

Текст · 4 августа 2021, 08:01

Максим Литавин, Михаил Зеленский, Александр Ердников, Дмитрий Кузнецов, Светлана Рейдерис Дмитриев,

Минздрав отказался рассказывать о своем закрытом коронавирусном регистре. Мы сделали это за него

«Медуза», вместе с которой мы продолжали наше исследование, внесена Минюстом в реестр «иностранных средств массовой информации, выполняющих функции иностранного агента». Поэтому в соответствии с приказом Роскомнадзора мы обязаны предварить публикацию следующим уведомлением: **«ДАННОЕ СООБЩЕНИЕ (МАТЕРИАЛ) СОЗДАНО И (ИЛИ) РАСПРОСТРАНЕНО ИНОСТРАННЫМ СРЕДСТВОМ МАССОВОЙ ИНФОРМАЦИИ, ВЫПОЛНЯЮЩИМ ФУНКЦИИ ИНОСТРАННОГО АГЕНТА, И (ИЛИ) РОССИЙСКИМ ЮРИДИЧЕСКИМ ЛИЦОМ, ВЫПОЛНЯЮЩИМ ФУНКЦИИ ИНОСТРАННОГО АГЕНТА»-** . Размер шрифта этого уведомления также регламентирован приказом Роскомнадзора.

В первом тексте мы рассказывали, каким образом коронавирусные сертификаты на «Госуслугах» позволили понять, как ведется учет в закрытых регистрах Минздрава. Мы писали, что нашли в регистре для больных коронавирусом 29 млн записей.

Минздрав назвал наши выводы «не более чем фантазиями», вопросы к данным возникли и у независимых аналитиков — например, о том, действительно ли нумерация в регистре последовательна, нет ли там большого числа пустых

записей и возможна ли перезапись удаленных номеров. И о том, попадает ли в реестр именно коронавирус.

В основе первой публикации лежало ограниченное число собранных у читателей сертификатов, которые не позволяли ответить на эти вопросы. На этот раз мы применили другой метод, который позволяет ответить на претензии Минздрава и снять основные вопросы к качеству наших данных.

Закрытый регистр, в который можно подсмотреть

По словам знакомого с устройством информационного ресурса источника «Медузы», доступа к Федеральному регистру больных коронавирусной инфекцией^[1] нет даже у технических специалистов Минздрава; полные данные могут получить только члены правительства и руководства администрации президента.

Тем не менее, мы нашли способ собрать статистику по электронным сертификатам, выданным тем, кто переболел, вакцинировался или просто сдал ПЦР-тест.

Это стало возможным из-за специфики сайта «Госуслуг». Если перейти по ссылке из QR-кода сертификата и открыть HTML-код страницы, одна из ссылок будет вести на техническую страницу «Госуслуг». В адресной строке можно подставить другой номер сертификата — и тогда сервер ответит, существует сертификат с таким номером или нет. Никаких данных из него доступно не будет (для этого надо знать уникальный код безопасности).

Изучая сертификаты этим методом, мы выяснили, что к концу июля в коронавирусных регистрах Минздрава и Роспотребнадзора^[2] было:

- 30 миллионов 270 тысяч записей о заболевших;
- 35 миллионов 780 тысяч записей о вакцинированных;
- 13 миллионов 580 тысяч записей о ПЦР-тестах^[3];

Мы дополнительно подтвердили, что уникальный номер из регистра при попадании на «Госуслуги» становится частью номера сертификата — без каких бы то ни было искажений. Петербурженка^[4]

переболела коронавирусом в феврале, но сертификат получила не сразу. Для этого в июле ей пришлось сначала обратиться в поликлинику, а потом и написать жалобу городским властям. В одном из ответов ей сообщили, что «данные о перенесенном заболевании в ЕПГУ^[5] внесены корректно» и

назвали ее уникальный номер регистровой записи (УНРЗ) — таким термином обозначают записи в регистрах федеральных ведомств. Когда через неделю сертификат все же появился на «Госуслугах», его номер полностью совпал с УНРЗ.

Полученные нами данные о числе записей в регистрах совпали с выводами по собранным вручную сертификатам. Вот как на основе нашего массива данных выглядит накопление порядковых номеров заболевших с момента начала эпидемии в России.

