

Всероссийский конкурс исследовательских и проектных работ школьников
«Высший пилотаж»

Интернет-запросы как индикатор здоровья городского населения

Исследовательская работа

Направление «Развитие государства и общества»

Автор: Хорошун Анна Куватовна,
учащейся 11ИЗкласса,
Лицей НИУ ВШЭ (г. Москва)

2024 г.

Содержание

Введение.....	3
Обзор современных подходов мониторинга состояния здоровья населения.....	4
Отслеживание контактов.....	4
Онлайн-физиологические измерения.....	4
Интернет-запросы	5
Краудсорсинг-наблюдение.....	6
Поиск данных с помощью искусственного интеллекта	6
Цель исследования:	7
Материалы и методы исследования	8
Анализ полученных результатов	9
Анализ интернет-запросов, связанных с заболеваниями	10
Анализ корреляции интернет-запросов, связанных с заболеваниями и соответствующими симптомами.	11
Оценка зависимости интернет-запросов от реальных случаев заболеваемости (на примере COVID-19)	12
Выводы.....	13
Предлагаемые рекомендации.....	14
Список источников.	14
Приложение А	16
Приложение Б.....	17
Приложение В	19
Приложение Г	24
Приложение Д	25
Приложение Е.....	26

Введение

Инфекционные заболевания создают существенные медицинские и социальные проблемы во всем мире. Бактерии и вирусы могут приводить к быстрому распространению какой-либо инфекционной болезни человека в мире (эпидемии), о чем свидетельствует коронавирусная инфекция COVID-19. По официальным данным Всемирной организации здравоохранения (ВОЗ) на 29 сентября 2024 года число умерших от коронавируса в мире составило 70 79 925 человек [1], зарегистрировано 777 126 421 подтвержденных случаев в 229 странах и территориях. В Российской Федерации было зарегистрировано 23 014 969 подтвержденных случаев заболевания COVID-19, из них 400 023 с летальным исходом¹. Прежде значительный вклад в общемировую заболеваемость и смертность внесли и многие другие инфекционные заболевания. Примерами может послужить пандемия свиного гриппа в 2009 году, пандемия, вызванная коронавирусом ближневосточного респираторного синдрома, а также эпидемия вирусной лихорадки Эбола в Западной Африке.

Своевременный контроль за инфекционными заболеваниями необходим для планирования целенаправленных мер профилактики здоровья населения. Традиционно данные об инфекционных заболеваниях собираются органами здравоохранения в виде положительных результатов лабораторных анализов, а также в виде данных о госпитализации и смерти. Однако такой подход требует значительных усилий, а также делает невозможным последовательный сбор данных в национальном или региональном масштабе, например, вследствие задержки в предоставлении отчетности. При этом важно отметить, что госпитализация и смертность охватывают в первую очередь тяжелые случаи и не учитывают заболевания, протекающие без выраженной симптоматики или вовсе бессимптомно. Из-за этих ограничений эти данные могут не отражать истинную частоту заболевания.

В последние годы появились новые источники данных, дополняющие традиционный контроль за инфекционными заболеваниями. Наблюдение с использованием новых типов входных данных может иметь преимущества по сравнению с традиционными источниками данных, поскольку они могут обеспечить своевременность и часто являются менее ресурсозатратными. Особенно показательным было использование новых технологий контроля инфекций в период пандемии COVID-19.

¹ <https://coronavirus-monitor.info/country/russia/>

Обзор современных подходов мониторинга состояния здоровья населения²

Отслеживание контактов

Приложения для смартфонов для цифрового отслеживания контактов стали инновационным подходом, использованным при пандемии COVID-19. Эти приложения предназначены для автоматической идентификации и быстрого уведомления о лицах, которые были в тесном непосредственном контакте с людьми, у которых был положительный результат теста на SARS-CoV-2. Такие приложения были развернуты во многих странах в очень короткие сроки для поддержки или дополнения традиционных систем отслеживания контактов. На острове Уайт было проведено пилотное развертывание приложения Test and Trace [2]. Результаты показали, что эпидемию на острове Уайт удалось быстро и эффективно контролировать после запуска Test and Trace. Позже использование приложений для отслеживания контактов в Швейцарии, Англии и Уэльсе предоставило дополнительные доказательства полезности цифрового отслеживания контактов [3,4]. Исследование швейцарского приложения показало, что цифровое отслеживание контактов может быть эффективным для выявления контактов, с лицами, которые затем получили положительный результат исследования на COVID-19, аналогичным обычному отслеживанию контактов выходными данными.

Онлайн-физиологические измерения

На ранней стадии пандемии COVID-19, инструменты для измерения температуры в режиме реального времени, такие как интеллектуальный термометр, были протестированы в качестве потенциальных инструментов для отслеживания COVID-19. В начале 2020 года данные умных термометров использовались для прогнозирования роста количества случаев (горячих точек) COVID-19 в различных штатах США путем сопоставления аномально высоких уровней гриппа с подтвержденными случаями COVID-19 [5]. Используя данные из геопространственной сети термометров, охватывающей более миллиона пользователей в США, авторы выявляли аномалии, создавая точные прогнозы сезонных гриппоподобных заболеваний (ГПЗ) для конкретных округов с момента, предшествующего потенциальной вспышке и в режиме реального времени сравнивая фактические данные с ожидаемыми. Было показано, что аномалии сильно коррелируют с количеством случаев COVID-19 и могут обеспечить систему раннего предупреждения для обнаружения эпицентров вспышек.

Данные со смарт-часов и трекеров Fitbit применялись также для выявления случаев COVID-19 [6]. Авторы проанализировали физиологические данные и данные об активности

² Для поиска научной литературы использовалась база данных Национальной библиотеки медицины Национального центра Биотехнологической информации (<https://www.ncbi.nlm.nih.gov/>).

32 человек, инфицированных COVID-19, и обнаружили, что у 26 из них (81%) наблюдались изменение частоты сердечных сокращений, количества ежедневных шагов или времени сна. Используя данные умных часов, было показано, что 63% случаев COVID-19 могли быть обнаружены до появления симптомов в режиме реального времени с помощью двухуровневой системы предупреждения, основанной на возникновении значительных повышений частоты сердечных сокращений в покое по сравнению с индивидуальным исходным уровнем.

Эти результаты показывают, что данные, собранные с датчиков, могут постепенно увеличивать точность модели, основанные только на симптомах, для различения людей с симптомами COVID-19 и COVID-19-отрицательных, что может позволить точнее идентифицировать вирус до того, как произойдет дальнейшее распространение. Такая стратегия может дополнять тестирование на вирусы, которое, как правило, представляет собой одноразовый или нечастый анализ проб.

Интернет-запросы

Во время COVID-19 поисковые запросы оказались полезным источником данных для моделей, генерирующих оценки заболеваемости [7]. Идея, которая находится в основе этого подхода, была достаточно проста. Когда люди болеют они часто ищут в Интернете информацию, связанную с их инфекцией. В отличие от моделей ГПЗ, модели COVID-19 нельзя было обучать на существующих данных за многие годы. Тем не менее, они были настолько полезны, что некоторые агентства общественного здравоохранения (например, Общественное здравоохранение Англии) включили их в еженедельные отчеты о эпиднадзоре за COVID-19 [8].

Google сделал оценки объема поиска для любого заданного запроса общедоступными через Google Trends, который стал важным источником данных для исследований в области цифрового эпидемиологического надзора за общественным здоровьем. Например, использование данных из 36 стран показало, что Google Trends полезен для выявления повторяющихся сезонных вспышек ветряной оспы (ветрянки), а также для прогнозирования их масштабов и периодов [9].

В качестве альтернативы данным поисковых запросов Google была предложена статистика использования общедоступных журналов Википедии. Исследование 2014 г. продемонстрировало, что журналы доступа к Википедии можно использовать в качестве входных данных для модели для оценки активности гриппоподобных заболеваний (ГПЗ) [10]. В документе использовались заранее определенные термины для выбора соответствующих статей Википедии, и в качестве варианта использования он

ограничивался гриппом в США. Более поздняя работа в том же году [11] расширила предыдущее исследование, продемонстрировав полезность журналов доступа к Википедии для нескольких заболеваний на нескольких языках, сочетая предварительно определенный выбор терминов с этапом фильтрации на основе модели, который уменьшил количество релевантных статей до 10. Из 14 комбинаций болезни и локализации модель работала хорошо в 8 случаях.

Также в качестве альтернативы Google Trends, можно использовать сервис «Яндекс.Подбор слов» – общедоступную онлайн-систему отслеживания поисковых запросов по неделям в поисковой системе «Яндекс». Например, были проанализированы запросы в «Яндекс» в России в период с 10.08.2020 по 28.11.2021 [12]. Было показано, что поисковые термины, связанные с симптомами, тестированием и лекарствами, тесно коррелируют с регистрируемыми случаями COVID-19 в России. В другой работе [13] было показано, что во время «первого» пика заболеваемости COVID-19 в России (апрель–май 2020 г.) рост количества запросов, связанных с обонянием, опережал рост новых случаев на 1–2 недели в 23 из 59 регионов России.

