

УДК: [004.94+519.7]:[614.4+616-002.5]

Сравнение схем моделирования естественного течения туберкулёза органов дыхания

Авилов К.К.^{*1,3}, Романюха А.А.^{1,3}, Белиловский Е.М.², Борисов С.Е.²

¹*Институт вычислительной математики им. Г.И. Марчука Российской академии наук,
г. Москва, Российская Федерация*

²*ГБУЗ «Московский научно-практический центр борьбы с туберкулёзом ДЗМ»,
г. Москва, Российская Федерация*

³*Московский государственный университет им. М.В. Ломоносова, г. Москва,
Российская Федерация*

Аннотация. При построении математических моделей распространения и контроля туберкулёза органов дыхания в качестве признака тяжести болезни часто используется наличие или отсутствие бактериовыделения. Существуют две основные схемы моделирования течения туберкулёза: более распространённая ”параллельная” – предполагает, что у определенной доли заболевших бактериовыделение развивается крайне быстро и затем сохраняется в течение всей болезни; и реже используемая ”последовательная” – предполагает, что туберкулёз всегда начинается со стадии без бактериовыделения, а затем некоторая часть больных прогрессирует в стадию с бактериовыделением. В данной статье производится сравнение этих двух схем моделирования на основе качества приближения ими реальных данных по г. Москва, 2013–2018 гг., содержащих информацию о сроках и результатах предыдущего флюорографического обследования для выявленных больных туберкулёзом органов дыхания, а также о их состоянии на момент выявления. Такие данные дают ограничение на длительность болезни, а потому позволяют судить о динамике развития бактериовыделения при нелеченном туберкулёзе. Мы создали агентную модель с реалистичными профилями смертности и прохождения флюорографии и с эквивалентом традиционной компартментной подмодели течения туберкулёза на основе линейных дифференциальных уравнений. По результатам расчётов по этой модели последовательная схема моделирования оказалась более близкой к реальности. Вместе с тем, в силу плохого приближения реальных данных был сделан вывод о необходимости доработки как подмодели выявления, так и подмодели естественного течения туберкулёза органов дыхания.

Ключевые слова: *математическая модель, туберкулёз, схема патогенеза, бактериовыделение, реальные данные, флюорография, агентное моделирование.*

ВВЕДЕНИЕ

Математические модели распространения и контроля инфекционных заболеваний широко применяются для обработки эпидемиологических данных, прогнозирования развития эпидемической ситуации и оптимизации противоэпидемических мероприятий. [1] В частности, математические модели распространения туберкулёза

*kkavilov@gmail.com

(ТБ) использовались начиная с 1960-ых годов (см. обзор [2]).

В большинстве случаев модели эпидемического процесса являются компартментными моделями, опирающимися на разделение популяции на несколько дискретных групп по отношению к рассматриваемой инфекции и болезни (напр., неинфицированные, больные, резистентные, вакцинированные, лечимые и т.п.) и описание течения эпидемического процесса в терминах изменения численностей этих групп за счёт перехода индивидов между группами в соответствии с происходящими с ними эпидемиологическими событиями. Наиболее известным примером такой модели является т.н. SIR-модель (susceptible-infected-recovered, чувствительный-инфицированный-излечившийся). [1]

Схема и законы переходов индивидов между группами-состояниями должны, очевидно, отражать реальную схему течения эпидемических процессов, что позволит математической модели адекватно усваивать реальные данные и строить корректные прогнозы. Для простых моделей типа SIR или для моделей распространения очень хорошо изученных инфекций общая схема математической модели, как правило, не вызывает сомнений. Однако в случае моделирования туберкулёза, отличающегося сложным и недостаточно изученным течением латентной инфекции и нелеченной болезни, используются различные схемы описания этих процессов.

В данной работе будет проведено сравнение двух наиболее распространенных схем моделирования течения нелеченного активного туберкулёза органов дыхания, использующих наличие или отсутствие бактериовыделения у больных в качестве признака разных стадий или форм туберкулёзной болезни. Предсказания двух типов моделей будут сравниваться с реальными данными по больным ТБ, выявленным в г. Москве в 2013–2018 гг., с учётом времени и результатов предыдущего флюорографического обследования. Отметим, что прямое наблюдение за нелеченными больными туберкулёзом невозможно в современных условиях, т.к. после выявления болезни немедленно начинается лечение. Поэтому реконструкция законов течения нелеченного туберкулёза у жителя современного мегаполиса возможна только на основе анализа косвенных данных при помощи математических моделей.

Краткие сведения о туберкулёзе органов дыхания

Туберкулёз – опасное инфекционное заболевание, вызываемое микобактериями *M. tuberculosis complex* (МБТ). [3] Инфицирование возбудителями ТБ обычно происходит воздушно-капельным путём. После инфицирования в подавляющем большинстве случаев инфекция переходит в латентное состояние, сохраняющееся, как считается, до конца жизни индивида (при условии отсутствия специфического лечения). У небольшой доли инфицированных (порядка 10 % в течение всей жизни) происходит активация инфекции и развивается активный туберкулёз. [4] У большинства заболевших (порядка 95 %) активный туберкулёз поражает лёгкие и другие органы дыхания, что может привести к выделению микобактерий и передаче их воздушно-капельным путём. Остальные заболевшие (порядка 5 %) имеют внелёгочный туберкулёз, который значительно менее эпидемиологически значим, т.к. значительно реже приводит к бактериовыделению и, в результате этого, к новым инфицированиям. [3]

Клинические формы туберкулёза органов дыхания (ТОД) многообразны, но на практике часто используются два признака, в целом описывающие состояние больного – бактериовыделение (наличие бактериовыделения – МБТ+, отсутствие – МБТ-) и распад легочной ткани (КВ+ и КВ- соответственно). Наличие как бактериовыделения, так и распада означает более тяжёлое состояние больного, а в первом случае – и его повышенную эпидемиологическую опасность. Большая тяжесть состояния больного увеличивает и его риск гибели от ТБ, и вероятность диагностирования его болезни (т. н.

выявления больного).

Динамика развития бактериовыделения при ТОД весьма вариативна: существуют как скоротечные формы ТБ, когда развитие болезни до начала бактериовыделения занимает несколько недель, так и хронические формы, когда болезнь длится многие годы и десятилетия, по-видимому, без развития бактериовыделения. [3] Средняя продолжительность ТОД составляет 2–3 года. [5] В самом начале болезни все больные являются МБТ- и КВ-. [6]

Выявление больных ТОД подразделяется на активное и пассивное. Активное выявление – это выявление при медицинском осмотре на ТБ (в подавляющем большинстве случаев – при флюорографическом (ФГ) обследовании), проведенном по инициативе медицинских служб. Пассивное выявление – это выявление при обращении с жалобами, когда плохое самочувствие больного, вызванное туберкулёзом, заставило его обратиться за медицинской помощью.

Схемы моделирования туберкулёза органов дыхания

При моделировании популяционной динамики ТОД с учётом бактериовыделения наиболее часто используются две полярные схемы:

1. Параллельная схема (рис. 1,а): наличие бактериовыделения считается определённым с самого начала болезни, что соответствует предположению об очень быстром развитии бактериовыделения у тех больных, которые должны развить его. В некоторых модификациях параллельной модели допускается постепенное развитие бактериовыделения у МБТ- больных, но обычно скорость этого процесса крайне мала, а потому он не оказывает заметного влияния на общую картину. Примеры такого рода моделей – работы [7, 8].
2. Последовательная схема (рис. 1,б): все больные в начале болезни считаются МБТ-, а затем с равной удельной вероятностью развивают бактериовыделение. Примеры такого рода моделей – работы [9, 10, 11, 12]. Отметим, что последовательная схема встречается в значительно меньшем количестве научных работ, чем параллельная.

