

Федеральная служба по надзору в сфере защиты прав  
потребителей и благополучия человека  
Федеральное бюджетное учреждение науки  
«ЦЕНТРАЛЬНЫЙ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ИНСТИТУТ  
ЭПИДЕМИОЛОГИИ»  
(ФБУН ЦНИИ Эпидемиологии Роспотребнадзора)

*На правах рукописи*

Монахова Ангелина Андреевна

**СОВЕРШЕНСТВОВАНИЕ ЭПИДЕМИОЛОГИЧЕСКОГО НАДЗОРА  
ЗА НОВОЙ КОРОНАВИРУСНОЙ ИНФЕКЦИЕЙ (COVID-19)  
НА ОСНОВЕ ЦИФРОВЫХ ТЕХНОЛОГИЙ УЧЕТА И АНАЛИЗА  
ДАННЫХ МОЛЕКУЛЯРНО-БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ**

3.2.2. Эпидемиология

Диссертация  
на соискание ученой степени  
кандидата медицинских наук

**Научный руководитель:**  
академик РАН,  
доктор медицинских наук, профессор  
Акимкин Василий Геннадьевич

Москва — 2026

ВВЕДЕНИЕ.....	5
ГЛАВА I. ОБЗОР ЛИТЕРАТУРЫ.....	18
1.1. Современная эпидемиологическая характеристика заболеваемости новой коронавирусной инфекцией (COVID-19) в Российской Федерации .....	18
1.2. Использование современных цифровых технологий для сбора, анализа и хранения данных молекулярно-биологических исследований .....	26
1.3. Система эпидемиологического надзора за новой коронавирусной инфекцией (COVID-19) в Российской Федерации .....	35
ГЛАВА II. МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ .....	42
2.1. Материалы исследования .....	42
2.2. Методы исследования.....	45
ГЛАВА III. ЭПИДЕМИОЛОГИЧЕСКИЙ АНАЛИЗ ЗАБОЛЕВАЕМОСТИ НАСЕЛЕНИЯ НОВОЙ КОРОНАВИРУСНОЙ ИНФЕКЦИЕЙ (COVID-19) В РОССИЙСКОЙ ФЕДЕРАЦИИ В 2020–2023 гг. С ИСПОЛЬЗОВАНИЕМ МЕТОДА КЛАСТЕРИЗАЦИИ ИНТЕНСИВНЫХ ПОКАЗАТЕЛЕЙ.....	50
3.1. Анализ динамики уровня заболеваемости новой коронавирусной инфекцией (COVID-19) в Российской Федерации в 2020–2023 гг. ....	51
3.2. Кластеризация интенсивных дневных показателей заболеваемости новой коронавирусной инфекции (COVID-19) в Российской Федерации в 2020-2023 гг. ....	61
ГЛАВА IV. ОЦЕНКА ВЛИЯНИЯ СОЦИАЛЬНО-ДЕМОГРАФИЧЕСКИХ И ИНФРАСТРУКТУРНЫХ ХАРАКТЕРИСТИК, ВЛИЯЮЩИХ НА УРОВЕНЬ И СТРУКТУРУ ЗАБОЛЕВАЕМОСТИ НОВОЙ КОРОНАВИРУСНОЙ ИНФЕКЦИЕЙ (COVID-19) В СУБЪЕКТАХ РОССИЙСКОЙ ФЕДЕРАЦИИ	79

4.1. Корреляционный анализ факторов, влияющих на заболеваемость населения субъектов Российской Федерации новой коронавирусной инфекцией (COVID-19) .....	80
4.2. Анализ факторов, влияющих на заболеваемость населения субъектов Российской Федерации новой коронавирусной инфекцией (COVID-19), с использованием метода машинного обучения.....	88
4.3. Сравнение двух аналитических подходов для оценки влияния факторов на уровень и структуру заболеваемости населения субъектов Российской Федерации новой коронавирусной инфекцией (COVID-19) (традиционный корреляционный анализ и метод машинного обучения) .....	91
ГЛАВА V. ВОЗМОЖНОСТЬ ПРИМЕНЕНИЯ ЦИФРОВЫХ ТЕХНОЛОГИЙ УЧЕТА РЕЗУЛЬТАТОВ И АНАЛИЗА ДАННЫХ МОЛЕКУЛЯРНО-БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ НАСЕЛЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ ДЛЯ ПОВЫШЕНИЯ ИНФОРМАТИВНОСТИ АНАЛИЗА ЭПИДЕМИОЛОГИЧЕСКОЙ ОБСТАНОВКИ ПО НОВОЙ КОРОНАВИРУСНОЙ ИНФЕКЦИИ (COVID-19).....	93
5.1. Цифровая программная Платформа агрегирования результатов лабораторных исследований (SOLAR) как инструмент эпидемиологического надзора за новой коронавирусной инфекцией (COVID-19) .....	94
5.2. Использование данных о результатах ПЦР-исследований по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2) для оперативного и ретроспективного эпидемиологического анализа заболеваемости новой коронавирусной инфекцией (COVID-19) .....	102
ГЛАВА VI. СОВЕРШЕНСТВОВАНИЕ ЭПИДЕМИОЛОГИЧЕСКОГО НАДЗОРА ЗА НОВОЙ КОРОНАВИРУСНОЙ ИНФЕКЦИЕЙ (COVID-19) НА ОСНОВЕ ЦИФРОВЫХ ТЕХНОЛОГИЙ УЧЕТА И АНАЛИЗА ДАННЫХ МОЛЕКУЛЯРНО-БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ .....	114
ЗАКЛЮЧЕНИЕ .....	125

ВЫВОДЫ.....	134
ПРАКТИЧЕСКИЕ РЕКОМЕНДАЦИИ.....	136
ПЕРСПЕКТИВЫ ДАЛЬНЕЙШЕЙ РАЗРАБОТКИ ТЕМЫ .....	137
СПИСОК СОКРАЩЕНИЙ.....	138
СПИСОК ЛИТЕРАТУРЫ.....	139

## ВВЕДЕНИЕ

### Актуальность темы исследования

Пандемия новой коронавирусной инфекции (COVID-19) ярко продемонстрировала разрушительные последствия массовых инфекционных болезней, уносящих миллионы человеческих жизней и подрывающих международную экономику. Пандемия не только обнажила проблемы системы здравоохранения большинства стран, но и явилась импульсом научного прогресса во многих областях медицинских и биологических наук [Попова А.Ю., 2021]. Глобальное распространение новой коронавирусной инфекции (COVID-19) способствовало развитию цифровых технологий в эпидемиологии и ускорило создание программных платформ и внедрение молекулярно-биологических исследований в практику общественного здравоохранения [Акимкин В.Г., 2025].

Молекулярно-биологические исследования представляют собой методы научной деятельности, сосредоточенной на изучении структуры и функции молекул, таких как ДНК, РНК и белки, и их роли в жизнедеятельности клеток и организма в целом. Данные исследования позволяют глубже понять механизмы заболеваний на молекулярном уровне, что открывает возможности для разработки новых методов диагностики, лечения и профилактики [Альбертс Б., 2013; Акимкин В.Г., 2025].

Применение полимеразной цепной реакции (ПЦР) усовершенствовало процесс диагностики инфекционных болезней. ПЦР позволяет проводить раннюю и дифференциальную диагностику, а также идентифицировать патогены с высокой специфичностью и чувствительностью. Благодаря возможности быстро масштабировать данную методику, массовая ПЦР-диагностика стала эффективным инструментом для выявления возбудителей инфекционных болезней, что особенно актуально в условиях пандемий и эпидемий, когда важна быстрая реакция системы здравоохранения

на распространение инфекции и снижение риска контактирования с потенциально инфицированными лицами.

Платформа SOLAR (от англ. *System Of Laboratory Aggregation Results* — Платформа агрегирования результатов лабораторных исследований), представляющая собой централизованную базу данных результатов ПЦР-исследований для выявления возбудителя новой коронавирусной инфекции (SARS-CoV-2), была разработана и внедрена в ФБУН Центральный НИИ Эпидемиологии Роспотребнадзора в рамках исполнения Постановления Правительства РФ №452 от 27 марта 2021 г. Целью создания Платформы явилось оперативное информирование граждан о результатах лабораторных исследований.

По состоянию на 1 февраля 2026 г. в систему загружено более 200 млн результатов ПЦР-исследований, поступающих от более чем 2 000 лабораторий различной ведомственной принадлежности. Передача данных осуществляется по защищенным каналам связи ФБУН Центральный НИИ Эпидемиологии Роспотребнадзора, откуда они перенаправляются на сервер портала Госуслуг. Таким образом, Платформа SOLAR представляет собой важный инструмент цифровой трансформации эпидемиологического надзора, обеспечивая централизованный сбор и анализ данных о распространении новой коронавирусной инфекции (COVID-19).

Учение об эпидемиологическом надзоре является одним из базовых в теории и практике эпидемиологической науки. Отечественные ученые (Громашевский Л.В., 1941; Беляков В.Д., 1986; Покровский В.И., 1987; Черкасский Б.Л., 1988) в XX веке сформулировали основополагающие теории, определяющие методологию эпидемиологического надзора на современном этапе развития научного познания.

В соответствии с теорией саморегуляции паразитарных систем академика Белякова В.Д., основу развития эпидемического процесса составляют фазовые изменения гетерогенности биологических свойств взаимодействующих популяций возбудителя и человека, базирующиеся на

взаимосвязанном процессе саморегуляции. При этом важное регулирующее значение имеют социальные и природные факторы. Согласно данной теории, эпидемиологический надзор рассматривается как информационно-аналитическая и управленческая система, предполагающая динамическую оценку тенденций развития эпидемического процесса в пространстве и времени, обеспечивающая своевременное вмешательство в его ход с целью снижения заболеваемости населения и ликвидации отдельных инфекций [Беляков В.Д., 1987; Акимкин В.Г., 2024].

Важными элементами системы эпидемиологического надзора за инфекционной заболеваемостью является постановка эпидемиологического диагноза и формирование эпидемиологического прогноза, включающие оценку эпидемиологической ситуации, ее причин и тенденций развития. В качестве начального этапа эпидемиологической диагностики рассматривается оценка проявлений эпидемического процесса по выявлению территорий риска, групп и времени риска, т.е. изучение пространственно-временной структуры и динамики заболеваемости населения [Дубоделов Д.В., 2024].

В современных реалиях информационные технологии шагнули далеко вперед и способны решать сложные задачи эпидемиологического анализа с применением цифровых платформенных решений. В последние годы значительно увеличился объем данных доступных для эпидемиологического анализа, что позволяет успешнее использовать имитационные математические модели для анализа и прогнозирования динамики эпидемического процесса, а также делает актуальным создание эффективных алгоритмов поиска значимых показателей, блоков интерпретации и обработки данных [Михеева И.В., 2026].

В связи с этим необходимы новые подходы, поскольку активное развитие цифровых технологий и цифровизации в целом, а также постоянно усложняющиеся задачи, возникшие перед системой эпидемиологического надзора в период пандемии новой коронавирусной инфекции (COVID-19), создали предпосылки для стремительного развития цифровой трансформации в эпидемиологии, что подчеркивает актуальность рассматриваемой проблемы.

Значительная разнородность по плотности населения, географическим и экономическим условиям субъектов Российской Федерации обусловила необходимость организации работы по осуществлению эпидемиологического анализа с использованием методов кластеризации [Дубоделов Д.В., 2024].

В настоящее время обсуждаются перспективные направления цифровизации, связанные с анализом и структурированием данных с использованием математических методов и подходов, а также с разработкой интеллектуальных алгоритмов. Широкое распространение получили методы машинного обучения (*англ. machine learning*), основанные на машинной обработке подготовленных статистических данных для выявления закономерностей и построения на их основе алгоритмов, позволяющих системе искусственного интеллекта формулировать обоснованные выводы на основе поступающей информации [Гаврилов Д. В., 2021]. Такие подходы повышают эффективность прогнозирования рисков за счет применения масштабных хранилищ данных, независимой идентификации новых предикторов риска и сложных взаимосвязей между ними [Дубоделов Д. В., 2024].

Совершенствование системы эпидемиологического надзора на основе цифровизации является важной научной задачей современной эпидемиологии. Пандемия новой коронавирусной инфекции (COVID-19) подчеркнула значимость цифровой трансформации в системе эпидемиологического надзора как важного дополнения к традиционным методам эпидемиологического анализа данных [Акимкин В.Г., 2023].

Таким образом, современные цифровые технологии открывают новые горизонты для оперативного и точного принятия управленческих решений, направленных на минимизацию рисков для здоровья населения, что делает их незаменимыми в системе эпидемиологического надзора. Кроме того, цифровые технологии могут помочь в поиске инновационных методологических решений по совершенствованию системы эпидемиологического надзора, создают новые возможности для управления

эпидемиологической ситуацией и повышают адаптивность системы здравоохранения к меняющимся условиям эпидемиологической обстановки.

### **Степень разработанности темы исследования**

Последствия пандемии новой коронавирусной инфекции (COVID-19) остаются важным предметом интереса эпидемиологической науки. Учитывая это, в разных странах мира, в том числе и Российской Федерации, происходит усовершенствование научных инструментов эпидемиологического надзора за распространением инфекции. Так, в Российской Федерации разработаны научные основы и реализован геномный эпидемиологический надзор за возбудителями инфекционных болезней. Созданы три программные цифровые Платформы, анализирующие значительные объемы данных (*англ. Big Data*): российская система агрегации данных VGARus, SOLAR и аналитическая платформа для анализа и прогнозирования эпидемиологической обстановки по заболеваемости новой коронавирусной инфекцией (COVID-19) [Акимкин В.Г., 2025].

Также, предложен и разработан метод кластеризации показателей заболеваемости за многолетний период, помогающий повысить информативность и объективность результатов эпидемиологического анализа данных [Дубоделов Д.В., 2024].

Использование результатов ПЦР-исследований в системе эпидемиологического надзора стало ключевым элементом в борьбе с инфекционными болезнями. ПЦР-исследования позволяют обнаруживать патогены на ранних стадиях заболевания, даже до появления клинических симптомов, что дает возможность своевременно начать лечение и предотвратить распространение инфекции. Кроме того, ПЦР-методы дают возможность с высокой точностью различать патогены, что особенно актуально при наличии схожих клинических симптомов (например, гриппа и

новой коронавирусной инфекции (COVID-19), что способствует правильному выбору тактики лечения [Акимкин В.Г., 2025].

В рамках эпидемиологического надзора ПЦР-исследования используются для отслеживания мутаций патогенов, что важно для оценки их вирулентности и возможности формирования новых штаммов [Акимкин В.Г., 2021].

Таким образом, актуальность изучения и внедрения в систему эпидемиологического надзора цифровых технологий анализа результатов молекулярно-биологических исследований и математических моделей развития эпидемического процесса на их основе не вызывает сомнения и является важным направлением его совершенствования.

### **Цель исследования**

Совершенствование эпидемиологического надзора за новой коронавирусной инфекцией (COVID-19) на основе цифровых технологий учета и анализа данных молекулярно-биологических исследований.

### **Задачи исследования**

1. Провести ретроспективный эпидемиологический анализ заболеваемости новой коронавирусной инфекцией (COVID-19) населения Российской Федерации за 2020–2023 гг. с использованием метода кластеризации субъектов в различные периоды пандемии.

2. Оценить влияние социально-демографических и инфраструктурных характеристик на уровень и динамику заболеваемости новой коронавирусной инфекцией (COVID-19) в субъектах Российской Федерации.

3. Изучить возможности применения системных цифровых технологий учета результатов и анализа данных молекулярно-биологических

исследований населения Российской Федерации для повышения информативности анализа эпидемиологической обстановки по новой коронавирусной инфекции (COVID-19).

4. Предложить научно обоснованные пути совершенствования системы эпидемиологического надзора за новой коронавирусной инфекцией (COVID-19) на основе внедрения цифровых технологий учета и анализа результатов молекулярно-биологических исследований.

### **Научная новизна исследования**

Углублены научные знания, характеризующие распространение новой коронавирусной инфекции (COVID-19) на территории Российской Федерации в различные периоды эпидемии. Проведен детальный эпидемиологический анализ динамики уровня и структуры заболеваемости населения Российской Федерации новой коронавирусной инфекцией (COVID-19) в период 2020–2023 гг. с использованием метода кластеризации интенсивных показателей заболеваемости.

Показана роль деперсонифицированных сведений ПЦР-исследований по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2) для проведения оперативного анализа заболеваемости новой коронавирусной инфекцией (COVID-19) на основе детальной оценки факторов риска и характеристик проявлений эпидемического процесса в различные периоды эпидемии в целях краткосрочного прогнозирования эпидемиологической обстановки по заболеваемости новой коронавирусной инфекцией (COVID-19).

Определена научная обоснованность и эффективность применения методов машинного обучения для проведения ретроспективного эпидемиологического анализа заболеваемости и выявления факторов, влияющих на уровень и динамику эпидемического процесса новой коронавирусной инфекции (COVID-19).

Научно обосновано совершенствование эпидемиологического надзора за новой коронавирусной инфекцией (COVID-19), базирующееся на применении цифровых технологий для обработки и анализа данных молекулярно-биологических исследований.

### **Теоретическая и практическая значимость исследования**

Диссертационное исследование позволяет оценить эффективность практического использования цифровых платформенных решений для оперативного и ретроспективного анализа заболеваемости новой коронавирусной инфекцией (COVID-19) в конкретных регионах и Российской Федерации в целом.

Результаты научного исследования подчеркивают роль молекулярно-биологических исследований в объективизации эпидемиологической информации и оперативности ее получения с целью принятия своевременных управленческих решений.

Показано, что внедрение современных цифровых инструментов позволяет существенно повысить качество и скорость анализа молекулярно-биологических данных, что способствует более точному прогнозированию эпидемиологических тенденций и принятию научно обоснованных управленческих решений на различных уровнях деятельности системы здравоохранения.

Научные данные, полученные в результате диссертационного исследования создают научную основу для использования прогностических моделей в интересах решения будущих эпидемиологических задач в отношении аэрозольных антропонозов с высоким эпидемическим и пандемическим потенциалом распространения возбудителя.

## **Методология и методы исследования**

Цель исследования определила последовательность этапов научной работы, каждый из которых опирался на применение соответствующих научных методов. Работа представляет собой комплексное исследование, включающее не только классические эпидемиологические подходы, но и математические методы, позволяющие глубже рассмотреть механизмы развития эпидемического процесса. На этапе разработки дизайна исследования применялся комплексный подход к оценке возможных источников систематических и случайных ошибок.

В основе эпидемиологического метода применена классическая схема: описательный эпидемиологический анализ распространенности и заболеваемости с определением факторов, влияющих на развитие эпидемического процесса новой коронавирусной инфекции (COVID-19), и последующий аналитический этап для установления причинно-следственных связей. Для этого использованы как дескриптивные статистические методы, так и аналитические методы, включающие методы машинного обучения.

Обработка полученных данных осуществлялась с помощью Microsoft Excel. Проведен корреляционный анализ для оценки связей между разными переменными. Все полученные результаты систематизированы и представлены в главах диссертации.

## **Положения, выносимые на защиту**

1. Применение метода кластеризации субъектов Российской Федерации по уровню заболеваемости населения новой коронавирусной инфекцией (COVID-19) при проведении ретроспективного эпидемиологического анализа, а также анализ интенсивных показателей заболеваемости среди установленных кластеров позволяет выявить

особенности проявлений эпидемического процесса новой коронавирусной инфекции (COVID-19).

2. Уровень заболеваемости и динамика показателей, характеризующие проявления эпидемического процесса новой коронавирусной инфекции (COVID-19) в субъектах Российской Федерации, имели выраженную корреляционную зависимость от доли городского населения, плотности населения региона, интенсивности экономического развития (количества активных предприятий), показателей активности миграционных процессов (коэффициента миграционного прироста).

3. Использование цифровой Платформы агрегирования результатов лабораторных исследований SOLAR (System of laboratory aggregation results) и современных цифровых технологий позволяет существенно объективизировать и оптимизировать процессы эпидемиологического анализа данных, обеспечить краткосрочное прогнозирование заболеваемости новой коронавирусной инфекцией (COVID-19).

4. Внедрение в практику эпидемиологического надзора платформ цифровизации данных молекулярно-биологических исследований и инструментов машинного обучения позволяют совершенствовать подходы к решению эпидемиологических задач.

### **Личный вклад автора в получении результатов**

Автор принимал непосредственное участие в разработке концепции исследования, формулировке цели, задач и логики построения диссертационной работы. Автором осуществлен анализ эпидемиологической информации, полученной из федеральных информационных систем и статистических отчетов.

Сбор, систематизация и подготовка исходных данных для эпидемиологического анализа выполнены автором в полном объеме. Проведен ретроспективный эпидемиологический анализ заболеваемости новой

коронавирусной инфекцией (COVID-19) в субъектах Российской Федерации. Автором осуществлена кластеризация территорий по эпидемиологическим показателям, проведен корреляционный анализ между принадлежностью к кластерам и демографическими характеристиками. Выполнено построение модели на основе метода машинного обучения (Random Forest) для оценки значимости факторов и сравнение ее результатов с другими подходами.

Автор лично сформулировал предложения по совершенствованию системы эпидемиологического надзора за новой коронавирусной инфекцией (COVID-19), основанные на применении цифровых платформ, результатов молекулярно-биологических данных и технологий автоматизированного анализа. Все аналитические расчеты, интерпретация результатов, подготовка графического материала, формулировка выводов и написание текста диссертации выполнены автором самостоятельно.

### **Внедрение в практику**

На основе результатов диссертационного исследования получены:

1. Свидетельство о государственной регистрации программы для ЭВМ №2026661550 от 22 апреля 2026 г. «EpidSmart — модуль COVID, версия 2.0», предназначенная для оценки эпидемиологической обстановки по новой коронавирусной инфекции (COVID-19), проведения оперативного и ретроспективного эпидемиологического анализа с выявлением групп, территорий и факторов риска (заявка №2026619201 от 03.04.2026 г.).

2. Свидетельство о государственной регистрации базы данных №2026621597 от 09 апреля 2026 г. «EpidSmart — модуль COVID, версия 2.0», предназначенная для обеспечения функционирования программного обеспечения «EpidSmart — модуль COVID, версия 2.0». Содержит информацию о проявлениях эпидемического процесса новой коронавирусной инфекции (COVID-19) в различных возрастных группах населения субъектов

Российской Федерации, демографические показатели, результаты молекулярно-генетического мониторинга за возбудителем (SARS-CoV-2).

Материалы диссертационного исследования внедрены в образовательный процесс на кафедре эпидемиологии с курсами молекулярной диагностики и дезинфектологии ФБУН ЦНИИ Эпидемиологии Роспотребнадзора.

### **Степень достоверности и апробация результатов**

Достоверность результатов исследования, ключевых выводов и рекомендаций обеспечена путем аналитического обобщения данных, полученных из открытых и проверенных источников научной литературы. В диссертационной работе применены общепризнанные методы сбора и обработки официальной статистической информации. Представлены масштабные эпидемиологические и молекулярно-биологические исследования, использующие современные методы эпидемиологического, молекулярно-биологического, математического и статистического анализа.

Основные положения и результаты исследования представлены и обсуждены на Конгрессах и научно-практических конференциях различных уровней:

- Научно-практическая конференция молодых ученых и специалистов ФБУН ЦНИИ Эпидемиологии Роспотребнадзора «Проблемы эпидемиологии, терапии и лабораторной диагностики инфекционных болезней — 2024» (29–30 мая 2024 г., г. Москва);
- VI Международная научно-практическая конференция по вопросам противодействия инфекционным заболеваниям (11–12 декабря 2025 г., г. Санкт-Петербург);
- Всероссийский Конгресс с международным «Эпидемиология — 2025» (15–16 октября 2025 г., г. Москва);

- V Ежегодной конференции по инфекционным болезням «Покровские чтения» (27–28 октября 2025 г., г. Москва);
- Всероссийский Конгресс с международным «Эпидемиология — 2026» (16–17 апреля 2026 г., г. Москва).

В окончательном виде диссертационная работа апробирована и рекомендована к защите на заседании апробационной комиссии ФБУН ЦНИИ Эпидемиологии Роспотребнадзора (протокол №113 от 14 апреля 2026 г.).

### **Соответствие диссертации паспорту научной специальности**

Научные положения диссертации соответствуют паспорту специальности 3.2.2. «Эпидемиология». Результаты проведенного исследования соответствуют областям исследований: пунктам 2, 5 и 6 паспорта специальности «Эпидемиология».

### **Публикации**

По теме диссертации опубликовано **12** научных работ, в том числе **4** статьи в изданиях, рекомендованных ВАК Российской Федерации для публикации основных научных результатов диссертации по специальности 3.2.2. «Эпидемиология».

### **Структура и объем диссертации**

Диссертационное исследование представлено на 152 листах, состоит из Введения, 6 глав (обзора литературы; главы «Материалы и методы исследования»; 4 глав собственных исследований), заключения, выводов и списка литературы. Диссертация иллюстрирована 8 таблицами и 11 рисунками. Список литературы содержит 120 источников, в том числе 51 — на русском языке и 69 — на английском языке.

## ГЛАВА I. ОБЗОР ЛИТЕРАТУРЫ

### 1.1. Современная эпидемиологическая характеристика заболеваемости новой коронавирусной инфекцией (COVID-19) в Российской Федерации

Анализ начального этапа распространения новой коронавирусной инфекции (COVID-19) в мире требует детального рассмотрения того, как возник и развивался эпидемический процесс, поскольку именно первые недели и месяцы определили дальнейшую динамику пандемии, характер глобального реагирования и масштабы последствий. Первые сообщения о случаях неясной пневмонии были зарегистрированы в г. Ухань в конце декабря 2019 г., когда медицинские службы Китая зарегистрировали группу пациентов с тяжелым поражением дыхательных путей. На тот момент природа заболевания была неизвестна, а эпидемиологические проявления не позволяли установить источник инфекции [99, 119].

Уже в первые дни января 2020 г. китайские специалисты выделили новый коронавирус, получивший обозначение SARS-CoV-2 (*англ. Severe acute respiratory syndrome coronavirus 2*). Его особенности, включая высокую способность к передаче и наличие бессимптомных форм, привели к тому, что инфекция начала стремительно распространяться не только внутри Китая, но и за его пределами. Международная мобильность населения, активные транспортные потоки и отсутствие иммунитета у людей создали условия для ускоренного распространения инфекции, что вскоре привело к появлению случаев в странах Азии, Европы и Северной Америки [52, 86, 113].

К середине февраля 2020 г. мир столкнулся с первыми локальными вспышками за пределами Китая, наиболее крупные из которых были зарегистрированы в Южной Корее, Италии и Иране. Эти страны стали первыми очагами масштабного распространения новой коронавирусной инфекции (COVID-19), продемонстрировав, что вирус обладает способностью формировать крупные кластеры заболевания и быстро выходить из-под

контроля при отсутствии своевременных ограничительных мер. Ситуация осложнялась тем, что на раннем этапе медицинские системы большинства государств не были готовы к такому уровню нагрузки, а знания о клинических особенностях новой инфекции оставались ограниченными [86, 97].

Именно в этот период стали очевидными ключевые характеристики возбудителя новой коронавирусной инфекции (SARS-CoV-2), включая значительную вариабельность симптомов, способность передаваться до появления первых признаков болезни и потенциал для развития тяжелых осложнений у отдельных групп населения. 11 марта 2020 г. ВОЗ официально объявила вспышку новой коронавирусной инфекции (COVID-19) пандемией, подчеркнув, что угрозу представляет не только сам вирус, но и скорость его распространения. Далее эпидемический процесс приобрел глобальный масштаб, а новая коронавирусная инфекция (COVID-19) стала фактором, определившим социальную, экономическую и медицинскую повестку практически всех государств мира [57, 87, 118].

Первые сообщения о пневмониях неясной этиологии в г. Ухань (КНР) появились в декабре 2019 г., в январе 2020 г. опубликованы полные геномные последовательности возбудителя новой коронавирусной инфекции (SARS-CoV-2) [99, 119].

Определение генома стало ключевым моментом для начала разработки диагностических тестов, средств профилактики и дальнейшего изучения особенностей нового патогена. Возбудитель новой коронавирусной инфекции (SARS-CoV-2) относится к семейству *Coronaviridae* — группе РНК-содержащих вирусов, известных своей способностью вызывать как легкие респираторные заболевания, так и тяжелые инфекции, подобные атипичной пневмонии (SARS) и ближневосточному респираторному синдрому (MERS) [53, 111].

Одной из критически важных особенностей нового коронавируса стала его высокая контагиозность, обусловленная как эффективным механизмом прикрепления к клеткам, так и возможностью передачи инфекции от

бессимптомных носителей. На ранних этапах пандемии это создало значительные трудности для контроля распространения заболевания, поскольку многие инфицированные продолжали вести привычный образ жизни, не подозревая о том, что могут быть источником инфекции. В совокупности эти характеристики возбудителя новой коронавирусной инфекции (SARS-CoV-2) определили стремительный рост заболеваемости в первые месяцы 2020 г. и дальнейшее глобальное развитие эпидемического процесса [57, 98, 105].

В Российской Федерации первый случай зарегистрирован в конце января 2020 г., с марта началось повсеместное распространение новой коронавирусной инфекции (COVID-19). Первые случаи новой коронавирусной инфекции (COVID-19) в Российской Федерации были зарегистрированы в конце января 2020 г. у граждан, прибывших из КНР [2]. Выявление зараженных происходило преимущественно в ходе активного санитарно-карантинного контроля в аэропортах, где проводился осмотр прибывающих пассажиров, измерение температуры и сбор эпидемиологического анамнеза. Инфицированные были оперативно изолированы, а контактные лица поставлены под медицинское наблюдение, что позволило на начальном этапе предотвратить локальное распространение инфекции [2, 15].

Реагирование службы на регистрацию первых случаев включало быстрое развертывание многоуровневой системы эпидемиологического надзора. Роспотребнадзор ввел режим повышенной готовности, усилил лабораторную диагностику и распространил методические рекомендации по выявлению и ведению пациентов с подозрением на новую коронавирусную инфекцию (COVID-19). В регионах создавались специализированные инфекционные стационары, готовились койко-места и обеспечивались условия для строгой изоляции. С марта 2020 г., по мере появления очагов местной передачи, меры были расширены до масштабных ограничений,

направленных на предотвращение повсеместного распространения инфекции [36, 37, 45].

Эпидемиологический анализ заболеваемости новой коронавирусной инфекции (COVID-19) в Российской Федерации включает комплексную оценку динамики, структуры и пространственного распределения заболеваемости, а также факторов, определяющих особенности эпидемического процесса. Эпидемический процесс характеризовался несколькими выраженными подъемами, связанными с доминированием сменяющих друг друга генетических вариантов вируса (Alpha, Delta, Omicron SARS-CoV-2 и др.). Периоды подъема сменялись фазами относительной стабилизации и снижения показателей заболеваемости, что явилось отражением эффективности проводимых противоэпидемических мероприятий [3, 4].

Эпидемический процесс новой коронавирусной инфекции (COVID-19) характеризовался неравномерным распространением возбудителя, чередованием волн заболеваемости и высокой вариативностью темпов заражения в разных регионах [34, 41]. Рост числа случаев определялся появлением новых вариантов вируса, сезонностью, уровнем коллективного иммунитета и мерами противодействия. Для новой коронавирусной инфекции (COVID-19) типична стадийность: чередование периодов подъема, когда отмечались резкие подъемы заболеваемости, и спадов, наступавших после введения ограничений или увеличения охвата вакцинацией.

Осенью 2020 г. в странах Северного полушария начался второй крупный подъем заболеваемости новой коронавирусной инфекцией (COVID-19). К концу декабря 2020 г. в мире зарегистрировано около 79,2 млн случаев заболевания и 1 754 493 летальных исхода. Второй период оказался более тяжелым по уровню смертности по сравнению с первым, особенно в странах Европы, Индии и США. В Российской Федерации к 26 декабря 2020 г. общее число подтвержденных случаев заболевания новой коронавирусной инфекцией (COVID-19) превысило 3,02 млн, число летальных исходов

составило 54 226 [66]. В декабре 2020 г. одобрены первые вакцины (в том числе российская «Спутник V»), однако массовая вакцинация развернута в начале 2021 г. [43].

Отмечена высокая летальность среди пожилых и лиц из групп риска (пациентов с хроническими заболеваниями сердечно-сосудистой системы, сахарным диабетом, ожирением, хроническими заболеваниями легких, онкологическими и иммунодефицитными состояниями) во многих странах системы здравоохранения функционировали в условиях серьезной перегрузки. Данный период считается одним из самых тяжелых: в конце 2020 г. темпы роста заболеваемости и смертности от новой коронавирусной инфекции (COVID-19) превосходили показатели предыдущего периода [8, 70, 82, 88].

Летом 2021 г. доминирующим вариантом стал Delta SARS-CoV-2, характеризующийся высокой контагиозностью и тяжелым течением. К 19 декабря 2021 г. во всем мире зарегистрировано более 273 млн случаев новой коронавирусной инфекции (COVID-19) и 5,3 млн летальных исходов [89, 116]. Во многих странах подъем заболеваемости, связанный с распространением варианта Delta SARS-CoV-2, начался уже в середине года, что сопровождалось ускорением темпов вакцинации: к концу 2021 г. примерно половина населения планеты получила хотя бы одну дозу вакцины от новой коронавирусной инфекции (COVID-19) [65].