Но важнее, что новые данные позволили нам ответить на возражения Минздрава и возникшие у наших читателей сомнения.

Что, если это единая база данных, где смешаны переболевшие, вакцинированные и кто-то еще?

В своем единственном комментарии после нашей публикации Минздрав утверждает, что «информационный ресурс» — один на всех (ответ процитирован в газете «Коммерсант», в канале *Life Shot* его можно прочесть полностью), туда заносят «заболевших *CoViD-19*, пневмониями, вакцинированных и другие категории записей». И всего в этом ресурсе «более 39 миллионов» записей.

Из собранных данных следует, что это точно не какая-то одна большая таблица на все случаи, а независимые регистры^[6]. Потому что во всех трех проверенных нами типах сертификатов попадаются одинаковые порядковые номера, например:

— 7000 0000 **1304 6765** — сертификат ПЦР-теста, скорее всего^[7], сделанного в середине июля 2021 года.

— 8661 1220 **1304 6765** — сертификат переболевшего, диагноз которому поставили в декабре 2020 года^[8] в Свердловской области^[9].

— 9490 0000 **1304 6765** — сертификат о вакцинации в Магаданской области^[10], который по своему порядковому номеру относится к началу мая 2021 года.

Неясно, из чего складывается названное Минздравом число 39 миллионов. Как отмечает аналитик открытых данных Александр Драган, оно не сходится даже с официальными данными о привитых и заболевших — только они дают в сумме более 43 миллионов записей.

Что, если номера в этом регистре — не порядковые?

«Уникальный номер регистровой записи не является сквозным порядковым номером — алгоритм его формирования нелинейный и учитывает несколько факторов», — говорится в комментарии Минздрава. Похожий вопрос возник и у читателей, подробно он сформулирован, к примеру, в посте аналитика открытых данных Александра Драгана.

В коде сертификата переболевшего всего восемь уникальных цифр^[11], обозначающих порядковый номер, поэтому мы случайным образом проверили большое количество номеров от 0000 0000 до 9999 9999. Вот результат распределения найденных номеров по месяцам (месяц указан отдельно от номера во второй группе цифр сертификата).

Как видно из графика, номеров больше 30 миллионов среди доступных сертификатов нет. Номера меньше 10 миллионов практически недоступны — если человек переболел больше полугода назад, его сертификат уже недействителен и пропадает с «Госуслуг».

Несмотря на это, мы нашли единичные старые сертификаты, в том числе за апрель 2020 года, когда эпидемия только начиналась — например, доступен московский сертификат с номером 8771 **0420 0017 5725**. Мы предполагаем, что в регистровых записях таких пациентов допустили какую-то ошибку^[12], из-за которой они считаются «вечно» переболевшими — и сертификат остается действительным.

Редкие отклонения номеров в большую сторону объясняются поздним внесением записи в регистр — так переболевший, например, в апреле получал номер, по порядку соответствующий июню.

Свидетельства того, что записи идут по порядку, есть и в документации к коронавирусному и другим регистрам. На сайте петербургского МИАЦ в разделе о работе с федеральным регистром по *COVID-19* есть таблица ([xlsx](#)) со структурой записей в этой базе данных. В ней указано, что тип строки *unrz* (уникальный номер регистрационной записи) — это *bigserial*. Это числовой [тип](#) данных в системах управления базами данных — например, в PostgreSQL^[13] — который [автоматически](#) генерирует последовательные числовые значения для каждой новой записи.

Версию с использованием порядковых номеров подтверждает и [видеоинструкция](#) Минздрава к ковидному регистру — там приведены три последовательных^[14] номера УНРЗ. В [pdf-файле](#) с инструкцией первые два из этих номеров используют для объединения дублей. Можно предположить, что их просто создали один за другим,

чтобы продемонстрировать доступный функционал. Во время исследования мы не раз находили сертификаты с идущими друг за другом номерами.

Представитель Минздрава на просьбу корреспондента «Медузы» рассказать о принципе нумерации в регистре ответил вопросом:

— Вас же не волнует, как ГИБДД выдает номера?

— Но машины не лежат под ИВЛ.

— Вам интересно, как работает генератор случайных чисел?