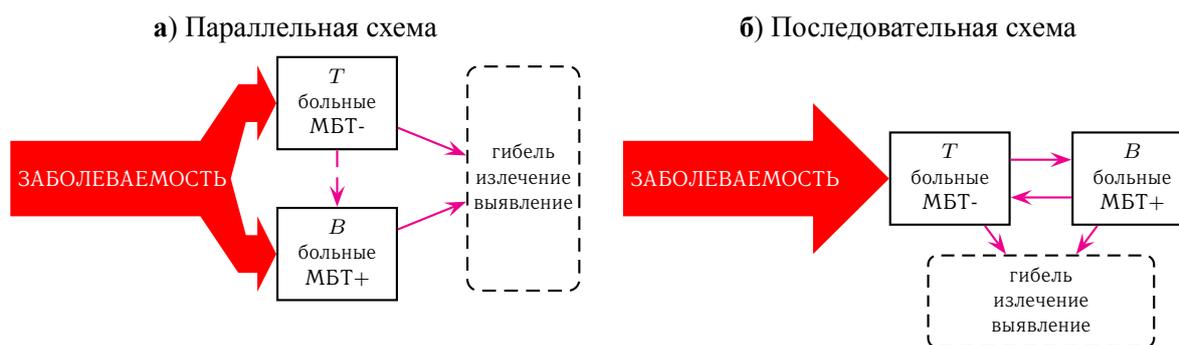


Рис. 1. Схемы описания течения туберкулёза органов дыхания на популяционном уровне: а) параллельная схема, б) последовательная схема. МБТ- – больные без бактериовыделения, МБТ+ – больные с бактериовыделением. Стрелками показаны переходы между группами.

Реальные данные и основной принцип сравнения моделей

В этом исследовании мы использовали базу данных по больным ТБ, выявленным в г. Москва в 2000–2018 гг., предоставленную ГБУЗ «МНПЦБТ ДЗМ». В записях этой

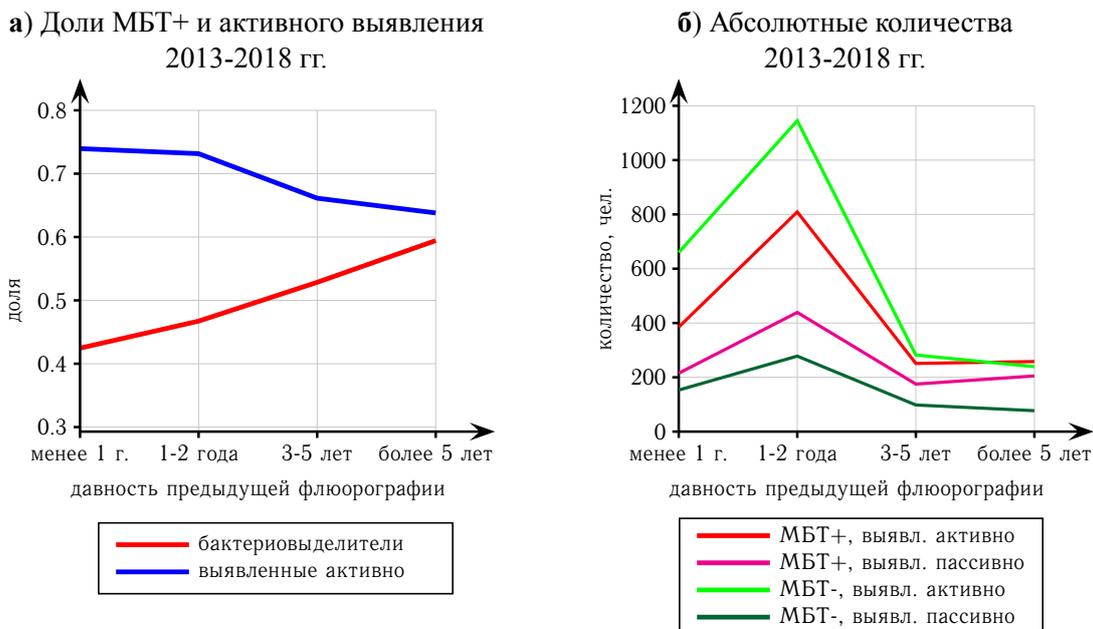


Рис. 2. Данные по выявлению больных туберкулёзом органов дыхания (ТОД) в г. Москва за 2013–2018 гг. в зависимости от давности предыдущего флюорографического обследования (постоянное население, возраст 15+ лет, впервые выявленный ТОД, здоровые на предыдущем обследовании): **а)** доли больных с бактериовыделением (МБТ+) и активно выявленных, **б)** абсолютные количества выявленных больных с и без бактериовыделения (МБТ+ и МБТ- соответственно) с разделением по типу выявления.

базы имеются сведения о множестве параметров больных на момент выявления, включая наличие бактериовыделения, метод выявления (активное или пассивное) и информацию о предыдущем флюорографическом исследовании больного (дискретно заданную давность такого исследования относительно момента выявления и результаты этого исследования, проверенные при экспертном пересмотре снимков).

Главной информацией, воспроизведение которой математическими моделями будет проверяться, является зависимость доли МБТ+ больных среди выявленных от давности предыдущей ФГ для тех, кто был здоров по данным предыдущей ФГ – см. рисунок 2 (критерии отбора больных: постоянное население, возраст от 15 лет и старше, впервые выявленный ТОД, здоров на предыдущей ФГ). Эта зависимость доли МБТ+ среди выявленных больных от времени после предыдущей ФГ не может быть непосредственно использована как мера скорости нарастания доли МБТ+ среди всех нелеченных (и еще невыявленных) больных, поскольку выявленные больные не включают в себя тех, кто умер или спонтанно излечился до момента потенциального выявления, а также из-за неоднозначности связи момента развития ТОД и момента предыдущей ФГ. Поэтому необходимо построение совместных моделей естественного течения ТОД и процессов выявления больных и сравнение их предсказаний с реальными данными. В случае, если удастся построить модель, достаточно хорошо приближающую такие данные, можно будет утверждать, что посредством ее настройки на данные будут оценены параметры течения нелеченного ТОД.