Несмотря на сохранявшуюся высокую летальность при Delta-варианте, с накоплением коллективного иммунитета и расширением охвата вакцинацией отмечалось снижение доли тяжелых случаев заболевания новой коронавирусной инфекцией (COVID-19), госпитализаций и летальных исходов. Так, по данным Центров по контролю и профилактике заболеваний США (CDC), уровень смертности среди госпитализированных снизился с 15,1 % в период циркуляции варианта Delta SARS-CoV-2 до 4,9 % во время распространения варианта Omicron SARS-CoV-2 [93]. В Российской Федерации осенью 2021 г. наблюдался выраженный четвертый период эпидемии новой коронавирусной инфекции (COVID-19): суточный прирост

числа заболевших обновлял негативные рекорды, а показатели госпитализаций и смертности значительно возросли [66].

С конца 2021 г. началось повсеместное распространение варианта Omicron SARS-CoV-2, обладающего высокой контагиозностью и преимущественным распространением легких и среднетяжелых форм заболевания новой коронавирусной инфекцией (COVID-19) [64, 103]. По данным ВОЗ, к 8 ноября 2022 г. в мире зарегистрировано около 630 млн лабораторно подтвержденных случаев новой коронавирусной инфекции (COVID-19) и 6,6 млн летальных исходов. В ряде стран в данный период отмечались максимальные суточные показатели заболеваемости новой коронавирусной инфекцией (COVID-19) за все время пандемии [117]. Несмотря на высокий уровень инфицирования, общая летальность снизилась по сравнению с предыдущими периодами. Что связывают с формированием популяционного (коллективного) иммунитета, обеспеченного как вакцинацией, так и перенесенным ранее заболеванием [100]. В Российской Федерации активное распространение варианта Omicron SARS-CoV-2 в начале 2022 г. сопровождалось значительным увеличением числа случаев заражения. Однако, уровень госпитализаций и летальных исходов оставался ниже, чем в предыдущие эпидемические периоды [42, 50].

С начала 2023 г. в большинстве регионов мира наблюдалась устойчивая тенденция к снижению заболеваемости новой коронавирусной инфекцией (COVID-19). Согласно данным ВОЗ, по состоянию на 7 января 2024 г. общее число лабораторно подтвержденных случаев составило более 774 млн, а количество зарегистрированных летальных исходов приблизилось к 7 млн [67]. В конце 2023 г. на фоне циркуляции новых подвариантов возбудителя новой коронавирусной инфекции (SARS-CoV-2), в том числе JN.1, в некоторых странах отмечалось локальное увеличение числа госпитализаций [114]. Тем не менее, повсеместное внедрение бустерной иммунизации и наличие сформировавшегося популяционного иммунитета способствовали ограничению развития тяжелых форм заболевания новой коронавирусной

инфекцией (COVID-19). В 2023–2024 гг. новая коронавирусная инфекция (COVID-19) в большинстве случаев протекала в легкой или среднетяжелой форме, нагрузка на стационарное звено оставалась на управляемом уровне [74, 100].

Динамика эпидемического процесса новой коронавирусной инфекции (COVID-19) определялась сочетанием факторов: степенью коллективного иммунитета, охватом населения вакцинацией, характером вирусных мутаций, соблюдением мер профилактики (масочный режим, социальное дистанцирование), а также качеством диагностики и эпидемиологического надзора. Введение масштабного ПЦР-обследования и активное использование молекулярно-генетических методов позволили своевременно выявлять новые варианты вируса и корректировать противоэпидемические меры [49, 81, 104].

В Российской Федерации эпидемический процесс новой коронавирусной инфекции (COVID-19) развивался неоднородно, сопровождаясь несколькими выраженными эпидемическими периодами. Первые значимые подъемы заболеваемости регистрировались осенью 2020 г., затем в середине 2021 г., когда распространялся вариант Delta SARS-CoV-2. Самый высокий рост заболеваемости наблюдался в начале 2022 г. под воздействием варианта Omicron SARS-CoV-2, который отличался высокой контагиозностью. Динамика была тесно связана с региональными особенностями, уровнем тестирования, вакцинацией и соблюдением профилактических мер, что приводило к различиям в интенсивности эпидемии между субъектами Российской Федерации [3, 4].

С течением времени уровень заболеваемости стал снижаться благодаря формированию смешанного иммунитета, включающего постинфекционный и поствакцинальный компоненты. Увеличившийся охват вакцинацией, улучшение методов диагностики и адаптация системы здравоохранения позволили уменьшить число тяжелых случаев и летальных исходов. Тем не менее на протяжении всего периода пандемии динамика оставалась

нестабильной, реагируя на сезонные колебания и появление новых вариантов вируса.

В Российской Федерации с начала пандемии зарегистрировано около 23 млн подтвержденных случаев новой коронавирусной инфекции (COVID-19) и свыше 400 тыс. смертей. Наиболее интенсивные эпидемические периоды приходились на осень 2020 г., лето–осень 2021 г. (период доминирования Дельта SARS-CoV-2) и январь–февраль 2022 г., когда Омикрон SARS-CoV-2 вызвал резкий подъем заболеваемости до 200 тыс. случаев в сутки. При этом тяжелые формы и смертность в период Омикрон SARS-CoV-2 были ниже примерно в 2–3 раза по сравнению с эпидемическими периодами 2021 г.

К 2023 г. уровень заболеваемости снизился более чем на 80% по сравнению с пиковыми значениями, а число госпитализаций уменьшилось почти вдвое. Вакцинация охватила около 60% населения, был создан широкий доступ к тестированию и лечению.

В период первого эпидемического подъема заболеваемости новой коронавирусной инфекцией (COVID-19) (апрель – июнь 2020 г.) в Российской Федерации общее число зарегистрированных случаев к концу июня достигло 647 849, количество летальных исходов — 9 320. Количество госпитализаций увеличивались стремительно, во многих странах наблюдалась перегрузка отделений интенсивной терапии, что потребовало масштабного развертывания дополнительных коек и мобилизации медицинских ресурсов [2, 6].

Наибольшая заболеваемость новой коронавирусной инфекцией (COVID-19) отмечалась среди лиц трудоспособного возраста (30–49 лет), что связано с высокой социальной мобильностью данной группы. Однако, самые высокие уровни госпитализаций и летальности приходились на пожилых граждан старше 60 лет и пациентов с хроническими заболеваниями. В детских возрастных группах в целом регистрировались менее тяжелые формы заболевания и более низкие показатели заболеваемости [29].

Кроме того, заболеваемость существенно различалась по регионам: наиболее высокие уровни регистрировались в крупных мегаполисах (г. Москва, Московская область, г. Санкт-Петербург), где выше плотность населения и миграционная активность [41]. В регионах с преобладанием сельского населения темпы распространения инфекции были ниже, однако там нередко наблюдались сложности с доступом к медицинской помощи [91, 120].

Эпидемический процесс новой коронавирусной инфекции (COVID-19) в стране носил волнообразный характер и отличался региональной неоднородностью показателей, что обусловлено различиями в мерах реагирования органов государственной власти и систем здравоохранения на распространение инфекции, а также уровнем социально-экономического развития и плотности населения, а также скоростью смены доминирующих вариантов возбудителя новой коронавирусной инфекции (SARS-CoV-2) [26, 34, 51].

## **1.2. Использование современных цифровых технологий для сбора, анализа и хранения данных молекулярно-биологических исследований**

Молекулярно-биологические исследования — это изучение структуры, свойств и взаимодействий биомолекул, таких как ДНК, РНК, белков и метаболитов, с целью понимания процессов, лежащих в основе функционирования клеток и организмов.

Опыт использования ПЦР и секвенирования в борьбе с эпидемиями формировался задолго до пандемии новой коронавирусной инфекции (COVID-19), однако именно последние события показали, насколько критически важными стали эти методы для глобальной системы биобезопасности. Международная практика демонстрирует, что ПЦР давно применяется как основной инструмент быстрой диагностики инфекций: еще в 1990-е годы она стала стандартом для выявления ВИЧ, туберкулеза, гепатитов

и ряда особо опасных вирусов. Вспышки атипичной пневмонии SARS в 2002–2003 гг. и пандемия свиного гриппа H1N1 в 2009 г. доказали роль ПЦР как метода оперативного выявления зараженных, что позволило ускорить карантинные меры и ограничить распространение заболеваний. Секвенирование, в свою очередь, развивалось как инструмент молекулярной эпидемиологии: оно применялось для отслеживания мутаций вирусов гриппа, мониторинга резистентности микроорганизмов и выявления путей передачи инфекций, включая вспышки лихорадки Эболы в 2014–2016 гг., когда впервые было развернуто мобильное полевое секвенирование [12, 14].

В Российской Федерации до пандемии новой коронавирусной инфекции (COVID-19) ПЦР-диагностика также была хорошо развита и широко использовалась для скрининга ВИЧ, гепатитов, туберкулеза и в ветеринарии. Уже в 2010-е гг. существовала сеть лабораторий Роспотребнадзора, оснащенных оборудованием для ПЦР, а отечественные производители выпускали сертифицированные тест-системы. Однако масштабное применение секвенирования в практической эпидемиологии до 2020 г. было ограниченным: технология использовалась главным образом в научных центрах и отдельных проектах по изучению гриппа, паразитарных заболеваний и генетического разнообразия патогенов [14, 27].

В мировой практике существует широкий спектр технологий для сбора, анализа и хранения медицинских данных, и их развитие значительно ускорилось во время пандемии. До пандемии основой цифровой медицины были электронные медицинские карты, лабораторные информационные системы и стандарты обмена данными вроде HL7 (*англ. Health Level Seven*) и FHIR (*англ. Fast Healthcare Interoperability Resources*). Данные системы обеспечивали регистрацию результатов обследований, работу диагностических лабораторий и интеграцию данных между учреждениями. Использовались также аналитические инструменты на базе статистических пакетов, больших данных (*англ. Big Data*) и машинного обучения (*англ. Machine Learning*), однако их применение было ограничено в основном

научными и крупными клиническими центрами. Данные хранились в локальных больничных базах, национальных регистрах и специализированных биомедицинских репозиториях [58, 59].

Пандемия новой коронавирусной инфекции (COVID-19) вызвала резкое расширение и ускоренное внедрение цифровых технологий в медицине. Системы телемедицины, мобильные приложения для мониторинга симптомов и централизованные платформы учета результатов тестирования стали стандартом. Значительно возросла роль облачных инфраструктур, позволяющих оперативно обрабатывать и распределять огромные массивы данных. Параллельно развивались аналитические модели, предназначенные для прогнозирования эпидемиологических тенденций, и инструменты геномного надзора, необходимые для отслеживания новых вариантов вируса [75].

Отдельное направление составляет работа с молекулярно-биологическими данными. До пандемии такие данные обрабатывались преимущественно в исследовательских лабораториях с использованием специализированных LIMS (*англ. Laboratory Information Management System*), платформ для анализа геномов и международных биорепозиториях вроде GenBank [106]. Во время пандемии их использование стало массовым, благодаря широкому внедрению технологий высокопроизводительного секвенирования и глобальных облачных платформ для биоанализа. Особое значение приобрели международные базы данных, такие как GISAID [112], которые обеспечивали оперативный обмен генетическими последовательностями возбудителя новой коронавирусной инфекции (SARS-CoV-2). На этой основе были созданы национальные и международные системы геномного эпиднадзора, ставшие ключевым инструментом в борьбе с пандемией.

Таким образом, пандемия стала важным технологическим рубежом, ускорив переход к более интегрированным, масштабируемым и

автоматизированным системам управления медицинскими и молекулярно-биологическими данными.

К ключевым инструментам цифровой эпидемиологии относятся методы машинного обучения, анализ данных реального мира (*англ. Real-World Data, RWD*) и технологии обработки больших данных [14], позволяющие значительно ускорить процессы сбора, структурирования и анализа эпидемиологической информации, обеспечивая проверку гипотез и принятие решений в режиме, приближенном к реальному времени. Тем не менее, широкомасштабное применение подобных инструментов требует дальнейшей научной апробации и стандартизации.

Одним из значимых компонентов цифрового эпидемиологического надзора стали онлайн-платформы мониторинга заболеваемости. В первые месяцы 2020 г. в ряде стран развернуты информационные панели (*англ. dashboards*), предоставляющие в открытом доступе регулярно обновляемую статистику по числу случаев заболевания новой коронавирусной инфекцией (COVID-19), количеству госпитализаций, летальности, объемам ПЦР-обследования и охвату вакцинацией. Одним из наиболее известных международных ресурсов стала платформа *Our World in Data*, аккумулирующая данные по большинству стран мира [95]. Подобные решения способствовали повышению прозрачности эпидемиологического надзора и расширению информированности населения. В Российской Федерации аналогичную функцию выполнял портал «Стопкоронавирус.рф» [30], обеспечивавший ежедневное обновление показателей с региональной детализацией.

Для организации внутриведомственного мониторинга результатов ПЦР-тестирования населения по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2) в Российской Федерации Федеральной службой по надзору в сфере защиты прав потребителей и благополучия человека в сжатые сроки разработана и внедрена цифровая платформа **SOLAR** (*англ. SARS-CoV-2 Online Laboratory Analysis Resource*)

[47]. Указанная система агрегирует результаты лабораторного тестирования на выявление возбудителя новой коронавирусной инфекции (SARS-CoV-2) в единую базу, доступную специалистам санитарно-эпидемиологической службы. Использование платформы SOLAR повысило оперативность и прозрачность предоставления стандартизированных данных, что способствовало более эффективному взаимодействию между организациями здравоохранения, научными учреждениями и специалистами по эпидемиологическому надзору [47].

Функциональность платформы обеспечила доступ к актуальной информации о санитарно-эпидемиологической ситуации на всех уровнях управления, что упростило аналитическую обработку данных и координацию противоэпидемических мероприятий. В ФБУН Центральный НИИ Эпидемиологии Роспотребнадзора также разработаны две цифровые системы: VGARus (*англ. Virus Genome Aggregator of Russia*) — платформа агрегирования результатов расшифровок генома возбудителей инфекционных и паразитарных заболеваний, и аналитическая платформа EpidSmart, предназначенная для проведения эпидемиологического анализа [47]. Таким образом, сформирована многоуровневая цифровая инфраструктура, обеспечивающая комплексный подход к управлению эпидемиологической ситуацией.

Цифровые технологии занимают центральное место в современной медицине и эпидемиологии, так как именно они позволяют эффективно работать с большими объемами данных, возникающими при генетических, микробиологических и биохимических исследованиях. Молекулярно-биологические исследования являются источником данных о структуре ДНК возбудителей, мутациях, генетических детерминантах резистентности к антимикробным препаратам, а также об изменениях в микробиоте и биомаркерах заболеваний. Эти данные используются для диагностики, мониторинга и прогнозирования развития инфекционных и неинфекционных процессов.

Основой для применения цифровых технологий является накопление больших массивов разнородных данных — геномных, протеомных, метагеномных и клинико-эпидемиологических. Все эти данные хранятся в специализированных базах данных которые обеспечивают безопасное хранение и быстрый доступ к информации. Для анализа применяются методы биоинформатики, алгоритмы машинного обучения и искусственного интеллекта, позволяющие выявлять закономерности, прогнозировать появление новых штаммов возбудителей, оценивать риск вспышек и разрабатывать персонализированные стратегии профилактики и лечения.

На международном уровне широко известны платформы типа Nextstrain [94], использующая геномные данные для отслеживания эволюции вирусов и построения прогнозов распространения инфекций, и проект GISAID [112], предоставляющий открытую базу данных по генетическим последовательностям вирусов, включая возбудителя новой коронавирусной инфекции (SARS-CoV-2). Эти системы работают на основе постоянного обновления данных о возбудителях и применяют алгоритмы математического моделирования и машинного обучения для построения динамических прогнозов.

Таким образом, цифровые технологии обеспечивают не только хранение и структурирование молекулярно-биологических данных, но и их глубокий аналитический разбор, позволяя на основании данных о заболеваемости и результатах лабораторных исследований строить научно обоснованные прогнозы, необходимые для контроля и предупреждения распространения заболеваний.

Особое значение в процессе эпидемиологического надзора за новой коронавирусной инфекцией (COVID-19) приобрели технологии анализа больших данных (*англ. Big Data*). Источниками информации выступали не только медицинские регистры, но и данные из других секторов: операторы связи предоставляли обезличенные сведения о перемещении населения, а интернет-компании — информацию о поисковых запросах и активности

пользователей в социальных сетях. В исследовании К.Т. Момыналиева и В. Г. Акимкина [25] показано, что динамика поисковых запросов, связанных с симптомами новой коронавирусной инфекции (COVID-19) в системе Google, коррелировала с последующим увеличением заболеваемости в Российской Федерации, что позволяет рассматривать такие показатели как потенциальный инструмент раннего эпидемиологического предупреждения.

Дополнительно, анализ данных мобильной связи использовался для оценки соблюдения ограничительных мер и прогнозирования эффективности введенных или отмененных карантинных мероприятий. Также началось внедрение методов искусственного интеллекта для выявления закономерностей в эпидемиологических данных.

В Российской Федерации применялись цифровые пропуска и мобильные приложения для контроля соблюдения карантинных мер (в частности, «Социальный мониторинг», использовавшийся в г. Москве), а также сервисы дистанционного наблюдения за состоянием здоровья граждан [5]. Цифровые технологии также активно использовались в телемедицине: с целью снижения нагрузки на поликлинические учреждения развернуты онлайн-платформы для удаленного консультирования пациентов с легкой формой заболевания новой коронавирусной инфекцией (COVID-19) [7]. Все указанные решения не только способствовали сдерживанию распространения инфекции, но и позволили собирать ценные массивы данных — о клинических проявлениях заболевания, соблюдении противоэпидемических рекомендаций, побочных эффектах вакцинации и других аспектах, представляющих интерес для эпидемиологов.

Несмотря на очевидные преимущества, процесс цифровизации эпидемиологического надзора сопровождался рядом значимых вызовов. Во-первых, остро встал вопрос обеспечения защиты персональных данных и этической допустимости использования информации. Массовое внедрение приложений для отслеживания контактов вызвало широкую дискуссию о

соотношении между эффективностью надзорных механизмов и соблюдением прав граждан на конфиденциальность.

Во-вторых, проявилась неоднородность технической готовности различных регионов и государств, что обозначило проблему так называемой «цифровой пропасти». Для полноценной реализации потенциала цифровых решений необходимы квалифицированные кадры и соответствующие вычислительные ресурсы, которые в ряде случаев оказываются недостаточными [46].

Кроме того, отсутствие интеграции между многочисленными цифровыми решениями и ведомственными базами данных стало серьезным ограничением для эффективного функционирования здравоохранения. В целях устранения данной проблемы с 2021 г. предприняты меры по созданию единой цифровой инфраструктуры здравоохранения. Согласно отчету Минздрава РФ за 2023 г., запущен модуль, посвященный новой коронавирусной инфекции (COVID-19), в рамках Единой государственной информационной системы в сфере здравоохранения (ЕГИСЗ), что позволило обеспечить связь между данными медицинских организаций, лабораторий и органов управления [18].

Схожие инициативы реализуются и на международном уровне. Так, Европейский центр профилактики и контроля заболеваний (ECDC) в 2022 г. опубликовал руководство по применению цифровых инструментов для мониторинга заболевания новой коронавирусной инфекцией (COVID-19), направленное на унификацию подходов государств Европейского Союза к использованию мобильных приложений и электронных ресурсов в целях эпидемиологического надзора [78].

Развитие цифровых технологий существенно модернизировало систему эпидемиологического надзора за новой коронавирусной инфекцией (COVID-19), повысив ее оперативность, масштабируемость и аналитические возможности. Применение цифровых решений обеспечило проведение мониторинга в режиме, близком к реальному времени (вплоть до ежедневной

и почасовой оценки ситуации), а также реализацию более адресных и обоснованных мер реагирования.

Несмотря на значительный прогресс достигнутый в процессе цифровизации здравоохранения и санитарно-эпидемиологического надзора остаются нерешенными ряд задач. Необходима техническая и организационная интеграция информационных систем, обеспечивающая совместимость и унификацию данных. Также важно предусмотреть надежные механизмы долговременного хранения и архивации эпидемиологических данных с возможностью последующего ретроспективного анализа.

Отдельной проблемой остается цифровое неравенство — ограниченный доступ к современным технологическим решениям в отдельных регионах, особенно в сельской местности и труднодоступных территориях [46].

В условиях пандемии новой коронавирусной инфекции (COVID-19) отмечалось активное межведомственное взаимодействие организаций здравоохранения, научных учреждений и органов исполнительной власти, направленное на обеспечение оперативного обмена данными о случаях инфицирования, результатах ПЦР-обследования, госпитализациях и летальных исходах (по данным Роспотребнадзора, 2020–2022 гг.). Значительные объемы разнородной информации потребовали разработки специализированных цифровых инструментов для их централизованного агрегирования, систематизации и аналитической обработки [31].

К 2021 г. в Российской Федерации ежедневно проводились сотни тысяч лабораторных исследований, формируя масштабный поток эпидемиологически значимой информации, необходимой для оперативной оценки текущей обстановки и принятия управленческих решений в системе здравоохранения. Параллельно с этим осуществлялась разработка и апробация альтернативных методов лабораторной диагностики, включая изотермическую амплификацию нуклеиновых кислот (LAMP) что

способствовало формированию основ для развития молекулярно-генетического мониторинга в эпидемиологической практике [1].

### **1.3. Система эпидемиологического надзора за новой коронавирусной инфекцией (COVID-19) в Российской Федерации**

Система эпидемиологического надзора за новой коронавирусной инфекцией (COVID-19) представляет собой совокупность взаимосвязанных мероприятий, направленных на систематический сбор, обработку, интерпретацию и использование информации о заболеваемости, распространенности факторов риска, а также об эффективности профилактических и противоэпидемических мер. Указанная система обеспечивает научно обоснованное планирование и реализацию санитарно-эпидемиологических мероприятий, включая принятие управленческих решений, направленных на предупреждение, ограничение и ликвидацию инфекционной заболеваемости. В рамках санитарно-эпидемиологического надзора реализуется взаимодействие с органами санитарного контроля, что формирует интегрированную модель обеспечения охраны здоровья населения.

В международной практике структура системы эпидемиологического надзора регламентирована ВОЗ и включает глобальный, региональный и национальный уровни. Их согласованное функционирование осуществляется на основе Международных медико-санитарных правил (ММСП), обеспечивая координацию усилий государств по предупреждению трансграничного распространения инфекций.

Пандемия новой коронавирусной инфекции (COVID-19) стала масштабным испытанием для систем эпидемиологического надзора и потребовала их срочной адаптации к новым условиям. Существенную роль в глобальной координации сыграл регулярный обмен эпидемиологическими данными между странами. Впервые национальные системы

эпидемиологического надзора были объединены в транснациональное информационное пространство, в рамках которого страны ежедневно передавали сведения о заболеваемости в соответствии с Международными медико-санитарными правилами.

Система эпидемиологического надзора за новой коронавирусной инфекцией (COVID-19) представляет собой многоуровневый комплекс мероприятий, направленных на постоянный сбор, анализ и интерпретацию информации о распространении заболевания, выявлении источников инфекции и оценке эффективности профилактических мер. Основой системы является информационно-аналитическая подсистема, обеспечивающая сбор данных о заболеваемости, госпитализациях, летальности, анализ динамики эпидемического процесса и формирование прогнозов для принятия управленческих решений.

Пандемия новой коронавирусной инфекции (COVID-19) потребовала оптимизации механизмов эпидемиологического надзора, способных оперативно учитывать изменения эпидемиологической обстановки. Современные подходы к организации эпидемиологического надзора предполагают реализацию стратегии упреждающего реагирования, интеграцию данных с разных уровней (федерального, регионального и муниципального) и формирование единой информационной платформы [3, 29].

В период пандемии новой коронавирусной инфекции (COVID-19) подходы к организации эпидемиологического надзора трансформировались от классической модели к интегрированной системе мониторинга в реальном времени. Подобная система базируется на использовании данных лабораторной диагностики, цифровой отчетности и научного анализа, что обеспечило более эффективное выявление очагов распространения инфекции и групп населения с повышенным эпидемиологическим риском. Вместе с тем пандемия новой коронавирусной инфекции (COVID-19) выявила ряд системных ограничений. Одной из ключевых нерешенных задач остается

полноценная интеграция данных из разрозненных источников в единую информационную платформу, особенно в условиях функциональной перегрузки системы здравоохранения.

Кроме того, в условиях пандемии новой коронавирусной инфекции (COVID-19) отмечена значительная неоднородность организационно-технической готовности систем эпидемиологического надзора: в отдельных странах и регионах наблюдались различия в уровне технического обеспечения, кадровых ресурсах и стандартах сбора и представления информации, что приводило к временным и структурным искажениям при интерпретации данных [77]. Указанные дефициты стали стимулом к внедрению новых цифровых и аналитических технологий, направленных на повышение устойчивости систем эпидемиологического надзора, о чем более подробно изложено в последующих главах диссертационного исследования.

Развитие молекулярно-генетического мониторинга стало необходимым ответом на высокую генетическую изменчивость возбудителя новой коронавирусной инфекции (COVID-19). Прогресс в области технологий высокопроизводительного секвенирования геномов позволил выполнять полное прочтение вирусного генома в течение нескольких часов, что ранее было технически недоступно [1, 2, 29]. В январе 2020 г. китайскими исследователями опубликованы первые полные геномные последовательности возбудителя новой коронавирусной инфекции (SARS-CoV-2), выделенного у пациентов с пневмонией в г. Ухань (КНР) [60]. Полученные данные подтвердили принадлежность вируса к роду бетакоронавирусов и стали основой для оперативной разработки ПЦР-тест-систем и вакцин.

Значимую роль в международной координации секвенирования сыграло расширение функциональных возможностей глобальной базы данных GISAID [112], первоначально созданной для обмена информацией о вирусах гриппа. Благодаря оперативной загрузке геномных последовательностей возбудителя новой коронавирусной инфекции (COVID-19) исследователями со всего мира,

к середине 2021 г. в базе представлено более 2,5 млн последовательностей, охватывающих свыше 185 стран.

В 2020 г. предложена динамическая система классификации генетических линий возбудителя новой коронавирусной инфекции (COVID-19) — PANGO (англ. *Phylogenetic Assignment of Named Global Outbreak Lineages*) [102], обеспечивающая оперативный обмен данными о новых мутациях вируса. Введены понятия вариантов, вызывающих обеспокоенность (VOC, Variant of Concern), включая такие линии, как Alpha, Beta, Gamma, Delta и Omicron SARS-CoV-2. Данные наименования стали общепринятыми в эпидемиологической и научной практике. Развитие молекулярно-генетического мониторинга обеспечило своевременное обнаружение новых вариантов возбудителя новой коронавирусной инфекции (COVID-19). Так, в конце 2020 г. в Великобритании выявлен высоко контагиозный вариант B.1.1.7 (Alpha), в Южно-Африканской Республике — B.1.351 (Beta), в Бразилии — P.1 (Gamma) [56, 73, 84], что подчеркивает критическую значимость геномного мониторинга для раннего выявления эпидемиологически значимых мутаций.

В рамках настоящего исследования важно пояснить разницу между геномным надзором и молекулярно-генетическим мониторингом. Геномный надзор используется в более широком, эпидемиологическом контексте. Он включает систематическое наблюдение, анализ и интерпретацию данных о геномных изменениях возбудителя во времени и пространстве. Цель — информировать органы здравоохранения о появлении, распространении и эволюции патогенов. Геномный эпидемиологический надзор является частью системы эпидемиологического надзора и направлен на поддержание готовности к вспышкам и адаптацию противоэпидемических мер [83].

Молекулярно-генетический мониторинг более узкое понятие и отражает техническую сторону работы — проведение лабораторных исследований (исследованию генетического материала возбудителя методами ПЦР, секвенирования). То есть мониторинг — инструмент получения

первичных данных о структуре генома, а надзор — система их интерпретации и практического применения. Использование обоих терминов оправдано, поскольку они обозначают взаимосвязанные, но разные уровни одной системы контроля за возбудителями инфекций.

Регулярное секвенирование образцов биологического материала от пациентов позволило идентифицировать ключевые мутации, влияющие на уровень контагиозности и вирулентности до того, как соответствующие варианты возбудителя новой коронавирусной инфекции (COVID-19) становились преобладающими в популяции [63, 115]. Молекулярно-генетический мониторинг сыграл определяющую роль в обеспечении своевременного эпидемиологического ответа на появление вариантов вируса, требующих наблюдения. В этой связи ВОЗ призвала страны нарастить мощности секвенирования и аналитической обработки данных для обеспечения раннего обнаружения новых вариантов возбудителя новой коронавирусной инфекции (COVID-19) [85].

В Российской Федерации формирование системы геномного эпидемиологического надзора происходило поэтапно. На начальном этапе пандемии объемы секвенирования были ограничены, однако к 2021 г. при участии Роспотребнадзора и научных учреждений создана распределенная сеть лабораторий, осуществляющих расшифровку полных геномов возбудителя новой коронавирусной инфекции (COVID-19). В качестве национального ресурса сформирована отечественная база данных VGARus (Платформа агрегирования результатов расшифровок генома возбудителей инфекционных и паразитарных заболеваний), агрегирующая генетическую информацию о циркулирующих вариантах вируса на территории Российской Федерации [47].

Согласно данным, представленным В.Г. Акимкиным [1, 4], в различные периоды эпидемии новой коронавирусной инфекции (COVID-19) в Российской Федерации последовательно доминировали следующие варианты возбудителя новой коронавирусной инфекции (COVID-19): уханьский (весна

2020 г.), B.1.1.7 (начало 2021 г.), Delta (вторая половина 2021 г.) и Omicron (с начала 2022 г.). Каждый эпидемический подъем имел свои особенности: например, вариант Delta SARS-CoV-2 ассоциировался с более тяжелым течением заболевания и повышенным уровнем летальности [3], тогда как Omicron SARS-CoV-2 характеризовался значительно большей контагиозностью при сравнительно низком уровне смертности. Геномный эпидемиологический надзор позволил подтвердить данные различия на молекулярном уровне, а также отследить влияние смены доминирующих вариантов возбудителя новой коронавирусной инфекции (SARS-CoV-2) на структуру заболеваемости [84].

ВОЗ в 2022 г. представила десятилетнюю стратегию глобального геномного эпидемиологического надзора (2022–2032 гг.), ориентированную на сохранение и развитие соответствующей инфраструктуры на случай потенциальных пандемий в будущем [85].

Таким образом, анализ литературных источников по обозначенным проблемам дает возможность сделать следующие выводы:

1. Уровень и динамика заболеваемости населения новой коронавирусной инфекцией (COVID-19) подробно описаны в работах отечественных и зарубежных авторов. Однако, остается недостаточно изученным вопрос о степени влияния отдельных социально-экономических факторов на разнородность показателей заболеваемости в субъектах Российской Федерации.

2. Развитие современных цифровых технологий сделало возможным использование в интересах санитарно-эпидемиологического надзора результатов массового ПЦР-обследования населения по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2), однако в настоящее время отсутствуют убедительные данные о возможности использования таких данных при проведении оперативного и ретроспективного эпидемиологического анализа.

3. Большинство существующих методов математического моделирования и прогнозирования динамики эпидемического процесса новой коронавирусной инфекции (COVID-19) в качестве исходных данных принимают сведения о количестве заболевших, при этом отсутствуют сведения о влиянии использования данных о результатах массового ПЦР-обследования по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2) на точность прогнозных моделей.

4. Существует необходимость дополнительного обобщения и систематизации опыта использования молекулярно-биологических исследований в системе эпидемиологического надзора за новой коронавирусной инфекцией (COVID-19), что подчеркивает важность настоящего научного исследования.

## ГЛАВА II. МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

Диссертационное исследование проведено в 2024–2026 гг. на базе Федерального бюджетного учреждения науки «Центральный научно-исследовательский институт эпидемиологии» Федеральной службы по надзору в сфере защиты прав потребителей и благополучия человека в рамках НИР АААА–А21–121011890131–4 «Научно-методическое обоснование системы эпидемиологического надзора и профилактики заболеваний, вызываемых SARS-подобными коронавирусами» (2021–2025 гг.) и НИР 1025030500172–4–3.3.9 «Совершенствование системы эпидемиологического надзора за инфекционными болезнями в Российской Федерации на основе современных цифровых технологий» (2026–2030 гг.).

В рамках проведенного диссертационного исследования использованы эпидемиологический, молекулярно-биологический, математический и статистический методы исследования.

### 2.1. Материалы исследования

Выполнен ретроспективный эпидемиологический анализ заболеваемости новой коронавирусной инфекции (COVID-19) на территории Российской Федерации за период 2020–2023 гг. с использованием отчетной формы № 1035 «Мониторинг о количестве заболевших коронавирусной инфекцией, в том числе внебольничными пневмониями, и летальных исходов» и формы №970 «Информация о случаях инфекционных заболеваний у лиц с подозрением на новую коронавирусную инфекцию».

Для оценки динамики циркулирующих вариантов возбудителя новой коронавирусной инфекции (SARS-CoV-2) использованы данные из Российской базы данных VGARus (Российской платформы для агрегирования информации о геномах вирусов (*англ. Virus Genome Aggregator of Russia*), созданного в рамках выполнения Постановлений Правительства Российской

Федерации от 23 марта 2021 г. № 448 «Об утверждении Временного порядка предоставления данных расшифровки генома возбудителя новой коронавирусной инфекции (COVID-19)». Для оценки значимости результатов молекулярно-биологических исследований и построения прогностических моделей проведен анализ данных российской информационной системы «SOLAR», созданной в рамках выполнения Постановления Правительства Российской Федерации от 27.03.2021 № 452 «Об обеспечении уведомления физических лиц о результатах исследований на наличие возбудителя новой коронавирусной инфекции (COVID-19) с использованием федеральной государственной информационной системы «Единый портал государственных и муниципальных услуг (функций)» и обмена информацией о результатах таких исследований», а также сведений Федеральной службы государственной статистики (таблица 1).