Краудсорсинг-наблюдение

Краудсорсинг предполагает сбор информации от большой группы людей. С ростом использования Интернета эти методы теперь используют цифровые платформы. Для эпидемиологического надзора за инфекционными заболеваниями исследования часто собирают данные о самосообщениях об инфицировании с помощью приложений или онлайн-опросов.

Онлайн-опросы в целях эпидемиологического надзора начались в 2003 г. с проведения Большого опроса по гриппу (de Grote Griepmeting), в ходе которого еженедельно собирались данные о симптомах гриппа у участников. Это опросы выявили сходство между сезонным гриппом, измеренным в исследовании, и данными, представленными врачами общей практики [14]. Отслеживание гриппа с помощью онлайн-опросов быстро расширилось: системы отслеживания гриппа теперь созданы в США, Португалии, Европе и Австралии. С момента появления COVID-19 многие из этих систем были адаптированы для включения эпидемиологического надзора за COVID-19 [15,16].

Поиск данных с помощью искусственного интеллекта

Сочетание искусственного интеллекта (ИИ) и информации, доступной из неофициальных источников, является еще одним методом эпидемиологического надзора за инфекционными заболеваниями. Эти методы позволяют анализировать и интерпретировать большие наборы данных, полученные из различных источников, таких как онлайн-новости

и СМИ, экспертные блоги, достоверные официальные предупреждения (например, ВОЗ), клинические отчеты, социальные сети и данные авиакомпаний. Наиболее распространенные типы применяемых методов ИИ включают методы, основанные на знаниях (knowledge-based, KB), обработку естественного языка (natural language processing, NLP) и машинное обучение (machine learning, ML).

Глобальная информационная сеть общественного здравоохранения (GPHIN 9) была первым цифровым инструментом, обеспечивающим раннее предупреждение о заболеваниях во всем мире [17] и обеспечила раннее выявление атипичной пневмонии в 2002 г [18]. Система автоматически извлекает данные из медиа-агрегаторов, таких как Google News, других онлайн-новостей, организаций здравоохранения и экспертных блогов, которые анализируются с использованием методов KB, NLP и ML.

ProMED-mail — это система, созданная Международным обществом инфекционных заболеваний в 1994 году для выявления необычных событий, затрагивающих здоровье людей, животных и растений [19, 20]. Система выявила несколько серьезных эпидемий [21]. Ежедневно ProMED-mail выводит в среднем шесть сообщений. HealthMap была одной из первых систем, предупредивших о вспышке COVID-19 за 11 дней до того, как ВОЗ подтвердила вспышку заболевания [22].

казанные способы сбора и анализа данных имеют несколько преимуществ по сравнению с традиционными методами. Ключевым преимуществом является своевременность, поскольку они часто полагаются на автоматический сбор данных. Они также могут охватывать более широкие слои населения, более локализованные районы и группы населения, и, таким образом, данные, собранные с помощью этих систем, могут быть более полными. Эти системы также требуют меньшего участия человека для работы и, следовательно, менее затратные. Когда все аспекты цифровых систем, включая сбор, анализ и представление данных, интегрированы, увеличивается ее гибкость.

Изучив современные подходы мониторинга состояния здоровья населения, я решила исследовать возможность использования интернет-запросов для контроля здоровья населения, как наиболее простой и доступный инструмент для мониторинга.

Цель исследования:

Оценка возможности использования интернет-запросов, связанных с инфекционными заболеваниями для контроля здоровья населения.

Для достижения поставленной цели были поставлены следующие задачи:

1. Выбор поисковых платформ для анализа интернет-запросов.
2. Определение региона исследования, для которого необходимо провести оценку возможности использования интернет-запросов, связанных с инфекционными заболеваниями для контроля здоровья населения.
3. Определение инфекционных заболеваний, об эпидемии которых сообщалось в СМИ. Определение соответствующих симптомов.
4. Создание базы данных поисковых запросов указанных симптомов за выбранный период с интервалом в одну неделю
5. Анализ интернет запросов, связанных с заболеваниями и симптомами.

Актуальность работы связана с разработкой новых подходов к прогнозированию эпидемий и количества людей, которые могут заболеть, с целью планирования объёма необходимой медицинской помощи и принятия мер по профилактике заболевания.

Материалы и методы исследования

1. *Выбор поисковых платформ для анализа интернет-запросов.*

Для анализа данных были выбраны наиболее популярные поисковые платформы по данным Яндекс.Радар (<https://radar.yandex.ru/search>). 65.12% всех поисковых запросов в России обрабатывает Яндекс (каждый шестой пользователь). Google занимает второе место (34.46%), Mail.ru на третьем (0.07%), Rambler на четвертом (0.04%). Соответственно, для анализа интернет-запросов были выбраны сервисы Google Trends и Яндекс Вордстат. Описание указанных сервисов представлено в **Приложении А**.

2. *Определение региона исследования, для которого необходимо провести оценку возможности использования интернет-запросов, связанных с инфекционными заболеваниями для контроля здоровья населения.*

Исходная предпосылка использование интернет-запросов связана с тем, что люди, страдающие заболеванием, будут так или иначе обращаться к онлайн-ресурсам (например, поиску, Википедии и т. д.) об этом конкретном заболевании. Однако для наличия сигнала в поисковых запросах необходимо, чтобы заболевшие люди имели доступ к этим онлайн-ресурсам. Как уже отмечалось, заболевшие люди должны иметь доступ к этим онлайн-ресурсам, тот есть должна быть высокий уровень информатизации. Поэтому для выявления регионов с высоким уровнем информатизации был использован рейтинг регионов России за 2023 год (<https://ria.ru/20231225/itogi-1917517698.html?ysclid=m5s62c4aid43368123>). Рейтинг основан на «социально-экономическом положении, состоянии рынка труда,

качестве жизни, материальном благополучии населения, научно-техническом развитии и приверженности жителей здоровому образу жизни». На основании проведенного анализа был выбран город Москва, который занимает первое место среди регионов. Следует отметить, что по данным Росстата за 2022 год, доля не пользующихся интернетом в Москве — 4,8%. При этом в целом по стране этот показатель составляет 10,9% (<https://3dnews.ru/1099470/rosstat-109-rossiyan-sovsem-ne-polzuyutsya-internetom>).

3. *Определение инфекционных заболеваний, об эпидемии которых сообщалось в СМИ. Определение соответствующих симптомов.*

Для решения данной задачи, был проанализированы новости, связанные с инфекционными заболеваниями за период 2023 по настоящее время. По результатам проведенного анализа для мониторинга интернет запросов были выбраны следующие заболевания и симптомы: корь - температура, сыпь, коклюш - кашель, температура, насморк, грипп - кашель, насморк, температура, краснуха - сыпь, менингит - тошнота (рвота). Дополнительно в качестве контроля были выбраны: ветрянка - сыпь, полиомиелит - понос (диарея), тошнота. Проанализированные новости указаны в **Приложении Б**.

4. Создание базы данных поисковых запросов указанных симптомов за период 01.01.2023 по 05.01.2025 с интервалом в одну неделю.

Созданная база данных доступна по адресу https://drive.google.com/drive/folders/1UG7VYrvhLUlSt7w79y0c9sVnS7CG8_kY?usp=sharing

5. Анализ интернет-запросов, связанных с заболеваниями.

Для сравнения количества запросов по каждому заболеванию и соответствующему симптому построены графики в программе Excel. Для определения взаимосвязи между запросами, связанными с симптомами и заболеванием рассчитывали коэффициент корреляции. Корреляция между переменными более 0,7 считали высокой. Для поиска возможной связи между количеством интернет-запросов и количеством заболевших использовали встроенную функцию Excel: «линия тренда». В качестве меры зависимости заболеваемости от интернет-запросов использовали коэффициент детерминации (r^2). Чем ближе значение коэффициента детерминации к 1, тем сильнее зависимость. В связи с отсутствием точных данных по количеству заболевших по времени (дни, недели) для выбранных заболеваний, были использованы данные по количеству заболевших для COVID-19, полученные с помощью сервиса Yandex Datalens: Коронавирус.

Анализ полученных результатов

Анализ интернет-запросов, связанных с заболеваниями

На рисунке 1 представлены данные по динамике запросов, связанных с заболеваниями в период с 01.01.2023 по 05.01.2025 в гор. Москве.