— То есть регистр по такому принципу работает?

— Присылайте запрос.

Ответ на запрос, отправленный за две недели до выхода этого текста, редакция так и не получила. «Медиазона» также спросила Минцифры о том, сколько всего было выдано сертификатов на «Госуслугах» и по какой причине они могут быть недоступны. В ведомстве обещали ответить, но несколько раз переносили сроки и к моменту выхода статьи так ничего и не сказали.

Что, если номера порядковые, но в самой базе есть пропуски?

Александр Драган и другие специалисты предположили, что в регистре Минздрава возможны пропуски — в силу технических особенностей построения баз данных. Это означало бы, что даже если номер 30 миллионов 270 тысяч действительно

последний по порядку, то в самой базе не 30 миллионов записей с указанием больных, а гораздо меньше.

Чтобы проверить эту версию, мы более подробно изучили диапазон номеров, выданных за последние полгода: от 10 до 30 миллионов. Как видно из общего графика, больших дыр в миллионы или даже десятки тысяч записей в нумерации нет^[15].

Embed Wrapper Not Realised
Embed Wrapper Not Realised
Embed Comment Not Realised

Можно было бы предположить, что в нумерации нет намеренных пропусков, но есть квоты^[16] — например, для регионов. Тогда если Москве выделен диапазон с 13 до 14 миллионов, никакой другой регион в этот диапазон своих больных записать не может. Проверив две тысячи номеров по порядку, мы выяснили, что таких квот в регистре тоже нет.

Embed Wrapper Not Realised
Embed Wrapper Not Realised
Embed Comment Not Realised

При этом в регистре, как и в любой базе данных, из-за ошибок действительно могут возникать пустоты в последовательных идентификаторах. Например, в [документации PostgreSQL](#) указано, что порядковый номер для записи — в нашем случае для УНРЗ — будет считаться задействованным, даже если данные внести не удалось.

Такое может иногда происходить из-за какой-то случайной ошибки, либо из-за постоянной ошибки

в коде программы, которая осуществляет запись в регистр. Во втором случае пропуски будут появляться регулярно, например, каждая вторая запись будет пропущена или число пропущенных строк подряд будет совпадать с числом подряд записанных строк. В ходе исследования регистра мы не нашли подобных паттернов, пропуски носят случайный характер.

Пустоты могут возникать и при удалении «мусорных» записей, которые неизбежно будут появляться во время работы — система рассчитана на 44 тысячи пользователей^[17], следует из двух контрактов с Ростехом на доработку коронавирусного регистра. Работавший с регистром врач из Красноярского края рассказал «Медиазоне», что в 2020 году на совещаниях по работе с системой в их регионе говорилось, что некоторые пациенты по ошибке «задваивались» — например, когда при переводе человека из одной больницы в другую его могли внести в систему повторно. По документации к актуальной версии такого происходить не должно: при внесении данных об уже упоминавшемся человеке пользователю в первую очередь предложат обновить существующую запись.

Оценить точное количество появившихся по подобным причинам дыр невозможно без доступа к регистру. Внести ясность тут может только Минздрав.

А что с перезаписью удаленных номеров?

Во время подготовки первого текста осведомленный источник сказал «Медузе», что в случае удаления ошибочно внесенной записи ее порядковый номер

освобождается, и его могут выдать для записи нового случая. Это означало бы, что в регистре вообще нет свободных номеров — даже на месте ошибочной записи, которая была удалена при перепроверке.

Собранные нами статистические данные о сертификатах не позволяют подтвердить это утверждение.

Чтобы выяснить, не происходит ли все же перезапись пациентов, надо собрать для большого количества записей не только уникальные номера (они у нас есть), но и даты внесения в регистр. Сравнив номера на соседних датах можно было бы понять, бывают отскоки в нумерации из-за перезаписи или нет.

Более того, нашим методом возможно проверить только то, доступен определенный номер сертификата или нет — остальные данные защищены кодом безопасности, который невозможно подобрать. Но даже если бы у нас был доступ ко всей информации из всех сертификатов, мы бы не смогли проверить, как скажут номера — дело в том, что дата внесения в регистр (то есть дата постановки диагноза) вообще не передается из Минздрава в «Госуслуги» для создания сертификата переболевшего (для него нужна только дата выздоровления).