Таблица 1 — Материалы исследования

Источник данных	Содержание данных	Объем сведений для анализа
Форма отчетности Роспотребнадзора №1035 «Мониторинг количества заболевших коронавирусной инфекцией, в том числе внебольничными пневмониями, и летальных исходов»	Мониторинг о количестве заболевших коронавирусной инфекцией, в том числе внебольничными пневмониями, и летальных исходов	<b>1 349</b> отчетных форм ( <b>19 млн</b> показателей)
Форма отчетности Роспотребнадзора №970 «Информация о случаях инфекционных заболеваний у лиц с подозрением на новую коронавирусную инфекцию»	Информация о случаях инфекционных заболеваний у лиц с подозрением на новую коронавирусную инфекцию	<b>796</b> отчетных форм ( <b>17 млн</b> случаев заболевания)
Данные Платформы агрегирования результатов лабораторных исследований SOLAR (System of laboratory aggregation results)	Результаты ПЦР-исследований на наличие возбудителя новой коронавирусной инфекции (SARS-CoV-2)	<b>Более 200 млн</b> исследований
Российская база данных VGARus (Российская платформа для агрегирования информации о геномах вирусов (англ. <i>Virus Genome Aggregator of Russia</i> ))	Сведения о вариантах возбудителя новой коронавирусной инфекции (SARS-CoV-2)	<b>298 тыс.</b> результатов секвенирования ( <b>194 тыс.</b> полных геномов и <b>104 тыс.</b> фрагментов генома возбудителя)
Данные Федеральной службы государственной статистики	Инфраструктурные и социально-демографические сведения	<b>Более 10 тыс.</b> показателей

## 2.2. Методы исследования

### Эпидемиологический метод

В рамках эпидемиологического метода осуществлялась регистрация новых случаев инфекции, госпитализаций, выздоровлений и летальных исходов, что позволяло отслеживать динамику заболеваемости, выделять периоды ее подъема и снижения, а также сравнивать эпидемиологические показатели между регионами. Значимое внимание уделялось анализу пространственно-временных закономерностей — различий в распространении инфекции по территории, в том числе между городскими и сельскими населенными пунктами, а также учету сезонных колебаний и климатических факторов, влияющих на эпидемический процесс.

В рамках исследования применены описательно-оценочный и аналитический методы эпидемиологического анализа. На основании официальных данных отчетной формы №1035 «Мониторинг о количестве заболевших коронавирусной инфекцией, в том числе внебольничными пневмониями, и летальных исходов» произведен ретроспективный анализ динамики заболеваемости населения субъектов Российской Федерации новой коронавирусной инфекцией (COVID-19) за период 2020–2023 гг. Рассчитаны интенсивные показатели (уровни заболеваемости на 100 тыс. населения) и экстенсивные показатели (распределение случаев по временным и территориальным характеристикам).

Проведено сопоставление уровня заболеваемости новой коронавирусной инфекцией (COVID-19) по полу, возрастным группам и типу населенного пункта (городские и сельские территории) в субъектах Российской Федерации за 2020–2023 гг. Целью данного анализа явилось выявление различий в эпидемиологической ситуации между указанными группами, а также потенциальных факторов, влияющих на интенсивность распространения инфекции.

Методология проведенного эпидемиологического анализа опиралась на классические эпидемиологические подходы, описанные в трудах российских ученых конца XX – начала XXI века: Беляков В.Д. [9], Черкасский Б.Л. [48] и Покровский В.И. [35], а также на современные методы статистического анализа больших данных.

### **Молекулярно-биологический метод**

Для подтверждения случаев новой коронавирусной инфекции (COVID-19) в 2022–2023 гг. использован метод полимеразной цепной реакции (ПЦР), обеспечивающий выявление РНК возбудителя даже в незначительных концентрациях (использованы данные о более 200 млн исследований) (табл.1). ПЦР-диагностика оставалась основным методом лабораторного подтверждения инфицирования, широко применялась в амбулаторной и стационарной практике и регламентировалась нормативными актами.

Также учтены результаты изотермической амплификации нуклеиновых кислот, в частности LAMP (*англ. loop-mediated isothermal amplification*), позволяющие сократить время анализа за счет проведения реакции при постоянной температуре без применения сложного оборудования, и методы секвенирования нового поколения (NGS), направленные на определение полной нуклеотидной последовательности генома возбудителя новой коронавирусной инфекции (SARS-CoV-2). Полученные геномные данные (более 298 тыс. результатов секвенирования) (табл.1) использовались для отслеживания мутаций, появления новых вариантов вируса (например, Alpha, Delta, Omicron SARS-CoV-2), и для оценки возможного влияния данных вариантов на эпидемиологические показатели, включая тяжесть течения инфекции и эффективность вакцинации.

В настоящем исследовании проанализированы данные ПЦР-диагностики новой коронавирусной инфекции (COVID-19) за 2022–2023 гг. на основе данных федеральной платформы SOLAR [47], обеспечивающей централизованный сбор результатов ПЦР-обследования по

выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2). К системе подключено более 2 000 организаций из 85 субъектов Российской Федерации. В едином реестре аккумулировано свыше 200 млн результатов, с ежедневным добавлением около 30 тыс. новых записей (в период 2020–2023 гг.) (табл.1).

### **Математический метод**

*Метод кластеризации (k-means).* Использован для кластеризации временных рядов ежедневной заболеваемости. Алгоритм k-means можно описать в несколько шагов: 1) выбор числа кластеров (k). На первом этапе необходимо определить количество кластеров, на которые будут разделены данные. Этот параметр задают вручную, и его правильный выбор напрямую влияет на качество кластеризации; 2) инициализация центроидов. Алгоритм случайным образом выбирает k начальных точек, называемых центроидами. Эти точки служат временными центрами кластеров; 3) назначение объектов кластерам. Каждый объект в наборе данных назначается к кластеру, центроид которого находится ближе всего; 4) обновление центроидов. После назначения объектов кластерам вычисляют новые центроиды. Каждый центроид перемещается в среднюю точку всех объектов, принадлежащих его кластеру; 5) повторение. Шаги 3 и 4 повторяют до тех пор, пока центроиды не перестанут значительно изменяться, что свидетельствует о достижении сходимости. В некоторых реализациях алгоритм также может быть завершен при достижении заданного числа итераций. Метод реализован с использованием библиотеки Matplotlib (версия 3.8.0) [92].

*Методы анализа линейных зависимостей.* В рамках исследования использованы методы анализа линейных зависимостей для оценки связи между отдельными социально-демографическими и инфраструктурными характеристиками субъектов Российской Федерации и уровнем заболеваемости новой коронавирусной инфекцией (COVID-19). Рассчитан коэффициент корреляции Пирсона, позволяющий количественно оценить

степень линейной зависимости между двумя количественными признаками. Значения коэффициента варьируют от  $-1$  до  $+1$  и характеризуют направление и силу взаимосвязи. В расчетах проанализированы показатели плотности населения, уровня урбанизации и обеспеченности транспортной инфраструктурой в сопоставлении с интенсивными показателями заболеваемости новой коронавирусной инфекцией (COVID-19) в период 2020–2023 гг. Оценку силы корреляционной связи проводили по шкале Чеддока (таблица 2).

Таблица 2 — Оценка силы корреляционной связи и значимости по шкале Чеддока [32]

Значения коэффициента корреляции	Характеристика тесноты корреляционной связи
0–0,1	очень слабая
0,1–0,3	слабая
0,3–0,5	средняя
0,5–0,7	умеренно сильная
0,7–0,9	высокая
0,9–0,99	очень высокая

*Метод машинного обучения «Random Forest».* В рамках исследования применен метод «Random Forest» — ансамблевый алгоритм машинного обучения на основе «решающих деревьев». Данный метод позволяет выявлять нелинейные зависимости и учитывать сложные взаимодействия между множеством факторов. Алгоритм формирует совокупность «решающих деревьев», каждое из которых обучается на случайной выборке данных и случайном подмножестве признаков, после чего прогнозы усредняются. Метод использован для анализа влияния социально-демографических и инфраструктурных показателей (уровня урбанизации, плотности населения,

доли пожилого населения, количества ПЦР-исследований по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2) и др.) на уровень заболеваемости новой коронавирусной инфекцией (COVID-19) по субъектам Российской Федерации в 2020–2023 гг.

### **Статистический метод**

Для обработки полученных данных применен комплекс методов статистического анализа: описательная статистика, графический анализ, а также методы параметрической и непараметрической статистики. Также использованы приемы математического моделирования и анализа временных рядов с задействованием библиотек SciPy [107], Pandas [96] и Statsmodels [108].

При изучении динамики заболеваемости новой коронавирусной инфекцией (COVID-19) населения Российской Федерации в 2020–2023 гг. использованы как дневные, так и недельные интенсивные показатели числа зарегистрированных случаев. Сезонная декомпозиция временных рядов произведена на основе ежедневных данных с применением метода локально взвешенной линейной регрессии (LOESS), реализованного в модуле statsmodels.tsa.seasonal [108]. Для дополнительного анализа сезонных колебаний регистрируемой заболеваемости построены графики автокорреляционной функции с использованием библиотеки Matplotlib (версия 3.8.0) [92]. Доверительный интервал (95% ДИ) рассчитывался с использованием точного метода Клоппера-Пирсона (онлайн калькулятор EPITOOLS [79]). Расчет стандартного отклонения осуществлен при помощи программы Microsoft Office Excel.

**ГЛАВА III. ЭПИДЕМИОЛОГИЧЕСКИЙ АНАЛИЗ  
ЗАБОЛЕВАЕМОСТИ НАСЕЛЕНИЯ НОВОЙ КОРОНАВИРУСНОЙ  
ИНФЕКЦИЕЙ (COVID-19) В РОССИЙСКОЙ ФЕДЕРАЦИИ  
В 2020–2023 гг. С ИСПОЛЬЗОВАНИЕМ МЕТОДА КЛАСТЕРИЗАЦИИ  
ИНТЕНСИВНЫХ ПОКАЗАТЕЛЕЙ**

Изучение количественных характеристик и динамических показателей эпидемического процесса является базовым элементом системы эпидемиологического надзора. Комплексный анализ уровня заболеваемости, темпов распространения инфекции и структурных особенностей эпидемического процесса позволяет сформировать объективную картину эпидемиологической обстановки и выявить ключевые детерминанты, влияющие на развитие эпидемического процесса.

В данной Главе для получения данных о Мониторинге количества заболевших новой коронавирусной инфекцией (COVID-19), в том числе внебольничными пневмониями, и летальных исходов использована форма отчетности Роспотребнадзора №1035 «Мониторинг количества заболевших коронавирусной инфекцией, в том числе внебольничными пневмониями, и летальных исходов». Для получения информации о случаях инфекционных заболеваний у лиц с подозрением на новую коронавирусную инфекцию применена форма отчетности Роспотребнадзора №970 «Информация о случаях инфекционных заболеваний у лиц с подозрением на новую коронавирусную инфекцию». Сведения о вариантах возбудителя новой коронавирусной инфекции (SARS-CoV-2) получены из Российской базы данных VGARus (Российская платформа для агрегирования информации о геномах вирусов (*англ. Virus Genome Aggregator of Russia*)). Демографические сведения для расчетов показателей взяты из данных Федеральной службы государственной статистики.

### **3.1. Анализ динамики уровня заболеваемости новой коронавирусной инфекцией (COVID-19) в Российской Федерации в 2020–2023 гг.**

В рамках данного диссертационного исследования проведен анализ уровня и динамики заболеваемости новой коронавирусной инфекцией (COVID-19) в Российской Федерации в период с 2020 по 2023 гг. За исследуемый период на территории Российской Федерации зарегистрировано более 23 млн подтвержденных случаев новой коронавирусной инфекции (COVID-19).

В период с 2020 по 2023 гг. в Российской Федерации прослеживались восемь эпидемических периодов распространения новой коронавирусной инфекции (COVID-19), различавшихся по интенсивности заболеваемости, длительности и особенностям циркуляции генетических вариантов возбудителя новой коронавирусной инфекции (SARS-CoV-2). На рисунке 3.1 отражена динамика недельных показателей заболеваемости, демонстрирующая чередование периодов подъема и снижения уровня заболеваемости, что обуславливалось как биологическими особенностями возбудителя, так и комплексом реализуемых противоэпидемических мероприятий [44]. Разделение эпидемического процесса новой коронавирусной инфекции (COVID-19) на условные периоды позволило учитывать специфику каждого из них при проведении последующего анализа (таблица 3.1).

**Первый эпидемический подъем** новой коронавирусной инфекции (COVID-19), пришедшийся на март – август 2020 г., характеризовался началом активной циркуляции вируса среди населения, не имеющего иммунитета. На фоне отсутствия естественной иммунной прослойки и ограниченности данных о механизмах передачи инфекции был введен наиболее строгий комплекс противоэпидемических мер, включавший режим самоизоляции, ограничение работы предприятий, перевод отдельных категорий организаций на дистанционный режим работы, а также временное сокращение социальных

контактов. Максимальное недельное значение заболеваемости зафиксировано 4 мая 2020 г. (на 19-й неделе), достигнув 51,41 случая на 100 тыс. населения. Реализация мер, действовавших в период с конца марта по май 2020 г., привела к замедлению распространения инфекции, снижению нагрузки на систему здравоохранения и постепенной стабилизации эпидемической ситуации. После частичной отмены ограничений заболеваемость продолжила снижаться, и к августу 2020 г. показатель составил минимум 23,83 случая на 100 тыс. населения.

С началом учебного года, на 37-й неделе 2020 г. (7 сентября), отмечен **второй эпидемический подъем**. Возобновление очного режима обучения, увеличение интенсивности контактов в организованных коллективах и влияние сезонных факторов сопровождалось ускорением распространения новой коронавирусной инфекции (COVID-19). В октябре–январе 2020 г. в большинстве субъектов Российской Федерации введены дополнительные ограничения: перевод отдельных образовательных организаций на дистанционный режим работы, корректировка графиков работы предприятий и другие меры, направленные на снижение мобильности населения. Максимальное недельное значение заболеваемости во втором периоде пришлось на 52-ю неделю 2020 г. (21 декабря) и составило 136,69 случая на 100 тыс. населения. В дальнейшем наблюдалась тенденция к снижению интенсивности эпидемического процесса новой коронавирусной инфекции (COVID-19), и к марту 2021 г. недельный показатель стабилизировался на уровне минимум 41,15 случая на 100 тыс. населения.

Параллельно с изменением интенсивности заболеваемости новой коронавирусной инфекцией (COVID-19) происходила смена доминирующих генетических вариантов возбудителя новой коронавирусной инфекции (SARS-CoV-2). В начале 2021 г. среди исследованных образцов биологического материала начали выявляться единичные случаи варианта Alpha SARS-CoV-2, однако его распространенность оставалась низкой и не оказывала существенного влияния на эпидемическую ситуацию.

Принципиальные эпидемиологические изменения начались весной 2021 г., когда в апреле был впервые идентифицирован вариант Delta SARS-CoV-2, отличавшийся более высокой трансмиссивностью. Уже в мае 2021 г. его доля среди циркулирующих вариантов возбудителя новой коронавирусной инфекции (SARS-CoV-2) достигла 38%, а в июне 2021 г. он стал доминирующим.

С мая 2021 г. регистрировался **третий эпидемический подъем** заболеваемости новой коронавирусной инфекцией (COVID-19), напрямую связанный с распространением варианта Delta SARS-CoV-2. Интенсивность проявлений эпидемического процесса сопровождалась ростом недельных показателей заболеваемости новой коронавирусной инфекцией (COVID-19): на 29-й неделе 2021 г. (12 июля) показатель достиг максимума 118,8 случая на 100 тыс. населения. В условиях нарастающей нагрузки на систему здравоохранения большинство субъектов страны с начала июня 2021 г. введены дополнительные ограничения, направленные на сдерживание распространения инфекции. К середине августа – началу сентября 2021 г. недельная заболеваемость новой коронавирусной инфекцией (COVID-19) стабилизировалась на минимальном уровне 91,32 случая на 100 тыс. населения.

В сентябре 2021 г. начался **четвертый эпидемический подъем**. Его развитие также обусловлено доминированием варианта Delta SARS-CoV-2, действием сезонных факторов и увеличением интенсивности социальных контактов после начала учебного года. В октябре–ноябре 2021 г. в ряде субъектов Российской Федерации введены дополнительные ограничения: перевод отдельных образовательных организаций на дистанционный формат деятельности, приостановка массовых мероприятий, усиление контроля за соблюдением масочного режима. Максимальный недельный показатель заболеваемости зарегистрирован на 45-й неделе 2021 г. (1 ноября) и составил 191,87 случая на 100 тыс. населения. К концу декабря 2021 г. эпидемический

процесс новой коронавирусной инфекции (COVID-19) стабилизировался на уровне минимум 85,33 случая на 100 тыс. населения.

В январе 2022 г. отмечен **пятый эпидемический подъем**, обусловленный активным распространением варианта Omicron SARS-CoV-2. Данный подъем сопровождался резким увеличением числа новых случаев, и на 7-й неделе 2022 г. (14 февраля) зарегистрирован максимальный недельный показатель заболеваемости новой коронавирусной инфекцией (COVID-19) за весь период наблюдений — 905 случаев на 100 тыс. населения.

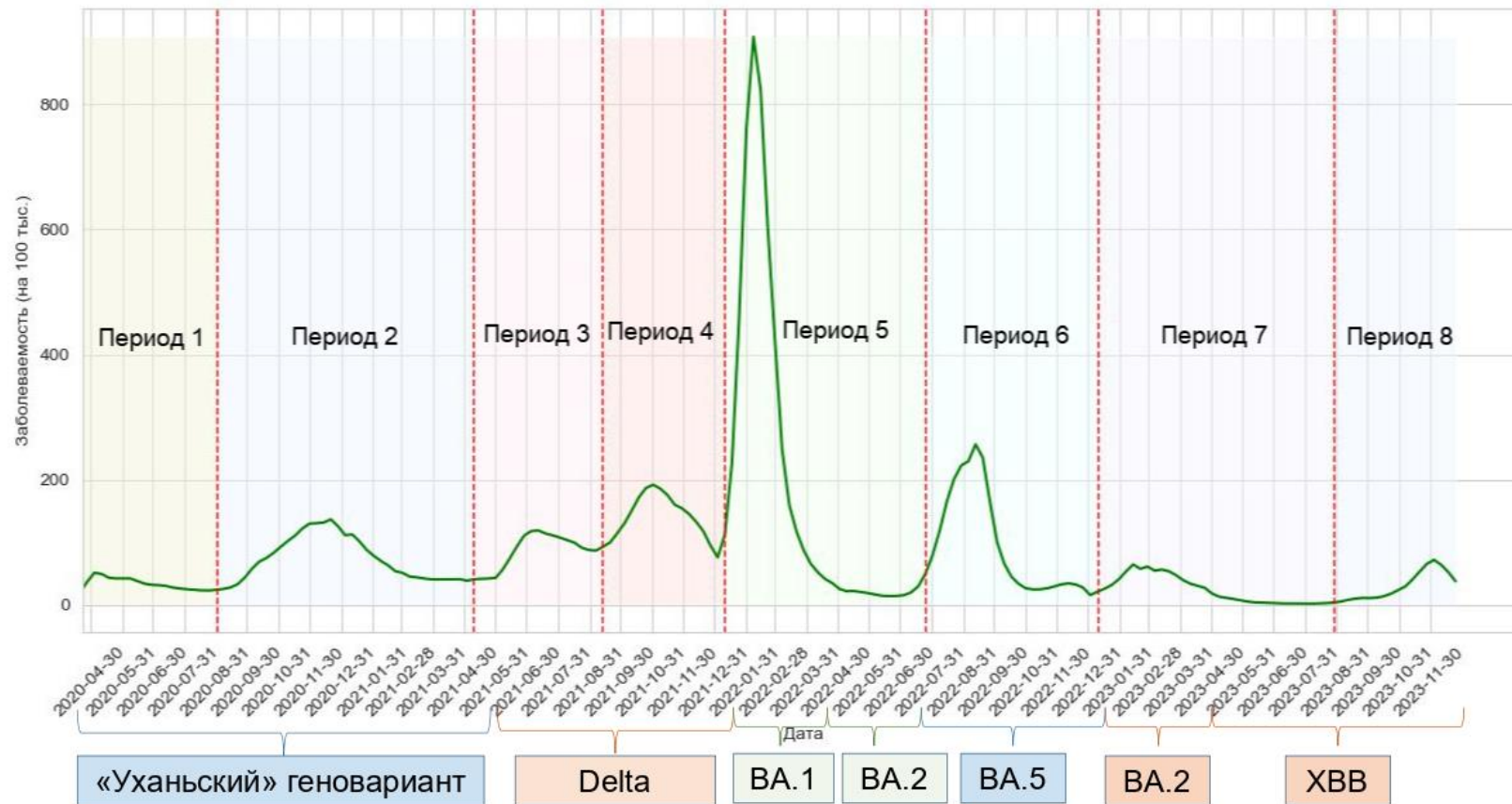
Существенный рост заболеваемости в конце января – феврале 2022 г. сопровождался изменениями возрастной структуры заболевших, что связано с высокой ролью детского населения в распространении варианта Omicron SARS-CoV-2. На этой фазе эпидемического процесса доминировала генетическая линия BA.1 Omicron SARS-CoV-2, которая в конце февраля – начале марта 2022 г. была вытеснена линией BA.2 Omicron SARS-CoV-2. Смена доминирующего варианта происходила на фоне устойчивого снижения заболеваемости, и уже в начале мая 2022 г. средний недельный показатель составил  $21,75 \pm 1,36$  случая на 100 тыс. населения.

**Шестой подъем заболеваемости** зарегистрирован в августе 2022 г. и связан с распространением линии BA.5 Omicron SARS-CoV-2. Максимальные показатели зарегистрированы на 38-й неделе 2022 г. (12 сентября), что совпало с влиянием сезонных факторов. С ноября 2022 г. отмечалась стабилизация эпидемического процесса, а средний недельный показатель составил  $26,36 \pm 1,94$  случая на 100 тыс. населения.

С середины января 2023 г. формировался **седьмой эпидемический подъем** заболеваемости новой коронавирусной инфекцией (COVID-19), обусловленный доминированием генетической линии XBB SARS-CoV-2. Наибольший недельный показатель заболеваемости при этом зарегистрирован на 8-й неделе 2023 г. (13 февраля) и составил 64,64 случая на 100 тыс. населения. С конца мая 2023 г. эпидемический процесс новой коронавирусной

инфекции (COVID-19) стабилизировался, а средний недельный показатель снизился до  $2,90 \pm 0,53$  случая на 100 тыс. населения.

**Восьмой подъем заболеваемости** новой коронавирусной инфекцией (COVID-19) зарегистрирован в сентябре 2023 г., что совпало с началом учебного года и активизацией сезонных факторов, способствующих распространению инфекции нового варианта возбудителя новой коронавирусной инфекции (SARS-CoV-2). Пик заболеваемости пришелся на 50-ю неделю 2023 г. (11 декабря 2023 г.) и составил 71,95 случая на 100 тыс. населения (рис. 3.1.).



**Рисунок 3.1** — Динамика недельных показателей заболеваемости новой коронавирусной инфекции (COVID-19) среди населения Российской Федерации в период 2020–2023 гг., рассчитанных на 100 тыс. населения

**Таблица 3.1** — Характеристика периодов эпидемического процесса новой коронавирусной инфекции (COVID-19) на территории Российской Федерации в 2020–2023 гг. с учетом распределения субъектов по кластерам

Номер периода	Начало периода	Конец периода	Длительность (неделя)	Недельные показатели заболеваемости по Российской Федерации (мин., макс., среднее) на 100 тыс.	Доминирующий вариант SARS-CoV-2	Кластер 3 (высокий уровень заболеваемости)		Кластер 2 (преимущественно высокий уровень заболеваемости)		Кластер 1 (преимущественно низкий уровень заболеваемости)		Кластер 0 (низкий уровень заболеваемости)	
						Кол-во субъектов	Средний недельный показатель заболеваемости (на 100 тыс)	Кол-во субъектов	Средний недельный показатель заболеваемости (на 100 тыс)	Кол-во субъектов	Средний недельный показатель заболеваемости (на 100 тыс)	Кол-во субъектов	Средний недельный показатель заболеваемости (на 100 тыс)
1	-	31.08.2020	-	макс - 51,41, среднее - 33,28±4,14	Уханьский	2	87,88±13,85	7	47,34±4,13	19	37,77±2,92	34	19,53±0,87
2	01.09.2020	09.05.2021	35	мин - 25,41, макс - 136,69, среднее - 74,28±11,74	Уханьский	1	217,14±51,72	4	181,23±2,56	23	80,96±3,33	34	46,70±1,50
3	10.05.2021	12.09.2021	18	мин - 40,90, макс - 118,80, среднее - 85,63±13,22	Delta	1	283,99±77,07	2	189,97±29,75	21	109,24±6,16	38	58,36±2,51
4	13.09.2021	09.01.2022	17	мин - 75,96, макс - 191,87, среднее - 140,04±17,32	Delta	1	538,08±69,98	9	242,60±14,50	18	167,15±6,52	34	106,25±3,28
5	10.01.2022	25.07.2022	29	мин - 14,00, макс - 907,46, среднее - 186,19 ± 97,55	BA.1, BA.2 (Omicron)	1	502,47±249,72	15	260,31±38,94	28	219,26±24,52	18	125,89±17,07
6	26.07.2022	09.01.2023	24	мин - 15,66, макс - 256,47, среднее - 42,83 ± 33,41	BA.6 (Omicron)	1	162,66±18,82	12	93,11±9,75	18	93,00±9,65	21	60,87±7,73
7	10.01.2023	28.08.2023	33	мин - 1,98, макс - 64,64, среднее - 23,47 ± 7,57	BA.2, XBB (Omicron)	7	51,03±6,95	1	41,36±10,54	16	29,23±2,41	38	16,72±0,97
8	29.08.2023	-	-	мин - 5,43, макс - 71,95, среднее - 31,00 ± 10,87	XBB (Omicron)	2	64,76±19,01	8	58,47±8,48	19	35,38±3,82	33	17,58±1,38

Динамика эпидемического процесса новой коронавирусной инфекции (COVID-19) в Российской Федерации в 2020–2023 гг. носила волнообразный характер с подъемами заболеваемости различной продолжительности и интенсивности. Учитывая существенные различия периодов заболеваемости, а также представленные различия в структуре циркулирующих генетических вариантов возбудителя новой коронавирусной инфекции (SARS-CoV-2), целесообразно проведение раздельного анализа уровня и динамики заболеваемости и факторов, оказывающих влияние на эпидемический процесс новой коронавирусной инфекции (COVID-19) для каждого из восьми периодов раздельно.

Показатель заболеваемости новой коронавирусной инфекции (COVID-19) всего населения Российской Федерации представляет собой среднее значение и не позволяет оценить насколько сильно варьируют уровни заболеваемости в разных регионах. Вариативность показателя заболеваемости новой коронавирусной инфекцией (COVID-19) в субъектах Российской Федерации в различные периоды пандемии показана в таблице 3.2.

**Таблица 3.2** — Распределение суточных показателей заболеваемости населения субъектов Российской Федерации в разные периоды пандемии новой коронавирусной инфекции (COVID-19)

Показатель на 100 тыс. населения					
Номер периода и даты	Минимальное значение	Максимальное значение	Среднее значение	Стандартное отклонение	Доверительный интервал (95%)
1. 01.03.2020 – 31.08.2020	0,69	18,46	5,06	3,34	5,06±0,71
2. 01.09.2020 – 09.05.2021	1,46	31,09	10,78	6,16	10,78±1,31
3. 10.05.2021 – 12.09.2021	1,08	40,57	12,20	6,28	12,20±1,34
4. 13.09.2021 – 09.01.2022	3,78	76,87	20,87	9,92	20,87±2,11

Показатель на 100 тыс. населения					
Номер периода и даты	Минимальное значение	Максимальное значение	Среднее значение	Стандартное отклонение	Доверительный интервал (95%)
5. 10.01.2022 – 25.07.2022	4,33	71,43	28,20	12,2	28,20±2,70
6. 26.07.2022 – 09.01.2022	0,93	31,49	12,41	5,60	12,41±1,19
7. 10.01.2023 – 28.08.2023	0,11	9,51	3,14	1,84	3,14±0,39
8. 29.08.2023 – 17.12.2023	0,17	12,55	3,95	2,57	3,95±0,55

Из данных, приведенных в таблице 3.2 видно, что в каждом из анализируемых периодов показатели заболеваемости новой коронавирусной инфекцией (COVID-19) распределены неравномерно относительно среднего значения. В некоторых регионах они превышали уровень 75-го перцентиля более чем на полтора межквартильных интервала. Такие высокие показатели заболеваемости без учета контекста и особенностей проведения эпидемиологического анализа при классическом варианте исследования могли бы быть отнесены к выбросам с последующим исключением из базы данных, однако, в нашем исследовании использованы данные о заболеваемости, скорректированные с учетом случайных отклонений.

В каждом из рассматриваемых периодов также существовали регионы с гораздо более низким уровнем заболеваемости по сравнению со средним значением. Однако, эти регионы не были выделены как «выбросы» («выскакивающие величины») из-за смещения распределения показателей заболеваемости среди субъектов Российской Федерации в большую сторону. Выявление таких регионов также является важной частью эпидемиологического анализа.

На наш взгляд, изучение распределения регионов Российской Федерации по показателю заболеваемости является основой для оценки интенсивности действия факторов, обуславливающих динамику эпидемического процесса новой коронавирусной инфекции (COVID-19) на

территории всей Российской Федерации, и изучения эффективности проводимого комплекса противоэпидемических мероприятий, что в свою очередь является предварительным и необходимым этапом работы по изучению влияния различных факторов на эпидемический процесс новой коронавирусной инфекции (COVID-19).

### **3.2. Кластеризация интенсивных дневных показателей заболеваемости новой коронавирусной инфекции (COVID-19) в Российской Федерации в 2020-2023 гг.**

Для всестороннего изучения динамики заболеваемости населения новой коронавирусной инфекцией (COVID-19) и влияния различных факторов на эпидемический процесс в различных субъектах Российской Федерации в разные периоды пандемии применен метод кластерного анализа. Кластерный анализ является статистическим методом, использующимся для группировки объектов (в данном случае — регионов) на однородные группы (кластеры) по схожим признакам. В рамках настоящего исследования регионы объединялись в кластеры на основе значений суточной заболеваемости новой коронавирусной инфекцией (COVID-19). Подобный подход позволил выявить территориальные различия в динамике эпидемического процесса новой коронавирусной инфекции (COVID-19) и классифицировать регионы по уровню и характеру распространения инфекции.

Поскольку результаты кластерного анализа существенно зависят от качества исходных данных, предварительно нами проведена оценка полноты временных рядов по данным дневной заболеваемости новой коронавирусной инфекцией (COVID-19) по каждому субъекту Российской Федерации. Выявлены регионы, в которых доля пропущенных значений превышала 1% от общего числа дней наблюдения. Данные субъекты исключены из основного кластерного анализа и объединены в отдельную группу — кластер 4. На рисунке 3.2 представлена доля пропущенных значений по каждому из таких регионов. В кластер 4 вошли 23 субъекта Российской Федерации, доля пропусков в которых варьировала от 1 до 17 %. Наиболее высокие значения пропусков наблюдались в Чукотском автономном округе, Республике Тыва и Ненецком автономном округе, что ограничивает возможность использования данных по заболеваемости населения новой коронавирусной инфекцией (COVID-19) для корректного сопоставления с остальными субъектами.

Для кластеризации дневных интенсивных показателей заболеваемости новой коронавирусной инфекцией (COVID-19) применен метод К-средних. В основе данного метода лежит объединение временных рядов по степени их сходства; при этом расстояние между рядом и центром кластера минимизируется на каждом шаге итерационного расчета. Такой подход устраняет необходимость заранее задавать пороговые уровни заболеваемости для классификации субъектов, позволяя алгоритму самостоятельно формировать группы на основе анализа структуры данных.

Кластеризация в данном исследовании охватывает период с 01 января 2020 г. по 01 января 2024 г. В результате нами выделено четыре кластера (0–3) (таблица 3.1), характеризующиеся различной интенсивностью эпидемического процесса новой коронавирусной инфекции (COVID-19).

**Кластер 0** включает субъекты с устойчиво низким уровнем заболеваемости (средний недельный показатель заболеваемости колебался от  $16,72 \pm 0,97$  в 7 периоде до  $125,89 \pm 17,07$  на 100 тыс. населения в 5 периоде), в которых дневные показатели на протяжении всего периода наблюдений оставались значительно ниже среднероссийских значений.

