , t – текущий день от начала эпидемии

$s \sim g_{i-i}$ g_{i-i} – Распределение генерационных интервалов - распределение времени между последовательными инфицированиями в цепочках передачи. $f_g(s)$ - ФПВ

$$\pi_s = \int_{s-1}^s f_{g_{i-i}}(\tau) d\tau$$

$\Lambda_t = \sum_{s=1}^t \pi_{t-s} I_{t-s}$ Инфекционный потенциал – число активных носителей

$E[I_t] = \Lambda_t R_t$ Ожидаемое число новых инфекций в момент времени t , где R_t – репродукционное число

$$I_t \sim Pois(\Lambda_t R_t)$$

Параметры модели

Распределение генерационных интервалов:

$$s \sim g_{i-i} = \text{Gamma}(\text{shape} = 2.6, \text{rate} = 0.4) \quad \text{Flaxman et. al. 2020}$$

α - shape
 ϑ – scale
rate:
 $\beta = 1 / \vartheta$

Распределение задержки от выявления заболевания до учета в статистике:

$$\tau_{o-r} \sim g_{o-r} = \text{Gamma}(\text{shape} = 5.48, \text{rate} = 0.77) \quad \text{Оценено для г. Москвы}$$

Распределение инкубационных периодов:

$$\tau_{i-o} \sim g_{i-o} = \text{Gamma}(\text{shape} = 1.352, \text{rate} = 0.265)$$

$$\tau_{i-r} = \tau_{i-o} + \tau_{o-r}$$

Собираем все вместе

$O_i, i = 1..t$ Число новых случаев заболевания зарегистрированных в день i , t – текущий день от начала эпидемии

$$\pi_\tau = \int_{\tau-1}^{\tau} g_{i-i}(s) ds$$

$$\Lambda_i = \sum_{s=1}^i \pi_s I_{i-s}, i = 1, 2, \dots, t$$

- инфекционный потенциал – число активных носителей в момент i

$$I_i = \Lambda_i R_i$$

– репродукционное число

Задержка от инфицирования до попадания в статистику:

$$f_{i-r}(\tau) = \int_0^{\tau} g_{o-r}(y) g_{i-o}(\tau - y) dy$$

$$\xi_\tau = F_{i-r}(\tau) - F_{i-r}(\tau - 1), \tau = 1, \dots, t$$

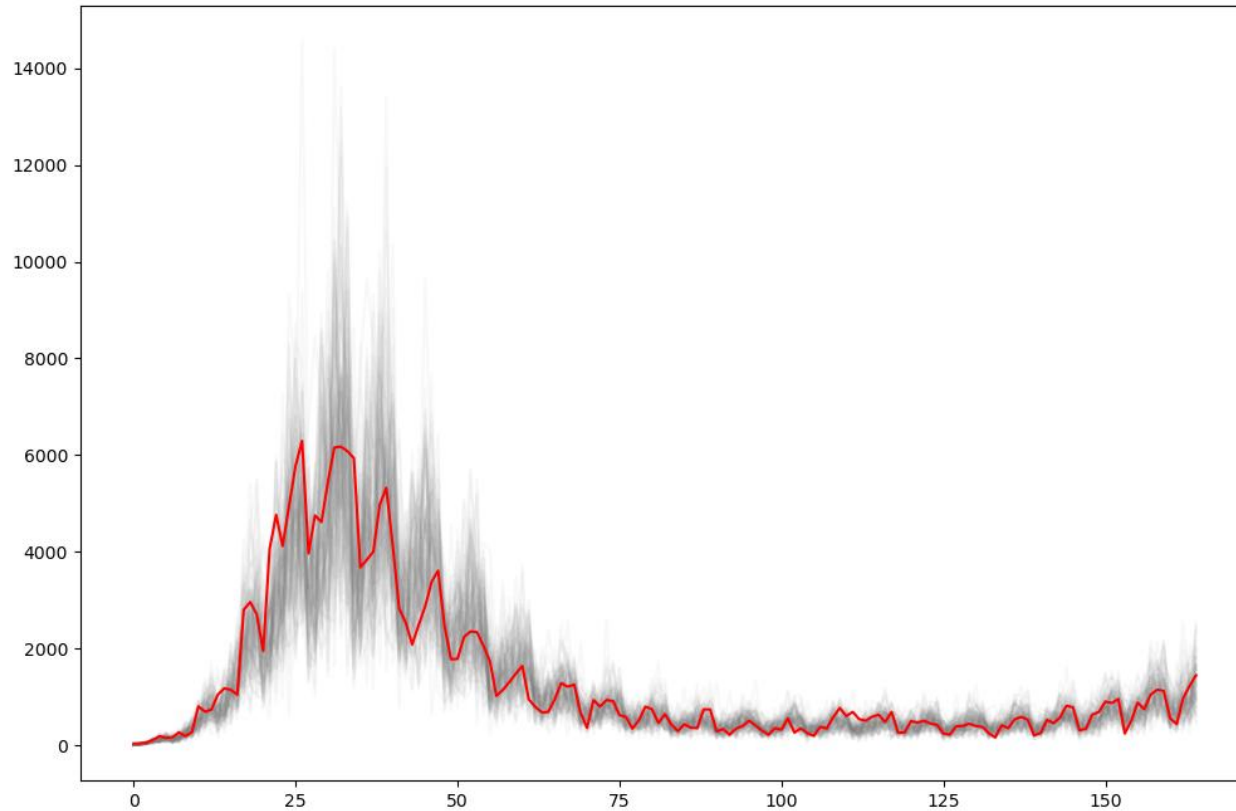
$$E[O_i] = \lambda_i = \omega_{[i \bmod 7]} \sum_{\tau=0}^i I_{i-\tau} \xi_\tau$$