Поставить точку в вопросе о теоретически возможной перезаписи возможно только с полным доступом к регистру. А он есть только у Минздрава.

Большая часть сертификатов переболевших недоступна на «Госуслугах»

На наших графиках видно, что далеко не все проверенные номера сертификатов оказались доступны — хотя мы подставляли в них все возможные регионы и месяцы. Среди номеров от 10 до 30 миллионов, которые добавили в регистр за последние полгода, мы можем собрать через «Госуслуги» только 10,26%^[18] — на остальные номера сервер отвечает, что сертификат недоступен.

Есть несколько причин, по которым у перенесшего коронавирус может не быть сертификата на «Госуслугах»:

— если человек переболел, но потом привился — вакцинация «гасит» сертификат о болезни;

— подтвержденная запись на «Госуслугах», необходимая для получения любого вида сертификатов, есть только у половины россиян — это 78,5 млн человек (таких подтвержденных записей точно нет у детей без паспорта);

— возможны ошибки при вводе СНИЛС в регистр (тогда запись не сможет прикрепиться к «Госуслугам»), таких случаев только официально до 20%;

— существуют служебные ограничения для сотрудников правоохранительных органов и военнослужащих, их данные тоже не попадают на «Госуслуги»;

Но все перечисленные проблемы не помешали 70% вакцинированных получить свои сертификаты на «Госуслугах» — мы проверили это, изучая номера из реестра привившихся. Спад в конце графика можно объяснить тем, что номер присваивают после первой дозы (в момент появления дневника самонаблюдений), а сертификат становится доступен только после полной вакцинации — 21 день спустя. Большинство привитых после отметки в 25 миллионов еще просто не завершили вакцинацию.

Кроме того, в июне в Москве начали массово прививать иностранцев, у которых может не быть учетной записи на «Госуслугах». Для их вакцинации специально переделали спортивную арену «Лужники». Как сообщили «Медузе» в московском департаменте здравоохранения, вакцинировали уже 170 тысяч иностранцев. Помимо «Госуслуг», они могут получить электронный сертификат на портале мэрии или по смс.

Почему переболевшими считают не всех попавших в регистр?

Другой причиной малого количества видимых сертификатов может быть принцип их отбора при синхронизации «Госуслуг» с регистром. В итоговом отчете за 2020 год Минздрав писал-

, что именно коронавирусный регистр «служил основанием для принятия управленческих решений»: в него «оперативно вносилась информация о случаях заболеваний, лечении и исходах пациентов». Вероятно, это связано с тем, что специфика регистра позволяет отслеживать не только положительные ПЦР-тесты, но и подозрительные случаи и тяжесть течения болезни у пациентов.

В регистр заболевших *COVID-19* вписывают не только тех, у кого тест показал наличие коронавируса, но и больных с диагнозами из группы риска. Например, туда в обязательном порядке вносят всех больных с пневмонией. Возможно попадание в регистр и людей с ОРВИ и другими респираторными заболеваниями.

Подобные критерии используют и в других странах. Например, с точки зрения американских Центров по контролю и профилактике заболеваний (CDC), симптомы ОРВИ или пневмонии (необязательно подтвержденные снимками) у контактных пациентов или людей из групп риска — это уже основание рассматривать случай как возможный коронавирус и включать его в публичную статистику. В России требования жестче: клинически подтвержденным (без положительного теста) считается только случай с зафиксированными изменениями в легких (КТ) и прямым контактом с заболевшим (либо человек, недавно вернувшийся из-за рубежа). Всемирная организация здравоохранения тоже призывает опираться на КТ, рентген или УЗИ в сомнительных случаях.

Об этом говорят и медики, которые работали с регистром. На уровне больниц доля внесенных в реестр нековидных пациентов минимальна, рассказал «Медиазоне» красноярский врач^[19]: «В инфекционном стационаре у нас было несколько сотен человек, нековидных было один-два-пять из шестисот, предположим». По его данным, в поликлиниках в регистр заносят людей с подтвержденным ПЦР или с ковидной пневмонией: «Я думаю, вклад ОРЗ [в регистр] минимален и на поликлиническом звене». При этом медик подчеркивает, что его мнение про поликлиники «теоретическое, а не практическое».