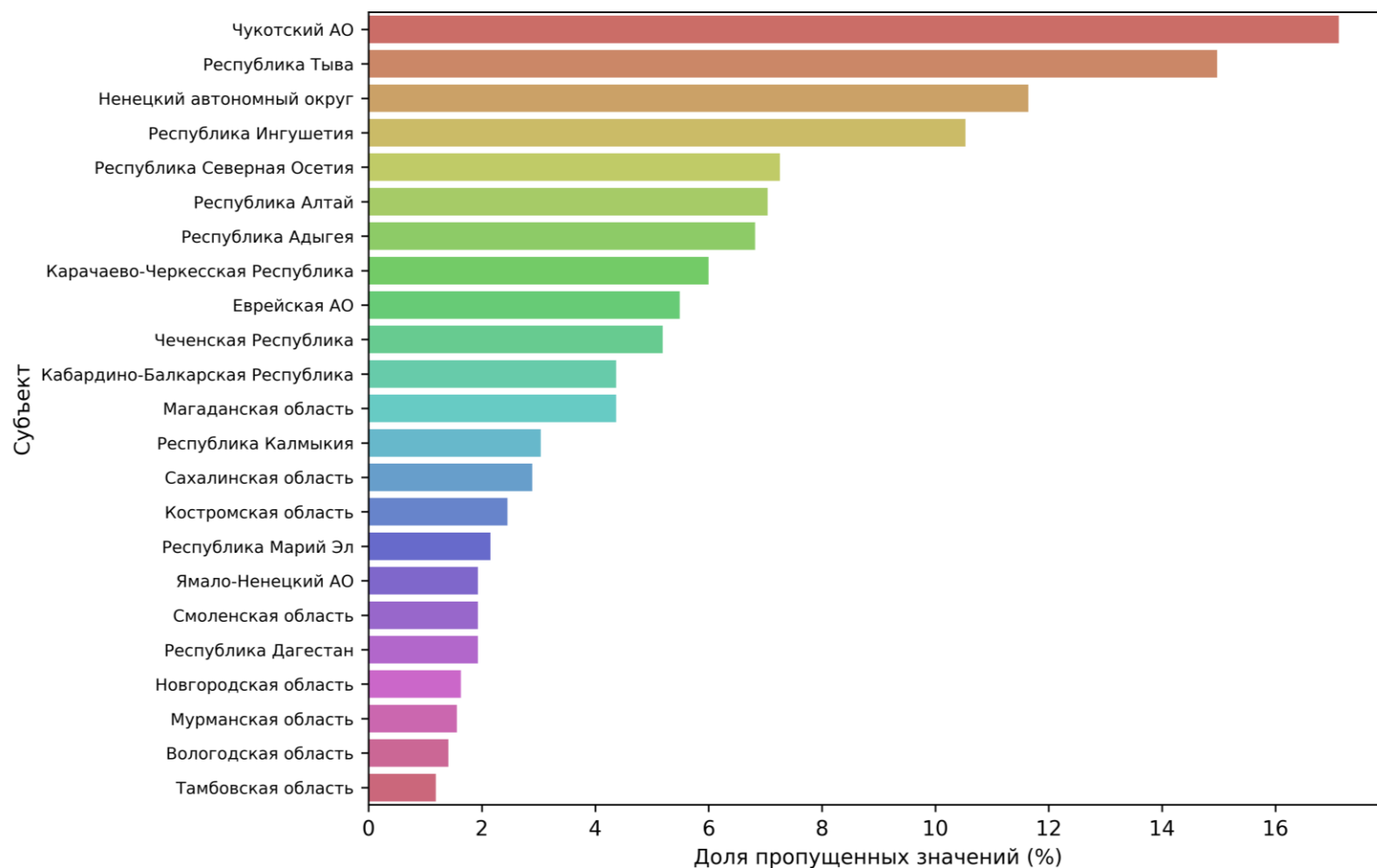
В регионах, отнесенных к **кластеру 1**, заболеваемость имела преимущественно низкий уровень (средний недельный показатель заболеваемости колебался от  $29,23 \pm 2,41$  в 7 периоде до  $219,26 \pm 24,52$  на 100 тыс. населения в 5 периоде), однако для них были характерны отдельные периоды умеренного роста интенсивности эпидемического процесса новой коронавирусной инфекции (COVID-19).

В **кластер 2** объединены субъекты, демонстрировавшие преимущественно высокий уровень заболеваемости населения новой коронавирусной инфекцией (COVID-19) (средний недельный показатель заболеваемости колебался от  $41,36 \pm 10,54$  в 7 периоде до  $260,31 \pm 38,94$  на 100 тыс. населения в 5 периоде). Для данного кластера характерна выраженная динамика подъемов на фоне смены доминирующих генетических вариантов возбудителя новой коронавирусной инфекции (SARS-CoV-2).

В кластер 3 отнесены субъекты Российской Федерации, характеризующиеся стабильно высоким уровнем заболеваемости населения новой коронавирусной инфекцией (COVID-19) на протяжении всего анализируемого периода с 2020 по 2023 гг. (средний недельный показатель заболеваемости колебался от  $51,03 \pm 6,95$  в 7 периоде до  $538,08 \pm 69,98$  на 100 тыс. населения в 4 периоде). Для данных регионов отмечены выраженные подъемы интенсивности эпидемического процесса, а также высокая чувствительность к смене генетических вариантов возбудителя новой коронавирусной инфекции (SARS-CoV-2), что приводило к более заметным колебаниям дневных и недельных показателей заболеваемости.

В кластер 4, как отмечалось ранее, включены субъекты, имеющие значительную долю пропущенных значений ежедневной заболеваемости. Данные регионы выделены в отдельную группу, поскольку неполная характеристика данных не позволяет корректно сравнивать их динамику с остальными субъектами и исключены из дальнейшего анализа.

Рассмотрим результаты кластеризации субъектов Российской Федерации по заболеваемости населения новой коронавирусной инфекцией (COVID-19) в каждом из эпидемических периодов, выделенных выше. Метод кластеризации позволяет сопоставлять субъекты Российской Федерации по сходству проявлений эпидемического процесса новой коронавирусной инфекции (COVID-19) в динамике, а также выявлять группы регионов, в которых развитие эпидемии имеет аналогичные характеристики. Такой подход является частным случаем применения машинного обучения, в котором степень сходства временных рядов определяется алгоритмом самостоятельно и не требует предварительного введения экспертных границ для группирования данных.



**Рисунок 3.2** — Доля пропущенных дневных показателей заболеваемости новой коронавирусной инфекцией (COVID-19) среди субъектов Российской Федерации, исключенных из кластерного анализа

На рисунке 3.3 представлены средние суточные значения интенсивного показателя заболеваемости новой коронавирусной инфекцией (COVID-19) в каждом кластере по периодам пандемии 2020–2023 гг. Данные на рисунке 3.3 согласуются с недельными показателями, приведенными в таблице 3.1, и демонстрируют характерные различия в динамике эпидемического процесса между кластерами.

В первые пять эпидемических периодов наблюдался постепенный рост среднего показателя заболеваемости новой коронавирусной инфекцией (COVID-19) во всех кластерах, что связано с последовательной сменой доминирующих генетических вариантов возбудителя новой коронавирусной инфекции (SARS-CoV-2), расширением масштабов эпидемического процесса и увеличением интенсивности распространения инфекции в совокупном населении. Особенно выраженная динамика отмечена в кластере 3, где средние показатели в периодах 4 и 5 достигали максимальных значений ( $538,08 \pm 69,98$  и  $502,47 \pm 249,72$  на 100 тыс. населения, соответственно) и существенно (в 2–5 раз,  $p < 0,01$ ) превышали уровни заболеваемости в других кластерах. В кластере 2 также наблюдались высокие значения ( $242,60 \pm 14,50$  и  $260,31 \pm 38,94$  на 100 тыс. населения в 4 и 5 периоды, соответственно), отражающие значительную нагрузку, связанную с активным распространением вариантов Delta и Omicron SARS-CoV-2.

Начиная с шестого периода пандемии, средний уровень заболеваемости начал снижаться во всех кластерах. Данный процесс связан с уменьшением числа восприимчивых лиц в результате перенесенной новой коронавирусной инфекции (COVID-19), вакцинации и ограничительными мероприятиями, реализованными в субъектах Российской Федерации. Наиболее значимое снижение отмечено в кластерах 0 и 1, где эпидемический процесс имел менее выраженный характер. В кластере 3 снижение также было отчетливым, однако уровень заболеваемости оставался выше, чем в других группах ( $51,03 \pm 6,95$  и  $64,76 \pm 19,01$  на 100 тыс. населения в 7 и 8 периодах,

соответственно), что подтверждает устойчивую предрасположенность данных регионов к более интенсивному распространению среди населения возбудителя новой коронавирусной инфекции (SARS-CoV-2).

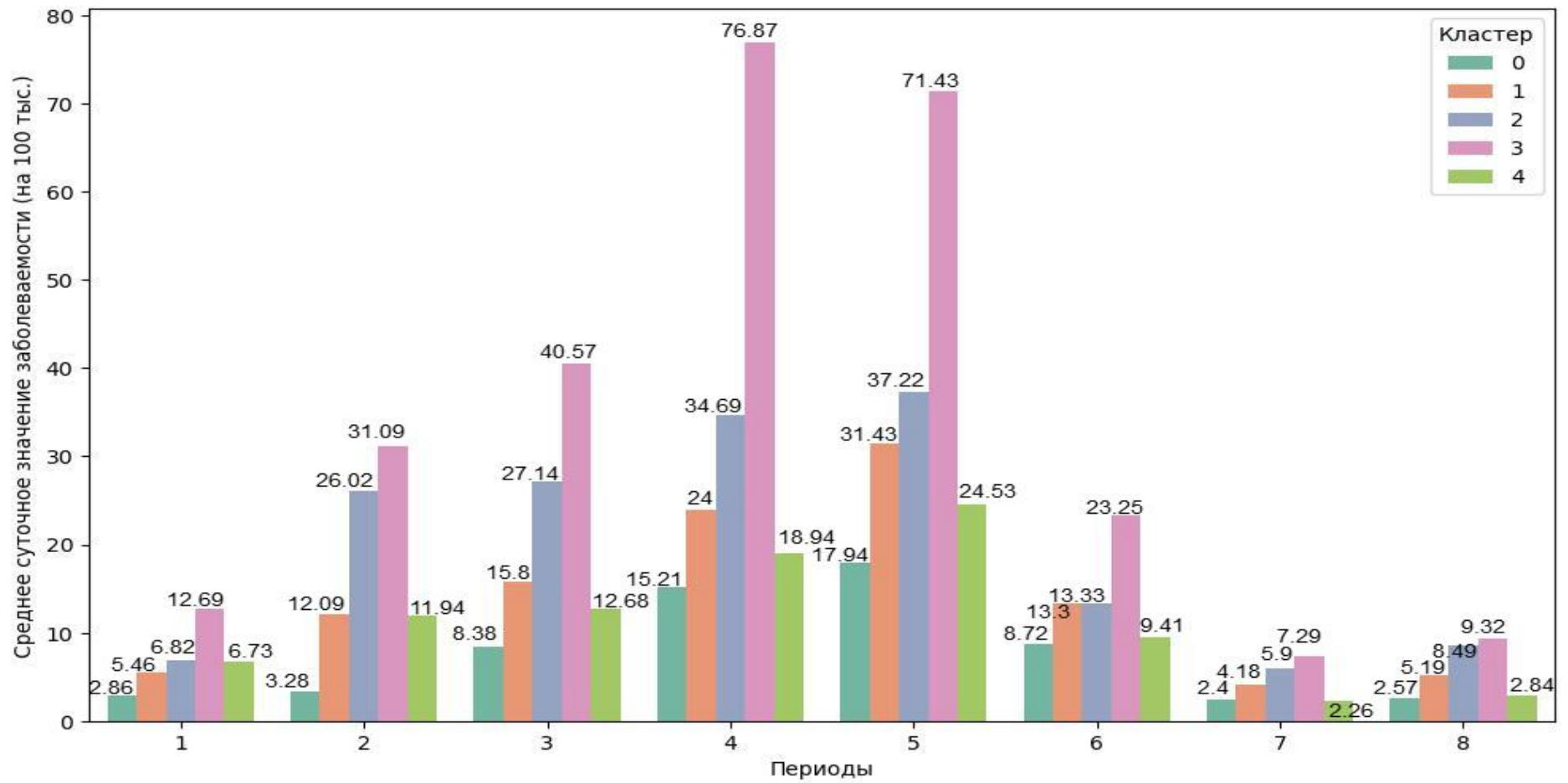
Рассмотрим возможность использовать описание средних показателей заболеваемости новой коронавирусной инфекцией (COVID-9) и дисперсии среди кластеров в различные периоды пандемии как дополнительного способа изучения уровня и динамики эпидемического процесса, его преимущества и недостатки перед описанием уровня и динамики среди совокупного населения при классическом подходе.

Дисперсия внутри каждого эпидемического периода будет показывать насколько интенсивное влияние на динамику эпидемического процесса оказывают различные факторы, в том числе характер и эффективность применяемых в каждом регионе и стране в целом противоэпидемических мероприятий, что невозможно оценить при анализе динамики заболеваемости совокупного населения. Также, величина разброса показателей заболеваемости новой коронавирусной инфекцией (COVID-19) в каждом из эпидемических периодов может указывать на различия в диагностике и учете случаев заболевания в каждом из регионов.

Таким образом, динамика средних суточных значений интенсивного показателя заболеваемости новой коронавирусной инфекцией (COVID-19) в кластерах не только отражает общую эпидемиологическую ситуацию в Российской Федерации в 2020-2023 гг., но и позволяет более детально проследить территориальные различия и специфику развития эпидемического процесса на уровне субъектов (рисунок 3.3, таблица 3.1).

Различия между среднесуточными показателями заболеваемости новой коронавирусной инфекцией (COVID-19) в субъектах, относящихся к кластерам 0 и 1, оказалась наиболее выраженной во втором эпидемическом периоде, когда значения в кластере 1 ( $12,09 \pm 0,17$  на 100 тыс. населения) превышали значения кластера 0 ( $3,28 \pm 0,08$  на 100 тыс. населения) в 3,7 раза ( $p < 0,01$ ). В шестом периоде различия были менее значительными ( $8,72 \pm 0,33$  и

13,3±0,53 на 100 тыс. населения в 0 и 1 кластере, соответственно) и различались в 1,5 раза ( $p < 0,01$ ), что свидетельствует об аналогичных тенденциях в динамике эпидемического процесса в регионах с минимальными уровнями заболеваемости новой коронавирусной инфекцией (COVID-19) (рисунок 3.3, таблица 3.1).



**Рисунок 3.3** — Динамика суточных интенсивных показателей заболеваемости новой коронавирусной инфекцией (COVID-19) в кластерах в каждом из периодов пандемии 2020–2023 гг.

Сравнение кластеров 1 и 2 показало, что в периодах 1, 6, 7 и 8 различия между средними суточными показателями заболеваемости новой коронавирусной инфекцией (COVID-19) были минимальными. Наиболее близкие значения зарегистрированы в шестом периоде —  $13,30 \pm 0,53$  и  $13,33 \pm 0,54$  на 100 тыс. населения, соответственно (различия — около 0,2%), тогда как в восьмом периоде различие достигало 63% —  $5,19 \pm 0,23$  и  $8,49 \pm 0,48$  на 100 тыс. населения, соответственно (рисунок 3.3, таблица 3.1).

В периодах 2, 3, 4 и 5 различия между кластерами 1 и 2 возрастали, что отражает неоднородность распространения доминирующих генетических вариантов возбудителя новой коронавирусной инфекции (SARS-CoV-2) среди населения различных регионов Российской Федерации. В частности, во втором периоде показатели кластера 2 ( $26,02 \pm 1,25$  на 100 тыс. населения) превышали значения кластера 1 ( $12,09 \pm 0,17$  на 100 тыс. населения) почти в 2 раза ( $p < 0,01$ ), что совпадает с активным ростом заболеваемости новой коронавирусной инфекцией (COVID-19) осенью 2020 г. (рисунок 3.3, таблица 3.1).

Различия между кластерами 2 и 3, где объединены регионы со средним и высоким уровнями заболеваемости, также были существенными. В восьмом периоде зафиксирован минимальный разрыв (9,77%), тогда как в четвертом периоде значения кластера 3 ( $76,87 \pm 3,73$  на 100 тыс. населения) превышали показатели кластера 2 ( $34,69 \pm 0,80$  на 100 тыс. населения) почти в 2 раза ( $p < 0,01$ ), что подтверждает устойчивое эпидемиологическое неблагополучие в субъектах, относящихся к кластеру 3, где на протяжении нескольких периодов наблюдались выраженные эпидемические подъемы заболеваемости новой коронавирусной инфекцией (COVID-19) (рисунок 3.3, таблица 3.1).

Как отмечалось ранее, субъекты Российской Федерации, в которых доля пропущенных ежедневных значений превышала 1%, были объединены в кластер 4. Вследствие отсутствия значительного объема данных

сопоставление динамики заболеваемости в данной группе с другими кластерами не проводилось.

Распределение субъектов Российской Федерации по кластерам в каждом из восьми эпидемических периодов представлено в таблице 3.3, где регионы упорядочены по убыванию суммы номеров кластеров. Такой подход позволяет выявить территории Российской Федерации, в которых уровень заболеваемости новой коронавирусной инфекцией (COVID-19) в анализируемый период был наиболее высокий. Согласно проведенному анализу, наиболее интенсивное развитие эпидемического процесса новой коронавирусной инфекцией (COVID-19) отмечалось в Хабаровском крае, г. Санкт-Петербурге и г. Москве — регионах, которые неоднократно относились к кластерам с повышенным уровнем заболеваемости.

Определение регионов с наиболее высокими уровнями заболеваемости новой коронавирусной инфекцией (COVID-19) играет важную роль в выявлении территорий с наибольшими рисками распространения инфекции. Однако, оценка суммарных значений за длительный период может искажаться кратковременными подъемами заболеваемости или возможными ошибками регистрации новых случаев новой коронавирусной инфекции (COVID-19). В этой связи особую значимость приобретает анализ переходов регионов между кластерами в разные периоды пандемии. Изменение кластерной принадлежности отражает динамику эпидемиологической ситуации и позволяет оценить эффективность реализуемых противоэпидемических мероприятий в субъектах Российской Федерации.

**Таблица 3.3** — Распределения субъектов Российской Федерации по кластерам интенсивных суточных показателей заболеваемости новой коронавирусной инфекцией (COVID-19) в различные периоды пандемии с 2020 по 2023 гг.

Субъект	Период								Сумма по кластерам (кластерный индекс)
	1	2	3	4	5	6	7	8	
Хабаровский край	3	3	3	3	3	3	2	2	22
г. Санкт-Петербург	1	2	2	2	2	3	3	2	17
г. Москва	1	2	2	2	0	3	3	2	15
Республика Карелия	0	2	1	2	2	3	3	1	14
Архангельская область	2	1	1	2	2	3	1	1	13
Республика Саха (Якутия)	1	0	1	1	2	3	1	3	12
Ханты-Мансийский АО	3	1	0	0	1	3	1	3	12
Воронежская область	1	1	1	1	2	1	3	1	11
Псковская область	2	2	1	2	2	1	0	1	11
г. Севастополь	0	1	1	2	2	2	1	2	11
Республика Коми	2	1	1	2	1	2	1	0	10
Самарская область	0	0	0	1	1	3	3	2	10
Орловская область	2	1	0	2	2	1	0	1	9
Пензенская область	1	1	1	1	2	2	1	0	9
Ярославская область	1	1	1	0	2	2	1	1	9
Забайкальский край	0	1	1	1	1	2	1	1	8

Продолжение таблицы 3.3

Субъект	Период								Сумма по кластерам (кластерный индекс)
	1	2	3	4	5	6	7	8	
Калининградская область	0	1	1	1	1	2	1	1	8
Пермский край	0	0	0	1	1	3	1	2	8
Тверская область	0	1	1	1	1	0	3	1	8
Тюменская область	0	0	0	0	1	3	3	1	8
Удмуртская Республика	0	0	0	0	1	3	1	2	7
Ульяновская область	2	1	1	2	1	0	0	0	7
Амурская область	0	0	0	1	2	2	1	0	6
Иркутская область	2	0	0	0	1	2	0	1	6
Липецкая область	0	1	1	1	2	1	0	0	6
Нижегородская область	1	1	0	1	1	1	0	1	6
Рязанская область	1	0	0	0	2	1	1	1	6
Чувашская Республика	1	0	0	0	0	2	1	2	6
Брянская область	1	1	1	1	1	0	0	0	5
Ивановская область	1	1	0	0	1	1	0	1	5
Курская область	1	1	1	1	0	1	0	0	5
Астраханская область	1	1	1	1	0	0	0	0	4
Камчатский край	2	1	0	0	1	0	0	0	4
Курганская область	0	0	0	1	1	2	0	0	4

Продолжение таблицы 3.3

Субъект	Период								Сумма по кластерам (кластерный индекс)
	1	2	3	4	5	6	7	8	
Московская область	1	1	1	1	0	0	0	0	4
Республика Башкортостан	0	0	0	0	1	2	0	1	4
Свердловская область	1	0	0	0	1	2	0	0	4
Томская область	0	0	0	0	2	1	1	0	4
Владимирская область	0	0	0	0	1	1	0	1	3
Красноярский край	1	0	0	0	1	1	0	0	3
Республика Бурятия	0	1	1	0	1	0	0	0	3
Саратовская область	0	0	0	0	1	1	1	0	3
Алтайский край	0	0	0	0	1	1	0	0	2
Белгородская область	0	0	0	0	0	1	0	1	2
Волгоградская область	0	0	0	0	2	0	0	0	2
Калужская область	1	1	0	0	0	0	0	0	2
Кировская область	0	1	0	0	1	0	0	0	2
Ленинградская область	0	0	0	0	0	1	0	1	2
Новосибирская область	0	0	0	0	1	1	0	0	2
Омская область	0	0	1	1	0	0	0	0	2
Республика Крым	0	0	1	1	0	0	0	0	2
Республика Мордовия	1	0	0	0	0	1	0	0	2

Продолжение таблицы 3.3

Субъект	Период								Сумма по кластерам (кластерный индекс)
	1	2	3	4	5	6	7	8	
Тульская область	1	0	0	0	1	0	0	0	2
Челябинская область	0	0	0	0	1	1	0	0	2
Кемеровская область	0	0	0	0	0	0	0	1	1
Оренбургская область	0	0	0	0	1	0	0	0	1
Краснодарский край	0	0	0	0	0	0	0	0	0
Приморский край	0	0	0	0	0	0	0	0	0
Республика Татарстан	0	0	0	0	0	0	0	0	0
Республика Хакасия	0	0	0	0	0	0	0	0	0
Ростовская область	0	0	0	0	0	0	0	0	0
Ставропольский край	0	0	0	0	0	0	0	0	0

В кластер 3 включены субъекты Российской Федерации, в которых регистрировались наиболее высокие показатели заболеваемости. Структура этого кластера была относительно стабильной: г. Москва, г. Санкт-Петербург и Хабаровский край отнесены в группу наиболее неблагоприятных регионов в большинство периодов, а в отдельные этапы пандемии в состав данного кластера также входили Республика Карелия, Архангельская область, Ханты-Мансийский автономный округ и другие субъекты. Можно предположить, что наиболее значимые факторы, определившие принадлежность этих территорий к кластеру с наиболее интенсивным развитием эпидемического процесса новой коронавирусной инфекции (COVID-19), связаны с особенностями расселения, плотностью населения,

высокой степенью урбанизации и выраженной транспортной нагрузкой, и связанностью между собой большинства из них.

г. Москва и г. Санкт-Петербург представляют собой крупнейшие мегаполисы Российской Федерации, где почти все население проживает в городской среде, а плотность населения достигает одних из максимальных значений в стране, что формирует условия для интенсивного социального взаимодействия в транспорте, местах работы, учебных учреждениях и многоэтажной жилой застройке.

Хабаровский край, при меньшей населенности в целом, также характеризуется высокой долей городских жителей, сосредоточенных в крупных агломерациях, таких как г. Хабаровск и г. Комсомольск-на-Амуре, что делает условия для передачи инфекции схожими с мегаполисами. Существенную роль играет и статус данных территорий как ключевых транспортных узлов: г. Москва традиционно является главным международным авиаузлом и центром железнодорожной сети страны, через который до начала пандемии проходили десятки миллионов пассажиров ежегодно; г. Санкт-Петербург выполняет аналогичные функции крупного туристического, делового и портового центра; г. Хабаровск является важнейшим транспортным хабом Дальнего Востока, через который проходят авиационные и железнодорожные маршруты, связывающие регион с Китаем и другими странами. Подобная пространственная и транспортная структура обусловила раннее появление новых генетических вариантов возбудителя новой коронавирусной инфекции (SARS-CoV-2) именно в данных регионах, а также более ранние и выраженные эпидемические подъемы заболеваемости новой коронавирусной инфекцией (COVID-19).

В состав кластера 2 входили субъекты Российской Федерации, в которых уровень заболеваемости новой коронавирусной инфекцией (COVID-19) на протяжении значительной части пандемии был выше среднего ( $6,82 \pm 0,25$ ,  $26,02 \pm 1,25$ ,  $27,14 \pm 1,67$ ,  $34,69 \pm 0,80$ ,  $37,22 \pm 2,15$ ,  $13,33 \pm 0,54$ ,  $5,90 \pm 0,58$ ,  $8,49 \pm 0,48$  на 100 тыс. населения с 1 по 8 периоды, соответственно),

но не столь стабильно и экстремально высок, как в кластере 3 (рисунок 3.3). Данные территории характеризовались выраженными подъемами заболеваемости в отдельные периоды, формируя своего рода «вторую линию» распространения инфекции после регионов с максимальной эпидемической нагрузкой. Наиболее часто в кластер 2 попадали Московская и Ленинградская области, образующие плотное социально-экономическое пространство вокруг двух мегаполисов, а также крупные промышленные регионы Центральной Российской Федерации, такие как Нижегородская, Владимирская, Ярославская и Тульская области, значительная часть субъектов Поволжья и Урала, включая Самарскую, Свердловскую и Челябинскую области и Республику Башкортостан, а также ряд северо-западных и южных территорий, среди которых Архангельская, Псковская и Ростовская области. Динамика заболеваемости в кластерах 2 и 3 нередко была схожей: в отдельные периоды показатели заболеваемости практически совпадали, как, например, в восьмом эпидемическом периоде, когда различия между среднесуточными значениями кластеров 2 и 3 составляла лишь около 10% ( $8,49 \pm 0,48$  и  $9,32 \pm 1,16$  на 100 тыс. населения, соответственно), что указывает на сходство проявлений эпидемического процесса в этих группах регионов.

В отличие от регионов преимущественно относившихся к кластеру 2, субъекты, входившие в кластер 1, демонстрировали преимущественно низкие уровни заболеваемости на протяжении большей части наблюдаемого периода ( $5,46 \pm 0,16$ ,  $12,09 \pm 0,17$ ,  $15,80 \pm 0,33$ ,  $24,00 \pm 0,36$ ,  $31,43 \pm 1,34$ ,  $13,30 \pm 0,53$ ,  $4,18 \pm 0,14$ ,  $5,19 \pm 0,23$  на 100 тыс. населения с 1 по 8 периоды, соответственно). К этой группе относились Калужская, Брянская, Оренбургская, Омская, Волгоградская области, ряд регионов Поволжья и Сибири, включая Ульяновскую, Пензенскую и Томскую области, Республика Бурятия, а также некоторые северные территории, такие как Республика Коми и Республика Карелия, и субъекты Северо-Кавказского федерального округа. Относительно невысокие значения заболеваемости в этих регионах могут быть связаны с меньшей плотностью населения, особенностями расселения, более низкой

интенсивностью миграционных потоков или своевременным введением ограничительных мер.

Наиболее специфичным оказался кластер 0, включавший разнородную группу субъектов Российской Федерации, где на протяжении всех периодов наблюдений не отмечено резких и выраженных подъемов заболеваемости —  $2,86 \pm 0,05$ ,  $3,28 \pm 0,08$ ,  $8,38 \pm 0,14$ ,  $15,21 \pm 0,18$ ,  $17,94 \pm 0,94$ ,  $8,72 \pm 0,33$ ,  $2,4 \pm 0,06$ ,  $2,57 \pm 0,08$  на 100 тыс. населения с 1 по 8 периоды, соответственно. Для данных территорий характерна либо географическая изолированность, ограничивающая интенсивность внешних контактов, либо быстрое и эффективное реагирование региональных органов власти на угрозу распространения возбудителя новой коронавирусной инфекции (SARS-CoV-2), что позволило избежать значительных резких эпидемических подъемов заболеваемости новой коронавирусной инфекцией (COVID-19) в разные периоды пандемии.

Таким образом:

1. В динамике заболеваемости новой коронавирусной инфекцией (COVID-19) населения Российской Федерации в период 2020–2023 гг. наблюдалось восемь подъемов и спадов заболеваемости, каждый из которых характеризовался рядом особенностей, как: интенсивность эпидемического процесса, его продолжительность, доминирующий вариант возбудителя новой коронавирусной инфекции (SARS-CoV-2) и комплекс проводимых противоэпидемических мероприятий.

2. Динамика заболеваемости новой коронавирусной инфекцией (COVID-19) в период с апреля 2020 г. по январь 2022 г. имела тенденцию к росту. Наиболее высокий уровень заболеваемости новой коронавирусной инфекцией (COVID-19) зарегистрирован в пятом периоде пандемии в январе — феврале 2022 г. — 905 случаев на 100 тыс. населения, после чего наметилась тенденция к снижению уровня заболеваемости.

3. Анализ средних показателей и дисперсии заболеваемости населения Российской Федерации среди кластеров внутри каждого периода

позволяет выявить следующие особенности эпидемического процесса новой коронавирусной инфекции (COVID-19) в каждом из них: 1) интенсивность, 2) степень влияния различных факторов, 3) различия в диагностике и регистрации случаев заболевания, что свидетельствует о возможности использования метода кластеризации как одного из этапов эпидемиологического анализа заболеваемости.

4. Проведение кластерного анализа временных рядов заболеваемости новой коронавирусной инфекцией (COVID-19) в 62 субъектах Российской Федерации позволило выделить субъекты с наибольшей интенсивностью эпидемического процесса в период с 2020 по 2023 гг.

5. Распределение субъектов Российской Федерации по кластерам в различные периоды пандемии варьировало. Наибольшее различие средних значений заболеваемости между отдельными кластерами наблюдалось со второго по пятый период пандемии новой коронавирусной инфекции (COVID-19), что совпадает по времени с периодами пандемии, когда перечень и объем проводимых противоэпидемических мероприятий в отдельных субъектах существенно различался.

6. Кластерный анализ временных рядов заболеваемости новой коронавирусной инфекцией (COVID-19) субъектов Российской Федерации позволил выделить группы территорий, достоверно различающихся по интенсивности проявлений эпидемического процесса. Наиболее высокая интенсивность эпидемического процесса регистрировалась в кластере 3, включающем в различные периоды г. Москву, г. Санкт-Петербург, Хабаровский край, где средние показатели заболеваемости новой коронавирусной инфекцией (COVID-19) превышали общероссийские в среднем 2,8–4,5 раза ( $p < 0,01$ ).

#### **ГЛАВА IV. ОЦЕНКА ВЛИЯНИЯ СОЦИАЛЬНО-ДЕМОГРАФИЧЕСКИХ И ИНФРАСТРУКТУРНЫХ ХАРАКТЕРИСТИК, ВЛИЯЮЩИХ НА УРОВЕНЬ И СТРУКТУРУ ЗАБОЛЕВАЕМОСТИ НОВОЙ КОРОНАВИРУСНОЙ ИНФЕКЦИЕЙ (COVID-19) В СУБЪЕКТАХ РОССИЙСКОЙ ФЕДЕРАЦИИ**

Выявление факторов, оказывающих существенное влияние на эпидемический процесс, является критически важной компонентой эпидемиологического анализа. Понимание механизмов распространения инфекции и определение значимых факторов риска создает основу для разработки эффективных мер противодействия заболеванию. Поиск таких факторов является необходимым этапом для подготовки к построению прогнозных моделей, что в свою очередь необходимо для последующего планирования комплекса противоэпидемических мероприятий.

В данной Главе для получения данных о Мониторинге количества заболевших новой коронавирусной инфекцией (COVID-19), в том числе внебольничными пневмониями, и летальных исходов использована форма отчетности Роспотребнадзора №1035 «Мониторинг количества заболевших коронавирусной инфекцией, в том числе внебольничными пневмониями, и летальных исходов». Для получения информации о случаях инфекционных заболеваний у лиц с подозрением на новую коронавирусную инфекцию применена форма отчетности Роспотребнадзора №970 «Информация о случаях инфекционных заболеваний у лиц с подозрением на новую коронавирусную инфекцию». Результаты ПЦР-исследований на наличие возбудителя новой коронавирусной инфекции (SARS-CoV-2) получены из Платформы агрегирования результатов лабораторных исследований SOLAR (англ. *System of laboratory aggregation results*). Демографические сведения для расчетов показателей взяты из данных Федеральной службы государственной статистики.

#### **4.1. Корреляционный анализ факторов, влияющих на заболеваемость населения субъектов Российской Федерации новой коронавирусной инфекцией (COVID-19)**

С целью проверки гипотезы о влиянии социально-демографических и инфраструктурных характеристик на уровень и динамику заболеваемости новой коронавирусной инфекцией (COVID-19), а также на распределение субъектов Российской Федерации по кластерам, проведен корреляционный анализ между кластерной принадлежностью субъектов и рядом показателей, отражающих плотность населения, уровень урбанизации и транспортную доступность.

Собраны данные по 62 субъектам Российской Федерации, включающие социально-экономические, демографические и транспортные показатели, представленные в таблице 4.1. На основе результатов кластерного анализа введена метрика «Сумма по кластерам» — интегральный кластерный индекс, отражающий принадлежность региона к определенным кластерам. Данный показатель представляет собой сумму по кластерам за 8 исследуемых периодов по каждому субъекту Российской Федерации. Высокое значение данного индекса указывает, что регион попадал в кластеры с более высоким уровнем заболеваемости новой коронавирусной инфекцией (COVID-19).

На первом этапе проведен корреляционный анализ с расчетом коэффициента корреляции Пирсона между «Суммой по кластерам» и исследуемыми показателями. Так как все рассматриваемые переменные являются количественными (большой частью близкими к непрерывным), расчет коэффициентов Пирсона в данном случае вполне допустим.

**Таблица 4.1** — Анализ показателей субъектов Российской Федерации по факторам, оказывающим возможное влияние на распространение новой коронавирусной инфекции (COVID-19) в различные периоды пандемии с 2020 по 2023 г.