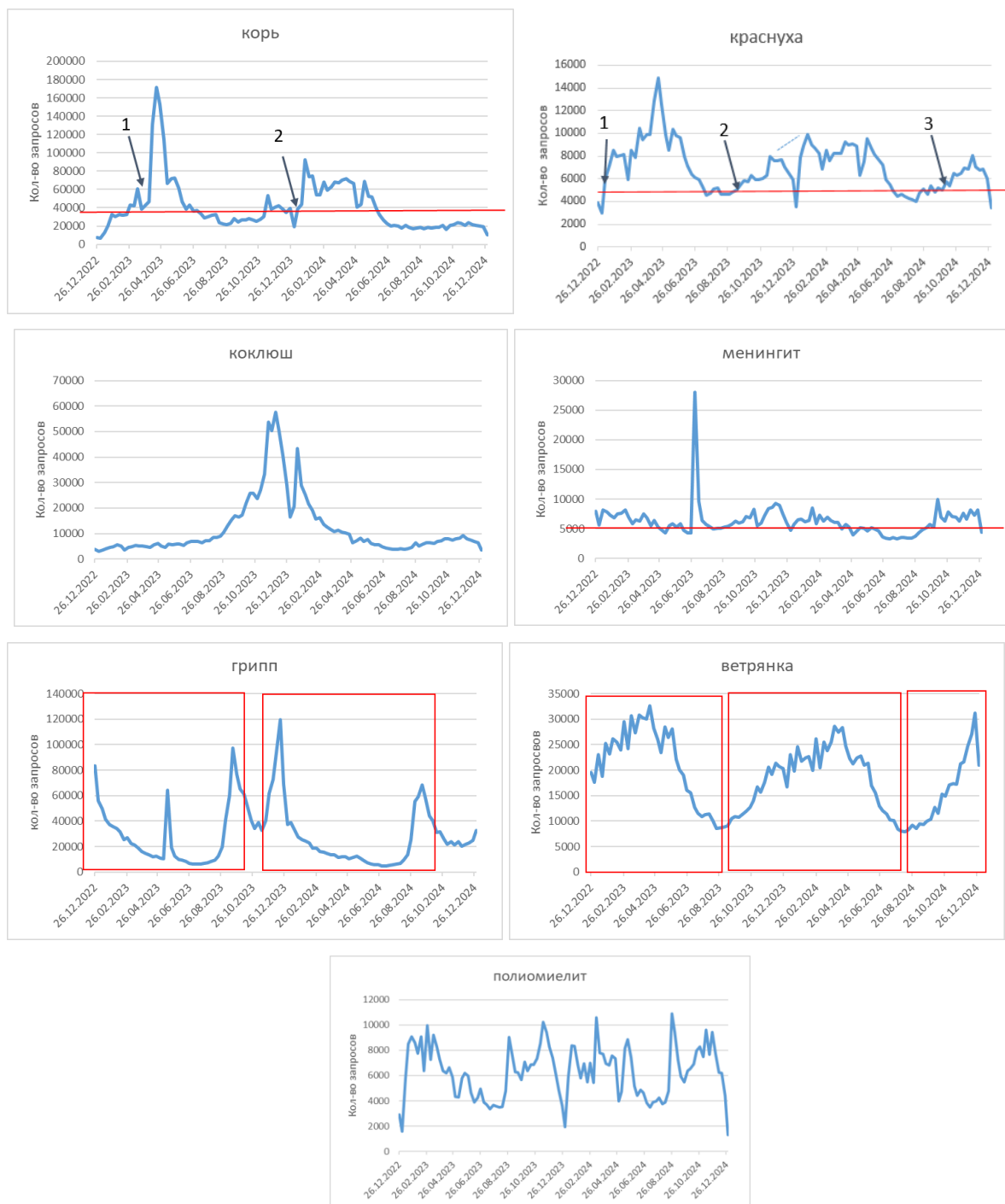


Рисунок 1. Динамика запросов, связанных с заболеваниями в период с 01.01.2023 по 05.01.2025 в гор. Москве.

Из представленных данных видно, что интерес к кори имел два пика с началом подъема 20.03.2023 (стрелка 1) и 08.01.2024 (стрелка 2). При этом известно, что студентов московского университета РГУ нефти и газа отправили на карантин из-за вспышки кори 12.04.2023 (<https://msk1.ru/text/incidents/2023/04/12/72214055/>). О повышении заболеваемости кори в Москве в 2024 году сведений обнаружить не удалось.

Интерес к краснухе имел три пика с начало подъема 09.01.2023 (стрелка 1), 11.09.2023 (стрелка 2) и 14.10.2024 (стрелка 3). При этом известно, «с июня 2024 года в Московской области, в т.ч. в Богородском г.о. и г.о. Электросталь, регистрируются случаи заболевания краснухой преимущественно среди взрослого населения, что свидетельствует об эпидемиологическом неблагополучии, т.к. за последние 5 лет наблюдения данная инфекция не регистрировалась». (<https://bogorodsky-okrug.ru/article/ob-epidemiologicheskoy-situatsii-po-krasnuhe-v-moskovskoj-oblasti-570473>).

Для коклюша интерес имел один пик с максимумом 57 тыс. запросов 04.12.2024 года с подъемом интереса к заболеванию с 21.08.2023 г. Всплеск заболеваемости коклюшом подтверждается публичными данными. Например, «по данным Росстата, опубликованным в статистическом сборнике «Социально-экономическое положение России», с января по октябрь 2023 года зафиксировано 27,8 тыс. случаев коклюша, что стало абсолютным рекордом последних лет» (<https://www.kommersant.ru/doc/6454690>).

Для запроса «менингит» зафиксирован всплеск интереса к заболеванию 03.07.2023 года. При этом 06 июля 2023, СМИ сообщили о смерти двоих заразившихся менингитом в Екатеринбурге (<https://www.vedomosti.ru/society/news/2023/07/06/984048-smi-soobschili-o-smerti-dvoih-zarazivshihsya-meningitom>). То есть интерес к заболеванию обусловлен новостным поводом.

Запросы «грипп» и «ветрянка» имеют периодический характер (выделено красными прямоугольниками) и носят сезонный характер, осень - зима.

Запрос «полиомиелит» имеет несколько пиков, но количество запросов не превышает 10 тыс. Каких-либо информационных поводов для поиска этого заболевания или новостей о вспышках данного заболевания обнаружить не удалось, что может говорить о некоем постоянном интересе пользователей без связи с новостями или заболеванием.

Анализ корреляции интернет-запросов, связанных с заболеваниями и соответствующими симптомами.

В результате проведенной работы были выявлены значимые корреляции между количеством запросов «коклюш» и «кашель», «грипп» и «кашель» в период с 01.01.2023 по 05.01.2025 (коэффициент корреляции $> 0,7$) по данным анализа данных Google trends (Таблица 1).

Таблица 1. Коэффициент корреляции между количеством запросов по заболеванию и соответствующими симптомами по данным Google Trends.

Заболевание/ симптом	Темпера- тура	Сыпь	Кашель	Насморк	Тошнота (рвота)	Понос (диарея)
-------------------------	------------------	------	--------	---------	--------------------	----------------

Корь	-0,05	0,34				
Коклюш	0,50		0,75	0,38		
Грипп	0,60		0,70	0,44		
Краснуха		0,23				
Менингит					-0,01 (0,0)	
Ветрянка		0,23				
Полиомиелит					-0,13 (0,0)	-0,19 (-0,20)

Однако, при анализе данных Yandex.Wordstat были выявлены дополнительно значимые корреляции между количеством запросов «корь» и «сыпь», «грипп» и «насморк» и «краснуха» и «сыпь» в период с 01.01.2023 по 05.01.2025 (коэффициент корреляции > 0,7) (Таблица 2).

Таблица 2. Коэффициент корреляции между количеством запросов по заболеванию и соответствующими симптомами по данным Yandex.Wordstat

Заболевание/ симптом	Температура	Сыпь	Кашель	Насморк	Тошнота (рвота)	Понос (диарея)
Корь	0,06	0,7				
Коклюш	0,62		0,76	0,5		
Грипп	0,57		0,79	0,7		
Краснуха		0,84				
Менингит					0,22 (0,23)	
Ветрянка		0,7				
Полиомиелит					0,42 (0,46)	0,08 (0,28)

Полученные данные позволяет заключить, что интерес к гриппу, коклюшу, краснухе и коре носит не случайный характер, так как пользователи также искали симптомы, которые характерны и для этих заболеваний.

Оценка зависимости интернет-запросов от реальных случаев заболеваемости (на примере COVID-19)

Для оценки зависимости интернет-запросов от реальных случаев заболеваемости (на примере COVID-19) предварительно были построены графики сравнения количества заболевших COVID-19 в Москве и количества запросов по каждому симптому их сравнения построены графики в программе Excel (**Приложение Г**). Для определения взаимосвязи между запросами людей в интернете и реальным количеством заболевших был посчитан коэффициент корреляции (Таблица 3).