Данные на рисунке 2 представлены за временной интервал 2013–2018 гг., поскольку, как показано на рисунке 3, доля активно выявленных больных существенно поменялась с 2000 по 2018 г., а рассматриваемые далее математические модели будут предполагать стационарную ситуацию по мероприятиям для выявления больных, что представляется

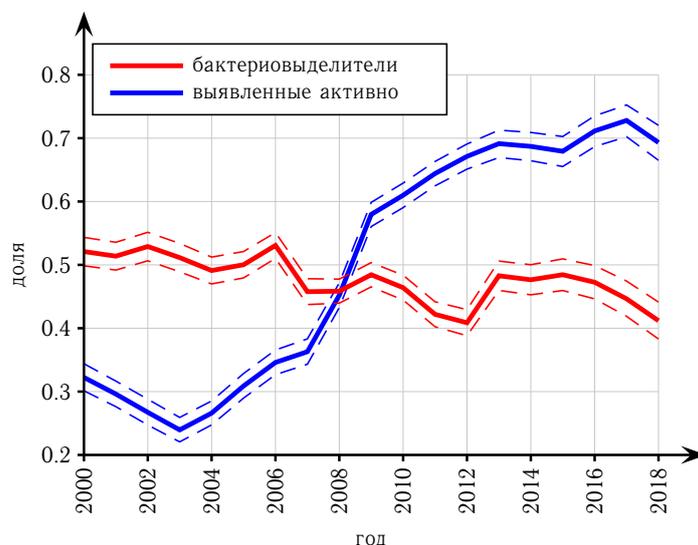


Рис. 3. Доли бактериовыделителей (МБТ+) и активно выявленных от всех выявленных больных ТОД в г. Москва в зависимости от года; пунктирные линии показывают 95 %-ые доверительные интервалы по биномиальному распределению.

правдоподобным для интервала 2013–2018 гг.

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Для исследования следствий двух подходов к моделированию течения нелеченного ТОД нами была построена математическая модель, включающая в себя оба подхода как частные случаи, а также позволяющая достаточно легко встраивать более сложные подмодели течения ТОД. Модель является симуляционной (агентной), в явном виде описывающей события, происходящие с больными с момента возникновения болезни и до выявления, самоизлечения или гибели.

Агентами в модели являются невыявленные больные ТОД. Агенты обладают следующими свойствами:

1. бактериовыделение (МБТ- или МБТ+),
2. время гибели от нетуберкулёзных причин (через это косвенно задаются возраст больного и скорость гибели от всех причин),
3. время предыдущего ФГ-обследования,
4. время следующего ФГ-обследования,
5. время до ближайшего события, связанного с течением болезни, и вид этого события (развитие бактериовыделения, прекращение бактериовыделения, самоизлечение, пассивное выявление, гибель от ТБ).

Схема переходов между состояниями агентов показана на рисунке 4. Параметры описаны в таблице 1. Агентная модель построена на принципе конкуренции независимых процессов. При переходе агента в новое состояние (МБТ- или МБТ+) рассчитываются времена до всех возможных в данном состоянии событий (перехода в другое состояние, гибели от ТБ, самоизлечения, пассивного выявления) и выбирается ближайшее по времени событие, которое запоминается вместе с его временем. Соответствующее событие происходит с агентом при наступлении времени ближайшего события. Процессы

гибели от нетуберкулёзных причин и активного выявления реализованы отдельно, за счёт проверки наступления времени гибели и времени следующего ФГ-обследования.

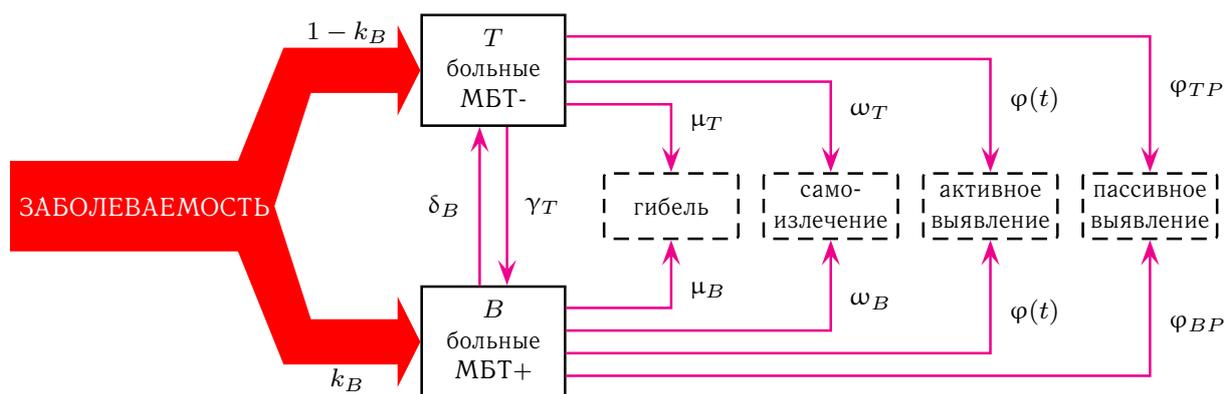


Рис. 4. Блок-схема математической модели. МБТ- – больные без бактериовыделения, МБТ+ – больные с бактериовыделением. Стрелками показаны переходы между группами. Параметры описаны в таблице 1.

Таблица 1. Параметры математической модели. Значения со звёздочкой означают базовые значения параметров, которые будут изменяться в процессе настройки модели

Обозначение	Описание	Знач-е в паралл. модели [7]	Знач-е в паралл. модели [8]	Знач-е в посл. модели [12]	Размерность
k_B	доля заболевших, сразу развивающих бактериовыделение	0.75	0.65	0	—
γ_T	скорость развития бактериовыделения у МБТ-больных	0	0.015	0.596	1/год
δ_B	скорость прекращения бактериовыделения у МБТ+больных	0	0	0.625	1/год
μ_T	скорость гибели от ТБ МБТ-больных	0.139	0.21	0.025	1/год
μ_B	скорость гибели от ТБ МБТ+больных	0.139	0.3	0.513	1/год
ω_T	скорость спонтанного самоизлечения МБТ-больных	0.058	0.2	0.217	1/год
ω_B	скорость спонтанного самоизлечения МБТ+больных	0.058	0.2	0	1/год
$\varphi(t)$	скорость прохождения флюорографических обследований, зависит от времени после предыдущего обследования		см. рис. 5*		1/год
φ_{TP}	скорость пассивного выявления МБТ-больных	0.15*	0.15*	0.15*	1/год
φ_{BP}	скорость пассивного выявления МБТ+больных	0.3*	0.3*	0.3*	1/год

При создании агента в модельной популяции (т. е. при развитии ТОД) производятся следующие операции:

1. Генерируется общая продолжительность жизни t_{life} по реальному профилю смертности (в его качестве взята смертность от всех причин для обоих полов в г. Москва за 2010 год [13], поскольку смертность от ТБ вносит малый вклад в общую смертность в г. Москва); определяется возраст A , в котором произошло развитие болезни (генерируется по равномерному распределению от 15 до 45 лет, поскольку именно в этих возрастах в российских условиях наблюдается основное количество

случаев ТОД; вид же распределения несущественен, т. к. в данной модели от возраста зависит только нетуберкулёзная смертность, а она мала в возрастах 15–45 лет), и запоминается время гибели от нетуберкулёзных причин

$$t_{death} = t_{life} - A.$$

В случае, если оказывается $t_{death} \leq 0$, продолжительность жизни t_{life} генерируется заново по тому же профилю смертности.