Субъект	Сумма по кластерам	Доля городского населения (%)	Количество активных предприятий	Количество вновь созданных предприятий	Коэффициент миграционного прироста на 10 тыс. человек	Плотность автомобильных дорог с твердым покрытием общего пользования на 1 000 кв. км	Плотность железнодорожных путей общего пользования км путей на 10 тыс. кв. км.	Авиационные перевозки (количество перевезенных пассажиров за год)	Плотность населения
Алтайский край	2	57,6	35 192	3 686	-64,7	217,2	93	519 743	14,08
Амурская область	6	67,90	10 611	891	-10,10	34,60	81	556 724	2,02
Архангельская область	13	77,3	13 638,99	653	-108,5	21,1	30	922 539	2,4
Астраханская область	4	65,09	24 528	1 338	-108,60	85,40	128	672 456	23,01
Белгородская область	2	65,5	24 144	2 576	61,6	734,5	258	468 672	57,9
Брянская область	5	69,59	11 998	749	-2,4	322,5	280	0	30,4
Владимирская область	3	77,8	19 280	973	31,7	356,6	316	0	36,7
Волгоградская область	2	77,20	58 029	4511	-2,70	147	143	1 214 216	21,6
Воронежская область	11	68,3	37 062,99	4 360	47,4	367,9	229	856 969	42,9
г. Севастополь	11	92,5	8 171	590	474,5	1168,2	0	0	508,3
Забайкальский край	8	68,8	7 945	618	-91,2	33,9	56	455 350	2,6
Ивановская область	5	81,80	20 614	2 300	-39,90	336,50	161	38 273	34,6
Иркутская область	6	77,9	39 562	3 218	-11,3	31,5	32	2 433 794	3,3
Калининградская область	8	76,9	31 845,99	1 270	133,19	524,3	442	2 370 157	63,9
Калужская область	2	75,2	16 362	1 153	43	333,9	288	119 667	34,4
Камчатский край	4	77,70	6 360	460	-113,70	4,50	0	757 698	0,7
Кемеровская область	1	86,40	28 456	2 065	-14	177,10	175	769 066	28,5

Продолжение таблицы 4.1

Субъект	Сумма по кластерам	Доля городского населения (%)	Количество активных предприятий	Количество вновь созданных предприятий	Коэффициент миграционного прироста на 10 тыс. человек	Плотность автомобильных дорог с твердым покрытием общего пользования на 1 000 кв. км	Плотность железнодорожных путей общего пользования км путей на 10 тыс. кв. км.	Авиационные перевозки (количество перевезенных пассажиров за год)	Плотность населения
Кировская область	2	77,20	19 700	1 220	-84,30	114,40	91	240 842	14,7
Краснодарский край	0	56,20	9 711	588	84,80	479,80	297	13 547 214	75,5
Красноярский край	3	78,90	48 415	3 526	-7,40	11,70	9	2 558 706	1,2
Курганская область	4	63,40	7 268	437	-68,70	134,19	104	0	14,9
Курская область	5	68,30	12 702	1 274	34,20	375,80	352	38 161	28,6
Ленинградская область	2	66,20	22 640	2 578	244,89	215,30	304	19 581 262	38,3
Липецкая область	6	63,40	14 474	1 423	35,90	545,49	315	0	35,1
г. Москва	15	100	568 079	68 172	80,50	2 547,89	1 921	101 692 403	5 098,1
Московская область	4	79	15 1697	15 174	191,10	798,90	490	101 692 403	185,1
Нижегородская область	6	79,70	63 923	5 720	10	300,09	158	1 114 056	42,7
Новосибирская область	2	79,50	81 682	8 140	42,40	114,80	85	6 571 396	16,5
Омская область	2	73,30	28 758	2 925	-76,30	99,20	52	1 348 505	12,6
Оренбургская область	1	59,60	20 795	1 472	-28,90	167,70	118	783 647	15
Орловская область	9	66,80	8 632	662	-13,70	384,50	241	0	34,9
Пензенская область	9	68,59	15 430	1 417	-42,60	290,60	191	186 175	29,8
Пермский край	8	75,80	45 095	3 464	-26	137,10	98	0	15,1
Приморский край	0	78,20	42 468	3 446	-5,70	92	95	3 079 344	11,9
Псковская область	11	70,70	8 912	646	4,90	304,20	197	0	12,4
Республика Башкортостан	4	61,80	60 237	7 234	5,09	309,80	102	3 556 533	28,2

Продолжение таблицы 4.1

Субъект	Сумма по кластерам	Доля городского населения (%)	Количество активных предприятий	Количество вновь созданных предприятий	Коэффициент миграционного прироста на 10 тыс. человек	Плотность автомобильных дорог с твердым покрытием общего пользования на 1 000 кв. км	Плотность железнодорожных путей общего пользования км путей на 10 тыс. кв. км.	Авиационные перевозки (количество перевезенных пассажиров за год)	Плотность населения
Республика Бурятия	3	59,30	14 367	1 329	9,50	26,80	35	0	2,8
Республика Карелия		79,20	13 424	799	-97,50	46,90	123	0	3,6
Республика Коми	10	77,50	10 815	553	-178	15,80	41	567 372	2
Республика Крым	2	50,80	88 200	7 471,99	69,90	495	255	0	79,5
Республика Мордовия	2	63,20	7 128	528	2,40	296,39	208	90 552	29
Республика Саха (Якутия)	12	66,09	13 486	1 200	1,10	4	3	1 083 920	0,3
Республика Татарстан	0	76,80	86 906	9135	36,60	448,20	129	828 182	57
Республика Хакасия	0	68,50	5 424	358	-6,20	92,60	108	209 772	6
Ростовская область	14	79,20	13 424	799	-97,50	46,90	123	0	3,6
Рязанская область	10	77,50	10 815	553	-178	15,80	41	0	2
Самарская область	2	50,80	88 200	7 471,99	69,90	495	255	2 999 252	79,5
г. Санкт-Петербург	2	63,20	7 128	528	2,40	296,39	208	19 581 262	29
Саратовская область	12	66,09	13 486	1200	1,10	4	3	327 968	0,3
Свердловская область	0	76,80	86 906	9 135	36,60	448,20	129	6 232 318	57
Ставропольский край	0	68,50	5 424	358	-6,20	92,60	108	2 938 314	6
Тверская область	8	76,20	21 060	1 569	2,20	252,30	215	0	15,9
Томская область	4	71,70	20 444	1 505	25,60	23,90	11	769 531	3,9
Тульская область	2	74,50	21 912	1 542	35,20	404,60	367	0	64,2
Тюменская область	8	80,50	61 940	4 378	44,50	15,60	17	2 039 007	13,7

Продолжение таблицы 4.1

Субъект	Сумма по кластерам	Доля городского населения (%)	Количество активных предприятий	Количество вновь созданных предприятий	Коэффициент миграционного прироста на 10 тыс. человек	Плотность автомобильных дорог с твердым покрытием общего пользования на 1 000 кв. км	Плотность железнодорожных путей общего пользования км путей на 10 тыс. кв. км.	Авиационные перевозки (количество перевезенных пассажиров за год)	Плотность населения
Удмуртская Республика		65,59	26 491	2435	-43,20	252,49	185	0	32,5
Ульяновская область	7	76,09	17 457	1 482	-28,50	254,09	187	344 616	30,7
Хабаровский край	7	65,59	26 491	2 435	-43,20	252,49	185	5 357 260	32,5
Ханты-Мансийский АО	7	76,09	17 457	1 482	-28,50	254,09	187	3 230 078	30,7
Челябинская область	2	82,70	63 484	4 707	6	240,20	203	1 713 532	38
Чувашская Республика	6	63,30	15 254	1 015	-24,30	432,60	230	394 586	61,2
Ярославская область	9	81,40	26 961	2 433	-7,80	275,70	180	39 079	36,9

Проведен корреляционный анализ с расчетом коэффициента корреляции Пирсона между интегральным кластерным индексом «Сумма по кластерам» и исследуемыми показателями. В таблице 4.2 представлены значения коэффициентов корреляции Пирсона между «Суммой по кластерам» и основными признаками, а также соответствующие р-значения для оценки статистической значимости выявленных зависимостей.

**Таблица 4.2** — Результаты корреляционного анализа между интегральным кластерным индексом «Сумма по кластерам» и исследуемыми показателями

Исследуемый показатель	Коэффициент корреляции Пирсона	р-значение	Сила корреляционной зависимости
Доля городского населения	+0,47	<0,01	Средняя положительная связь
Количество активных предприятий	+0,27	<0,05	Слабая положительная связь
Количество новых предприятий	+0,28	<0,05	Слабая положительная связь
Коэффициент миграционного прироста	-0,02	>0,05	Нет значимой связи
Плотность автодорог	+0,32	<0,05	Средняя положительная связь
Плотность железных дорог	+0,34	<0,01	Средняя положительная связь
Авиационные перевозки (количество перевезенных пассажиров)	+0,17	<0,01	Слабая положительная связь
Плотность населения	+0,39	<0,01	Средняя положительная связь

По принятой шкале интерпретации силы корреляции величина от 0 до 0,1 соответствует **очень слабой связи**, 0,1–0,3 — **слабой**, 0,3–0,5 — **средней**, 0,5–0,7 — **умеренно сильной**, 0,7–0,9 — **высокой**, и свыше 0,9 — **очень**

**высокой (почти функциональной).** Исходя из этого, можно сделать следующие выводы:

1. **Доля городского населения** положительно коррелирует с интегральным кластерным индексом ( $r = 0,47$ , связь средней силы). Более урбанизированные регионы попадают в «неблагополучные» кластеры. Вероятно, высокая урбанизация сопровождается другими факторами риска (большая плотность населения, интенсивные контакты).

2. **Плотность населения** также показывает положительную связь с интегральным кластерным индексом ( $r = 0,399$ , средняя). Регионы с высокой плотностью населения чаще имеют более высокие значения кластерного индекса (что совпадает с гипотезой о влиянии скученности населения на заболеваемость).

3. **Плотность транспортных сетей** (автомобильных и железных дорог) умеренно коррелирует с интегральным кластерным индексом ( $r = 0,32–0,34$ ). Более развитая инфраструктура дорог связывается с более высоким кластерным индексом, что может отражать большее распространение инфекции. Однако, установленные данные корреляции средней силы связи и требуют проведения дополнительных исследований для получения доказательной базы.

4. **Экономическая активность**, выраженная через количество действующих и вновь открытых предприятий, продемонстрировала слабую положительную корреляцию с интегральным кластерным индексом ( $r = 0,27$ ), что может свидетельствовать о том, что более экономически развитые регионы, обладающие большим числом предприятий, чаще характеризуются более высоким уровнем заболеваемости. Предположительно, такие территории привлекают трудоспособное население и активные социальные слои населения, способствующие более интенсивному распространению новой коронавирусной инфекции (COVID-19). Однако, выявленная связь оказалась статистически малозначима.

Коэффициент корреляции Пирсона отражает исключительно линейную зависимость. Если реальная связь имеет нелинейный или пороговый характер, значение  $r$  может оказаться близким к нулю, что может создать ошибочное впечатление об отсутствии зависимости.

#### 4.2. Анализ факторов, влияющих на заболеваемость населения субъектов Российской Федерации новой коронавирусной инфекцией (COVID-19), с использованием метода машинного обучения

Для выявления нелинейной зависимости между кластерным индексом и изучаемыми показателями использован метод машинного обучения — алгоритм «Random Forest» (модель на основе решающих деревьев, модель «дерево решений»), позволяющий вычислить величину кластерного индекса (целевой показатель при обучении модели) по совокупности признаков (социально-экономических, демографических, транспортных).

Существенные различия отдельных периодов пандемии новой коронавирусной инфекции (COVID-19) делают целесообразным проведение ретроспективного анализа и краткосрочного прогноза заболеваемости для каждого из периодов отдельно, из-за гетерогенности взаимодействия популяции возбудителя и человека.

Модель «дерева решений» искала нелинейные зависимости между входными факторами и суммой по кластерам. По итогам оценена **значимость признаков** — показатель, показывающий вклад каждого признака в значения целевого показателя. Значимость рассчитывалась по совокупному уменьшению неопределенности при использовании данного признака в разделениях «дерева». «Random Forest» автоматически выполняет отбор признаков, игнорируя избыточные или неинформативные переменные.

В таблице 4.3 приведены полученные значения значимости признаков (нормированы так, что сумма по всем признакам равна 1,0 или 100%).

**Таблица 4.3** — Значения значимости признаков, полученных при построении модели «дерева решений»

Признак	Значимость признака (доля вклада)
Доля городского населения	0,52 (52%)
Плотность населения (чел./км <sup>2</sup> )	0,28 (28%)
Количество активных предприятий	0,19 (19%)
Коэффициент миграционного прироста	0,01 (1%)
Количество вновь созданных предприятий	0 (0%)
Плотность автодорог (км/1000 км <sup>2</sup> )	0 (0%)
Плотность железных дорог (км/10 000 км <sup>2</sup> )	0 (0%)
Авиаперевозки (количество перевезенных пассажиров за год)	0 (0%)

Применение модели «Random Forest» позволило выделить четыре главных фактора, влияющих на величину интегрального кластерного индекса:

1. **Доля городского населения** оказалась самым значимым признаком (около 52 % совокупной важности), и полностью согласуется с результатами проведенного в п.4.1 «Корреляционный анализ факторов, влияющих на заболеваемость населения регионов Российской Федерации новой коронавирусной инфекцией (COVID-19)», который также выявил значимую положительную связь с интегральным кластерным индексом. Высокая доля городского населения существенно повышает риск распространения новой коронавирусной инфекции (COVID-19), вероятно, вследствие высокой интенсивности социальных контактов, нагрузки на систему здравоохранения и других факторов.

2. **Плотность населения** — второй по значимости признак (~28%), показатель которого высоко коррелировал с долей городского населения и числом предприятий (как показывал корреляционный анализ: например,  $r = 0,89$  между плотностью населения и числом предприятий, что является высоким показателем). Тем не менее, модель «Random Forest» распределила

важность между ними, подтверждая, что плотность населения напрямую влияет на уровень заболеваемости: в плотнонаселенных регионах распространение инфекций происходит существенно быстрее и интенсивнее.

3. **Количество активных предприятий** — третий по вкладу признак (~19%). Экономическая активность региона, связанная с численностью предприятий, очевидно сопряжена с количеством населения и урбанизацией, поэтому ее влияние могло частично отражать те же процессы (например, крупные экономические центры – это обычно мегаполисы с высокой плотностью населения и интенсивным транспортным трафиком). Тем не менее, модель «Random Forest» сочла этот признак менее информативным по сравнению с предыдущими. Вероятно, количество предприятий служит косвенным индикатором уровня экономического развития и деловой активности населения, что влияет на распространение инфекционного заболевания.

4. **Коэффициент миграционного прироста** — четвертый значимый фактор (~1%).

Отметим, что суммарная важность первых четырех признаков, а именно: доли городского населения, плотности населения, количества активных предприятий и коэффициента миграционного прироста составляет ~100%, а остальные факторы получили нулевое значение, что указывает на тот факт, что алгоритм машинного обучения счел их полностью нерелевантными при наличии лидирующих взаимосвязанных признаков.

### **4.3. Сравнение двух аналитических подходов для оценки влияния факторов на уровень и структуру заболеваемости населения субъектов Российской Федерации новой коронавирусной инфекцией (COVID-19) (традиционный корреляционный анализ и метод машинного обучения)**

Проведенный анализ позволяет осуществить сравнительную оценку традиционного корреляционного метода и модели, основанной на «решающих деревьях», в контексте выявления взаимосвязей.

Корреляция Пирсона измеряет только линейную степень связи между двумя переменными. «Решающие деревья» не предполагают заранее определенной формы зависимости между признаками и целевыми переменными. Они рекурсивно разделяют признаки на области, адаптируясь к сложным и нелинейным взаимосвязям между ними.

В нашем случае модель «дерево решений» выявила влияние коэффициента миграционного прироста на интегральный кластерный индекс (не проявившийся в корреляционном анализе). Для выявления сложных зависимостей применение модели «дерево решений» более эффективно, тогда как традиционный расчет корреляций может упустить важные эффекты.

Применение метода расчета корреляций выдает единственное число ( $r$ ) для пары переменных, отражающее силу и направление связи, что упрощает интерпретацию: знак говорит о положительной или отрицательной тенденции, величина — о силе (по указанной шкале).

В отличие от этого, значимость признаков в модели «дерево решений» представлена набором чисел (по одному на каждый признак), показывающих вклад в модель. Данные числа характеризуют относительную важность, но не имеют направления влияния. Например, высокая важность доли городского населения не указывает напрямую, повышает урбанизация заболеваемость или снижает. Кроме того, модель «решающее дерево» учитывает все признаки одновременно. Важность признака отражает его вклад с учетом остальных факторов, что дает более целостное представление о его роли в многомерной системе.

Таким образом, по итогам исследований, проведенных в данной главе, можно сделать следующие выводы:

1. Корреляционный анализ между кластерной принадлежностью субъектов и рядом показателей, отражающих плотность населения, уровень урбанизации и транспортной доступности показал наличие положительной корреляции интегрального кластерного индекса с долей городского населения ( $r = 0,47$ , связь средней силы), плотностью населения ( $r = 0,399$ , средняя), плотностью автомобильных и железных дорог ( $r = 0,32-0,34$ ), а также экономической активностью, выраженной посредством количества действующих и вновь открытых предприятий ( $r = 0,27$ ).

2. Использование метода машинного обучения машинного обучения — «Random Forest» позволило определить значимость показателей, отражающих плотность населения, уровень урбанизации и транспортной доступности на величину интегрального кластерного индекса субъекта. Такими показателями оказались: доля городского населения (около 52 % совокупной важности); плотность населения — второй по значимости признак (~28%); количество активных предприятий — третий по вкладу признак (~19%); коэффициент миграционного прироста — четвертый значимый фактор (~1%).

3. С учетом анализа результатов двух использованных методик показано, что уровень заболеваемости и характер распространения новой коронавирусной инфекции (COVID-19) в субъектах Российской Федерации зависит от степени урбанизации территорий, доли городского населения, плотности населения, интенсивности экономической активности и миграционных процессов.

**ГЛАВА V. ВОЗМОЖНОСТЬ ПРИМЕНЕНИЯ ЦИФРОВЫХ  
ТЕХНОЛОГИЙ УЧЕТА РЕЗУЛЬТАТОВ И АНАЛИЗА ДАННЫХ  
МОЛЕКУЛЯРНО-БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ  
НАСЕЛЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ ДЛЯ ПОВЫШЕНИЯ  
ИНФОРМАТИВНОСТИ АНАЛИЗА ЭПИДЕМИОЛОГИЧЕСКОЙ  
ОБСТАНОВКИ ПО НОВОЙ КОРОНАВИРУСНОЙ ИНФЕКЦИИ  
(COVID-19)**

В ГЛАВЕ III. «ЭПИДЕМИОЛОГИЧЕСКИЙ АНАЛИЗ ЗАБОЛЕВАЕМОСТИ НАСЕЛЕНИЯ НОВОЙ КОРОНАВИРУСНОЙ ИНФЕКЦИЕЙ (COVID-19) В РОССИЙСКОЙ ФЕДЕРАЦИИ В 2020–2023 гг. С ИСПОЛЬЗОВАНИЕМ МЕТОДА КЛАСТЕРИЗАЦИИ ИНТЕНСИВНЫХ ПОКАЗАТЕЛЕЙ», было показано, что существует ряд ограничений для осуществления эпидемиологического анализа заболеваемости: различия в динамике эпидемического процесса в разные периоды пандемии новой коронавирусной инфекции (COVID-19), значительные различия в уровне и тенденциях динамики заболеваемости в отдельных регионах Российской Федерации, пробелы в данных по регистрации заболеваемости в отдельных субъектах. Благодаря внедрению в систему эпидемиологического надзора цифровых технологий учета результатов и анализа данных молекулярно-биологических исследований, появилась возможность использовать агрегированные данные о результатах ПЦР-исследований по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-COV-2) как важного индикатора текущей эпидемиологической ситуации.

В данной главе приведены результаты функционирования Платформы агрегирования результатов лабораторных исследований «SOLAR» и анализ возможности использования агрегированных данных о результатах ПЦР-исследований для оценки эпидемиологической обстановки по заболеваемости новой коронавирусной инфекцией (COVID-19) на территории Российской Федерации в 2020–2023 гг.

## **5.1. Цифровая программная Платформа агрегирования результатов лабораторных исследований (SOLAR) как инструмент эпидемиологического надзора за новой коронавирусной инфекцией (COVID-19)**

Цифровая трансформация процессов анализа данных стала неотъемлемой частью эпидемиологического надзора, особенно в период пандемии [83] и явилась предпосылкой для разработки цифровой платформы SOLAR, представляющей важный инструмент для обработки и анализа данных молекулярно-биологических исследований. Платформа создана в рамках Постановления Правительства РФ №452 от 27 марта 2021 г. в ФБУН Центральный НИИ Эпидемиологии Роспотребнадзора с целью оперативного информирования граждан о результатах ПЦР-исследований, направленных по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2). К ноябрю 2021 г. завершена интеграция всех лабораторий различного ведомственного подчинения, функционирующих на территории Российской Федерации, в систему SOLAR.

Платформа SOLAR (схема представлена на рисунке 5.1.1) интегрирует данные, поступающие от более чем 2 000 лабораторий по всей Российской Федерации, и позволяет в реальном времени обрабатывать миллионы результатов ПЦР-исследований по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2). При помощи данной Платформы происходит быстрая передача данных в личные кабинеты пользователей через портал Госуслуг, при этом от момента загрузки результатов диагностической лабораторией до получения информации пользователем проходит не более 30 минут. Подобная скорость обработки и передачи данных крайне важна для принятия своевременных мер по борьбе с пандемией новой коронавирусной инфекции (COVID-19). Автоматизация процесса получения результатов ПЦР-исследований и их дальнейшая передача в медицинские информационные

системы упрощает управление данными и повышает точность эпидемиологического мониторинга.

В ФБУН ЦНИИ Эпидемиологии Роспотребнадзора разработаны программные интерфейсы (специализированные компьютерные программы) для обеспечения безопасной передачи персонифицированных сведений о лицах, в отношении которых проводилось исследование, а также и самих результатов исследований. Идентификация лиц, прошедших исследования, является важным аспектом для передачи данных в личный кабинет гражданина на портале государственных услуг, а также для проведения адресных противоэпидемических мероприятий.

В январе 2022 г. объем передаваемых, хранимых и обрабатываемых данных достигал 2 млн результатов в день, что потребовало создания распределенных баз данных и вычислительных компьютерных кластеров. Все каналы связи и оборудование, участвующие в передаче персональных данных, были приведены в соответствие с требованиями законодательства Российской Федерации по защите информации. Для анализа результатов ПЦР-исследований и организации противоэпидемических мероприятий в субъектах Российской Федерации создан информационный ресурс SOLAR, предоставляющий специалистам Роспотребнадзора авторизованный доступ к интерактивным дашбордам с аналитикой по обезличенным данным.

На 1 февраля 2026 г. на платформу загружено свыше 200 млн результатов ПЦР-исследований по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2). Данные результаты получены более чем от 2 000 лабораторий различной ведомственной принадлежности и переданы по защищенным каналам связи в ФБУН Центральный НИИ Эпидемиологии Роспотребнадзора. Затем они были загружены на сервер портала Госуслуг через интеграционную платформу Роспотребнадзора.

Ключевые характеристики работы платформы агрегирования результатов лабораторных исследований SOLAR включают следующие аспекты:

- полный цикл обработки информации — от момента внесения результатов в лабораторную информационную систему до поступления на портал Государственных услуг — осуществляется в автоматическом режиме, что существенно минимизирует риск операционных ошибок;

- высокая степень верификации получаемой информации обеспечивается многоступенчатой системой контроля качества при проведении ПЦР-исследования, что позволяет практически исключить диагностические ошибки как в сторону ложноположительных, так и ложноотрицательных результатов;

- временной промежуток между получением лабораторных данных до поступления результатов на интегративную платформу Роспотребнадзора и портал Государственных услуг сведен к минимально возможному значению;

- система интегрирует диагностические данные как из медицинских учреждений, куда пациенты обращаются за специализированной помощью, так и от лиц, проходящих профилактическое обследование в добровольном порядке, в том числе при отсутствии явных признаков патологии.

Использование платформы SOLAR позволяет осуществлять массовое выявление и учет заболевших, в том числе до появления клинических симптомов заболевания.

Данные, поступающие из SOLAR, впоследствии интегрируются в аналитическую платформу EpidSmart для комплексного эпидемиологического анализа совместно со сведениями о количестве зарегистрированных случаев заболевания, гендерно-возрастной структурой заболевших, сведениями о распределении заболевших по формам тяжести и данными молекулярно-генетического мониторинга за изменчивостью возбудителя.

На базе аналитической платформы EpidSmart проводится оперативный и ретроспективный эпидемиологический анализ заболеваемости новой коронавирусной инфекцией (COVID-19) населения Российской Федерации с учетом сведений о результатах ПЦР-исследований по выявлению РНК

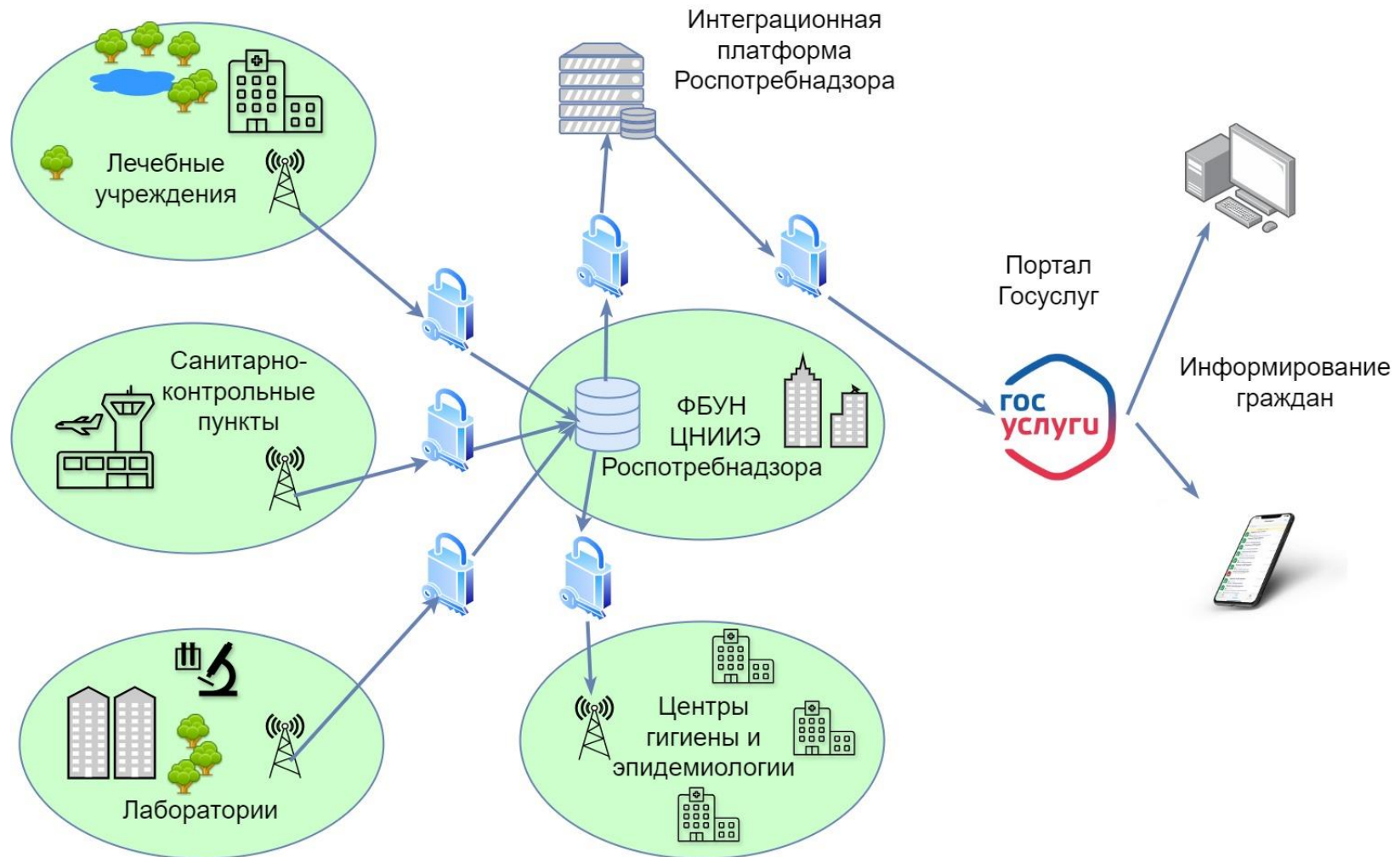
возбудителя новой коронавирусной инфекции (SARS-CoV-2). Данные визуализируются на интерактивных дашбордах.

Пример визуализации данных о результатах ПЦР-исследований по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2) поступающих из платформы SOLAR в аналитической платформе EpidSmart представлен на рисунке 5.1.2.

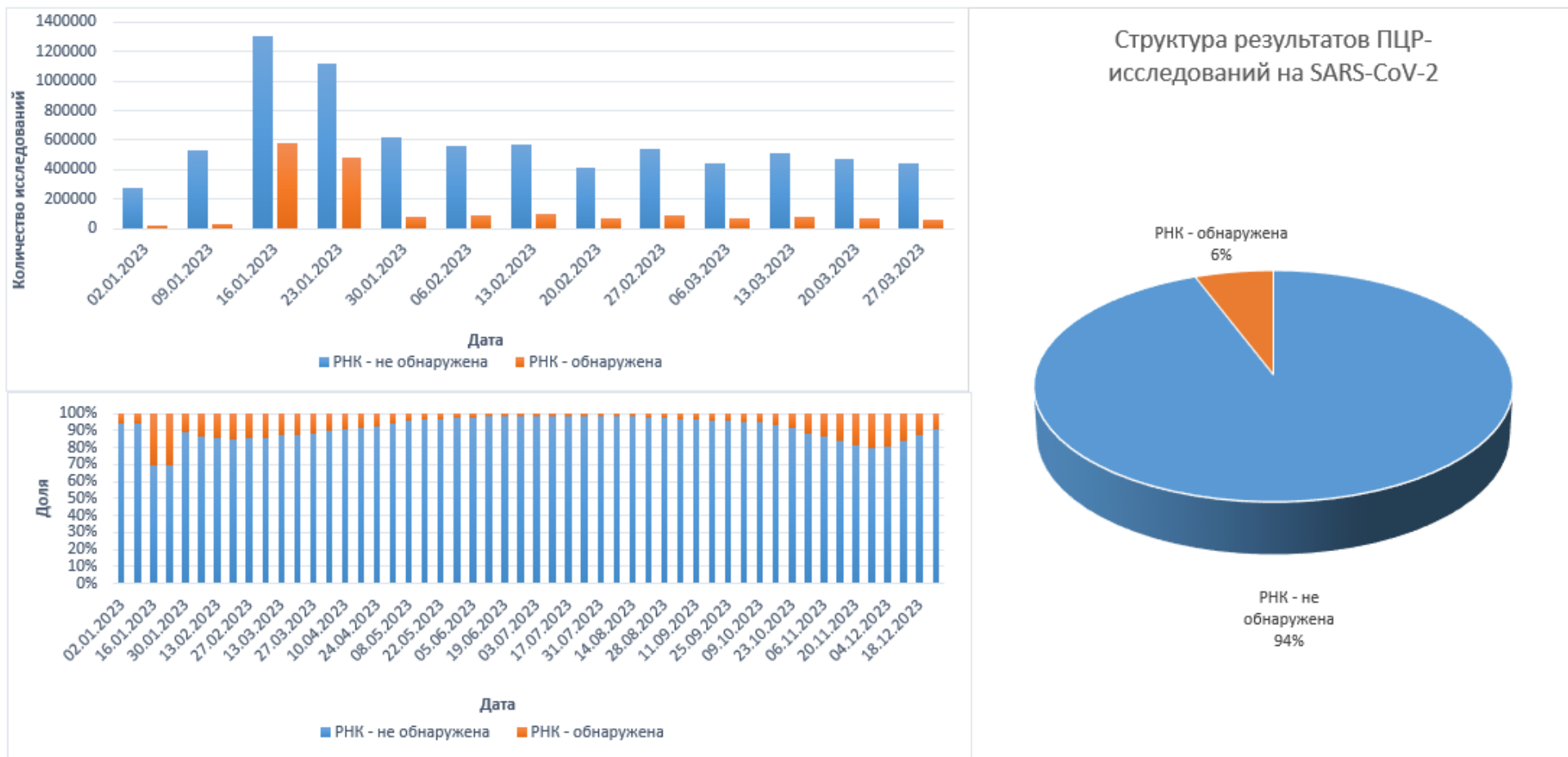
Таким образом, платформа SOLAR является инновационным и высокотехнологичным облачным решением для сбора и дальнейшего анализа данных о результатах ПЦР-исследований. Населению Российской Федерации Платформа SOLAR представляет возможность получить результаты ПЦР-исследования по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2) без необходимости посещения лечебно-профилактического учреждения, что в два и более раз снижает риск возможного заражения. Кроме того, повышение доступности результатов лабораторного обследования, которое достигается за счет функционирования платформы SOLAR опосредованно повышает долю охвата населения обследованием по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2), дает возможность каждому гражданину, не подвергая дополнительному риску заражения, обследовать себя и близких.

Информирование региональных органов Роспотребнадзора о деперсонифицированных результатах диагностического лабораторного обследования дает дополнительную возможность контролировать степень охвата населения обследованием. Так, в соответствии с постановлением Главного государственного врача Российской Федерации от 16 октября 2020 г. №31 «О дополнительных мерах по снижению рисков распространения COVID-19 в период сезонного подъема заболеваемости острыми респираторными вирусными инфекциями и гриппом» был установлен норматив обследования населения (не менее 150 исследований на 100 тыс. населения, среднесуточно за 7 дней), Платформа SOLAR стала важным инструментом контроля достижения этого целевого показателя.