Участковый врач^[20] сказала «Медиазоне», что вносит в регистр всех пациентов с подозрением на *COVID-19*; после этого человека направляют на ПЦР или КТ, где диагноз либо подтверждается, либо нет. Что происходит с записями дальше, она не знает, так как занимается только первичной диагностикой. По ее словам, на пиках заболеваемости коронавирусом практически в любом ОРВИ врачи подозревают *COVID-19*: «Легкий коронавирус от ОРВИ не отличается ровным счетом ничем. Поэтому могут как самовольно убрать из регистра, так и не убрать то, что убрать надо. <...> С учетом того, что легких все-таки большинство, их не отправляют на КТ, только ПЦР и симптомы. ПЦР правду показывает в 60% случаев^[21]. Все, кто ведет регистр, это понимают».

При этом по данным НИИ гриппа им. Смородинцева, который ежемесячно выделяет у тысяч пациентов

различные вирусы, с начала эпидемии в России грипп почти не распространялся (если не считать декабря и февраля); другие вирусы, вызывающие ОРВИ, иногда выявлялись чаще, чем коронавирус.

Таким образом, в теории и практике врачей есть много критериев для занесения пациента в коронавирусный реестр Минздрава. Но при передаче из реестра на «Госуслуги» учитываются только записи, в которые врач вписал *COVID-19* по лабораторным тестам^[22] или по симптомам и снимкам КТ^[23]. Никакой другой диагноз для выдачи сертификата не принимается.

И если врач в конкретном медицинском учреждении по каким-то причинам не проставил в регистре пометку о коронавирусе, то даже пациент с пневмонией из коронавирусного госпиталя не сможет получить сертификат переболевшего на «Госуслугах».

Многие больные коронавирусом могли вообще не попасть в регистр

После публикации первого текста читатель «Холода» рассказал редакции об интересном случае: «Я переболел в конце апреля, выздоровел в начале мая. Но я болел в маленьком городе и в маленькой больнице. На "Госуслугах" и других сайтах никакой информации о болезни не было, сертификата не было, хотя я официально получал больничный с диагнозом *COVID*».

14 июля читатель пошел в поликлинику, где секретарь главврача пообещала сразу внести его в регистр. Сертификат он получил уже через полтора часа. И

хотя дата в нем — апрель, порядковый номер был больше 29 миллионов, так как запись внесли только в июле.

Этот случай — точно не исключение. Мы нашли десятки сертификатов, чей порядковый номер слишком велик^[24] для месяца постановки диагноза.

Получается, что после объявления о необходимости получать QR-коды переболевшие коронавирусом люди массово пошли выяснять, почему им не выдали сертификаты.

И тогда их начали вносить в регистр со старыми месяцами (по дате постановки диагноза), но крупными порядковыми номерами. В отдельные

месяцы таких случаев — до 5%. Если предположить, что такой процент недоучета был на протяжении всей эпидемии, то в России может быть больше миллиона перенесших *COVID-19*, о которых нет никакой информации на федеральном уровне — даже в закрытых регистрах.

И что в итоге?

Если судить по собранным нами данным, только десятая часть из уже 30 миллионов записей в регистре Минздрава превращается в сертификаты переболевших — это около трех миллионов. С учетом того, что подтвержденные записи на «Госуслугах» есть лишь у половины россиян, можно считать, что эта цифра в грубом приближении совпадает с официальными сводками оперативного штаба (6,3 млн случаев на начало августа). Но в регистр не всегда заносят даже тех, кому точно ставили диагноз коронавирус.

Кроме того, оценка в несколько миллионов заболевших не согласуется ни с данными об избыточной смертности, ни с количеством официально госпитализированных коронавирусных пациентов (подробнее об этом — в тексте «Медузы»). А значит, сведения о количестве занесенных в регистр Минздрава пациентов — важный и, скорее всего, наиболее полный источник информации о реальных масштабах эпидемии. И номера сертификатов переболевших позволяют отслеживать его в реальном времени.