2. Генерируются времена предыдущего и следующего ФГ-обследований: время между ФГ-обследованиями t_{FG} генерируется по распределению f_{FG} , определяемому на основании заданной удельной скорости прохождения ФГ-обследования $\varphi(t)$ в зависимости от времени после предыдущего обследования t (базовое значение показано на рисунке 5, модифицированные значения описываются далее в уравнении (15); малое значение $\varphi_0(t)$ в первый год после предыдущего ФГ-обследования отражает стандарт проведения ФГ не чаще раза в год; снижение $\varphi(t)$ в последующие годы моделирует “вымывание” более регулярно проходящих ФГ индивидов):

$$t_{FG} \sim f_{FG}, \quad (1)$$

$$f_{FG}(t) = \frac{t\varphi(t)x(t)}{\int_0^\infty s\varphi(s)x(s)ds}, \quad (2)$$

$$x(t) = \exp\left(-\int_0^t \varphi(s)ds\right). \quad (3)$$

Затем генерируется случайная величина r , равномерно распределенная на $[0, 1]$, отражающая разделение времени t_{FG} между интервалами до момента выявления и после момента выявления

$$r = \text{rand}(0, 1). \quad (4)$$

В итоге время предыдущего ФГ-обследования t_{FGprev} и время следующего ФГ-обследования t_{FGnext} задаются как

$$t_{FGprev} = -rt_{FG}, \quad (5)$$

$$t_{FGnext} = (1 - r)t_{FG}, \quad (6)$$

где в качестве нулевого момента времени подразумевается момент развития болезни. Вывод формул (1–6) см. в Приложении.

3. С вероятностью k_B агент переходит в группу B (МБТ+ больные), иначе (с вероятностью $(1 - k_B)$) – в группу T (МБТ- больные).

При переходе агента в состояние T (МБТ- больной) производятся следующие операции:

1. Вычисляются времена до прогрессирования к МБТ+ стадии (t_{Tprog}), гибели от ТБ ($t_{T\mu}$), спонтанного самоизлечения (t_{Tcure}) и пассивного выявления (t_{Tpass}):

$$t_{Tprog} = -\frac{\ln(1 - \text{rand}(0, 1))}{\gamma_T}, \quad (7)$$

$$t_{T\mu} = -\frac{\ln(1 - \text{rand}(0, 1))}{\mu_T}, \quad (8)$$

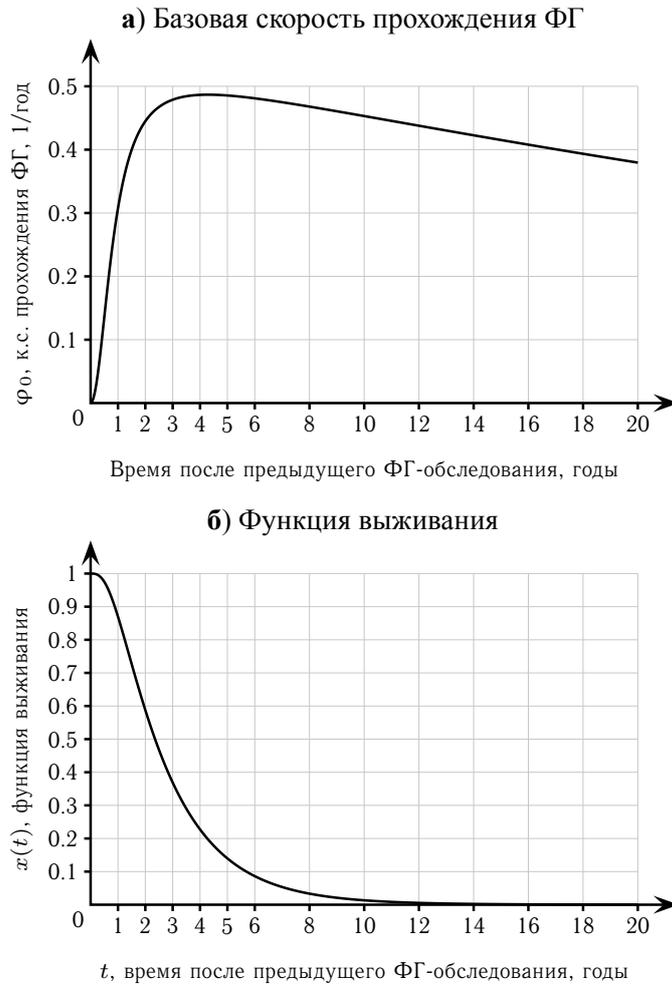


Рис. 5. Базовые параметры прохождения флюорографических обследований (ФГ): **а)** базовая скорость прохождения флюорографических обследований $\varphi_0(t)$ как функция времени после предыдущего обследования, **б)** соответствующая $\varphi_0(t)$ функция выживания, т. е. доля популяции, не прошедшая очередное ФГ-обследование в течение времени t после предыдущего обследования.

$$t_{Tcure} = -\frac{\ln(1 - \text{rand}(0, 1))}{\omega_T}, \quad (9)$$

$$t_{Tpass} = -\frac{\ln(1 - \text{rand}(0, 1))}{\varphi_{TP}}, \quad (10)$$

где $\text{rand}(0, 1)$ – это случайное число, равномерно распределенное на $[0, 1]$. Вывод формул (7–10) см. в Приложении.

2. В качестве следующего события для данного агента выбирается ближайшее:

$$[\text{event}] = \text{argmin}([t_{Tprog}, t_{T\mu}, t_{Tcure}, t_{Tpass}, t_{FGnext}, t_{death}]).$$

При переходе агента в состояние B (МБТ+ больной) производятся следующие операции:

1. Вычисляются времена до регрессирования к МБТ- стадии (t_{Bregr}), гибели от ТБ

($t_{B\mu}$), спонтанного самоизлечения (t_{Bcure}) и пассивного выявления (t_{Bpass}):

$$t_{Bregr} = -\frac{\ln(1 - \text{rand}(0, 1))}{\delta_B}, \quad (11)$$

$$t_{B\mu} = -\frac{\ln(1 - \text{rand}(0, 1))}{\mu_B}, \quad (12)$$

$$t_{Bcure} = -\frac{\ln(1 - \text{rand}(0, 1))}{\omega_B}, \quad (13)$$

$$t_{Bpass} = -\frac{\ln(1 - \text{rand}(0, 1))}{\varphi_{BP}}, \quad (14)$$

где $\text{rand}(0, 1)$ – это случайное число, равномерно распределенное на $[0, 1]$. Вывод формул (11–14) см. в Приложении.

2. В качестве следующего события для данного агента выбирается ближайшее:

$$[event] = \text{argmin}([t_{Bregr}, t_{B\mu}, t_{Bcure}, t_{Bpass}, t_{FGnext}, t_{death}])$$

При наступлении самоизлечения или гибели агента он изымается из модельной популяции. При наступлении времени следующего ФГ-обследования агент считается активно выявленным (т.е. предполагаются 100 %-ые чувствительность и специфичность ФГ). При наступлении времени выявления больного он заносится в список выявленных больных с фиксированием бактериовыделения на момент выявления (МБТ- или МБТ+) и обстоятельств выявления (активное или пассивное) и так же изымается из модельной популяции. Это также подразумевает предположение об идеальном тестировании больных на бактериовыделение в момент выявления.

Упрощающее допущение об идеальной точности ФГ-обследования можно обосновать тем, что общая скорость прохождения этого обследования φ будет свободным параметром при настройке модели (см. уравнение (15)) – это косвенно учитывает неидеальную чувствительность обследования, а также тем, что лица с положительным результатом ФГ затем проходят дополнительное обследование на ТОД и не попадают в базу данных как “здоровые на предыдущей ФГ” в случае ложноположительного результата – это компенсирует неидеальную специфичность ФГ-обследования. Идеальность тестирования на бактериовыделение в рамках модели можно интерпретировать как постулирование идентичности границы между модельными МБТ- и МБТ+ случаями и уровня чувствительности лабораторной диагностики бактериовыделения в г. Москве.