Кроме того, Платформа позволила проводить оперативную оценку эпидемиологической обстановки не только на основе сведений о динамике заболеваемости, но и с учетом данных о структуре и динамике результатов лабораторных исследований.



**Рисунок 5.1.1** — Принцип функционирования программной платформы SOLAR в различных системах федеральных органов исполнительной власти



**Рисунок 5.1.2** — Пример визуализации данных о результатах ПЦР-исследований по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2), поступающих из платформы SOLAR в аналитическую платформу EpidSmart

Данные поступающие из платформы SOLAR является ключевым компонентом в работе программного приложения «Путешествую без COVID-19» — флагманского проекта Фонда цифровых инициатив [38]. Данная программа обеспечила возможность гражданам пересекать границу со странами Евразийского экономического союза, предъявляя результаты ПЦР-обследования по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2) в виде QR-кода, отображаемого на экране смартфона, содержащего ссылку на информацию о дате, месте и результате ПЦР-обследования. Данный функционал обеспечил необходимые условия для планирования и осуществления санитарно-карантинных мероприятий в пунктах пропуска через государственную границу.

**5.2. Использование данных о результатах ПЦР-исследований по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2) для оперативного и ретроспективного эпидемиологического анализа заболеваемости новой коронавирусной инфекцией (COVID-19)**

С целью определения перспектив использования в рамках оперативного и ретроспективного эпидемиологического анализа заболеваемости деперсонифицированных данных, поступающих из программной платформы SOLAR, для оценки динамики и прогнозирования эпидемического процесса новой коронавирусной инфекции (COVID-19) среди населения Российской Федерации, решены две исследовательские задачи:

1. **Определение наличия и величины корреляции, а также временного интервала между динамикой официально зарегистрированных случаев заболеваемости новой коронавирусной инфекцией (COVID-19) совокупного населения Российской Федерации и динамикой количества ПЦР-исследований по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2) в 5–8 периоды пандемии (10.01.2022–17.12.2023 гг.).** Здесь и далее под *временным интервалом* мы понимаем общепринятое определение — это время, которое проходит между двумя событиями, что применительно к решаемой задаче означает величину смещения одного временного ряда относительно другого. Выбор периодов пандемии для изучения связан с началом полноценного функционирования платформы SOLAR к началу 5-го периода пандемии (с декабря 2021 г.) и ограничивается концом 2023 г., когда объем проводимых ПЦР-исследований по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2) значительно снизился.

2. **Определение наличия и величины корреляции, а также временного интервала между динамикой заболеваемости новой коронавирусной инфекцией (COVID-19) совокупного населения Российской Федерации и динамикой доли положительных результатов**

ПЦР-исследований по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2) в 5–8 периоды пандемии (10.01.2022–17.12.2023 гг.).

Для определения корреляции и временного интервала между динамикой заболеваемости новой коронавирусной инфекцией (COVID-19) совокупного населения Российской Федерации и динамикой количества ПЦР-исследований по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2) использованы сведения о количестве ежедневно регистрируемых новых случаев заболевания по данным отраслевых форм статистического наблюдения и программной платформы SOLAR за 5–8 периоды пандемии (10.01.2022–17.12.2023 гг.).

В 5-м периоде пандемии новой коронавирусной инфекции (COVID-19) максимальный коэффициент корреляции между количеством заболевших и количеством ПЦР-исследований по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2) получен при величине интервала, равном 9 дням, при этом он составил 0,93 (очень высокая связь по шкале Чеддока) (рисунок 5.2.1). Это означает, что динамика количества проведенных ПЦР-исследований опережала регистрацию новых случаев заболевания новой коронавирусной инфекцией (COVID-19) в среднем на 9 дней в 5-м периоде пандемии (10.01–25.07.2022 г.). При этом зависимость между анализируемыми показателями с высокой степенью точности (коэффициент детерминации равен 0,86, что соответствует высокой силе связи по шкале Чеддока) в этом периоде может быть описана линейной функцией (рисунок 5.2.2). Это дает возможность утверждать о наличии прямой линейной зависимости между показателем количества ПЦР-исследований и количеством заболевших в 5-м периоде пандемии (10.01–25.07.2022 гг.). Полученные результаты свидетельствуют о возможности использования показателя количества ПЦР-исследований в 5-м периоде пандемии для краткосрочного прогнозирования динамики регистрации новых случаев заболевания.

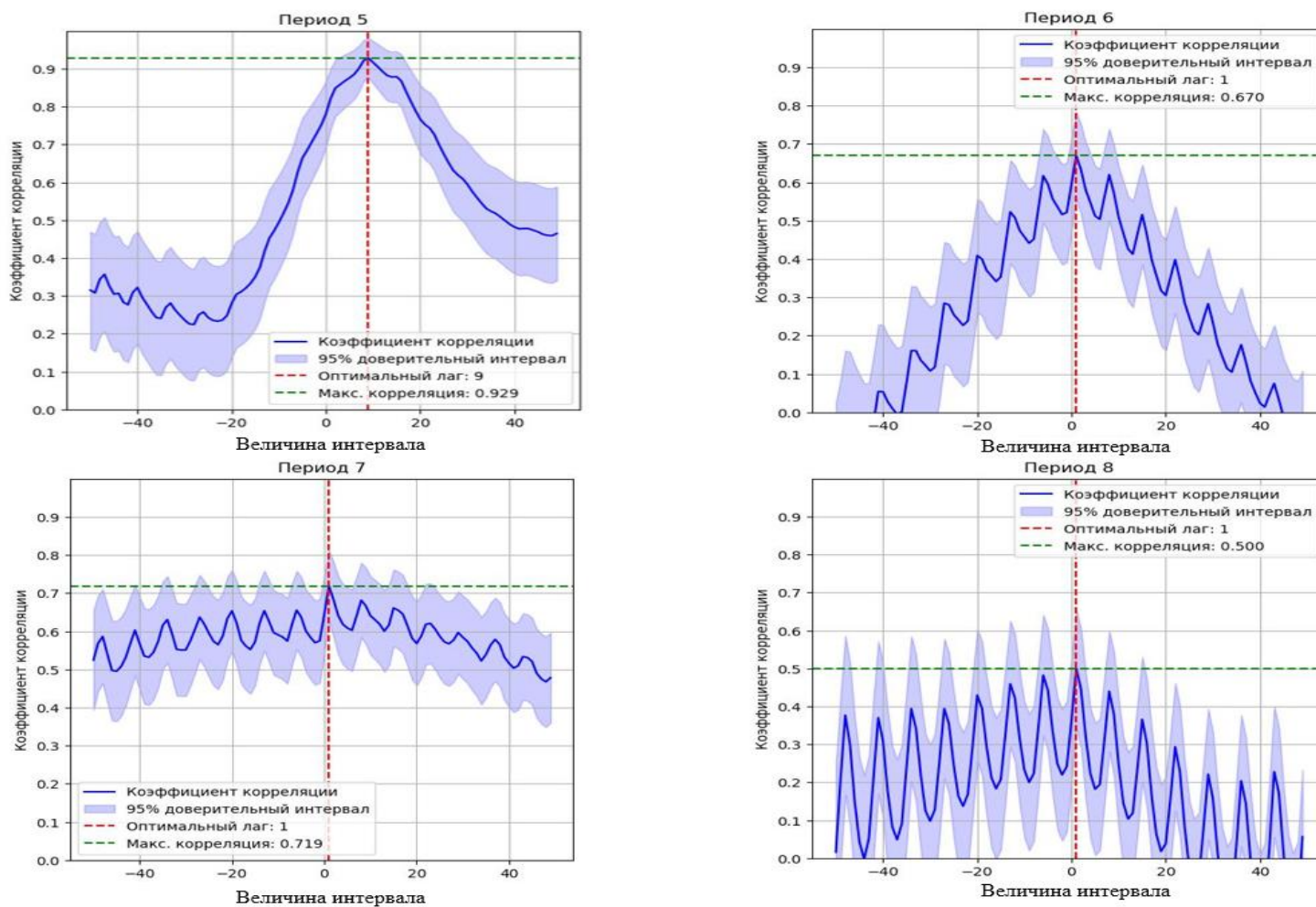
В 6–8 периодах (26.07.2022–17.12.2023 гг.) пандемии наблюдалась выраженная недельная цикличность в обоих анализируемых временных рядах, что нашло свое отражение на графиках экстремумов взаимной корреляции. Здесь и далее под термином экстремум мы используем общепринятое определение — значение непрерывной функции, являющееся ее максимумом или минимумом [33]. При этом максимальный коэффициент корреляции между динамикой количества заболевших и количеством ПЦР-исследований в 6–8 периодах получен при величине интервала, равном 1 дню, и составил 0,67 (умеренно сильная связь по шкале Чеддока), 0,72 (высокая связь по шкале Чеддока) и 0,50 (умеренно сильная связь по шкале Чеддока), соответственно. Как видно на коррелограммах, приведенных на рисунке 5.2.2, предполагать значимую корреляцию между анализируемыми временными рядами даже при оптимальном значении временного интервала не представляется возможным.

Поскольку количество проводимых исследований зависит не только от эпидемиологической обстановки, но и от большого количества других факторов, таких как принятые стандарты обследования больных и контактных лиц, а также возможностей лабораторной диагностической базы, целесообразным является **оценка взаимосвязи между динамикой регистрации новых случаев заболевания и динамикой доли положительных ПЦР-исследований** по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2), поскольку эти сведения также поступают из системы SOLAR.

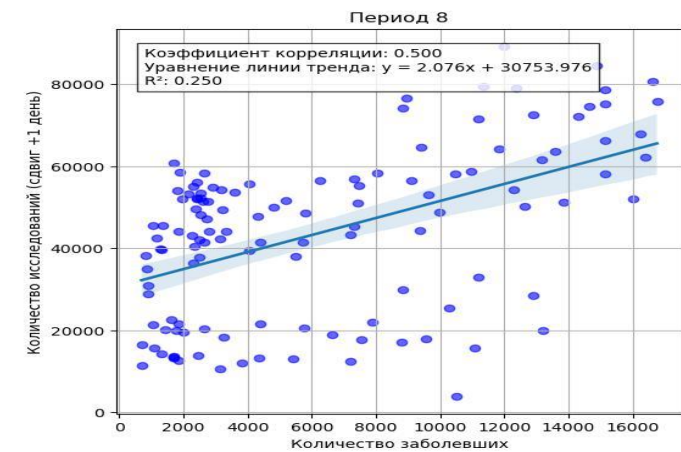
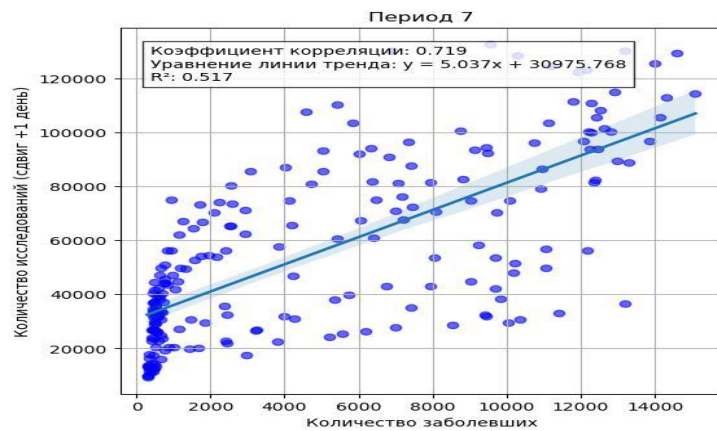
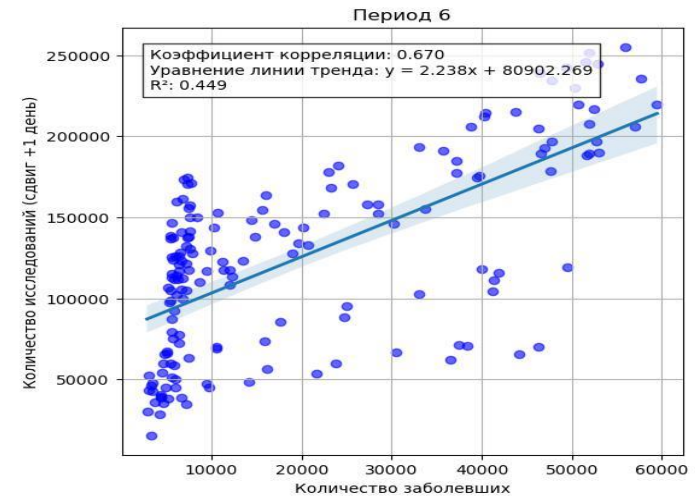
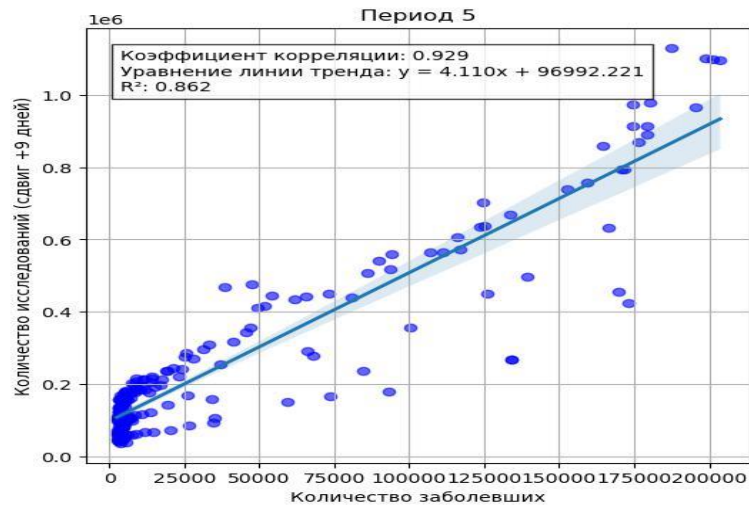
В 5-м периоде (10.01–25.07.2022 гг.) пандемии новой коронавирусной инфекции (COVID-19) максимальный коэффициент корреляции между динамикой количества заболевших и долей положительных результатов ПЦР-исследований по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2) получен при величине интервала, равном 6 дням, при этом он составил 0,97 (очень высокая связь по шкале Чеддока) (рисунок 5.2.3). Полученные значения свидетельствуют о том, что динамика доли положительных результатов опережала регистрацию новых случаев

заболевания новой коронавирусной инфекцией (COVID-19) в среднем на 6 дней в 5-м периоде пандемии (10.01–25.07.2022 гг.). При этом зависимость между анализируемыми показателями с высокой степенью точности (коэффициент детерминации равен 0,95 – очень высокая связь по шкале Чеддока) в этом периоде может быть описана линейной функцией (рисунок 5.2.4).

В 6-м периоде (26.07.2022–09.01.2023 гг.) пандемии максимальное значение коэффициента корреляции (0,98 – очень высокая связь по шкале Чеддока) между динамикой регистрации новых случаев заболевания новой коронавирусной инфекцией (COVID-19) и динамикой доли положительных результатов ПЦР-исследований по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2) получено при значении интервала, равном 8 дням (динамика доли положительных результатов опережает динамику регистрации новых случаев заболевания в среднем на 8 дней). Построенная линейная модель взаимосвязи временных рядов показала высокую точность, коэффициент детерминации составил 0,96 (очень высокая связь по шкале Чеддока).



**Рисунок 5.2.1** — Экстремум взаимной корреляции между динамикой заболеваемости новой коронавирусной инфекцией (COVID-19) и количеством ПЦР-исследований по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2) в 5-ом – 8 периодах пандемии (10.01.2022–17.12.2023 гг.)



**Рисунок 5.2.2** — Коррелограммы динамики заболеваемости новой коронавирусной инфекцией (COVID-19) и количества ПЦР-исследований по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2) в 5-ом – 8 периодах пандемии (10.01.2022–17.12.2023 гг.)

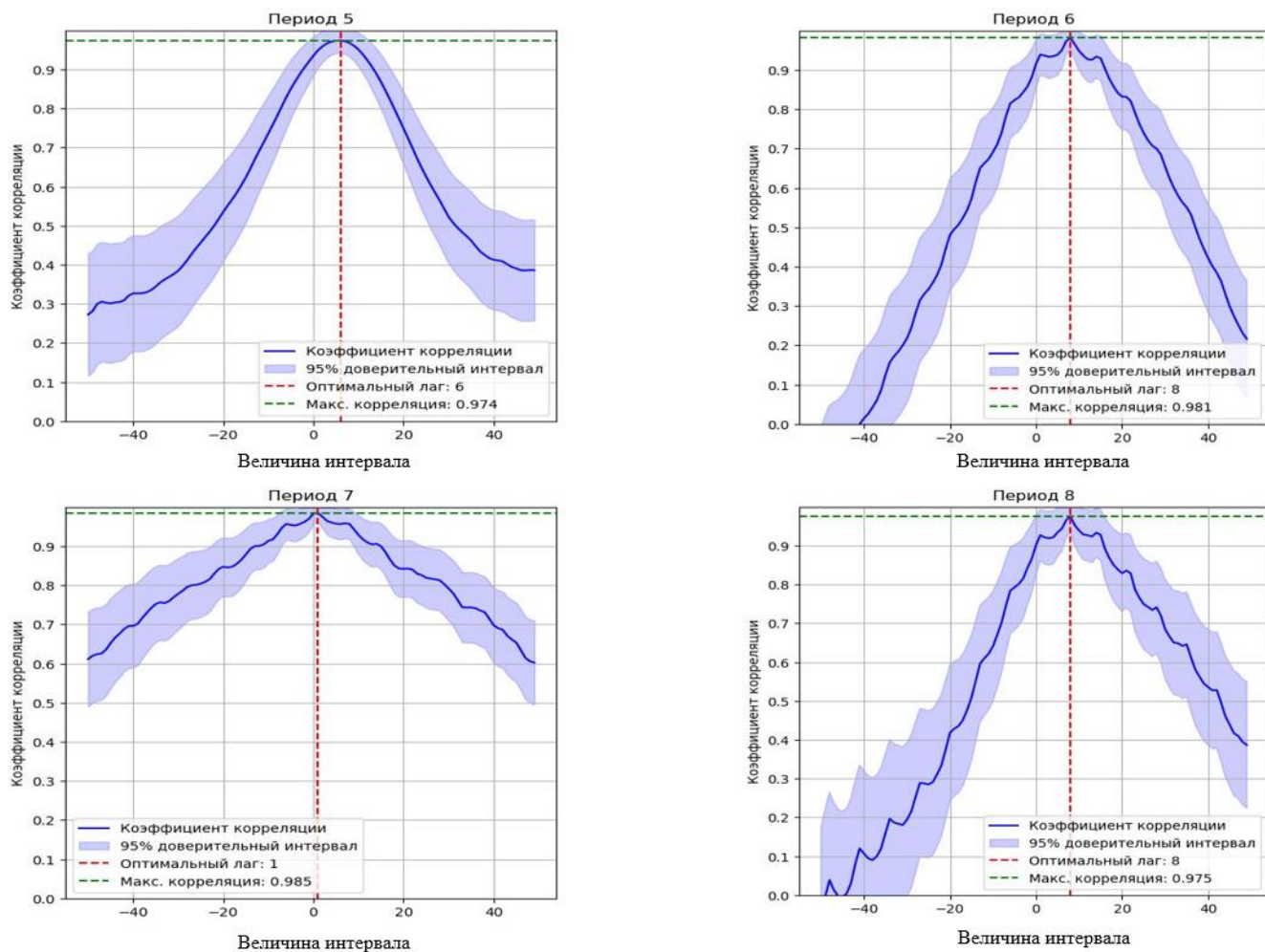
Полученные результаты свидетельствуют о возможности использования показателя доли положительных результатов ПЦР-исследований по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2) для краткосрочного прогнозирования динамики регистрации новых случаев заболевания в 6-м периоде пандемии (26.07.2022–09.01.2023 гг.).

В 7-м периоде пандемии (10.01–28.08.2023 гг.) максимальное значение коэффициента корреляции, равное 0,99 (очень высокая связь по шкале Чеддока), получено при величине временного интервала, равном одному дню, что свидетельствует о том, что динамика доли положительных результатов ПЦР-исследований по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2) опережала динамику регистрации новых случаев заболевания в среднем на 1 день. Такой показатель подтверждает высокую информативность доли положительных ПЦР-результатов, который можно использовать для оперативного прогнозирования заболеваемости, особенно в условиях высокой динамичности распространения вируса. Данный показатель имеет важное значение для более точного планирования противоэпидемических мероприятий, таких как временные ограничения или мобилизация медицинских ресурсов.

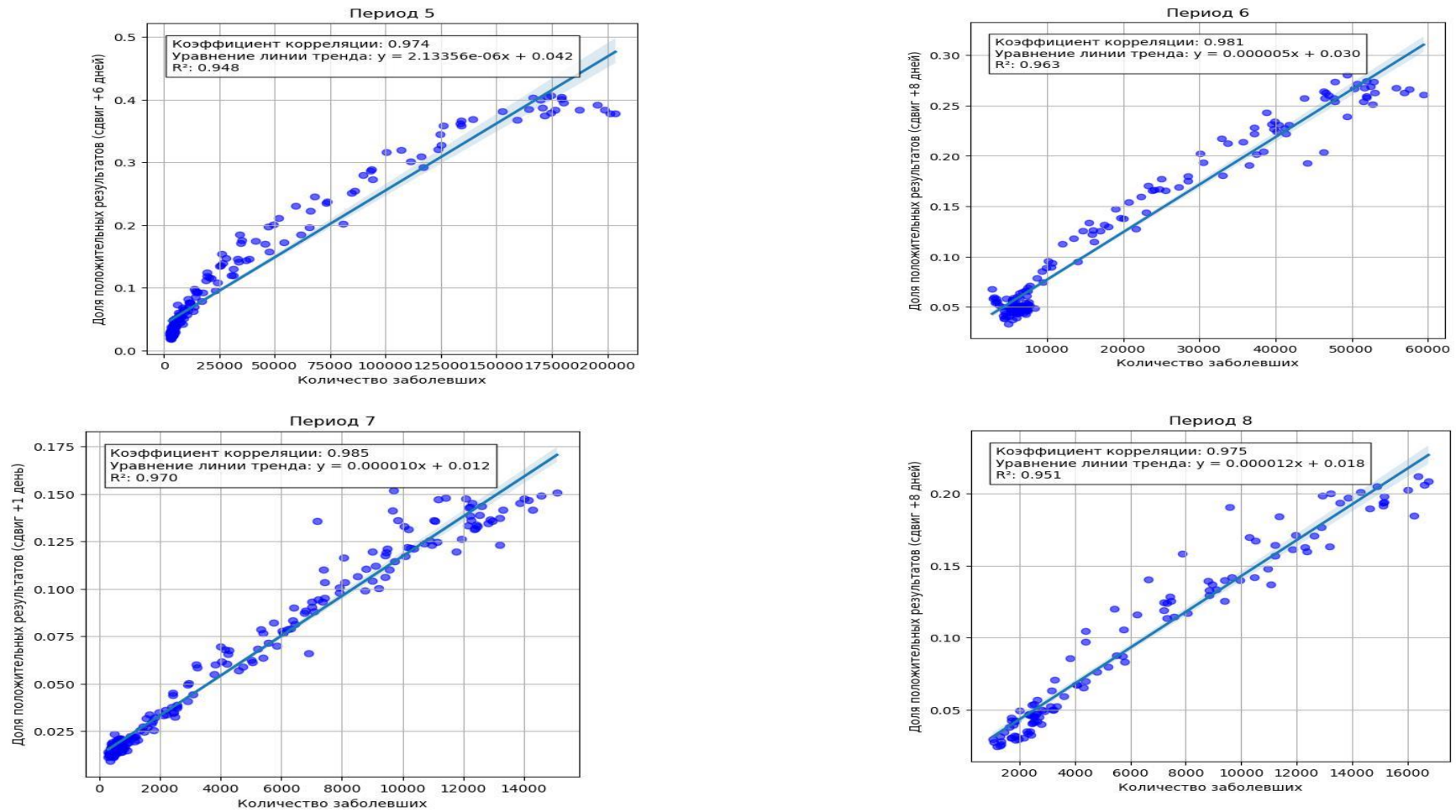
В 8-м периоде пандемии (29.08–17.12.2023 гг.) максимальный коэффициент корреляции между динамикой количества заболевших новой коронавирусной инфекцией (COVID-19) и долей положительных результатов ПЦР-исследований по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2) получен при величине интервала, равном 8 дням, при этом он составил 0,98 (очень высокая связь по шкале Чеддока). Полученные значения свидетельствуют о том, что динамика доли положительных результатов опережала регистрацию новых случаев заболевания в среднем на 8 дней в восьмом периоде пандемии. Этот результат подтверждает, что доля положительных результатов ПЦР-исследований по выявлению РНК

возбудителя новой коронавирусной инфекции (SARS-CoV-2) является чувствительным индикатором, который может прогнозировать рост заболеваемости с высокой точностью за несколько дней до того, как новые случаи заболевания будут зарегистрированы. При этом зависимость между анализируемыми показателями с высокой степенью точности (коэффициент детерминации равен 0,95 – очень высокая связь по шкале Чеддока) в этом периоде может быть описана линейной функцией, что также подтверждает высокую прогностическую способность данного показателя в эпидемиологических моделях.

Важной особенностью платформы SOLAR является включение в систему данных о результатах лабораторных исследований у лиц, не имеющих клинических проявлений заболевания и не обращавшихся за медицинской помощью. Интеграция результатов ПЦР-исследований лиц с бессимптомными формами инфекции в единый контур наблюдения обеспечила более полное представление о масштабах циркуляции возбудителя новой коронавирусной инфекции (SARS-CoV-2) и позволила выявлять потенциальных носителей вируса на более ранних этапах. Полученные результаты подчеркивают важность массового ПЦР-обследования как инструмента выявления скрытой циркуляции вируса и демонстрируют, что включение данных о бессимптомных носителях значительно повышает эффективность оперативного эпидемиологического анализа, позволяя корректнее и эффективнее оценивать реальные уровни заболеваемости и выстраивать более обоснованные стратегии профилактики и контроля распространения инфекции.



**Рисунок 5.2.3** — Экстремум взаимной корреляции между динамикой заболеваемости новой коронавирусной инфекцией (COVID-19) и долей положительных результатов ПЦР-исследований по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2) в 5-ом – 8 периодах пандемии (10.01.2022–17.12.2023 гг.)



**Рисунок 5.2.4** — Коррелограммы динамики заболеваемости новой коронавирусной инфекцией (COVID-19) и долей положительных результатов ПЦР-исследований по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2) в 5-ом – 8 периодах пандемии (10.01.2022–17.12.2023 гг.)

Таким образом, по итогам данной главы можно сделать следующие выводы:

1. Молекулярно-биологические методы, в частности, ПЦР-диагностика и секвенирование, играют ключевую роль в системе эпидемиологического надзора, обеспечивая высокую чувствительность, специфичность и оперативность выявления инфекционных агентов, в том числе возбудителя новой коронавирусной инфекции (SARS-CoV-2).

2. Программная платформа SOLAR, разработанная ФБУН Центральный НИИ Эпидемиологии Роспотребнадзора, стала эффективным инструментом для агрегации, обработки и анализа данных ПЦР-исследований по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2) среди населения Российской Федерации, обеспечив интеграцию более чем 2 000 лабораторий различной ведомственной принадлежности. В настоящее время на Платформу загружено свыше 200 млн результатов исследований.

3. Ключевые преимущества платформы SOLAR включают: полный автоматический цикл обработки информации — от внесения результатов в лабораторную систему до публикации на портале Госуслуг; многоступенчатую систему контроля качества, минимизирующую диагностические ошибки; сокращение временного промежутка между получением лабораторных данных и их публикацией; интеграцию данных лабораторных исследований как от медицинских учреждений, так и от лиц, проходящих профилактическое обследование.

4. Анализ данных, собранных с помощью платформы SOLAR, позволил выявить значимую корреляцию между динамикой заболеваемости новой коронавирусной инфекцией (COVID-19) и динамикой количества ПЦР-исследований, а также долей положительных результатов. В частности, в 5-м периоде пандемии (10.01–25.07.2022 гг.) зафиксирован максимальный коэффициент корреляции (0,93 – очень высокая связь по шкале Чеддока) при

временном интервале в 9 дней между количеством ПЦР-исследований и регистрацией новых случаев заболевания.

5. Показатель доли положительных результатов ПЦР-исследований продемонстрировал высокую прогностическую способность для краткосрочного прогнозирования (7–14 дней) динамики регистрации новых случаев заболевания. Максимальные коэффициенты корреляции в разные периоды пандемии достигали значений от 0,97 до 0,99 (очень высокая связь по шкале Чеддока) при различных величинах временного интервала, что подтверждает информативность этого показателя для оперативного планирования противоэпидемических мероприятий.

6. Интеграция данных о результатах лабораторных исследований у бессимптомных носителей вируса через платформу SOLAR позволила существенно повысить точность эпидемиологического мониторинга, обеспечить контроль степени охвата населения обследованием и оперативно оценивать эпидемиологическую обстановку на основе структуры и динамики лабораторных данных.

## **ГЛАВА VI. СОВЕРШЕНСТВОВАНИЕ ЭПИДЕМИОЛОГИЧЕСКОГО НАДЗОРА ЗА НОВОЙ КОРОНАВИРУСНОЙ ИНФЕКЦИЕЙ (COVID-19) НА ОСНОВЕ ЦИФРОВЫХ ТЕХНОЛОГИЙ УЧЕТА И АНАЛИЗА ДАННЫХ МОЛЕКУЛЯРНО-БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ**

Эпидемиологический надзор — это комплексная информационно-аналитическая система, обеспечивающая органы здравоохранения актуальными и достоверными данными, необходимыми для осуществления мероприятий по профилактике и снижению инфекционной заболеваемости среди населения. Данная система включает в себя сбор, анализ, интерпретацию и передачу информации о развитии эпидемического процесса, а также о факторах, влияющих на распространение инфекционных заболеваний.

К основным задачам эпидемиологического надзора относятся: своевременное выявление и регистрация инфекционных заболеваний; определение масштабов и структуры заболеваемости; выявление причин и условий, способствующих распространению инфекций; оценка эффективности проводимых профилактических и противоэпидемических мероприятий; прогнозирование эпидемиологической ситуации и разработка рекомендаций по ее улучшению.

Эпидемиологический надзор служит основой для разработки стратегии и совершенствования деятельности Федеральной службы по надзору в сфере защиты прав потребителей и благополучия человека, позволяя определять приоритетные направления работы, оптимизировать распределение ресурсов и повышать эффективность профилактических и противоэпидемических мероприятий.

Кроме того, данные, полученные в результате эпидемиологического надзора, используются для разработки и внедрения новых методов диагностики, лечения и профилактики инфекционных заболеваний, что в

конечном итоге, способствует снижению заболеваемости и улучшению качества жизни населения.

Эпидемиологический надзор тесно взаимосвязан с системой эпидемиологического контроля — комплексом профилактических и противоэпидемических мероприятий, направленных на предотвращение возникновения и распространения инфекционных заболеваний. Эпидемиологический контроль включает в себя организацию и проведение санитарно-эпидемиологических расследований, разработку и реализацию планов профилактических мероприятий, а также мониторинг и оценку их эффективности.

Таким образом, эпидемиологический надзор играет ключевую роль в обеспечении санитарно-эпидемиологического благополучия населения и снижении инфекционной заболеваемости. Его развитие и совершенствование являются важными задачами для органов здравоохранения и санитарно-противоэпидемической службы.

Схема функционирования системы эпидемиологического надзора, согласно Б.Л. Черкасскому (2006 г.), приведена на рисунке 6.1.



**Рисунок 6.1** — Схема функционирования системы эпидемиологического надзора согласно Б.Л. Черкасскому [39]

Важным этапом функционирования системы эпидемиологического надзора является сбор, хранение и передача информации. Данный процесс включает в себя не только сбор и хранение данных, но и их оперативную передачу между различными уровнями и структурами системы. Как правило, подразумевается сбор, хранение и передача информации о количестве заболевших, а также сведений эпидемиологического анамнеза, которые помогают выявить причины возникновения заболеваний, их распространение и принять меры по их предотвращению.

В современном мире, где эпидемиологические угрозы могут возникать внезапно и распространяться с высокой скоростью, роль информационных систем в эпидемиологическом надзоре становится все более значимой. Одним из примеров таких систем является цифровая Платформа SOLAR, представляющая собой программное решение, которое может быть рассмотрено как самостоятельный элемент информационной подсистемы эпидемиологического надзора.

Использование цифровой платформы SOLAR в системе эпидемиологического надзора за новой коронавирусной инфекцией (COVID-19) позволило существенно трансформировать информационную, диагностическую и управленческую подсистемы, обеспечив их интеграцию на основе единых цифровых решений и больших массивов данных молекулярно-биологических исследований.

Совершенствование **информационной подсистемы** эпидемиологического надзора связано, прежде всего, с централизацией и унификацией потоков данных о результатах ПЦР-исследований по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2). Платформа SOLAR обеспечила агрегирование деперсонифицированных сведений, поступающих из более чем 2 000 лабораторий различной ведомственной принадлежности, что позволило сформировать масштабный массив данных. В отличие от фрагментарного сбора информации, характерного для начального этапа пандемии, внедрение SOLAR обеспечило

централизованный сбор данных из всех субъектов Российской Федерации, минимизацию дублирования и снижение вероятности ошибок при передаче сведений.