Таблица 3. Коэффициент корреляции между количеством запросов и количеством заболевших COVID-19 за период 16.03.2020 – 02.01.2023 для Москвы

	Тошнота	Головная боль	Температура	Обоняние	Кашель
Кол-во заболевших	0.7	0.7	0.5	0.4	0.4

Из полученных данных видно, что коэффициент корреляции 0.7 между количеством запросов и количеством заболевших получен только для запросов «Тошнота» и «Головная боль». Далее были построены диаграммы рассеяния, построены линии тренда и определен коэффициент детерминации (**Приложение Д**). Коэффициент детерминации $R^2 = 0,67$ был получен только для функции ($y = 0,0003x^2 - 9,782x + 87566$) описывающей зависимость между количеством запросов «головная боль» и количеством заболевших COVID-19.

В связи с тем, что было несколько периодов подъемов заболевания COVID-19, связанных с разными штаммами этого вируса и по-разному проявлялись симптомы в разные периоды, был выбран более короткий период с 16.03.2020 по 15.03.2021, поскольку визуально видно и подтверждено расчётами коэффициента корреляции (Таблица 4), что количество запросов и заболевания имеют корреляцию.

Таблица 4. Коэффициент корреляции между количеством запросов и количеством заболевших COVID-19 за период 16.03.2020 по 15.03.2021.

	Тошнота	Головная боль	Температура	Обоняние	Кашель
Кол-во заболевших	0.81	0.76	0.7	0.96	0.5

Далее были построены диаграммы рассеяния, построены линии тренда и определен коэффициент детерминации для запросов, которые имели коэффициент корреляции с реальной заболеваемость COVID-19 более 0.7 (**Приложение Е**). Высокий коэффициент детерминации $R^2 = 0,92$ был получен для запроса «обоняние» и количеством заболевших COVID-19, которая описывается уравнением $y = 0,8472x - 3050,9$ и для запроса «тошнота» и количеством заболевших COVID-19, которая описывается уравнением $y = 6,31932x - 79669$ ($R^2 = 0,6607$) в период 16.03.2020 по 15.03.2021.

Выводы

1. Рост интернет-запросов «коклюш», «краснуха», «корь», «грипп», «ветрянка» связан с распространением этих заболеваний среди населения города Москва.

2. Интернет –запросы, связанные с симптомами заболеваний, коррелируют с поиском информации по соответствующему заболеванию, следовательно, интерес к гриппу, ветрянке коклюшу, краснухе и коре носит не случайный характер.
3. Между количеством интернет-запросов, связанных с поиском определенных симптомов заболевания и реальными случаями заражения COVID-19, существует зависимость, которая может быть выражена в форме уравнения. Например, в период с 16.03.2020 по 15.03.2021 связь количества заболевших COVID-19 количества запросов «обоняние» описывается уравнением $y=0,8472x - 3050,9$, а количество заболевших COVID-19 связано с количеством запросов «тошнота» и описывается уравнением $y=6,31932x - 79669$.

Предлагаемые рекомендации

1. Данные интернет-запросов могут быть использованы для прогноза вспышек инфекционных заболеваний.
2. Данные интернет-запросов могут быть использованы для расчета количества людей, которые могут заболеть, с целью планирования объёма необходимой медицинской помощи и принятия мер по профилактике заболевания. Необходимым условием является наличие записей по количеству заболевших людей день за днем.
3. Необходимо разработать алгоритмы, которые позволят корректно определить онлайн-активность, вызванную реальным заболеванием на местах.
4. Предлагается активно использовать поисковые интернет-системы для синдромного наблюдения, которая использует ежедневную статистику частоты поисковых запросов, в системе здравоохранения.

Список источников.

1. <https://data.who.int/dashboards/covid19/cases>
2. Kendall M, Milsom L, Abeler-Dörner L, et al. Epidemiological changes on the Isle of Wight after the launch of the NHS Test and Trace programme: a preliminary analysis. *Lancet Digit Health*. 2020;2(12):e658-e666. doi:10.1016/S2589-7500(20)30241-7
3. Wymant C, Ferretti L, Tsallis D, et al. The epidemiological impact of the NHS COVID-19 app. *Nature*. 2021;594(7863):408-412. doi:10.1038/s41586-021-03606-z
4. Salathé M, Althaus C, Anderegg N, et al. Early evidence of effectiveness of digital contact tracing for SARS-CoV-2 in Switzerland. *Swiss Med Wkly*. 2020;150:w20457. Published 2020 Dec 16. doi:10.4414/smw.2020.20457
5. Chamberlain SD, Singh I, Ariza CA, Daitch AL, Philips PB, Dalziel BD. Real-time detection of COVID-19 epicenters within the United States using a network of smart thermometers. *MedRxiv* (2020). Available from: 10.1101/2020.04.06.20039909

6. Mishra T, Wang M, Metwally AA, Bogu GK, Brooks AW, Bahmani A, et al. Pre-symptomatic detection of COVID-19 from smartwatch data. *Nat Biomed Eng.* (2020) 4:1208–20. 10.1038/s41551-020-00640-6
7. Lamos, V., Majumder, M.S., Yom-Tov, E. et al. Tracking COVID-19 using online search. *npj Digit. Med.* 4, 17 (2021). <https://doi.org/10.1038/s41746-021-00384-w>
8. <https://www.gov.uk/government/publications/sources-of-surveillance-data-for-influenza-covid-19-and-other-respiratory-viruses/sources-of-surveillance-data-for-influenza-covid-19-and-other-respiratory-viruse>
9. Bakker, Kevin M, Micaela Elvira Martinez-Bakker, Barbara Helm, and Tyler J Stevenson. 2016. “Digital Epidemiology Reveals Global Childhood Disease Seasonality and the Effects of Immunization.” *Proceedings of the National Academy of Sciences* 113 (24): 6689–94.
10. McIver, David J, and John S Brownstein. 2014. “Wikipedia Usage Estimates Prevalence of Influenza-Like Illness in the United States in Near Real-Time.” *PLoS Computational Biology* 10 (4): e1003581
11. Generous, Nicholas, Geoffrey Fairchild, Alina Deshpande, Sara Y Del Valle, and Reid Priedhorsky. 2014. “Global Disease Monitoring and Forecasting with Wikipedia.” *PLoS Computational Biology* 10 (11): e1003892.
12. Хорошун Д.К., Момыналиев К.Т., Воронин Е.М., Акимкин В.Г. Анализ поисковых запросов в «Яндексе», связанных с COVID-19 в Российской Федерации. *Медицинский алфавит.* 2022;(14):14-22. <https://doi.org/10.33667/2078-5631-2022-18-14-22>
13. Momynaliev K, Khoroshun D, Akimkin V. Online queries as a criterion for evaluating epidemiological status and effectiveness of COVID-19 control measures in Russia: results from Yandex. Wordstat analysis. *BMJ Open* 2022;12:e056716. doi:10.1136/ bmjopen-2021-056716
14. Koppeschaar CE, Colizza V, Guerri C, Turbelin C, Duggan J, Edmunds WJ, et al. Influenzanet: citizens among 10 countries collaborating to monitor influenza in Europe. *JMIR Public Health Surveill.* (2017) 3:e7429. 10.2196/publichealth.7429
15. Mccoll K, Debin M, Souty C, Guerri C, Turbelin C, Falchi A, et al. Are people optimistically biased about the risk of COVID-19 infection? Lessons from the first wave of the pandemic in Europe. *Int J Environ Res Public Health.* (2021) 19:436. 10.3390/ijerph19010436
16. Sullivan SG, Carlson S, Cheng AC, Chilver MB, Dwyer DE, Irwin M, et al. Where has all the influenza gone? The impact of COVID-19 on the circulation of influenza, other respiratory viruses, Australia, March to September 2020. *Eurosurveillance.* (2020) 25:2001847. 10.2807/1560-7917.ES.2020.25.47.2001847
17. Dion M, AbdelMalik P, Mawudeku A. Big data: big data and the global public health intelligence network (GPHIN). *Can Commun Dis Rep.* (2015) 41:209. 10.14745/ccdr.v41i09a02
18. Carter D, Stojanovic M, Hachey P, et al. Global Public Health Surveillance Using Media Reports: Redesigning GPHIN. *Stud Health Technol Inform.* 2020;270:843-847. doi:10.3233/SHTI200280
19. Yu VL, Madoff LC. ProMED-mail: an early warning system for emerging diseases. *Clin Infect Dis* 2004; 39: 227–232
20. Carrion M, Madoff LC. ProMED-mail: 22 years of digital surveillance of emerging infectious diseases. *Int Health* 2017; 9: 177–183.
21. Madoff L, Brownstein J. ProMED and HealthMap: Collaboration to improve emerging disease surveillance. *Int J Infect Dis* 2010; 14: e184
22. Cho A. AI systems aim to sniff out coronavirus outbreaks. *Science.* 2020;368(6493):810-811. doi:10.1126/science.368.6493.810

Краткое описание поисковых систем.