Таким образом, построенная агентная математическая модель фактически является Монте-Карло-реализацией модели, которую можно описать следующей системой дифференциальных уравнений:

$$\begin{aligned} \frac{dT}{dt} &= (1 - k_B)I - (\gamma_T + \mu_T + \omega_T + \varphi(t - t_{FGprev}) + \varphi_{TP} + \mu(t - t_{birth}))T + \delta_B B, \\ \frac{dB}{dt} &= k_B I + \gamma_T T - (\delta_B + \mu_B + \omega_B + \varphi(t - t_{FGprev}) + \varphi_{BP} + \mu(t - t_{birth}))B, \end{aligned}$$

где I – заболеваемость ТОД, t_{FGprev} – время предыдущего ФГ-обследования, $\mu(s)$ – смертность от нетуберкулёзных причин в зависимости от возраста s и t_{birth} – время рождения индивида.

Вышеописанная агентная модель была реализована в однопоточном (непараллельном) варианте и просчитывалась на настольной ЭВМ на базе процессора Intel Core

i7-8700. Несмотря на то, что эта модель, опирающаяся на принцип конкуренции независимых процессов и определение времён до событий, имела на порядки большую производительность, чем “наивная” модель с прямой симуляцией вероятностей событий на каждом шаге по времени, итоговый уровень производительности наложил серьезные ограничения на количество прогонов модели, используемых в вычислительных экспериментах. В дальнейшем предполагается создание параллельной версии агентной модели (в т. ч. с использованием GPGPU-вычислений), что позволит радикально увеличить производительность.

ВЫЧИСЛИТЕЛЬНЫЕ ЭКСПЕРИМЕНТЫ

Для сравнения параллельной и последовательной схем моделирования течения ТОД использовалась вышеописанная математическая модель с тремя наборами параметров (табл. 1): параллельной схеме соответствовали близкие друг к другу параметры из работ [7, 8], а последовательной – параметры из работы [12].

В первую очередь были произведены расчёты при базовых значениях параметров процесса выявления больных, указанных в таблице 1. Эти значения были выбраны на основе физического смысла задачи и соответствия по порядку величины доле активно выявленных среди всех выявленных больных. С каждым набором параметров производилось 2000 симуляций, в рамках каждой симуляции генерировалось 100 тыс. больных. Медианы и распределения долей МБТ+ больных для этих 2000 симуляций показаны на рисунке 6.

Затем была произведена приближительная настройка трёх вариантов математической модели на реальные данные за 2013-2018 гг. Свободными параметрами при настройке модели были параметры процесса выявления больных: скорости пассивного выявления МБТ- и МБТ+ больных – φ_{TP} и φ_{BP} , а также коэффициент a , масштабирующий скорость активного выявления $\varphi(t)$:

$$\varphi(t) = a\varphi_0(t), \quad (15)$$

где $\varphi_0(t)$ – базовая зависимость скорости активного выявления от времени после предыдущего ФГ-обследования (рис. 5). В силу недетерминированности решений агентной симуляции, применение стандартных алгоритмов оптимизации было невозможно. Поэтому приближительная настройка проводилась методом случайного поиска (Монте-Карло): многократно генерировались случайные наборы значений свободных параметров из заданного диапазона, вычислялись решения модели с каждым набором параметров и выбирались наборы параметров, давшие наилучшее в смысле средних квадратов приближение реальных данных решением модели. В качестве целевых реальных данных использовалась зависимость доли МБТ+ больных среди выявленных больных от времени после предыдущего ФГ-обследования (для больных, бывших здоровыми при предыдущем обследовании) – см. рисунок 2,а. Иначе говоря, минимизировалась функция

$$S^2 = \sum_{i=1}^4 \left(\frac{\Phi_i^B}{\Phi_i^T + \Phi_i^B} - \frac{\widehat{\Phi}_i^B}{\widehat{\Phi}_i^T + \widehat{\Phi}_i^B} \right)^2,$$

где Φ_i^B – суммарное количество модельных выявленных МБТ+ больных в i -ой группе по давности ФГ, Φ_i^T – аналогичное количество модельных МБТ- больных, а величины с “крышками” – это соответствующие величины в реальных данных.

Для каждого из трёх рассматриваемых вариантов математической модели производилось 4000 расчётов со случайными наборами свободных параметров, по