-ожидаемое число новых случаев в момент времени i

$$\omega \sim \text{Dirichlet}(k = 7, \alpha = (\alpha_1, \dots, \alpha_k))$$

$$O_i \sim \text{Pois}(\lambda_i)$$

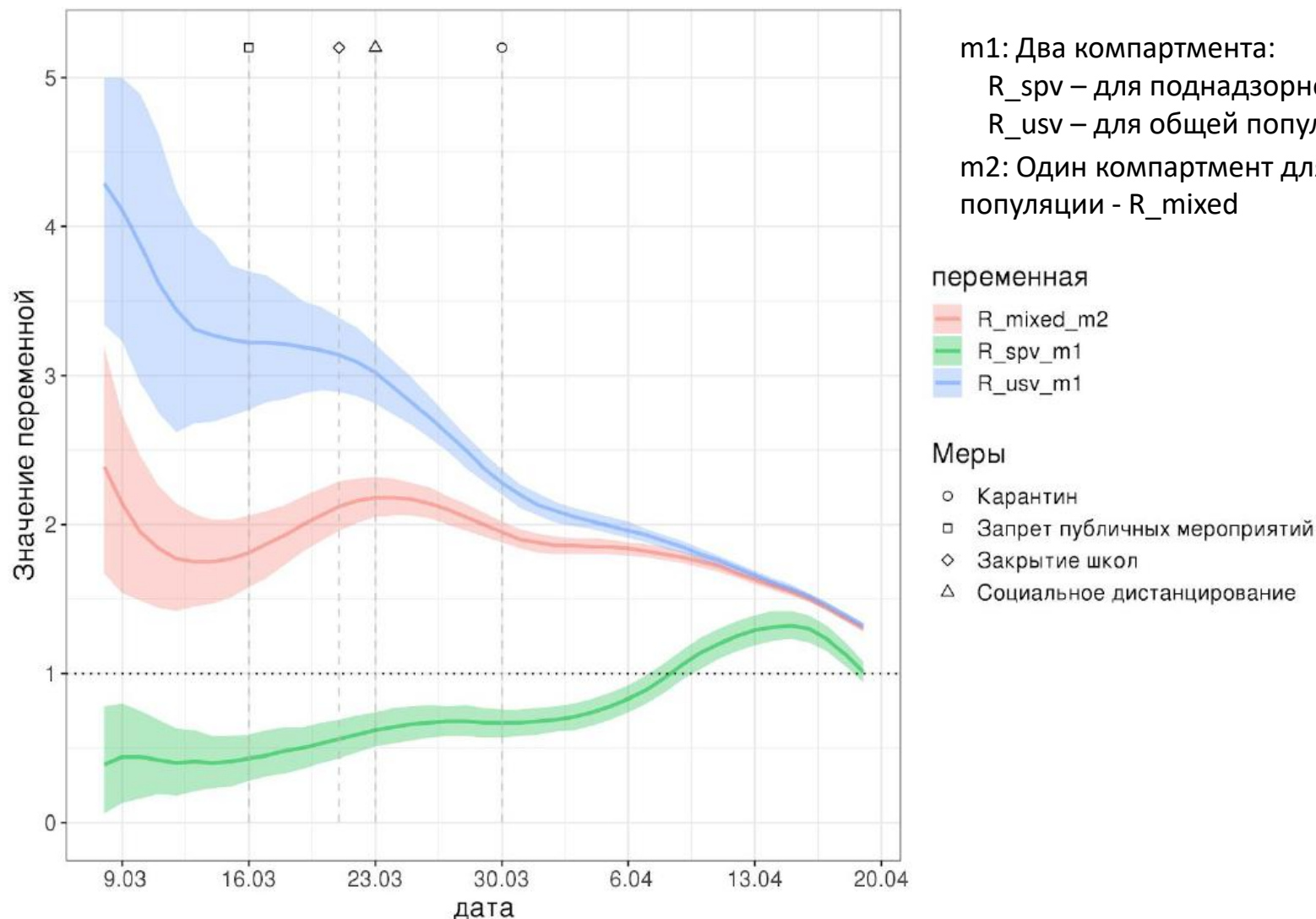
Эффект дня недели



$$E[O_i] = \lambda_i = \omega_{[i \bmod 7]} \sum_{\tau=0}^i I_{i-\tau} \xi_{\tau}$$

Снижение эффективности мер активного выявления по ходу развития эпидемии

Модели m1 и m2 одинаково хорошо описывают данные, но m2 – учитывает их структуру



Заключение

- Мы представили наш опыт оперативного анализа начальной фазы эпидемии SARS-CoV-2 в РФ:
 - Динамика выявленных случаев до 16 марта 2020 объяснялась, в основном, завозами из-за рубежа.
 - Активное выявление контактных лиц эффективно прерывало цепочки распространения инфекции до 08 апреля 2020
 - 16.03.2020 подтверждено начало внутренней циркуляции вируса
 - Оценена доля выявляемых случаев от общего числа инфицированных на 18.04.2020
 - Оценена динамика показателя R_t

Благодарности

- Плоскиревой А.А. – ЦНИИ Эпидемиологии Роспотребнадзора
- Федонину Г.Г. – ЦНИИ Эпидемиологии Роспотребнадзора
- Поповой А.В. – ЦНИИ Эпидемиологии Роспотребнадзора
- Базыкину Г.А. – «Сколтех»
- Сафиной К.Р. – «Сколтех»
- Гарушянц С.К. – ИППИ РАН