Google Trends представляет данные в относительном количестве. Сервис делит частоту запроса на общее число поисковых запросов, при этом учитывает конкретный регион и выбранный период. Исходя из этих данных, система оценивает уровень интереса по шкале от 0 до 100 баллов. Оценку 100 получает самый популярный за этот период запрос или тема — совокупность похожих запросов. Остальные распределяются в зависимости от отношения к этому числу. Оценка 0 означает, что запрос не популярен. Пример данных Google Trends представлен на рис.1.

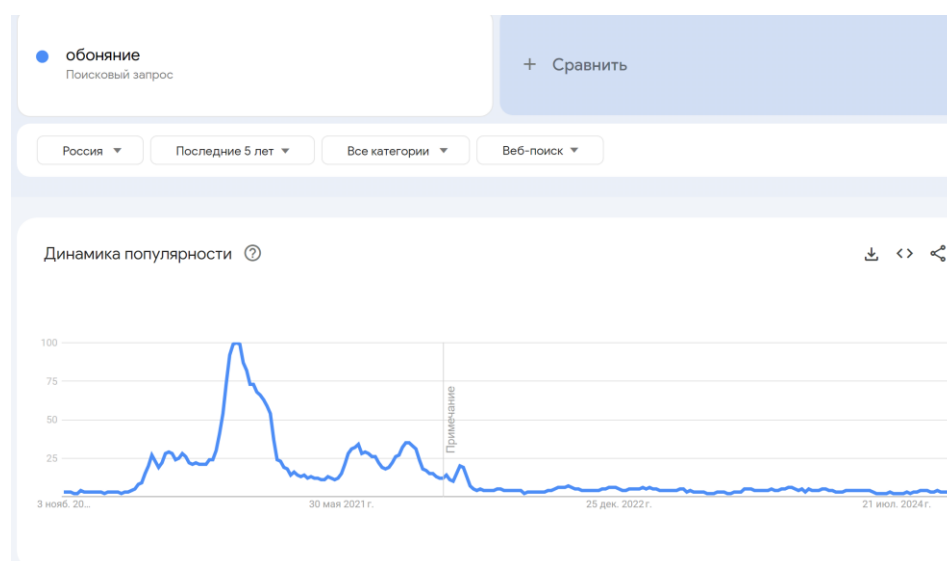


Рисунок 1. Пример данных Google Trends по запросу «обоняние»

Сервис Яндекс Вордстат позволяет получить абсолютное количество запросов по регионам (рис. 2).

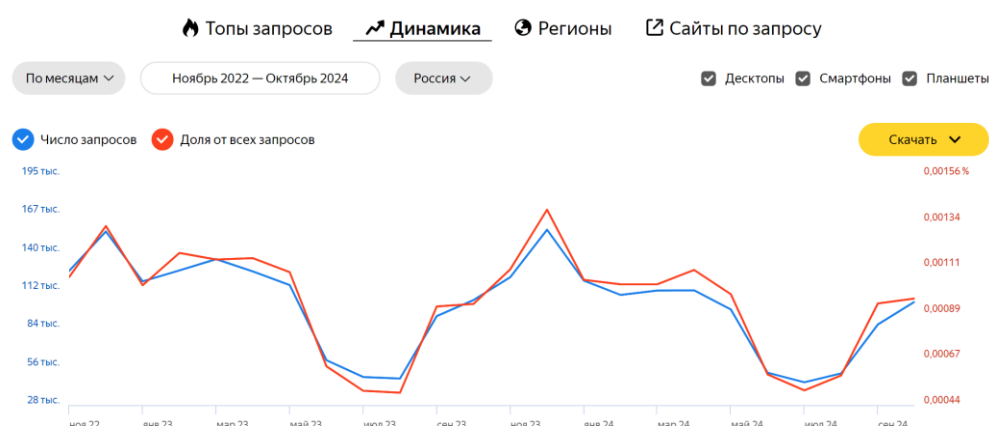


Рисунок 2. Пример данных Яндекс Вордстат по запросу «обоняние»

Список сообщений, которые были проанализированы для выбора заболеваний и симптом.

«За 10 месяцев 2023 г. в Москве зарегистрировано 5065 случаев **коклюша**, показатель заболеваемости составил 40,0 на 100 тыс. населения, что уже превышает годовой показатель 2022 г. в 5,3 раза» (<https://events.niioz.ru/event/vebinar-kokliush-u-detei-i-vzroslykh-sovremennye-problemy>).

«Рост заболеваемости **коклюшем** в России связан с неконтролируемыми потоками мигрантов, в особенности из стран ближнего зарубежья, и возросшей пропагандой от антипрививочников, пишет «Коммерсантъ» со ссылкой на результаты опроса о ситуации с коклюшем, который в декабре 2023 года провел профессиональный сервис «Справочник врача» среди 1316 респондентов. Такие ответы выбрали 30,2% и 34,5% врачей соответственно. Среди других причин — недостаточная осведомленность населения о важности вакцинации (12,8%)» (<https://www.forbes.ru/society/504399-vraci-ob-asnili-rost-zabolevaemosti-koklusem-pritokom-migrantov-i-antiprivivocnikami>).

«В 2023 году Россия столкнулась с заметным всплеском заболеваемости **коклюшем**, вызывая серьезные тревоги в медицинском сообществе. Вступив в 2023 год, эпидемия коклюша в России стала предметом внимания медицинского сообщества. Это обусловлено тем, что резкое увеличение заболеваемости в текущем году значительно превысило уровень предыдущих лет. При взгляде на статистику за весь 2022 год, где было зарегистрировано всего 3,1 тыс. случаев коклюша, в первые десять месяцев 2023 года этот показатель резко возрос, достигнув 27,8 тыс. случаев». ([Резкий рост заболеваемости коклюшем в России - ЦРБ БОРОВСКОГО РАЙОНА](#)).

«Ногинский территориальный отдел Управления Роспотребнадзора по Московской области сообщает об ухудшении эпидемиологической ситуации по заболеваемости **краснухой** в Московском регионе. С июня 2024 года в Московской области, в т.ч. в Богородском г.о. и г.о. Электросталь, регистрируются случаи заболевания краснухой преимущественно среди взрослого населения, что свидетельствует об эпидемиологическом неблагополучии, т.к. за последние 5 лет наблюдения данная инфекция не регистрировалась». (<https://bogorodsky-okrug.ru/article/ob-epidemiologicheskoy-situatsii-po-krasnuhe-v-moskovskoj-oblasti-570473>).

«Депздрав: в Москве ведется системная работа по вакцинации от кори. В СМИ распространяется информация о росте заболеваемости кори. В Москве мы действительно фиксируем рост спорадической заболеваемости и единичные очаги **кори** среди

непривитых, но в целом ситуация не является критической и находится на контроле как Департамента здравоохранения, так и Управления Роспотребнадзора по Москве. Практически все заболевающие – это не привитые против кори граждане, что и служит причиной заболевания». (<https://mosgorzdrav.ru/ru-RU/news/default/card/6990.html>).

«Но в этом году внимание к себе привлекло ведомство, о сводках которого с пугающими цифрами мы уже начали забывать. Речь идет о Роспотребнадзоре. Из его данных за девять месяцев года (медицинская статистика собирается небыстро и выходит с опозданием) следует, что заболеваемость **корью** в России в 2023 году выросла в 288 раз. В процентах выглядит еще более впечатляюще — прирост составил 28 732%. Несколько успокаивают разве что абсолютные числа — болезнь нашли у 8073 человек». (<https://www.rbc.ru/life/news/65859dd39a79473521852df4>).

«За три летних месяца 2023 года **менингит** диагностировали более чем у 500 жителей России, выяснили «Известия», собрав данные региональных органов здравоохранения. Большинство случаев — 345 — приходится на Саратовскую область. Заболеваемость там выросла на порядок по сравнению с прошлым годом. Эти цифры можно объяснить как улучшением систем диагностики, так и ослаблением иммунитета у населения после COVID-19, полагают эксперты. В Роспотребнадзоре ситуацию назвали ожидаемой». (<https://iz.ru/1567327/nataliia-bykadorova/massovoe-proiavlenie-v-rossii-za-letu-vyivili-bolee-500-sluchaev-meningita>).

«За прошедшее лето более чем у 500 человек выявили **менингит** энтеровирусной и бактериальной природы. Об том сообщает газета "Известия" со ссылкой на данные региональных органов здравоохранения. Так, большая часть зафиксированных случаев заболевания припала на Саратовскую область, где за первые полгода менингитом заболели 80 граждан, что в 13 раз больше, чем за аналогичный период прошлого года. Летом в регионе было зафиксировано 245 новых случаев заражения заболеванием». (https://www.m24.ru/news/obshchestvo/01092023/614532?utm_source=CopyBuf).