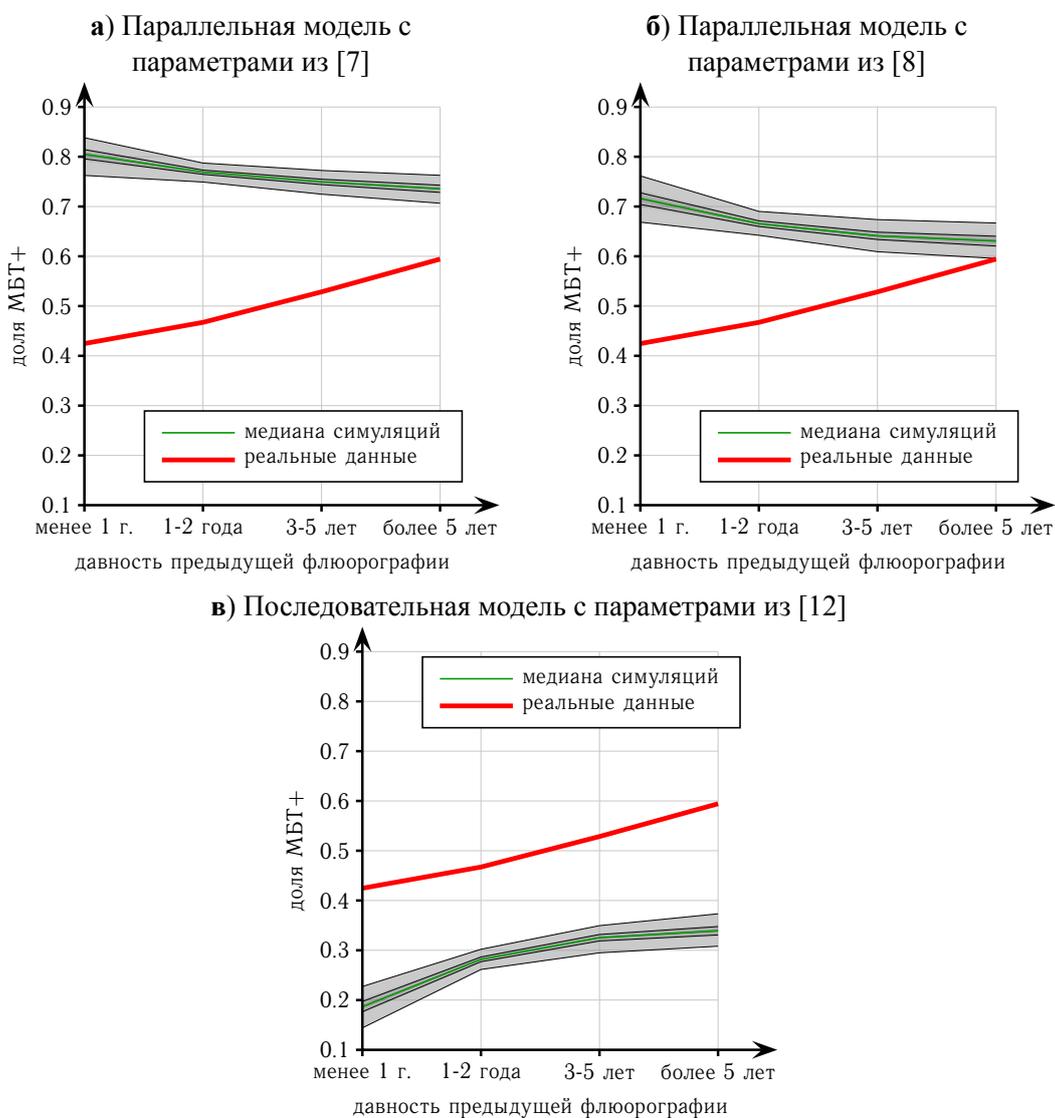


Рис. 6. Результаты расчётов трёх вариантов математической модели с базовыми значениями параметров выявления (табл. 1): доля больных с бактериовыделением (МБТ+) среди выявленных больных. Сравнение с реальными данными для г. Москва, 2013–2018 гг. Тонкими черными линиями показаны 0-, 20-, 80- и 100-ые процентиля распределений 2000 прогонов каждого варианта моделей: **а)** параллельная модель с параметрами из [7], **б)** параллельная модель с параметрами из [8], **в)** последовательная модель с параметрами из [12].

Таблица 2. Наилучшие найденные значения свободных параметров

Обозначение	Описание	Знач-е в паралл. модели [7]	Знач-е в паралл. модели [8]	Знач-е в посл. модели [12]	Расчётный диапазон	Размерность
a	масштабирующий коэффициент скорости активного выявления	0.0745	0.157	0.245	[0.06,6.6]	—
φ_{TP}	скорость пассивного выявления МБТ- больных	0.785	0.837	0.127	[0.01,1]	1/год
φ_{BP}	скорость пассивного выявления МБТ+ больных	0.353	0.344	0.996	[0.3,2]	1/год

100 тыс. больных в каждой симуляции. Наилучшие найденные значения свободных параметров приведены в таблице 2. Результаты настройки показаны на рисунке 7.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ ВЫЧИСЛИТЕЛЬНЫХ ЭКСПЕРИМЕНТОВ

Результаты расчётов как с базовыми значениями параметров выявления (рис. 6), так и с приблизительно-оптимальными значениями (рис. 7) с очевидностью показывают, что ни одна из трёх рассмотренных моделей течения нелеченного ТОД не даёт хорошего приближения к реальным данным. При этом различия результатов симуляций и реальных данных не только количественные, но и качественные: кривые зависимости доли МБТ+ больных от времени после предыдущего ФГ-обследования имеют выпуклость в разные стороны.

При этом отметим ряд наблюдений:

- При базовых значениях параметров выявления:
 - Модель с последовательной схемой воспроизводит рост доли МБТ+ больных с ростом времени после предыдущего ФГ-обследования.
 - Обе модели с параллельной схемой дают радикально противоречащее реальным данным падение этой доли (рис. 6).
- При параметрах выявления, полученных в результате приблизительной настройки:
 - Последовательная модель приближает реальные данные по доле МБТ+ больных по порядку величины (рис. 7,в) при достаточно правдоподобных значениях параметров, но не воспроизводит выпуклость данных вниз.
 - Обе параллельные модели хотя и воспроизводят рост доли МБТ+ больных с ростом времени после предыдущего ФГ-обследования (рис. 7,а и 7,б), но делают это лишь при “нефизичных” значениях параметров выявления, когда скорость пассивного выявления МБТ- больных значительно больше таковой для МБТ+ больных (табл. 2). Поскольку субъективное самочувствие МБТ+ больных несомненно в среднем хуже, чем у МБТ- больных, скорость пассивного выявления (т.е. выявления при обращении с жалобами на здоровье) должна быть выше у МБТ+ больных.
 - Модель с параметрами из работы [7] не позволяет количественно приблизить долю МБТ+ больных даже по порядку величины (рис. 7,а).
 - Все три модели дают оценки скорости активного выявления намного меньше ожидаемой ($a \ll 1$), что в итоге даёт общую долю активно выявленных больных намного меньше реальной и значительный сдвиг

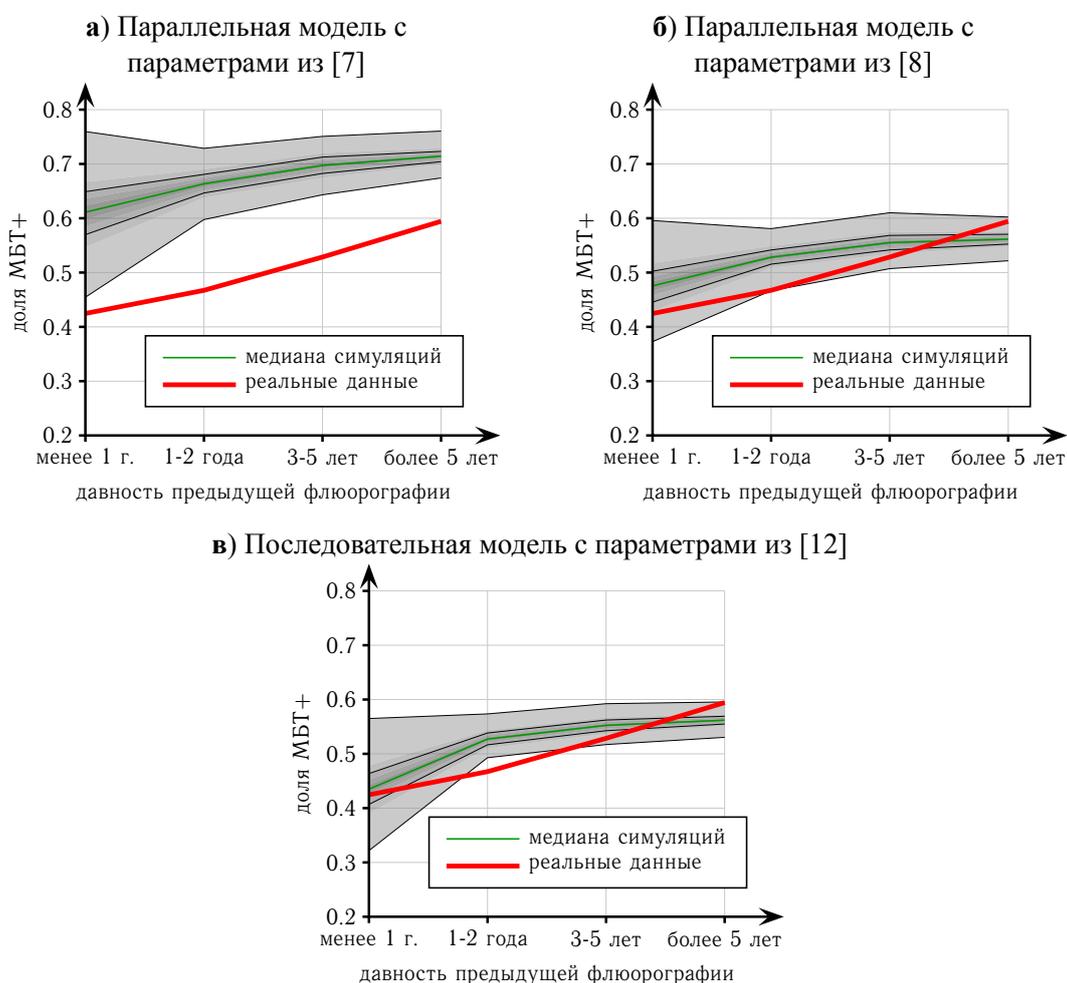


Рис. 7. Результаты расчётов трёх вариантов математической модели с наилучшими найденными значениями параметров выявления (табл. 2): доля больных с бактериовыделением (МБТ+) среди выявленных больных. Сравнение с реальными данными для г. Москва, 2013–2018 гг. Тонкими черными линиями показаны 0-, 20-, 80- и 100-ые процентиля распределений 2000 прогонов каждого варианта моделей: **а)** параллельная модель с параметрами из [7], **б)** параллельная модель с параметрами из [8], **в)** последовательная модель с параметрами из [12].