Текст и данные: Михаил Зеленский («Холод»), Максим Литаврин, Дмитрий Трещанин («Медиазона»), Дмитрий

Кузнец, Александр Ершов, Светлана Рейтер и Денис Дмитриев («Медуза»)

Инфографика: Кирилл Хейфец, Давид Френкель («Медиазона»)

Редакторы: Егор Сковорода, Давид Френкель («Медиазона»)

Embed Wrapper Not Realised

Embed Wrapper Not Realised

Embed Comment Not Realised

-
1. Федеральный регистр больных коронавирусной инфекцией (COVID-19) — это часть Единой государственной информационной системы в сфере здравоохранения (ЕГИСЗ), которую с 2019 года разрабатывает и поддерживает Ростех.
 2. Отвечает только за ПЦР.
 3. «Госуслуги» принимают как положительные, так и отрицательные тесты. Пока что их мало, поскольку лаборатории обязаны предоставлять данные с конца марта 2021 года, но в июле правительству пришлось ужесточить контроль — данные загружали далеко не все.
 4. Попросила не указывать ее имени. Ее переписка с ведомствами есть в распоряжении редакции.
 5. Единый портал госуслуг.
 6. Минздрав использует в своем ответе слово «ресурс», очевидно, отсылая к постановлению правительства о создании такой системы. Разные регистры — это части ресурса. Намеренно или нет, ведомство смешивает разные понятия в одно и только путает читателей.
 7. Исходя из сертификатов с близкими номерами, предоставленных нам читателями.
 8. Об этом свидетельствуют цифры 1220 в номере сертификата.
 9. В номере сертификата указан 66-й регион.

10. В номере сертификата указан 49-й регион.
11. Весь номер сертификата состоит из 16 цифр: первая цифра кодирует тип сертификата, например, «8» означает сертификат переболевшего, следующие две цифры отвечают за код региона, четвертая цифра — код получателя бюджетных средств (во всех найденными нами номерах это «1» — то есть, органы исполнительной власти субъектов Российской Федерации), затем идут месяц и год начала заболевания в формате 0821. Оставшиеся 8 цифр — уникальный номер сертификата.
12. По правилам «Госуслуги» не должны принимать в работу сертификаты без даты исхода заболевания.
13. О том, что весь ЕГИСЗ по крайней мере до 2016 года работал под управлением PostgreSQL, сказано в отчете Минздрава о создании системы. Логично предположить, что ПО менять не стали: в контракте на доработку коронавирусного регистра от 2020 года указано требование: программное обеспечение должно быть либо российским, либо свободно распространяемым. PostgreSQL подходит под вторую категорию.
14. 771042000013359, 771042000013360, 771042000013361.
15. Лишь в паре мест пропуск между найденными случайным образом номерами достигал нескольких десятков тысяч значений, средний размер промежутка — 2 409 значений, медиана — 939.
16. Это может быть разновидностью шардинга — когда одну базу разделяют на несколько частей, чтобы распределить нагрузку. Но в этом случае при создании новой записи должен быть определен критерий, по которому выбирается та или иная часть базы, например, номер региона.
17. 40 тысяч сотрудников медицинских организаций, 50 сотрудников Минздрава, 850 сотрудников региональных минздравов, 100 сотрудников федеральных органов исполнительной власти, 1000 сотрудников региональных органов; 2000 мест зарезервировано под доступ «иным лицам» по решению оперштаба.
18. Погрешность 0,21% — доверительный интервал (надежность оценки) 95%.
19. Попросил не указывать его имени.
20. Попросила не раскрывать ее имя и регион работы.

21. ПЦР-тесты, по данным Минздрава, дают ложноотрицательный результат в 30-40% случаев.

22. Код U07.1 по МКБ-10 («COVID-19, вирус идентифицирован»).

23. Код U07.2 по МКБ-10 («COVID-19, вирус не идентифицирован»).

Применяется в случаях, когда коронавирус диагностирован клинически и эпидемиологически, но лабораторные исследования неубедительны или недоступны.

24. Мы считаем отстающим сертификат, если его номер больше, чем минимальный номер, найденный два месяца спустя. Например, если в апреле не было сертификатов с номером меньше 23 миллионов, то любой февральский (двухмесячной давности) сертификат с номером больше 23 миллионов — отстающий.