«Заболеваемость **гриппом** и ОРВИ превысила эпидемиологические пороги на 20% в 35 городах России, заявил главный внештатный специалист Минздрава России по инфекционным болезням Владимир Чуланов, сообщили РБК в пресс-службе Минздрава. Превышение эпидемиологических порогов означает начало эпидемии. Согласно разработанной Роспотребнадзором методике вычисления эпидемических порогов используются статистические сведения о суммарной заболеваемости гриппом и ОРВИ за календарные недели каждого года в течение последних 5–10 лет». (<https://www.rbc.ru/rbcfreenews/6514197c9a79473959309467>).

Графики сравнение количества запросов по каждому заболеванию и соответствующему симптому

Рисунок 1. Сравнение динамики запросов «коклюш» и симптомов по данным Google trends (GT) и Yandex.Wordstat (YW). По оси ординат на графиках GT (слева) указано относительное количество запросов, по оси ординат на графиках YW (справа) указано абсолютное количество запросов.

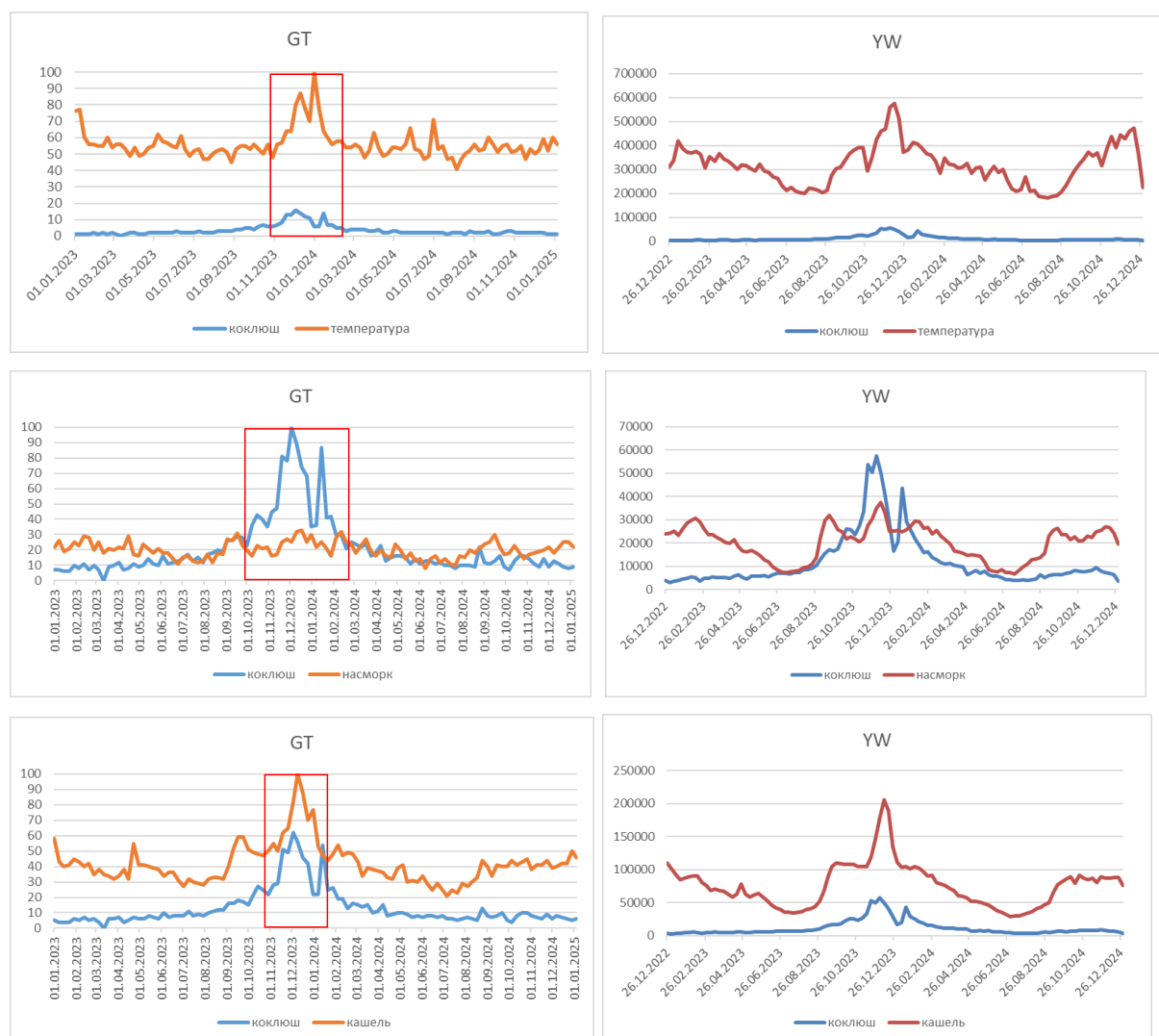


Рисунок 2. Сравнение динамики запросов «корь» и симптомов по данным Google trends (GT) Yandex.Wordstat (YW). По оси ординат на графиках GT (слева) указано относительное количество запросов, по оси ординат на графиках YW (справа) указано абсолютное количество запросов.

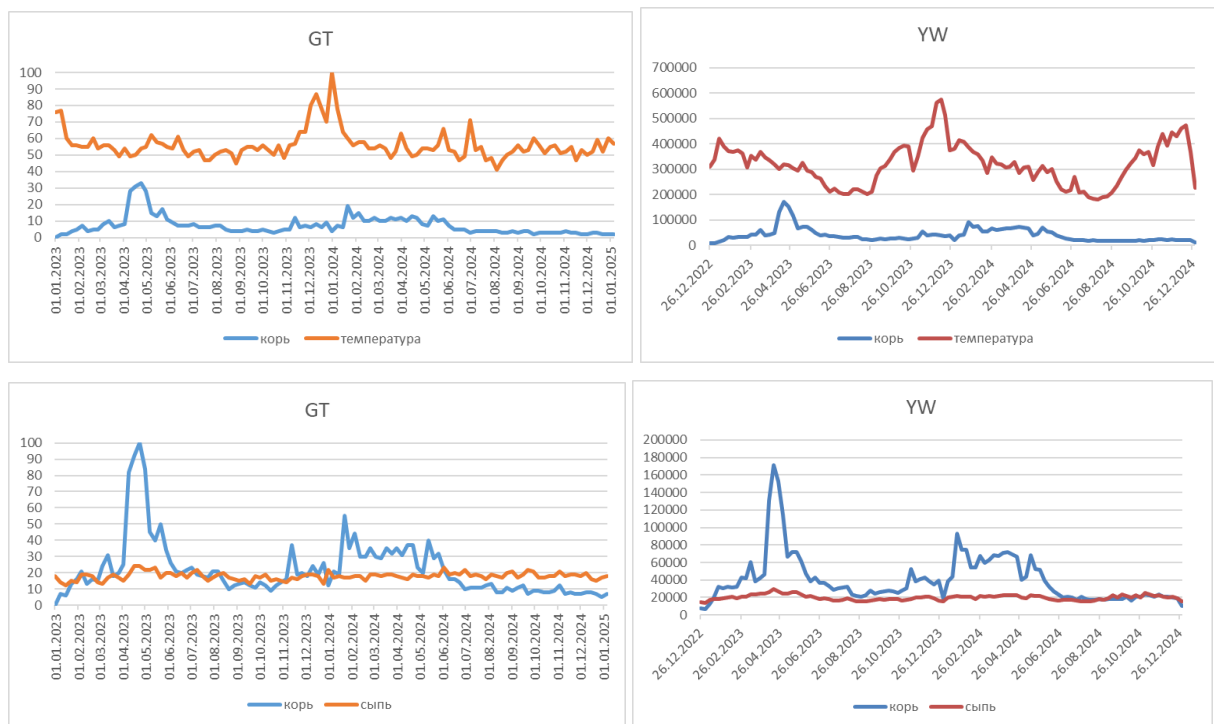


Рисунок 3. Сравнение динамики запросов «грипп» и симптомов по данным Google trends (GT) Yandex.Wordstat (YW). По оси ординат на графиках GT (слева) указано относительное количество запросов, по оси ординат на графиках YW (справа) указано абсолютное количество запросов.