распределения количества выявленных больных в сторону больших времён после предыдущего ФГ-обследования по сравнению с реальными данными (результаты не показаны). При этом последовательная модель даёт самую большую оценку скорости активного выявления.

Эти наблюдения позволяют сделать следующие выводы:

- Последовательная схема моделирования течения нелеченного ТОД ближе к реальности, чем параллельная.
- Подмодели как активного, так и пассивного выявления больных ТОД нуждаются в доработке и верификации.
- Коэффициенты (и, возможно, структуру) подмодели течения нелеченного ТОД следует оптимизировать на основе реальных данных. В качестве данных для настройки модели следует использовать количества выявленных больных с разделением одновременно по бактериовыделению и методу выявления (рис. 2,б). Ограниченная вычислительная производительность агентной модели не позволила использовать такой подход на текущем этапе исследований.

ЗАКЛЮЧЕНИЕ

Данная работа является первым этапом исследований по обратному инжинирингу количественной модели патогенеза туберкулёза по имеющимся данным для современного населения. Подавляющее большинство параметров и схем патогенеза нелеченного ТБ, используемых при математическом моделировании, опираются на данные доантибиотиковой эпохи (первая половина XX века), когда не существовало эффективных лекарств от туберкулёза, а потому наблюдение за фактически нелеченными больными в туберкулёзных санаториях было этически приемлемым. В современных условиях такое невозможно. Вместе с тем, параметры естественного течения туберкулёза для жителей современных мегаполисов могут сильно отличаться от таковых для населения начала XX века: патогенез туберкулёза зависит от качества питания [14, 15], уровня иммуносупрессии [15] (которая, в свою очередь, зависит в том числе от уровня стресса), воздействия загрязнения воздуха [15] и т.д. Все эти факторы могли значительно измениться за прошедший век вместе с изменением уклада жизни. Поэтому использующаяся в России и на постсоветском пространстве практика регулярных флюорографических осмотров всего населения предоставляет уникальный материал для оценки динамики туберкулёза в современных условиях, т. к. даёт пусть и косвенную, но актуальную и объективную информацию о течении нелеченной болезни от момента её развития и до момента выявления больного.

Полученные нами результаты свидетельствуют о том, что последовательная схема моделирования течения туберкулёза, использовавшаяся ранее немногочисленными зарубежными исследователями и нашим научным коллективом, ближе к реальности, чем параллельная схема, использовавшаяся в большинстве других работ (включая работы, на которых основываются рекомендации Всемирной Организации Здравоохранения). Вместе с тем, выявлена необходимость значительной корректировки использовавшихся ранее математических моделей и их параметров.

Работа выполнена при поддержке гранта РФФИ 18-01-00856 А.

ПРИЛОЖЕНИЕ 1.

Генерация параметров агентов по распределениям

В этом приложении обосновываются формулы (1)–(14), задающие генерацию параметров агентов, и описывается их численная реализация.

Рассмотрим подмодель, описывающую некую группу индивидов численностью $x(t)$ в момент времени t , из которой индивиды выбывают с удельной скоростью $\mu(t)$ ($0 \leq \mu(t) < \infty$, функция $\mu(t)$ – достаточно гладкая). Тогда их численность подчиняется уравнению

$$\frac{dx(t)}{dt} = -\mu(t)x(t). \quad (16)$$

Если принять начальное условие

$$x(0) = 1, \quad (17)$$

то $x(t)$ будет функцией выживания для этой подмодели. Решением задачи Коши (16)–(17) для $t \in [0, \infty)$ будет

$$x(t) = \exp\left(-\int_0^t \mu(s)ds\right). \quad (18)$$

Введем дополнительное предположение, что функция $\mu(t)$ такова, что

$$\lim_{t \rightarrow +\infty} x(t) = 0. \quad (19)$$

Например, это выполняется при предположении, что существуют такое t_0 и $\mu_{min} > 0$, что $\mu(t) > \mu_{min} > 0$ при $\forall t > t_0$.

Время покидания индивидами группы будет иметь плотность распределения

$$f_{t \text{ покид}}(t) = \frac{-\frac{dx(t)}{dt}}{-\int_0^{+\infty} \frac{dx(s)}{ds} ds} = -\frac{dx(t)}{dt} = \mu(t)x(t) \quad (20)$$

и, соответственно, функцию распределения

$$F_{t \text{ покид}}(t) = \frac{-\int_0^t \frac{dx(s)}{ds} ds}{-\int_0^{+\infty} \frac{dx(s)}{ds} ds} = 1 - x(t). \quad (21)$$

Для генерации реализации случайной величины, имеющей распределение (21) – то есть времени до события, происходящего с удельной скоростью $\mu(t)$ – необходимо решить уравнение $\text{rand}(0, 1) = F_{t \text{ покид}}(t)$, где $\text{rand}(0, 1)$ – это случайное число, равномерно распределенное на $[0, 1]$. При постоянной удельной скорости $\mu(t) = \mu = \text{const} > 0$ это приводит к решению

$$t = -\frac{\ln(1 - \text{rand}(0, 1))}{\mu}, \quad (22)$$

которое даёт формулы (7)–(14) в основном тексте статьи. По той же схеме обращения функции распределения $F_{t \text{ покид}}(t) = 1 - x(t)$ генерируется общая продолжительность жизни t_{life} создаваемых агентов.

Теперь рассмотрим подмодель с возвращением, в которой индивиды после наступления события, в предыдущей подмодели исключавшего индивида из популяции, возвращаются в популяцию в момент времени $t = 0$ (это соответствует процессу прохождения флюорографических обследований с удельной скоростью $\mu(t)$, где t – время после предыдущего обследования). Затем в случайный момент времени производится выбор случайного индивида из популяции (в случае основной модели – развитие

туберкулёзной болезни, которое, как предполагается, статистически независимо от прохождений флюорографических обследований).

Найдем распределение длительностей между обследованиями для случайно выбранного индивида. Плотность этого распределения $f_{FG}(t)$ должна быть пропорциональна $t f_{t \text{ покид}}(t)$, поскольку вероятность выбора индивида пропорциональна длине интервала между обследованиями t . Таким образом, производя необходимую нормировку, получаем плотность распределения

$$f_{FG}(t) = \frac{t\mu(t)x(t)}{\int_0^{+\infty} s\mu(s)x(s)ds}, \quad (23)$$

совпадающую с формулами (2)–(3) при замене $\mu(t)$ на $\varphi(t)$. Генерация времени между обследованиями для случайно выбранного индивида осуществляется обращением соответствующей функции распределения

$$F_{FG}(t) = \frac{\int_0^t s\mu(s)x(s)ds}{\int_0^{+\infty} s\mu(s)x(s)ds}. \quad (24)$$

Поскольку время начала болезни статистически независимо от прохождения флюорографических обследований, время начала болезни будет равномерно распределено на интервале между обследованиями. Поэтому генерация времени после предыдущего обследования и времени до следующего обследования на момент развития болезни осуществляется путём разделения ранее сгенерированного времени между обследованиями (t_{FG}) в пропорции, определяемой равномерно распределенным случайным числом $r = \text{rand}(0, 1)$:

$$t_{FGprev} = -rt_{FG}, \quad (25)$$

$$t_{FGnext} = (1 - r)t_{FG}, \quad (26)$$

где под $t = 0$ понимается момент развития болезни, а t_{FGprev} и t_{FGnext} являются временами предыдущего и следующего флюорографических обследований. Это соответствует формулам (4)–(6) в основном тексте.