**Диагностическая подсистема** эпидемиологического надзора, предусматривающая предэпидемическую диагностику, постановку эпидемиологического диагноза и прогнозирование эпидемиологической ситуации на основе оценки всех потоков информации, была существенно усовершенствована благодаря внедрению платформы SOLAR. Систематизация данных о результатах ПЦР-исследований по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2) способствовала более точному определению начала и завершения эпидемических подъемов, объективизации эпидемиологического диагноза и повышению достоверности расчета интенсивных показателей заболеваемости. Кроме того, интеграция данных SOLAR с аналитическими инструментами (в том числе методами кластеризации и машинного обучения) позволила выявлять скрытые закономерности развития эпидемического процесса, оценивать вклад различных факторов и формировать научно обоснованные прогнозы.

Совершенствование **управленческой подсистемы** выразилось в повышении оперативности и обоснованности принимаемых решений. Наличие актуальной информации в режиме, приближенном к реальному времени, обеспечило возможность быстрого реагирования на изменение эпидемиологической ситуации. Разработка и внедрение платформы SOLAR позволили получить инструмент для оценки эффективности проводимых противоэпидемических мероприятий, распределения ресурсов, а также планирования дополнительных мер в субъектах с неблагоприятной динамикой показателей.

Использование платформы SOLAR в сочетании с разработанной математической моделью краткосрочного прогноза позволило перейти от реактивной модели управления к проактивной, основанной на прогнозировании возможных сценариев развития эпидемического процесса.

Это создало условия для своевременного введения или смягчения ограничительных мероприятий, оптимизации объемов тестирования и повышения адресности профилактических мер.

Платформа SOLAR существенно дополняет сведения, поступающие от региональных Центров гигиены и эпидемиологии в субъектах Российской Федерации в виде экстренных извещений и отчетных форм. Она позволяет полностью автоматизировать процесс сбора и анализа данных, что ускоряет принятие решений и повышает эффективность эпидемиологического надзора. Кроме того, Платформа SOLAR может интегрироваться с другими информационными системами, что обеспечивает более широкий охват и глубину анализа эпидемиологической ситуации.

Платформа SOLAR представляет собой инновационное решение для автоматизации процесса сбора и обработки данных о результатах ПЦР-исследований на наличие РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2). Данная система позволяет значительно минимизировать вероятность возникновения технических ошибок, которые могут негативно сказаться на точности и своевременности получения информации. Использование данных Платформы для совершенствования системы эпидемиологического надзора показано на рисунке 6.2.

Одним из ключевых преимуществ платформы SOLAR является возможность передачи данных по защищенным каналам связи, что не только обеспечивает высокий уровень безопасности информации, но и существенно сокращает время поступления данных в организации Роспотребнадзора. Благодаря этому, специалисты могут оперативно получать доступ к актуальным сведениям, что способствует более эффективному принятию решений и проведению мероприятий по контролю и предотвращению распространения инфекции.

Хранение сведений о результатах лабораторных исследований на платформе SOLAR организовано в отдельной базе данных с использованием

облачных технологий, обеспечивая надежное и безопасное хранение информации, а также упрощает доступ к ней для анализа и обработки.

На этапе комплексного анализа данных о результатах лабораторных исследований и сведений о регистрации случаев заболеваний особенно важно формирование анализируемых параметров на единой платформе, что позволяет создать полную и всестороннюю картину эпидемиологической ситуации, выявить тенденции и закономерности, а также оптимизировать процесс принятия решений, на основе полученных данных.

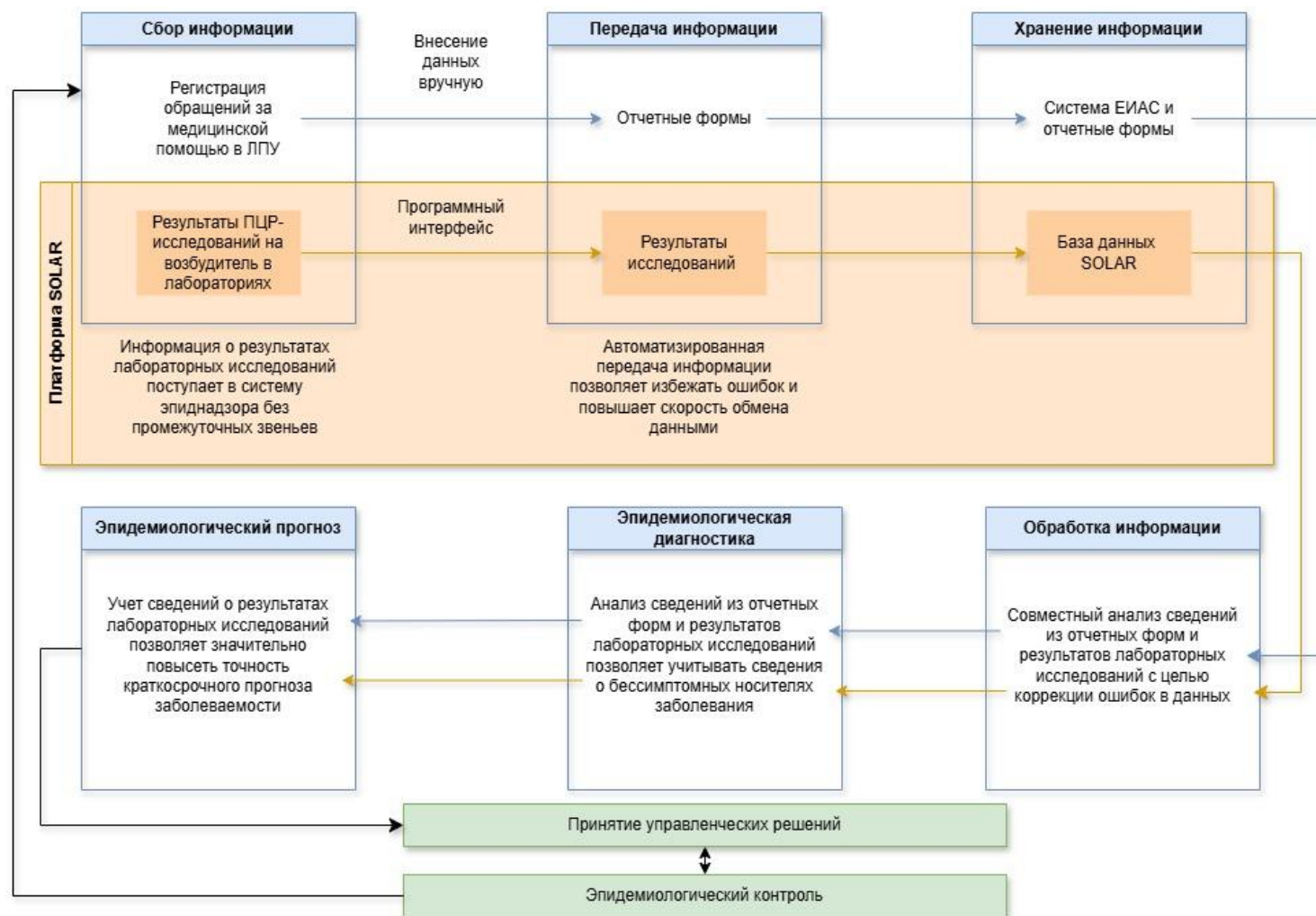
Платформа SOLAR также может быть интегрирована с другими информационными системами и ресурсами, что расширяет ее функциональность и возможности для анализа, позволяя учитывать дополнительные факторы и параметры, влияющие на эпидемиологическую ситуацию, и более точное прогнозирование развития событий.

На этапе проведения оперативного анализа эпидемиологической обстановки крайне важно использовать сведения о результатах ПЦР-исследований в качестве дополнительного и дублирующего источника данных о динамике эпидемического процесса. ПЦР (полимеразная цепная реакция) — это высокоточный метод молекулярной биологии, который позволяет выявлять наличие определенного генетического материала (в данном случае — вирусных частиц) в образцах биологического материала. Результаты ПЦР-исследований населения Российской Федерации являются важным индикатором текущей эпидемиологической ситуации, поскольку позволяют оперативно получать данные о количестве инфицированных лиц в регионах Российской Федерации и в стране в целом.

Использование сведений о результатах ПЦР-исследований особенно актуально в условиях динамично развивающейся эпидемиологической обстановки, когда необходимо оперативно реагировать на изменения в распространении инфекции, что подтверждается результатами исследований, описанных в *Главе V «ВОЗМОЖНОСТЬ ПРИМЕНЕНИЯ ЦИФРОВЫХ ТЕХНОЛОГИЙ УЧЕТА РЕЗУЛЬТАТОВ И АНАЛИЗА ДАННЫХ*

*МОЛЕКУЛЯРНО-БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ НАСЕЛЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ ДЛЯ ПОВЫШЕНИЯ ИНФОРМАТИВНОСТИ АНАЛИЗА ЭПИДЕМИОЛОГИЧЕСКОЙ ОБСТАНОВКИ ПО НОВОЙ КОРОНАВИРУСНОЙ ИНФЕКЦИИ (COVID-19)»,* которые демонстрируют эффективность и надежность данного подхода.

В процессе эпидемиологической диагностики для выявления регионов, наиболее неблагоприятных по новой коронавирусной инфекции (COVID-19), в течении нескольких периодов пандемии целесообразно применение метода кластеризации регионов по уровню заболеваемости. Метод позволяет произвести ранжирование субъектов Российской Федерации с учетом коррекции случайных «выбросов» в количестве зарегистрированных случаев заболевания новой коронавирусной инфекцией (COVID-19).



**Рисунок 6.2** — Использование данных Платформы SOLAR для совершенствования системы эпидемиологического надзора за новой коронавирусной инфекцией (COVID-19)

Подобные «выскакивающие величины» могут быть связаны с различными факторами, например, с особенностями системы здравоохранения в регионе или с особенностями проведения ПЦР-обследования по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2). Учет этих факторов позволяет получить более точную картину заболеваемости новой коронавирусной инфекцией (COVID-19) в конкретном субъекте.

Кластеризация представляет собой метод группировки объектов (в данном случае — регионов) на основе их сходства по определенным признакам. В контексте эпидемиологической диагностики кластеризация позволяет выявить группы регионов со схожими уровнями заболеваемости новой коронавирусной инфекцией (COVID-19). Проведенное исследование позволило выделить также и группы территорий, достоверно различающихся по интенсивности проявлений эпидемического процесса. Используемый нами метод кластеризации на основе машинного обучения позволяет учитывать сложные зависимости между признаками и автоматически определять наиболее сходные по эпидемиологическим параметрам группы регионов.

Использование данных Платформы SOLAR, а именно показателя доли положительных результатов ПЦР-исследований, продемонстрировал высокую прогностическую способность для краткосрочного прогнозирования динамики регистрации новых случаев заболевания. Максимальные коэффициенты корреляции в разные периоды пандемии достигали значений от 0,97 до 0,99 (очень высокая связь по шкале Чеддока) при различных величинах временного интервала, что подтверждает информативность этого показателя для оперативного планирования противоэпидемических мероприятий.

Таким образом, цифровизация эпидемиологического надзора за новой коронавирусной инфекцией (COVID-19) стала ключевым инструментом в

обеспечении своевременного контроля за распространением заболевания, повышении точности диагностики и ускорении принятия управленческих решений в условиях быстро меняющейся эпидемиологической ситуации. Пандемия новой коронавирусной инфекции показала роль и высокую значимость внедрения в систему эпидемиологического надзора информационных систем, таких как платформа SOLAR, которая может быть рассмотрена как самостоятельный элемент информационной подсистемы эпидемиологического надзора.

Внедрение Платформы SOLAR в информационную подсистему эпидемиологического надзора позволило автоматизировать процесс сбора и обработки данных о результатах ПЦР-исследований на наличие РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2), что позволило значительно минимизировать вероятность возникновения технических ошибок, субъективно увеличить оперативность, точность получаемой информации и скорость принятия комплекса управленческих решений для предотвращения распространения заболевания.

Проведенное исследование показало, что использование Платформы SOLAR в качестве элемента информационной подсистемы эпидемиологического надзора может значительно повысить его эффективность. Это достигается за счет более точного и оперативного сбора данных, а также за счет возможности более глубокого анализа и прогнозирования эпидемиологических тенденций. Доступ к большому объему данных результатов ПЦР-тестирования, собираемых аналитической платформой SOLAR, обеспечил возможность построения более реалистичных и адаптивных моделей, способных учитывать региональные особенности развития эпидемического процесса новой коронавирусной инфекции.

Цифровая трансформация представляет собой ключевое направление в современной эпидемиологии, обладающее значительным потенциалом с точки зрения улучшения параметров общественного здравоохранения. В условиях стремительного развития эпидемического процесса цифровые

решения создают основу мониторинга в режиме, приближенном к реальному времени, что существенно повышает чувствительность системы эпидемиологического надзора к ранним изменениям эпидемиологических показателей. Особое значение в условиях цифровой трансформации приобретает системное использование данных молекулярно-биологических исследований. Совершенствование системы эпидемиологического надзора за новой коронавирусной инфекцией (COVID-19) на основе цифровых технологий учета и анализа данных молекулярно-биологических исследований позволяет объективизировать эпидемиологический анализ, повысить точность расчета показателей и обеспечить научно обоснованное прогнозирование развития эпидемического процесса. Использование больших массивов стандартизированных лабораторных данных в сочетании с методами математического моделирования и машинного обучения формирует предпосылки для перехода к проактивной модели эпидемиологического надзора, ориентированной на предупреждение осложнения ситуации и своевременное принятие управленческих решений.

## ЗАКЛЮЧЕНИЕ

Пандемия новой коронавирусной инфекции (COVID-19) ускорила развитие цифровых технологий в эпидемиологии, способствовала созданию новых программных платформ и внедрению молекулярно-биологических методов в практику здравоохранения. Считающееся уже «золотым стандартом» диагностики инфекционных болезней, ПЦР-обследование позволяет точно выявлять патогены с высокой чувствительностью и специфичностью, что неоднократно показано в работах отечественных и зарубежных авторов [23, 71, 72, 83]. Масштабируемость данного молекулярно-биологического метода сделала его эффективным инструментом для борьбы с инфекциями, особенно в условиях пандемии, когда необходимо своевременное проведение эффективных противоэпидемических мероприятий. Введение масштабного ПЦР-обследования и активное использование молекулярно-генетических методов позволили своевременно выявлять новые варианты вируса и привели к накоплению значительного объема данных, имеющих различную структуру и источники соответствующих определению Big data (большие данные), что создало предпосылки для разработки новых, научно обоснованных подходов к анализу эпидемиологической обстановки с широким внедрением цифровых программных платформ.

В частности, проведение массового ПЦР-обследования населения Российской Федерации привело к накоплению значительного массива информации о результатах выявления РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2), что создало основу для разработки и внедрения в деятельность Роспотребнадзора специального программного обеспечения — платформы SOLAR (*System Of Laboratory Aggregation Results*). Внедрение данной цифровой Платформы позволило качественно улучшить возможности эпидемиологического надзора — ключевого направления эпидемиологической науки, за счет использования дополнительного

источника сведений о динамике эпидемического процесса новой коронавирусной инфекции (COVID-19).

С учетом беспрецедентного опыта получения таких сведений нами была принята попытка оценки информативности для оперативного анализа сведений поступающих из платформы SOLAR и потенциала их применения для краткосрочного прогнозирования уровня и динамики эпидемического процесса. Пандемия новой коронавирусной инфекции (COVID-19) подчеркнула необходимость цифровой трансформации в эпидемиологическом надзоре, в том числе за счет разработки и внедрения в практическую деятельность программных платформ, в том числе — SOLAR. Активное развитие технологий позволило решать сложные задачи эпидемиологического анализа с помощью цифровых платформ и методов машинного обучения. Данные подходы повышают эффективность прогнозирования рисков благодаря масштабным базам данных и выявлению сложных взаимосвязей, что неоднократно подчеркивали отечественные и зарубежные авторы [14, 47, 75].

Современные цифровые технологии позволяют принимать оперативные управленческие решения в интересах минимизации рисков для здоровья населения, способствуя поиску новых методологических решений, управлению эпидемиологической ситуацией и адаптации системы санитарно-эпидемиологического надзора к изменяющимся условиям [47]. Так, в исследовании Горяева Д.В. и соавторов [13] показано, что цифровизация обеспечивает оперативность, высокую точность и реализацию управленческих решений, что также продемонстрировано и в нашем исследовании. В исследовании Акимкина В.Г. и соавторов [23] показано как цифровизация в области эпидемиологии значительно ускорила благодаря развитию информационных технологий в период пандемии новой коронавирусной инфекции (COVID-19). В ФБУН ЦНИИ Эпидемиологии Роспотребнадзора разработаны три платформы: VGARus для сбора данных о геномах возбудителя новой коронавирусной инфекции (SARS-CoV-2), SOLAR

для сбора и анализа результатов молекулярно-биологического обследования населения Российской Федерации по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2) и EpidSmart модуль «COVID» для анализа эпидемиологической ситуации, связанной с новой коронавирусной инфекцией (COVID-19). Указанные инструменты играют важную роль в мониторинге распространения новой коронавирусной инфекции (COVID-19).

Одной из важнейших задач системы эпидемиологического надзора является определение территорий и групп риска по заболеваемости. Для решения этой задачи используется большой спектр математических инструментов. Нами показано, что уровень заболеваемости новой коронавирусной инфекцией (COVID-19) демонстрировал значительные региональные различия на территории Российской Федерации во все периоды пандемии. Наивысшие показатели заболеваемости регистрировались в крупных мегаполисах, включая г. Москву, Московскую область и г. Санкт-Петербург, что, в целом, обусловлено высокой плотностью населения и интенсивной миграционной активностью в данных регионах. В сельских районах динамика распространения инфекции характеризовалась более низкими темпами, однако, в этих областях часто возникали проблемы, связанные с ограниченным доступом к медицинским услугам, что объясняется различиями в перечне и сроках противоэпидемических мероприятий, использованием различных в стратегических подходов органами государственной власти и органами здравоохранения к противодействию распространению инфекции, а также вариациями в уровнях социально-экономического развития регионов, плотности населения и скорости распространения новых генетических вариантов возбудителя новой коронавирусной инфекции (SARS-CoV-2) и соотносится с результатами исследований других авторов [3, 26, 41, 51, 91, 120].

В динамике заболеваемости новой коронавирусной инфекцией (COVID-19) населения Российской Федерации в период 2020–2023 гг. отмечалось восемь эпидемических циклов заболеваемости,

характеризовавшихся особенностями проявлений эпидемического процесса и обусловленных спецификой доминирующих вариантов возбудителя новой коронавирусной инфекции (SARS-CoV-2). Несмотря на то, что циклический характер заболеваемости характерен для многих инфекционных болезней, для новой коронавирусной инфекции (COVID-19), на наш взгляд, наиболее четко прослеживается взаимосвязь между показателями заболеваемости и сменой доминирующих вариантов возбудителя, что отчасти, может быть связано с бурным развитием нового, наукоемкого направления в системе эпидемиологического надзора - молекулярно-генетического мониторинга. Нами показано, что наиболее высокий уровень заболеваемости новой коронавирусной инфекцией (COVID-19) зарегистрирован в пятом периоде пандемии в январе — феврале 2022 г. — 905 случаев на 100 тыс. населения, после чего наметилась тенденция к снижению уровня заболеваемости и полностью соотносится с данными, полученными в исследовании Акимкина В.Г. [3].

Для оценки влияния инфраструктурных и социально-демографических факторов на уровень и динамику заболеваемости новой коронавирусной инфекции (COVID-19) необходимо проведение анализа на уровне отдельных субъектов Российской Федерации. Решение данной задачи потребовало разработки подхода разделения субъектов на кластеры с учетом уровня и динамики заболеваемости в каждый из периодов пандемии. Применение кластеризации позволило избежать необходимости определения пороговых показателей для распределения регионов по группам, что привело бы к дополнительной субъективизации результатов такой группировки. Кроме того, применение стандартного алгоритма кластеризации по методу k-средних для разделения субъектов на группы позволяет унифицировать и в значительной степени автоматизировать аналитический процесс.

Проведенный нами кластерный анализ временных рядов заболеваемости новой коронавирусной инфекцией (COVID-19) позволил получить распределение субъектов Российской Федерации по кластерам в

различные периоды пандемии. Наибольшее различие средних значений заболеваемости между отдельными кластерами наблюдалось со второго по пятый период пандемии новой коронавирусной инфекции (COVID-19), что совпадает по времени с периодами пандемии, когда перечень и объем проводимых противоэпидемических мероприятий в отдельных субъектах существенно различался. Подобные методы кластеризации территории по заданным признакам и временным рядам также применяются в работах ряда отечественных [10, 16, 19] и зарубежных авторов [62, 101].

При этом установлено, что наиболее высокие показатели заболеваемости стабильно регистрировались в субъектах с высокой плотностью населения, значительной долей городского населения и высокой миграционной активностью. Полученные результаты согласуются с данными других исследований. Так, в работе Carteni A. (2020) посредством применения метода множественной регрессии показана прямая связь плотности населения, мобильности и числа случаев заболевания новой коронавирусной инфекцией [61]. В исследовании Santana C. (2023) показано, что мобильность населения, уровень урбанизации и экономическая активность являются ключевыми факторами распространения новой коронавирусной инфекции (COVID-19) в Великобритании [68].

В диссертационном исследовании продемонстрировано, что молекулярно-биологические методы, в частности, ПЦР-диагностика и секвенирование, играют ключевую роль в системе эпидемиологического надзора, обеспечивая высокую чувствительность, специфичность и оперативность выявления инфекционных агентов, в том числе возбудителя новой коронавирусной инфекции (SARS-CoV-2), что соотносится с данными исследователей, как отечественных [4, 23], так и зарубежных [69, 71].

В исследовании продемонстрирована также весомая роль интеграции данных о результатах лабораторных исследований у бессимптомных носителей вируса посредством платформы SOLAR, что существенно повысило точность эпидемиологического мониторинга, обеспечило контроль

охвата населения обследованием и оперативной оценки эпидемиологической обстановки на основе структуры и динамики лабораторных данных.

Разработанные в работе алгоритмы мониторинга коррелируют с данными масштабного исследования в Словакии, показанные в научной работе Pavelka M. (2021), где было доказано, что массовое экспресс-тестирование в сочетании с цифровым учетом позволяет снизить распространенность новой коронавирусной инфекции (COVID-19) на 70% за одну неделю [110]. Наш подход, основанный на оперативной передаче результатов ПЦР-исследований посредством платформы SOLAR, также подтверждает выводы о том, что скорость идентификации носителей в предсимптомной фазе, когда происходит до 44% вторичных заражений, согласно исследованию He X. et al. (2020 г.), имеет решающее значение для прерывания путей передачи инфекции [109].

Нами научно обосновано применение данных молекулярно-биологических исследований Платформы SOLAR для совершенствования эпидемиологического надзора с целью объективизации эпидемиологического анализа, повышения точности расчетов и прогнозирования динамики развития эпидемического процесса.

Важным результатом диссертационной работы является обоснование эффективности применения цифровых платформ для эпидемиологического надзора. Установлено, что использование Платформы SOLAR, агрегирующей результаты более чем 200 млн ПЦР-исследований, позволяет существенно повысить оперативность и достоверность анализа эпидемиологической ситуации, а также расширяет возможности для краткосрочного прогнозирования заболеваемости.

Данный вывод подтверждается результатами международных исследований. Так, согласно данным Alsunaidi SJ (2021 г.), использование цифровых платформ и больших данных позволяет значительно повысить эффективность мониторинга пандемии и обоснованность принимаемых управленческих решений [55]. В исследовании Jiao Z. системно обобщены

примеры применения больших данных и искусственного интеллекта в борьбе с эпидемиями, описаны практические случаи и проблемы с акцентом на профилактику и контроль эпидемий, показано, что большие данные и искусственный интеллект обладают значительным потенциалом для борьбы с новой коронавирусной инфекцией (COVID-19) [54].

Разработанная в диссертации архитектура интеграции в систему эпидемиологического надзора данных массового ПЦР-обследования населения Российской Федерации по выявлению возбудителя новой коронавирусной инфекции (SARS-CoV-2) посредством Платформы SOLAR, обладает свойством высокой масштабируемости. Полученные результаты доказывают, что созданный цифровой контур может быть адаптирован для мониторинга других инфекций с высоким эпидемическим и пандемическим потенциалом распространения возбудителя инфекции, что позволяет рассматривать предложенную в диссертационной работе систему как прототип универсального инструмента обеспечения биологической безопасности.

Диссертационная работа, выполненная с применением методологического подхода, основанного на принципах доказательной медицины и эпидемиологического анализа, подкрепила гипотезу о важности внедрения цифровых технологий в систему эпидемиологического надзора. Результаты проведенного исследования показывают, что цифровые технологии являются неотъемлемым и критически значимым инструментом для диагностики, мониторинга и предотвращения эпидемического распространения инфекционных заболеваний. В рамках работы проведена всесторонняя оценка эффективности использования информационных систем и аналитических инструментов в эпидемиологической практике, что позволило определить их потенциальные преимущества.

Диссертационное исследование **заполнило пробелы в научных данных**, раскрыв вопросы определения групп территорий, достоверно различающихся по интенсивности проявлений эпидемического процесса.

Наиболее высокая интенсивность эпидемического процесса регистрировалась в кластере, включающем в различные периоды г. Москву, г. Санкт-Петербург, Хабаровский край, и впервые показано, что средние показатели заболеваемости новой коронавирусной инфекцией (COVID-19) превышали общероссийские в среднем 2,8–4,5 раза ( $p < 0,01$ ).

При использовании данных Платформы SOLAR, нами впервые показано, что в 5-м периоде пандемии (10.01–25.07.2022 гг.) зафиксирован максимальный коэффициент корреляции (0,93 – очень высокая связь по шкале Чеддока) при временном интервале в 9 дней между количеством ПЦР-исследований и регистрацией новых случаев заболевания. Показатель доли положительных результатов ПЦР-исследований продемонстрировал высокую прогностическую способность в интересах краткосрочного прогнозирования (7–14 дней) динамики регистрации новых случаев заболевания. Максимальные коэффициенты корреляции в разные периоды пандемии достигали значений от 0,97 до 0,99 (очень высокая связь по шкале Чеддока) при различных величинах временного интервала, что подтверждает информативность этого показателя для оперативного планирования противоэпидемических мероприятий.

Цифровая трансформация представляет собой ключевое направление в современной эпидемиологии, обладающее значительным потенциалом с точки зрения улучшения параметров общественного здравоохранения. В условиях стремительного развития эпидемического процесса цифровые решения создают основу мониторинга в режиме, приближенном к реальному времени, что существенно повышает чувствительность системы эпидемиологического надзора к ранним изменениям эпидемиологических показателей.

Особое значение в условиях цифровой трансформации приобретает системное использование данных молекулярно-биологических исследований. Совершенствование системы эпидемиологического надзора за новой коронавирусной инфекцией (COVID-19) на основе цифровых технологий

учета и анализа данных молекулярно-биологических исследований позволяет объективизировать эпидемиологический анализ, повысить точность расчета показателей и обеспечить научно обоснованное прогнозирование развития эпидемического процесса. Использование цифровых платформ в комбинации с методами математического моделирования и технологиями машинного обучения формирует основу для осуществления постоянного мониторинга и прогнозирования потенциальных эпидемических опасностей на территории Российской Федерации и разработке стратегий для их предотвращения.

## ВЫВОДЫ

1. В динамике заболеваемости новой коронавирусной инфекцией (COVID-19) населения Российской Федерации в период 2020–2023 гг. отмечалось восемь эпидемических циклов заболеваемости, характеризовавшихся особенностями проявлений эпидемического процесса и обусловленных спецификой доминирующих вариантов возбудителя новой коронавирусной инфекции (SARS-CoV-2). Динамика заболеваемости новой коронавирусной инфекцией (COVID-19) в период с апреля 2020 г. по январь 2022 г. имела тенденцию к росту. Наиболее высокий уровень заболеваемости новой коронавирусной инфекцией (COVID-19) зарегистрирован в пятом периоде пандемии в январе — феврале 2022 г. — 905 случаев на 100 тыс. населения (вариант Омикрон SARS-CoV-2), после чего наметилась тенденция к снижению уровня заболеваемости.

2. Кластерный анализ временных рядов заболеваемости новой коронавирусной инфекцией (COVID-19), проведенный в 62 субъектах Российской Федерации, позволил выделить группы территорий достоверно различающиеся по интенсивности проявлений эпидемического процесса. Наиболее высокая интенсивность эпидемического процесса регистрировалась в кластере, включающем в различные периоды г. Москву, г. Санкт-Петербург, Хабаровский край, где средние показатели заболеваемости новой коронавирусной инфекцией (COVID-19) превышали общероссийские в среднем 2,8–4,5 раза ( $p < 0,01$ ).

3. На основе применения корреляционного анализа и использования метода машинного обучения установлено, что уровень заболеваемости и характер распространения новой коронавирусной инфекции (COVID-19) в субъектах Российской Федерации имел выраженную зависимость от доли городского населения, плотности населения региона, интенсивности экономического развития (количества активных предприятий), показателей активности миграционных процессов (коэффициента миграционного прироста).

4. Цифровая Платформа агрегирования результатов лабораторных исследований SOLAR (System of laboratory aggregation results) является эффективным инструментом для агрегации, обработки и анализа данных ПЦР-исследований по выявлению РНК возбудителя новой коронавирусной инфекции (SARS-CoV-2) среди населения Российской Федерации. Интеграция данных о результатах лабораторных исследований среди бессимптомных носителей вируса посредством Платформы SOLAR позволила обеспечить контроль уровня охвата населения ПЦР-обследованием и оперативную оценку эпидемиологической обстановки на основе показателей заболеваемости, структуры и динамики данных лабораторных исследований.

5. Совершенствование системы эпидемиологического надзора за новой коронавирусной инфекцией (COVID-19) на основе цифровизации технологий учета и анализа результатов молекулярно-биологических исследований позволяет существенно дополнять сведения о заболеваемости населения Российской Федерации, достигать высокой информативности о характере анализируемой эпидемиологической ситуации, обеспечивать достоверное краткосрочное прогнозирование эпидемиологической обстановки в интересах оперативности и адресности принимаемых управленческих решений.

## ПРАКТИЧЕСКИЕ РЕКОМЕНДАЦИИ

По результатам, полученным в данном диссертационном исследовании, при проведении ретроспективного эпидемиологического анализа нами рекомендуется применение метода кластеризации территорий по интенсивным показателям при изучении заболеваемости населения Российской Федерации новой коронавирусной инфекцией (COVID-19) для получения объективной сравнительной оценки развития эпидемического процесса в различных субъектах Российской Федерации.

Данные молекулярно-биологических методов исследования и кластеризация субъектов Российской Федерации по интенсивным показателям заболеваемости населения могут применяться в качестве индикатора эффективности проведения противоэпидемических мероприятий с учетом особенностей каждого субъекта Российской Федерации.

Для автоматизации рутинных процессов, проведения комплексного анализа эпидемиологических данных с использованием методов машинного обучения, а также надежного хранения и поиска информации о заболеваемости новой коронавирусной инфекцией (COVID-19) на территории Российской Федерации целесообразно использовать современные цифровые платформы, в том числе с агрегированными результатами молекулярно-биологических исследований (SOLAR).

## ПЕРСПЕКТИВЫ ДАЛЬНЕЙШЕЙ РАЗРАБОТКИ ТЕМЫ

Результаты, полученные в рамках диссертационного исследования, касающиеся применения кластерного анализа для изучения ретроспективных данных по заболеваемости новой коронавирусной инфекцией (COVID-19), открывают новые перспективы для углубленного анализа региональных особенностей эпидемического процесса и других инфекций с аэрозольным механизмом передачи, с высоким эпидемическим и пандемическим потенциалом распространения возбудителя. Внедрение математических методов моделирования и алгоритмов машинного обучения в данный контекст позволит значительно повысить точность прогнозирования эпидемической обстановки и эффективность управления эпидемиологическими рисками, что в свою очередь, будет способствовать разработке более целенаправленных и научно обоснованных стратегий эпидемиологического надзора.

Дополнение существующей структуры Платформы SOLAR данными молекулярно-биологических исследований, полученными при изучении заболеваемости другими инфекционными агентами, позволит выстроить высокотехнологичную систему эпидемиологического надзора, основанную на современных цифровых решениях.

**СПИСОК СОКРАЩЕНИЙ**

**ПЦР** — полимеразная цепная реакция

**SOLAR** (от англ. *System Of Laboratory Aggregation Results*) — Платформа агрегирования результатов лабораторных исследований)

**ВОЗ** — Всемирная Организация Здравоохранения

**SARS-CoV-2** (англ. *Severe acute respiratory syndrome coronavirus 2*) — коронавирус 2 типа, вызывающий тяжелый острый респираторный синдром

**COVID-19** (англ. *COronaVirus Disease 2019*) — тяжелая острая респираторная инфекция, вызываемая коронавирусом SARS-CoV-2 (2019-nCoV)

**VGARus** (англ. *Virus Genome Aggregator of Russia*) — платформа агрегирования результатов расшифровок генома возбудителей инфекционных и паразитарных заболеваний

## СПИСОК ЛИТЕРАТУРЫ

1. COVID-19 в России: эпидемиология и молекулярно-генетический мониторинг / В. Г. Акимкин, Т. А. Семенович, С. В. Углева [и др.] // Вестник Российской академии медицинских наук. – 2022. – Т. 77, № 4. – С. 254-260. – DOI 10.15690/vramn2121.
2. COVID-19: научно-практические аспекты борьбы с пандемией в Российской Федерации / А. Ю. Попова, Е. Б. Ежлова, В. Ю. Смоленский [и др.]. – Саратов: Общество с ограниченной ответственностью "Амирит", 2021. – 608 с. – ISBN 978-5-00140-901-4.
3. COVID-19: эволюция пандемии в России. Сообщение I: проявления эпидемического процесса COVID-19 / В. Г. Акимкин, А. Ю. Попова, А. А. Плоскирева [и др.] // Журнал микробиологии, эпидемиологии и иммунобиологии. – 2022. – Т. 99, № 3. – С. 269-286. – DOI 10.36233/0372-9311-276.
4. COVID-19: эволюция пандемии в России. Сообщение II: динамика циркуляции геновариантов вируса SARS-CoV-2 / В. Г. Акимкин, А. Ю. Попова, К. Ф. Хафизов [и др.] // Журнал микробиологии, эпидемиологии и иммунобиологии. – 2022. – Т. 99, № 4. – С. 381-396. – DOI 10.36233/0372-9311-295.
5. D-russia | Схема работы системы цифровых пропусков в Москве и области [Электронный ресурс] URL: <https://d-russia.ru/shema-raboty-sistemy-cifrovyyh-propuskov-v-moskve-i-oblasti.html> (дата обращения: 29.04.2026).
6. Yandex DataLens. Коронавирус: дашборд. [Электронный ресурс] URL: <https://datalens.ru/gallery/k0rxsq57niv46?preview=1> (дата обращения: 29.04.2026).
7. Анализ применения телемедицинских технологий в Российской Федерации в 2019-2022 гг. / М. Д. Лагутин, В. П. Чигрина, Д. А. Самофалов [и др.] // Проблемы социальной гигиены, здравоохранения и истории медицины. – 2023. – Т. 31, № 2. – С. 264-269. – DOI 10.32687/0869-866X-2023-31-2-264-269.
8. Аналитическая записка: Влияние COVID-19 на пожилых людей. [Электронный ресурс] URL: [https://www.un.org/sites/un2.un.org/files/2021/01/old\\_persons\\_russian.pdf](https://www.un.org/sites/un2.un.org/files/2021/01/old_persons_russian.pdf) (дата обращения: 29.04.2026).