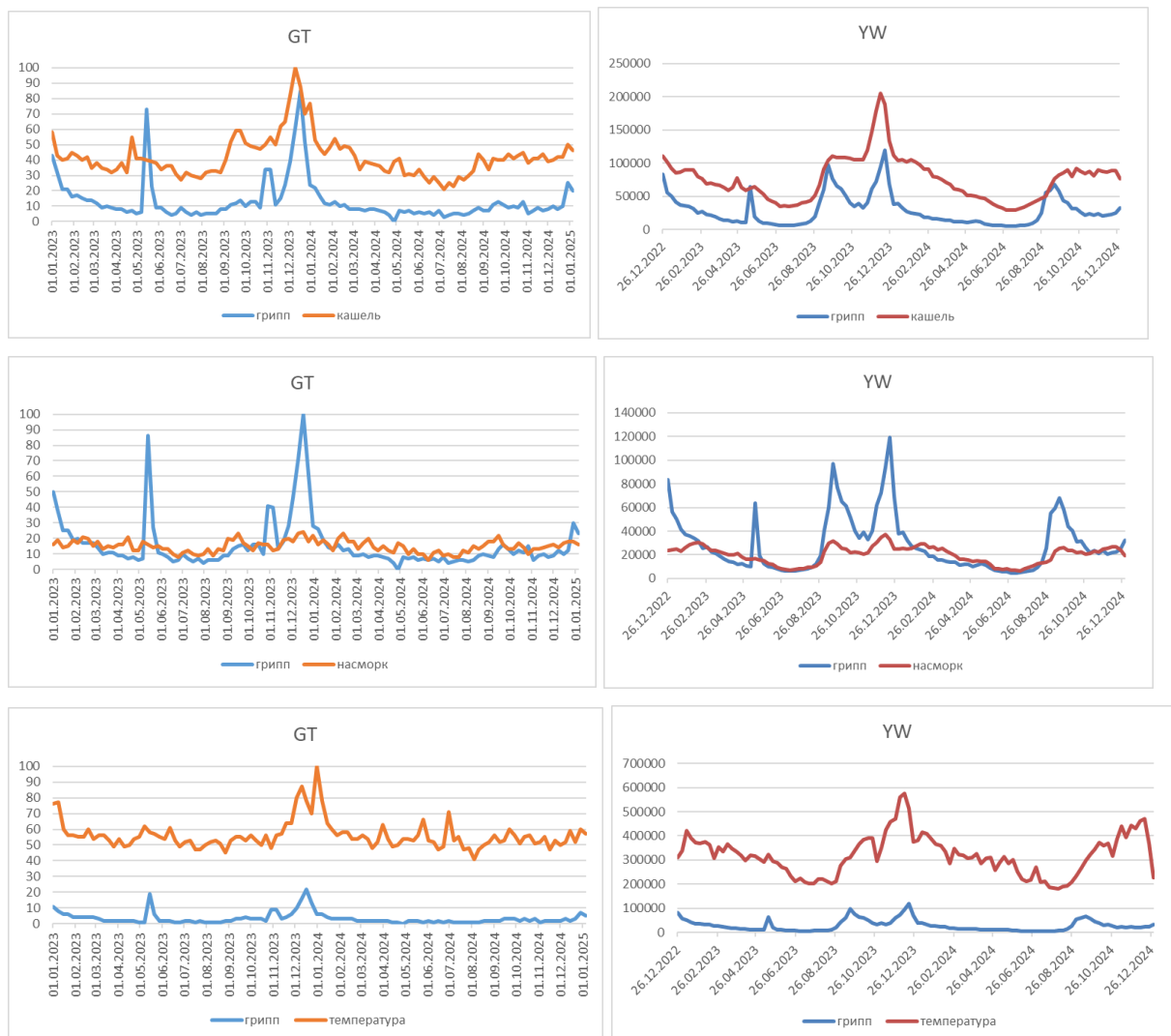


Рисунок 4. Сравнение динамики запросов «краснуха» и симптомов по данным Google trends (GT) Yandex.Wordstat (YW). По оси ординат на графиках GT (слева) указано относительное количество запросов, по оси ординат на графиках YW (справа) указано абсолютное количество запросов.

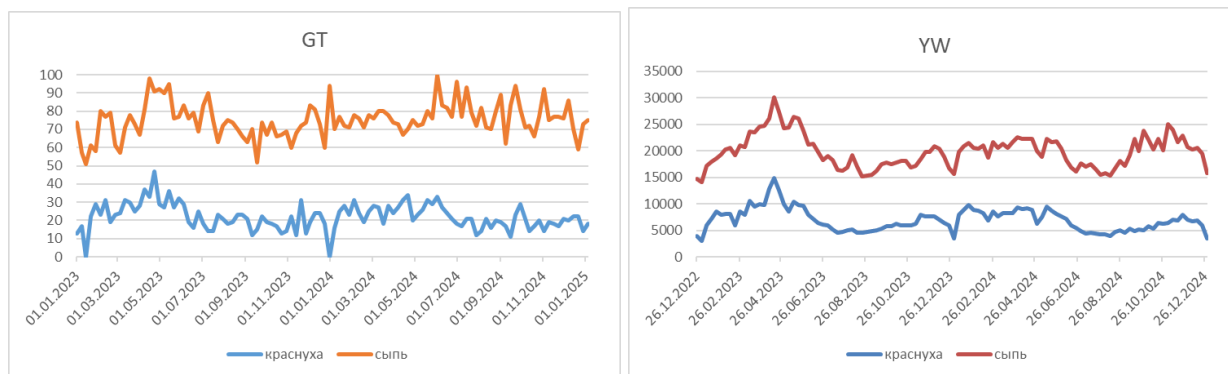


Рисунок 5. Сравнение динамики запросов «менингит» и симптомов по данным Google trends (GT) Yandex.Wordstat (YW). По оси ординат на графиках GT (слева) указано относительное количество запросов, по оси ординат на графиках YW (справа) указано абсолютное количество запросов.

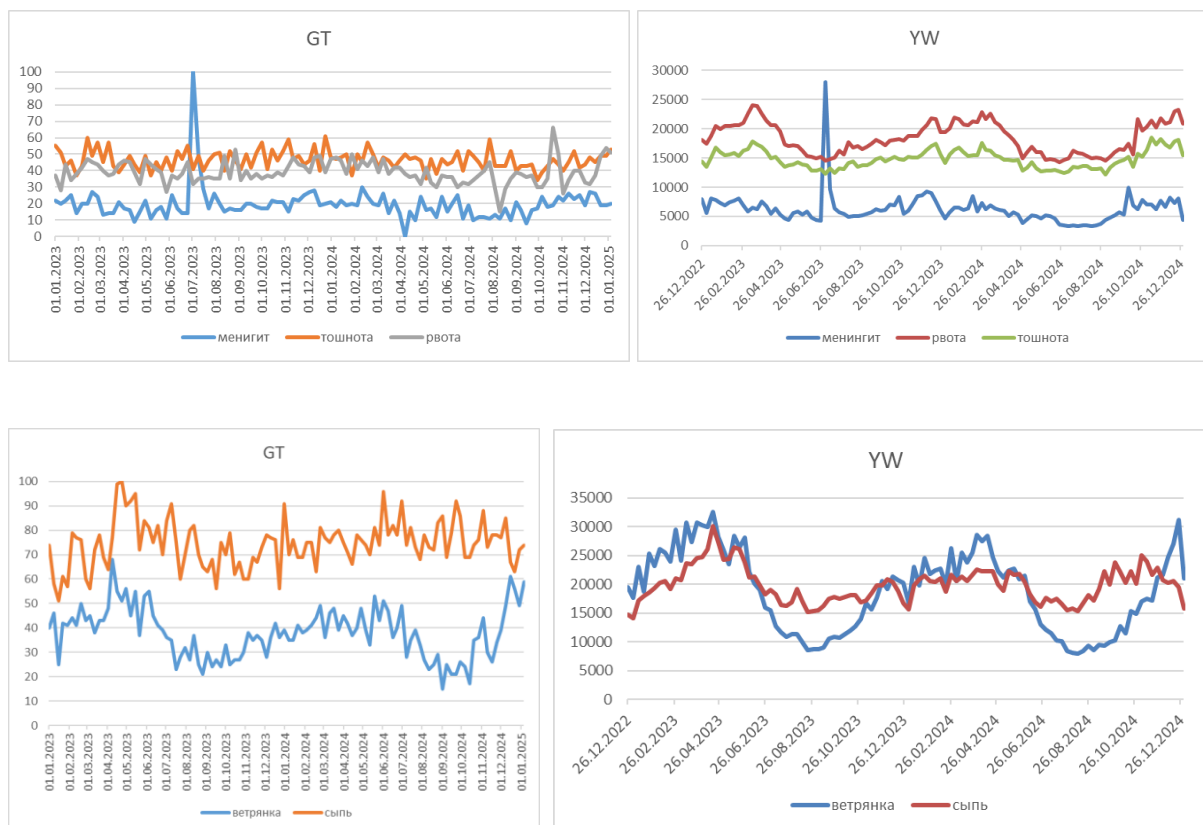


Рисунок 6. Сравнение динамики запросов «ветрянка» и симптомов по данным Google trends (GT) Yandex.Wordstat (YW). По оси ординат на графиках GT (слева) указано относительное количество запросов, по оси ординат на графиках YW (справа) указано абсолютное количество запросов.