При практической реализации моделей обращение функций распределения $F_{t \text{ покид}}$ и F_{FG} происходит численно, за счёт формирования таблиц поиска с регулярной сеткой.

Сначала формируется сетка по времени $\{t_i\}$, $i = 1, \dots, N$, покрывающая физически значимый интервал $[T_1, T_{end}]$ (т.е. $t_1 = T_1$, $t_N = T_{end}$), причем сетка может быть неравномерной с концентрацией в областях с большей значимостью. Затем вычисляются сеточные значения $x_i = x(t_i)$ – либо аналитически, либо за счёт численного интегрирования в формуле (18). Для единообразности вычислений присваивается значение $x_N = 0$, т.е. распределение усекается на T_{end} , что вносит малую ошибку, если T_{end} выбрано достаточно большим. После этого путём численного интегрирования строятся табличные функции $F_{t \text{ покид } i} = F_{t \text{ покид}}(t_i)$ и $F_{FG i} = F_{FG}(t_i)$, и для равномерной сетки $r_j = j\delta_r$, $j = 0, \dots, 1/\delta_r$, $\delta_r = 10^{-6}$ решаются уравнения $r_j = F_{t \text{ покид}}(t_{t \text{ покид } j}^*)$ и $r_j = F_{FG}(t_{FG j}^*)$ за счёт обращения кусочно-линейного представления соответствующих табличных функций. В итоге, для заданной реализации случайной величины $r = \text{rand}(0, 1)$ соответствующие ей значения $t_{t \text{ покид}}^*$ и t_{FG}^* вычисляются по линейной интерполяции между узлами с номерами $j_1 = \lfloor r/\delta_r \rfloor$ и $j_2 = j_1 + 1$ для табличных функций $t_{t \text{ покид } j}^*$ и $t_{FG j}^*$ соответственно.

При достаточном количестве узлов сетки по времени (N) и сетки по случайной величине ($1/\delta_r$) ошибки, вносимые таким приближением, оказываются много меньше

точности, с которой, с точки зрения физики задачи, известны обрабатываемые распределения.

СПИСОК ЛИТЕРАТУРЫ

1. Андерсон Р., Мэй Р. *Инфекционные болезни человека. Динамика и контроль*. Москва: Мир, “Научный мир”, 2004. 784 с. ISBN: 5-03-003552-4.
2. Авилов К.К., Романюха А.А. Математические модели распространения и контроля туберкулёза. *Математическая биология и биоинформатика*. 2007. Т. 2. № 2. С. 188–318. doi: [10.17537/2007.2.188](https://doi.org/10.17537/2007.2.188).
3. *Фтизиатрия: национальное руководство*. Под ред. М.И. Перельмана. М.: ГЭОТАР-Медиа, 2010. 512 с. ISBN: 978-5-9704-1232-9.
4. Литвинов В.И. *Латентная туберкулезная инфекция: свойства возбудителя; реакции макроорганизма; эпидемиология и диагностика (IGRA-тесты, Диаскинтест и другие подходы), лечение*. М.: Департамент здравоохранения г. Москвы, МНПЦ борьбы с туберкулезом, 2016. 194 с.
5. Tiemersma E.W., van der Werf M.J., Borgdorff M.W., Williams B.G., Nagelkerke N.J.D. Natural history of tuberculosis: duration and fatality of untreated pulmonary tuberculosis in HIV negative patients: a systematic review. *PLOS One*. 2011. V. 6. Article No. e17601. doi: [10.1371/journal.pone.0017601](https://doi.org/10.1371/journal.pone.0017601).
6. Dannenberg A.M. Jr. Pathophysiology: basic aspects. In: *Tuberculosis and non-tuberculous mycobacterial infections*. Ed. Schlossberg D. Philadelphia, PA, USA: WB Saunders, 1999. P. 17–47.
7. Blower S.M., McLean A.R., Porco T.C., Small P.M., Hopewell P.C., Sanchez M.A., Moss A.R. The intrinsic transmission dynamics of tuberculosis epidemics. *Nature Medicine*. 1995. V. 1. No. 8. P. 815–821. doi: [10.1038/nm0895-815](https://doi.org/10.1038/nm0895-815).
8. Baltussen R., Floyd K., Dye C. Cost effectiveness analysis of strategies for tuberculosis control in developing countries. *BMJ*. 2005. V. 331. No. 7529. P. 1364. doi: [10.1136/bmj.38645.660093.68](https://doi.org/10.1136/bmj.38645.660093.68).
9. Brogger S. Systems analysis in tuberculosis control: a model. *Am. Rev. Respir. Dis.* 1967. V. 95. P. 419–434.
10. Perelman M.I., Marchuk G.I., Borisov S.E., Kazennykh B.Ya., Avilov K.K., Karkach A.S., Romanyukha A.A. Tuberculosis epidemiology in Russia: the mathematical model and data analysis. *Russ. J. Numer. Anal. Math. Modelling*. 2004. V. 19. P. 305–314. doi: [10.1515/1569398041974905](https://doi.org/10.1515/1569398041974905).
11. Yaesoubi R., Cohen T. Identifying dynamic tuberculosis casefinding policies for HIV/TB coepidemics. *Proc. Natl. Acad. Sci. USA*. 2013. V. 110. P. 9457–9462. doi: [10.1073/pnas.1218770110](https://doi.org/10.1073/pnas.1218770110).
12. Avilov K.K., Romanyukha A.A., Borisov S.E., Belilovsky E.M., Nechaeva O.B., Karkach A.S. An approach to estimating tuberculosis incidence and case detection rate from routine notification data. *The International Journal of Tuberculosis and Lung Disease*. 2015. V. 19. N. 3. P. 288–294. doi: [10.5588/ijtld.14.0317](https://doi.org/10.5588/ijtld.14.0317).
13. *Российская база данных по рождаемости и смертности*. Центр демографических исследований Российской экономической школы, Москва (Россия). URL: http://demogr.nes.ru/index.php/ru/demogr_indicat/data (дата обращения: 20.08.2019).

14. WHO. *Guideline: Nutritional Care and Support for Patients With Tuberculosis*. Geneva: World Health Organization, 2013. ISBN: 978-92-4-150641-0.
15. Narasimhan P., Wood J., Macintyre C. R., Mathai D. Risk factors for tuberculosis. *Pulmonary medicine*. 2013. Article ID 828939. doi: [10.1155/2013/828939](https://doi.org/10.1155/2013/828939).

Рукопись поступила в редакцию 25.09.2019.
Переработанный вариант поступил 21.11.2019.
Дата опубликования 05.12.2019.