9. Беляков, В.Д. Эпидемиология: Учебник / В.Д. Беляков, Р.Х. Яфаев. – Москва, 1989. – 416 с. – ISBN 5-225-01513- 1.
10. Биотехнологии в геномном эпидемиологическом надзоре. Состояние и перспективы развития / В. Г. Акимкин, А. С. Черкашина, А. И. Тюменцев [и др.] // Эпидемиология и вакцинопрофилактика. – 2025. – Т. 24, № 3. – С. 4-13. – DOI 10.31631/2073-3046-2025-24-3-4-13.
11. Галушин, П. В. Кластерный анализ социально значимых заболеваний в Российской Федерации / П. В. Галушин, Е. Н. Галушина // Вестник НГУЭУ. – 2023. – № 1. – С. 169-183. – DOI 10.34020/2073-6495-2023-1-169-183.
12. Геномное секвенирование вируса SARS-CoV-2 Руководство по внедрению для обеспечения максимального позитивного воздействия на общественное здравоохранение 8 января 2021 г. [Электронный ресурс] URL: <https://iris.who.int/server/api/core/bitstreams/2e81b795-48db-4f65-abe3-da2d00e232fc/content> (дата обращения: 21.04.2026).
13. Горяев, Д. В. Цифровые технологии в задачах управления санитарно-эпидемиологической ситуацией на уровне субъекта федерации / Д. В. Горяев, И. В. Тихонова // Здравоохранение Российской Федерации. – 2022. – Т. 66, № 5. – С. 375-379. – DOI 10.47470/0044-197X-2022-66-5-375-379.
14. Денисов, Н. С. Тренды популяционных исследований: молекулярная и цифровая эпидемиология (обзор) / Н. С. Денисов, Е. М. Каменских, О. С. Федорова // Современные технологии в медицине. – 2022. – Т. 14, № 4. – С. 60-72. – DOI 10.17691/stm2022.14.4.07.
15. Информационный бюллетень о ситуации и принимаемых мерах по недопущению распространения заболеваний, вызванных новым коронавирусом. [Электронный ресурс] URL: [https://www.rospotrebnadzor.ru/about/info/news/news\\_details.php?ELEMENT\\_ID=14067](https://www.rospotrebnadzor.ru/about/info/news/news_details.php?ELEMENT_ID=14067) (дата обращения: 28.04.2026).
16. Киршин, И. А. Кластеризация субъектов Российской Федерации по ресурсной обеспеченности населения в сфере здравоохранения и использованию коечного фонда / И. А. Киршин, А. И. Касимова // Статистика и Экономика. – 2025. – Т. 22, № 5. – С. 30-40. – DOI 10.21686/2500-3925-2025-5-30-40.
17. Кластеризация стандартизованных показателей кумулятивной инцидентности за многолетний период как метод анализа пространственного

распределения случаев заболевания / Д. В. Дубоделов, С. В. Углева, Г. А. Гасанов [и др.] // Эпидемиология и инфекционные болезни. Актуальные вопросы. – 2024. – Т. 14, № 1. – С. 31-35. – DOI 10.18565/epidem.2023.14.1.31-5.

18. Минздрав России [Электронный ресурс] URL: <https://minzdrav.gov.ru/news/2024/04/20/21284-minzdrav-rossii-podvyol-itogi-raboty-za-2023-god> (дата обращения: 29.04.2026).

19. Михайлов, А. Ю. Влияние природно-климатических и социально-экономических факторов на заболеваемость и смертность от туберкулеза в России / А. Ю. Михайлов, Я. Ю. Панкова // Социальные аспекты здоровья населения. – 2024. – Т. 70, № S5. – DOI 10.21045/2071-5021-2024-70-5S-22.

20. Михеева, И. В. Проблемы цифровой трансформации аналитической подсистемы эпидемиологического надзора / И. В. Михеева, А. Н. Герасимов, В. Г. Акимкин // Эпидемиология и инфекционные болезни. Актуальные вопросы. – 2026. – Т. 16, № 1. – С. 6-11. – DOI 10.18565/epidem.2026.16.1.6-11.

21. Модель прогнозирования пандемии COVID-19 на основе машинного обучения в отдельных регионах Российской Федерации / Д. В. Гаврилов, Р. В. Абрамов, А. В. Кирилкина [и др.] // Фармакоэкономика. Современная фармакоэкономика и фармакоэпидемиология. – 2021. – Т. 14, № 3. – С. 342-356. – DOI 10.17749/2070-4909/farmakoeconomika.2021.108.

22. Молекулярная биология клетки: в 3-х томах. Т. I / Б. Альбертс, А. Джонсон, Д. Льюис и др.— М.–Ижевск: НИЦ «Регулярная и хаотическая динамика», Институт компьютерных исследований, 2013. — 808 стр. – ISBN 978-5-4344-0112-8.

23. Молекулярно-генетический мониторинг и технологии цифровой трансформации в современной эпидемиологии / В. Г. Акимкин, К. Ф. Хафизов, Д. В. Дубоделов [и др.] // Вестник Российской академии медицинских наук. – 2023. – Т. 78, № 4. – С. 363-369. – DOI 10.15690/vramn13672.

24. Молекулярные методы диагностики новой коронавирусной инфекции: сравнение петлевой изотермической амплификации и полимеразной цепной реакции / В. Г. Акимкин, В. В. Петров, К. В. Красовитов [и др.] // Вопросы вирусологии. – 2021. – Т. 66, № 6. – С. 417-424. – DOI 10.36233/0507-4088-86.

25. Момыналиев, К. Т. Анализ запросов динамики Google Trends в России в период пандемии коронавирусной инфекции как инструмент эпидемиологического надзора / К. Т. Момыналиев, В. Г. Акимкин // Эпидемиология и инфекционные болезни. Актуальные вопросы. – 2020. – Т. 10, № 4. – С. 33-37. – DOI 10.18565/epidem.2020.10.4.33-7.

26. Неоднородность параметров модифицированной SIR-модели волн эпидемического процесса COVID-19 в Российской Федерации / А. Ю. Попова, Н. В. Зайцева, В. Б. Алексеев [и др.] // Гигиена и санитария. – 2023. – Т. 102, № 8. – С. 740-749. – DOI 10.47470/0016-9900-2023-102-8-740-749.

27. О мерах по совершенствованию эпидемиологического надзора в части индикации возбудителей инфекционных болезней / Г. Г. Онищенко, Е. Б. Ежлова, Ю. В. Демина, А. А. Мельникова // Эпидемиология и инфекционные болезни. Актуальные вопросы. – 2013. – № 2. – С. 4-13.

28. О состоянии санитарно-эпидемиологического благополучия населения в Российской Федерации в 2022 году: Государственный доклад. М.: Федеральная служба по надзору в сфере защиты прав потребителей и благополучия человека, 2023. 368 с.

29. О состоянии санитарно-эпидемиологического благополучия населения в Российской Федерации в 2021 году: Государственный доклад. М.: Федеральная служба по надзору в сфере защиты прав потребителей и благополучия человека, 2022. 340 с.

30. Объясняем.рф | Стопкоронавирус [Электронный ресурс] URL: <https://xn--90aivcdt6dxbc.xn--p1ai/stopkoronavirus/> (дата обращения: 29.04.2026).

31. Организация масштабных многоцентровых клинических исследований на территории г. Москвы в период пандемии новой коронавирусной инфекции на основе современных цифровых решений / Е. С. Кучерявых, Ю. Н. Карпова, Я. С. Панарина [и др.] // Медицинские технологии. Оценка и выбор. – 2023. – № 3(45). – С. 82-92. – DOI 10.17116/medtech20234503182.

32. Орлов, А. А. Коэффициенты корреляции: шкала Чеддока и значимость / А. А. Орлов, А. И. Орлов // Контроллинг. – 2024. – № 4(94). – С. 28-37.

33. Основы математического анализа: учебник для вузов: в 2 ч. / В. А. Ильин, Э. Г. Позняк - 3-е изд. - М. : Наука. Физматлит, 2000. - (Курс высшей математики и математической физики; вып. 2). - ISBN 5-02-015230-7. Ч. 2. - 2000. - 447 с. - ISBN 5-02-015232-3.

34. Особенности эпидемического процесса COVID-19 в каждую из пяти волн заболеваемости в России / Л. С. Карпова, А. Б. Комиссаров, К. А. Столяров [и др.] // Эпидемиология и вакцинопрофилактика. – 2023. – Т. 22, № 2. – С. 23-36. – DOI 10.31631/2073-3046-2023-22-2-23-36.

35. Покровский, В. И. Описательное эпидемиологическое исследование: (ретроспективный эпидемиологический анализ) / В. И. Покровский, Н. Н. Филатов, И. П. Палтышев. – Москва: Санэпидмедиа, 2005. — 239 с.

36. Постановление Главного государственного санитарного врача РФ от 18.03.2020 N 7 "Об обеспечении режима изоляции в целях предотвращения распространения COVID-2019" (с изменениями и дополнениями) | ГАРАНТ [Электронный ресурс] URL: <https://base.garant.ru/73764449/> (дата обращения: 28.04.2026).

37. Приказ Министерства здравоохранения РФ от 19.03.2020 N 198н "О временном порядке организации работы медицинских организаций в целях реализации мер по профилактике и снижению рисков распространения новой коронавирусной инфекции COVID-19" (с изменениями и дополнениями) (документ не действует) | ГАРАНТ [Электронный ресурс] URL: <https://base.garant.ru/73769697/> (дата обращения: 28.04.2026).

38. Путешествую без COVID-19 [Электронный ресурс] URL: [https://fci.eabr.org/projects/travel\\_without\\_covid](https://fci.eabr.org/projects/travel_without_covid) (дата обращения: 29.04.2026)

39. Руководство по эпидемиологии инфекционных болезней [в 2 т.], Т.1 / Н.И. Брико, Г.Г. Онищенко, В.И. Покровский.- Москва: ООО "Издательство "Медицинское информационное агентство", 2019. – 880 с.- с. 465-468.

40. Саморегуляция паразитарных систем: молекулярно-генетические механизмы / В. Д. Беляков, Д. Б. Голубев, Г. Д. Каминский, В. В. Тец. – Ленинград: Медицина: Ленинградское отделение, 1987. – 239 с.

41. Сравнение первых трех волн пандемии COVID-19 в России (2020-2021 гг.) / Л. С. Карпова, К. А. Столяров, Н. М. Поповцева [и др.] //

Эпидемиология и вакцинопрофилактика. – 2022. – Т. 21, № 2. – С. 4-16. – DOI 10.31631/2073-3046-2022-21-2-4-16.

42. Сравнительный анализ структуры заболеваемости COVID-19 населения Саратовской области в периоды преобладания геновариантов Delta и Omicron / А. А. Зимирова, В. Е. Куклев, В. А. Сафронов, И. Н. Вяткин // Инфекция и иммунитет. – 2024. – Т. 14, № 1. – С. 181-186. – DOI 10.15789/2220-7619-CAP-17555.

43. Статистика вакцинации от коронавируса. [Электронный ресурс] URL: <https://gogov.ru/articles/covid-v-stats> (дата обращения: 28.04.2026).

44. Теория саморегуляции паразитарных систем и COVID-19 / В. Г. Акимкин, Т. А. Семенов, Д. В. Дубоделов [и др.] // Вестник Российской академии медицинских наук. – 2024. – Т. 79, № 1. – С. 33-41. – DOI 10.15690/vramn11607.

45. Указ Мэра Москвы от 05.03.2020 N 12-УМ "О введении режима повышенной готовности" (с изменениями и дополнениями) | ГАРАНТ [Электронный ресурс] URL: <https://base.garant.ru/73711482/> (дата обращения: 28.04.2026).

46. Указ Президента РФ от 09.05.2017 N 203 "О Стратегии развития информационного общества в Российской Федерации на 2017 - 2030 годы" | ГАРАНТ [Электронный ресурс] URL: <https://base.garant.ru/71670570/> (дата обращения: 28.04.2026).

47. Цифровые решения (vgarus, solar, «epidsmart - модуль COVID») в системе эпидемиологического надзора за новой коронавирусной инфекцией (COVID-19) / В. Г. Акимкин, Д. В. Дубоделов, А. С. Есьман [и др.] // Эпидемиология и инфекционные болезни. Актуальные вопросы. – 2025. – Т. 15, № 1. – С. 43-50. – DOI 10.18565/epidem.2025.15.1.43-50.

48. Черкасский, Б. Л. Руководство по общей эпидемиологии / Б. Л. Черкасский. — Москва, 2001. — 560 с.

49. Эпидемиологическая и экономическая эффективность массового ПЦР-обследования населения Российской Федерации в период эпидемии COVID-19 / В. Г. Акимкин, Д. В. Дубоделов, А. С. Есьман [и др.] // Эпидемиология и вакцинопрофилактика. – 2025. – Т. 24, № 6. – С. 28-35. – DOI 10.31631/2073-3046-2025-24-6-28-35.

50. Эпидемиологические особенности COVID-19 и характеристики пациентов в раннем периоде распространения варианта Омикрон SARS-CoV-2 в Российской Федерации / А. А. Фомичева, Н. Н. Пименов, С. В. Комарова [и др.] // Эпидемиология и вакцинопрофилактика. – 2024. – Т. 23, № 4. – С. 116-127. – DOI 10.31631/2073-3046-2024-23-4-116-127.

51. Эпидемиологические особенности коронавирусной инфекции COVID-19. Социальные факторы эпидемического процесса / Е. А. Медведева, Г. Г. Марьин, А. А. Плоскирева [и др.] // Эпидемиология и инфекционные болезни. Актуальные вопросы. – 2025. – Т. 15, № 2. – С. 21-27. – DOI 10.18565/epidem.2025.15.2.21-27.

52. A Novel Coronavirus from Patients with Pneumonia in China, 2019 / Na. Zhu, D. Zhang, W. Wang [et al.] // New England Journal of Medicine. – 2020. – Vol. 382, No. 8. – P. 727-733. – DOI 10.1056/nejmoa2001017.

53. A Novel Coronavirus Genome Identified in a Cluster of Pneumonia Cases — Wuhan, China 2019–2020 / W. Tan, X. Zhao, X. Ma [et al.] // China CDC Weekly. – 2020. – Vol. 2, No. 4. – P. 61-62. – DOI 10.46234/ccdcw2020.017.

54. Application of big data and artificial intelligence in epidemic surveillance and containment / Z. Jiao, H. Ji, Ju. Yan, X. Qi // Intelligent Medicine. – 2023. – Vol. 3, No. 1. – P. 36-43. – DOI 10.1016/j.imed.2022.10.003.

55. Applications of Big Data Analytics to Control COVID-19 Pandemic / Sh. J. Alsunaidi, A. M. Almuhaideb, N. M. Ibrahim [et al.] // Sensors. – 2021. – Vol. 21, No. 7. – P. 2282. – DOI 10.3390/s21072282.

56. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England / E. Volz, S. Mishra, M. Chand [et al.] // Nature. – 2021. – Vol. 593, No. 7858. – P. 266-269. – DOI 10.1038/s41586-021-03470-x.

57. Asymptomatic SARS-CoV-2 infection: A systematic review and meta-analysis / P. Sah, M. C. Fitzpatrick, C. F. Zimmer [et al.] // Proceedings of the National Academy of Sciences of the United States of America. – 2021. – Vol. 118, No. 34. – P. e2109229118. – DOI 10.1073/pnas.2109229118.

58. Bender D., Sartipi K. HL7 FHIR: An Agile and RESTful approach to healthcare information exchange // Proceedings of the 26th IEEE international symposium on computer-based medical systems. – IEEE, 2013. – С. 326-331.

59. Big data in healthcare: management, analysis and future prospects / S. Dash, S. K. Shakyawar, M. Sharma, S. Kaushik // *Journal of Big Data*. – 2019. – Vol. 6, No. 1. – P. 54. – DOI 10.1186/s40537-019-0217-0.
60. Burki, T. First shared SARS-CoV-2 genome: GISAID vs virological.org / T. Burki // *The Lancet Microbe*. – 2023. – Vol. 4, No. 6. – P. e395. – DOI 10.1016/s2666-5247(23)00133-7.
61. Carteni, A. How mobility habits influenced the spread of the COVID-19 pandemic: Results from the Italian case study / A. Carteni, L. Di Francesco, M. Martino // *Science of the Total Environment*. – 2020. – Vol. 741. – P. 140489. – DOI 10.1016/j.scitotenv.2020.140489.
62. Cerqueti, R. Combining rank-size and k-means for clustering countries over the COVID-19 new deaths per million / R. Cerqueti, V. Ficcadenti // *Chaos, Solitons & Fractals*. – 2022. – Vol. 158. – P. 111975. – DOI 10.1016/j.chaos.2022.111975.
63. Classification of Omicron (B.1.1.529): SARS-CoV-2 Variant of Concern [Электронный ресурс]. — URL: [https://www.who.int/news/item/26-11-2021-classification-of-omicron-\(b.1.1.529\)-sars-cov-2-variant-of-concern](https://www.who.int/news/item/26-11-2021-classification-of-omicron-(b.1.1.529)-sars-cov-2-variant-of-concern) (дата обращения: 29.04.2026).
64. Comparative Analysis of the Risks of Hospitalisation and Death Associated with SARS-CoV-2 Omicron (B.1.1.529) and Delta (B.1.617.2) Variants in England / T. Nyberg, N. M. Ferguson, S. G. Nash [et al.] // *Social Science Research Network*. – 2022. – DOI 10.2139/ssrn.4025932.
65. Coronavirus (COVID-19) Vaccinations - Our World in Data. [Электронный ресурс] URL: <https://ourworldindata.org/covid-vaccinations> (дата обращения: 29.04.2026).
66. COVID-19 cases | WHO COVID-19 dashboard [Электронный ресурс] URL: <https://data.who.int/dashboards/covid19/cases?n=c> (дата обращения: 28.04.2026).
67. COVID-19 epidemiological update – 19 January 2024. [Электронный ресурс] URL: <https://www.who.int/publications/m/item/covid-19-epidemiological-update---19-january-2024> (дата обращения: 29.04.2026).
68. COVID-19 is linked to changes in the time–space dimension of human mobility / C. Santana, F. Botta, H. Barbosa [et al.] // *Nature Human Behaviour*. – 2023. – Vol. 7, No. 10. – P. 1729-1739. – DOI 10.1038/s41562-023-01660-3.

69. COVID-19 Pandemic: from Molecular Biology, Pathogenesis, Detection, and Treatment to Global Societal Impact / S. Sood, J. Kumar, V. Aggarwal [et al.] // *Current Pharmacology Reports*. – 2020. – Vol. 6, No. 5. – P. 212-227. – DOI 10.1007/s40495-020-00229-2.

70. Current context: the COVID-19 pandemic and continuing challenges to global health. [Электронный ресурс] URL: <https://www.who.int/about/funding/invest-in-who/investment-case-2.0/challenges> (дата обращения: 29.04.2026).

71. Dastidar, M. G. Public health management during COVID-19 and applications of point-of-care based biomolecular detection approaches / M. G. Dastidar, S. Roy // *Environmental and Health Management of Novel Coronavirus Disease (COVID-19)*, 2021. – P. 345-378. – DOI 10.1016/B978-0-323-85780-2.00009-3.

72. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR / V. M. Corman, T. Bleicker, S. Brünink [et al.] // *Eurosurveillance*. – 2020. – Vol. 25, No. 3. – P. 2000045. – DOI 10.2807/1560-7917.ES.2020.25.3.2000045.

73. Detection of a SARS-CoV-2 variant of concern in South Africa / H. Tegally, E. Wilkinson, M. Giovanetti [et al.] // *Nature*. – 2021. – Vol. 592, No. 7854. – P. 438-443. – DOI 10.1038/s41586-021-03402-9.

74. Differences in the Clinical Course of COVID-19 in Patients Hospitalized in the 2023/2024 and 2024/2025 Seasons / R. Flisiak, D. Zarębska-Michaluk, M. Brzdęk [et al.] // *Journal of Clinical Medicine*. – 2025. – Vol. 14, No. 17. – P. 5992. – DOI 10.3390/jcm14175992.

75. Digital technologies in the public-health response to COVID-19 / J. Budd, B. S. Miller, E. M. Manning [et al.] // *Nature Medicine*. – 2020. – Vol. 26, No. 8. – P. 1183-1192. – DOI 10.1038/s41591-020-1011-4.

76. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia / Q. Li, X. Guan, P. Wu [et al.] // *New England Journal of Medicine*. – 2020. – Vol. 382, No. 13. – P. 1199-1207. – DOI 10.1056/nejmoa2001316.

77. ECDC [Электронный ресурс] URL: <https://www.ecdc.europa.eu/assets/ECDC-highlights-2022/2022-in-brief.html#:~:text=In%202022%2C%20ECDC%20delivered%2089%25,the%20end%20of%20the%20year.> (дата обращения: 24.04.2026).

78. ECDC Digital Roadmap 2022–2027 [Электронный ресурс] URL: [https://www.ecdc.europa.eu/sites/default/files/documents/ECDC-Digital-Roadmap-2022-2027\\_final\\_with\\_covers.pdf](https://www.ecdc.europa.eu/sites/default/files/documents/ECDC-Digital-Roadmap-2022-2027_final_with_covers.pdf) (дата обращения: 29.04.2026).

79. Epitools - Calculate Confidence Limits for a Sample Proportion [Электронный ресурс]. — URL: <https://epitools.ausvet.com.au/ciproportion> (дата обращения: 29.04.2026).

80. Estimation of COVID-19 outbreak size in Italy / A. R. Tuite, V. Ng, E. Rees, D. Fisman // *The Lancet Infectious Diseases*. – 2020. – Vol. 20, No. 5. – P. 537. – DOI 10.1016/s1473-3099(20)30227-9.

81. Evolutionary trajectory of SARS-CoV-2 genome shifts during widespread vaccination and emergence of Omicron variant / K. Gayvert, Sh. Mckay, W. K. Lim [et al.] // *npj Viruses*. – 2023. – Vol. 1, No. 1. – P. 5. – DOI 10.1038/s44298-023-00007-z.

82. Factors Associated with Mortality among Elderly People in the COVID-19 Pandemic (SARS-CoV-2): A Systematic Review and Meta-Analysis / V. P. Alves, F. G. Casemiro, B. G. De. Araujo [et al.] // *International Journal of Environmental Research and Public Health*. – 2021. – Vol. 18, No. 15. – P. 8008. – DOI 10.3390/ijerph18158008.

83. Genomic surveillance to combat COVID-19: challenges and opportunities / Ja. D. Robishaw, S. M. Alter, J. J. Solano [et al.] // *The Lancet Microbe*. – 2021. – Vol. 2, No. 9. – P. e481-e484. – DOI 10.1016/s2666-5247(21)00121-x.

84. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil / N. R. Faria, T. A. Mellan, Ch. Whittaker [et al.] // *Science*. – 2021. – Vol. 372, No. 6544. – P. 815-821. – DOI 10.1126/science.abh2644.

85. Global genomic surveillance strategy for pathogens with pandemic and epidemic potential 2022–2032 [Электронный ресурс]. — URL: <https://www.who.int/initiatives/genomic-surveillance-strategy> (дата обращения: 29.04.2026).

86. Hâncean, M. G. The impact of human mobility networks on the global spread of COVID-19 / M. G. Hâncean, M. Slavinec, M. Perc // *Journal of Complex Networks*. – 2021. – Vol. 8, No. 6. – DOI 10.1093/comnet/cnaa041.

87. Hu, J. The Clinical Characteristics and Risk Factors of Severe COVID-19 / J. Hu, Y. Wang // *Gerontology*. – 2021. – DOI 10.1159/000513400.

88. Impact of Using Health Information Systems (HIS) among Healthcare Professionals during Pandemic COVID-19 / I. D. Abdullah, N. A. Nik Ariffin, A. Mohd Yunus, S. Syamsuriansyah // *Environment-Behaviour Proceedings Journal*. – 2022. – Vol. 7, No. SI10. – P. 185-191. – DOI 10.21834/ebpj.v7isi10.4119.

89. Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021 / F. Campbell, B. Archer, H. Laurenson-Schafer [et al.] // *Eurosurveillance*. – 2021. – Vol. 26, No. 24. – DOI 10.2807/1560-7917.es.2021.26.24.2100509.

90. Kang, Yu. Ju. Characteristics of the COVID-19 Outbreak in Korea From the Mass Infection Perspective / Yu. Ju. Kang // *Journal of Preventive Medicine and Public Health*. – 2020. – Vol. 53, No. 3. – P. 168-170. – DOI 10.3961/jpmph.20.072.

91. Kravchenko, N. A. Spread of the COVID-19 In Russia: Regional Peculiarities / N. A. Kravchenko, A. I. Ivanova // *Regional Research of Russia*. – 2021. – Vol. 11, No. 4. – P. 428-434. – DOI 10.1134/S2079970521040080.

92. Matplotlib: Visualization with Python [Электронный ресурс]. — URL: <https://matplotlib.org> (дата обращения: 29.04.2026).

93. Mortality Risk Among Patients Hospitalized Primarily for COVID-19 During the Omicron and Delta Variant Pandemic Periods — United States, April 2020–June 2022 / S. Adjei, K. Hong, N. A. M. Molinari [et al.] // *MMWR: Morbidity & Mortality Weekly Report*. – 2022. – Vol. 71, No. 37. – P. 1182-1189. – DOI 10.15585/mmwr.mm7137a4.

94. Nextstrain. [Электронный ресурс] URL: <https://nextstrain.org> (дата обращения: 21.04.2026).

95. OurWorldinData.org. [Электронный ресурс] <https://ourworldindata.org/coronavirus> (дата обращения: 21.04.2026).

96. Pandas [Электронный ресурс]. — URL: <https://pandas.pydata.org/> (дата обращения: 29.04.2026).

97. Preliminary estimation of the novel coronavirus disease (COVID-19) cases in Iran: A modelling analysis based on overseas cases and air travel data / Z. Zhuang, Sh. Zhao, Q. Lin [et al.] // *International Journal of Infectious Diseases*. – 2020. – Vol. 94. – P. 29-31. – DOI 10.1016/j.ijid.2020.03.019.

98. Presymptomatic, asymptomatic and post-symptomatic transmission of SARS-CoV-2: joint British Infection Association (BIA), Healthcare Infection

Society (HIS), Infection Prevention Society (IPS) and Royal College of Pathologists (RCPath) guidance / M. A. Mugglestone, N. V. Ratnaraja, A. Bak [et al.] // *BMC Infectious Diseases*. – 2022. – Vol. 22, No. 1. – P. 1-15. – DOI 10.1186/s12879-022-07440-0.

99. Pro-MED. PRO/AH/EDR. Undiagnosed pneumonia – China (HU): Request For Information. Archive Number: 20191230.6864153. 30 Dec. 2019. [Электронный ресурс] URL: <https://promedmail.org/promed-post/?id=6864153> (дата обращения: 28.04.2026).

100. Protective effectiveness of previous SARS-CoV-2 infection and hybrid immunity against the omicron variant and severe disease: a systematic review and meta-regression / N. Bobrovitz, H. Ware, X. Ma [et al.] // *The Lancet Infectious Diseases*. – 2023. – Vol. 23, No. 5. – P. 556-567. – DOI 10.1016/s1473-3099(22)00801-5.

101. Rafflesia U, Rosadi D, Utami H, Sari DP. Robust k-means clustering for modeling the spread of coronavirus disease in Indonesia // U. Rafflesia, D. Rosadi, H. Utami, D.P. Sari. / *AIP Conference Proceedings*. – 2024, Vol. 3024, No. 1. – P. 030008. – DOI 10.1063/5.0204720.

102. Rambaut, A. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology / A. Rambaut, E. C. Holmes, Á. O'toole [et al.] // *Nature Microbiology*. – 2020. – Vol. 5, No. 11. – P. 1403-1407. – DOI 10.1038/s41564-020-0770-5.

103. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa / R. Viana, S. Moyo, D. G. Amoako [et al.] // *Nature*. – 2022. – DOI 10.1038/d41586-021-03832-5.

104. SARS-CoV-2 evolution on a dynamic immune landscape / N. A. Raharinirina, N. Gubela, D. Börnigen [et al.] // *Nature*. – 2025. – Vol. 639, No. 8053. – P. 196-204. – DOI 10.1038/s41586-024-08477-8.

105. SARS-CoV-2 Transmission From People Without COVID-19 Symptoms / M. A. Johansson, T. M. Quandelacy, S. Kada [et al.] // *JAMA Network Open*. – 2021. – Vol. 4, No. 1. – P. e2035057. – DOI 10.1001/jamanetworkopen.2020.35057.

106. Sayers, E. W. Database resources of the national center for biotechnology information / E. W. Sayers, E. E. Bolton, J. R. Brister [et al.] //

Nucleic Acids Research. – 2022. – Vol. 50, No. D1. – P. D20-D26. – DOI 10.1093/nar/gkab1112.

107. SciPy [Электронный ресурс]. — URL: <https://scipy.org/> (дата обращения: 29.04.2026).

108. Statsmodels | Statistical models, hypothesis tests, and data exploration [Электронный ресурс]. — URL: <https://www.statsmodels.org/stable/index.html> (дата обращения: 29.04.2026).

109. Temporal dynamics in viral shedding and transmissibility of COVID-19 / Xi. He, E. H. Y. Lau, P. Wu [et al.] // Nature Medicine. – 2020. – Vol. 26, No. 5. – P. 672-675. – DOI 10.1038/s41591-020-0869-5.

110. The impact of population-wide rapid antigen testing on SARS-CoV-2 prevalence in Slovakia / M. Pavelka, K. Van-Zandvoort, S. Abbott [et al.] // Science. – 2021. – Vol. 372, No. 6542. – P. 635-641. – DOI 10.1126/science.abf9648.

111. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. / A. E. Gorbalenya, S. C. Baker, R. S. Baric [et al.] // Nature Microbiology. – 2020. – Vol. 5, No. 4. – P. 536-544. – DOI 10.1038/s41564-020-0695-z.

112. Tracking of hCov-19 variants. [Электронный ресурс] URL: <https://www.gisaid.org/hcov19-variants/> (дата обращения: 21.04.2026).

113. Transmission risk of asymptomatic SARS-CoV-2 infection: a systematic review and meta-analysis / Ci. Zhang, Ch. Zhou, W. Xu [et al.] // Infectious Medicine. – 2023. – Vol. 2, No. 1. – P. 11-18. – DOI 10.1016/j.imj.2022.12.001.

114. Virological characteristics of the SARS-CoV-2 JN.1 variant / Yu. Kaku, K. Okumura, M. Padilla-Blanco [et al.] // The Lancet Infectious Diseases. – 2024. – Vol. 24, No. 2. – P. e82. – DOI 10.1016/s1473-3099(23)00813-7.

115. Wastewater sequencing reveals early cryptic SARS-CoV-2 variant transmission / S. Karthikeyan, J. I. Levy, P. De Hoff [et al.] // Nature. – 2022. – Vol. 609, No. 7925. – P. 101-108. – DOI 10.1038/s41586-022-05049-6.

116. Weekly epidemiological update on COVID-19 - 21 December 2021. [Электронный ресурс] URL: <https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---21-december-2021> (дата обращения: 29.04.2026).

117. Weekly epidemiological update on COVID-19 - 9 November 2022. [Электронный ресурс] URL: <https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---9-november-2022> (дата обращения: 29.04.2026).

118. WHO Declares COVID-19 a Pandemic / D. Cucinotta, M. Vanelli // *Acta bio-medica : Atenei Parmensis*. – 2020. – Vol. 91, No. 1. – P. 157-160. – DOI 10.23750/abm.v91i1.9397.

119. WHO. Pneumonia of unknown cause – China. [Электронный ресурс] URL: [www.who.int/csr/don/05-january-2020-pneumonia-of-unkown-cause-china/en/](http://www.who.int/csr/don/05-january-2020-pneumonia-of-unkown-cause-china/en/) (дата обращения: 28.04.2026).

120. Zemtsov, S. P. COVID-19: Spatial Dynamics and Diffusion Factors across Russian Regions / S. P. Zemtsov, V. L. Baburin // *Regional Research of Russia*. – 2020. – Vol. 10, No. 3. – P. 273-290. – DOI 10.1134/S2079970520030156.