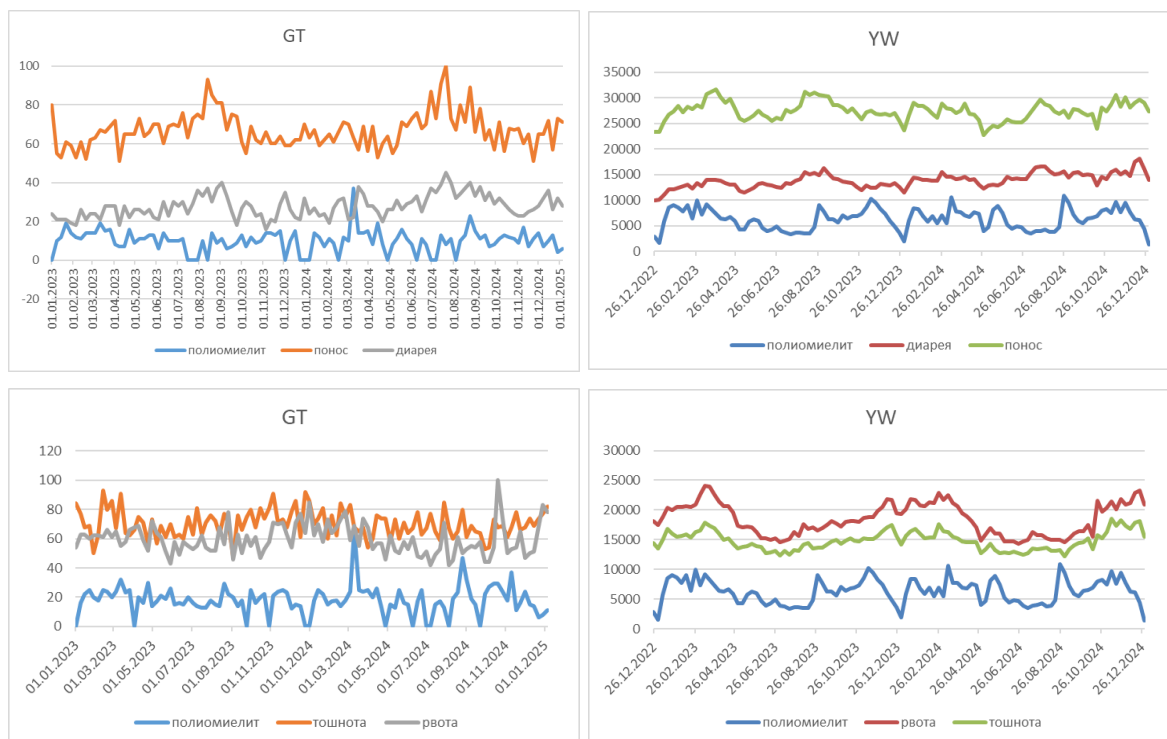
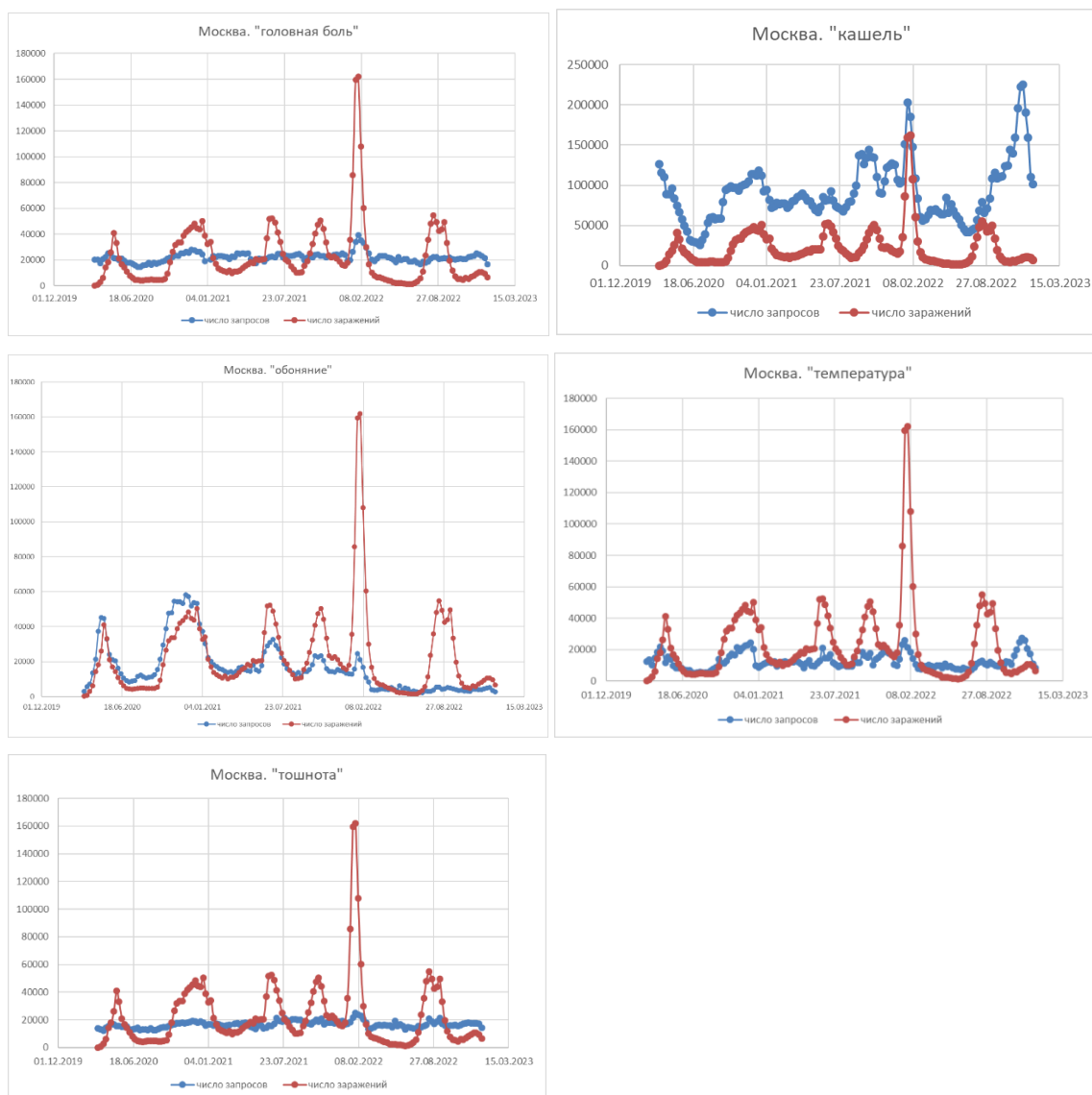
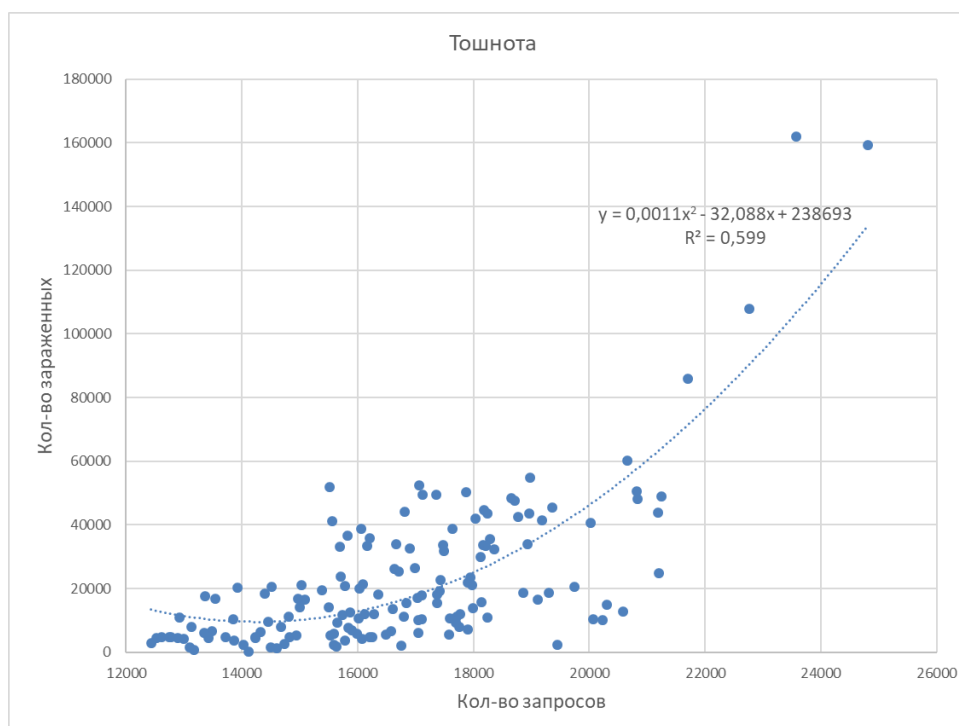
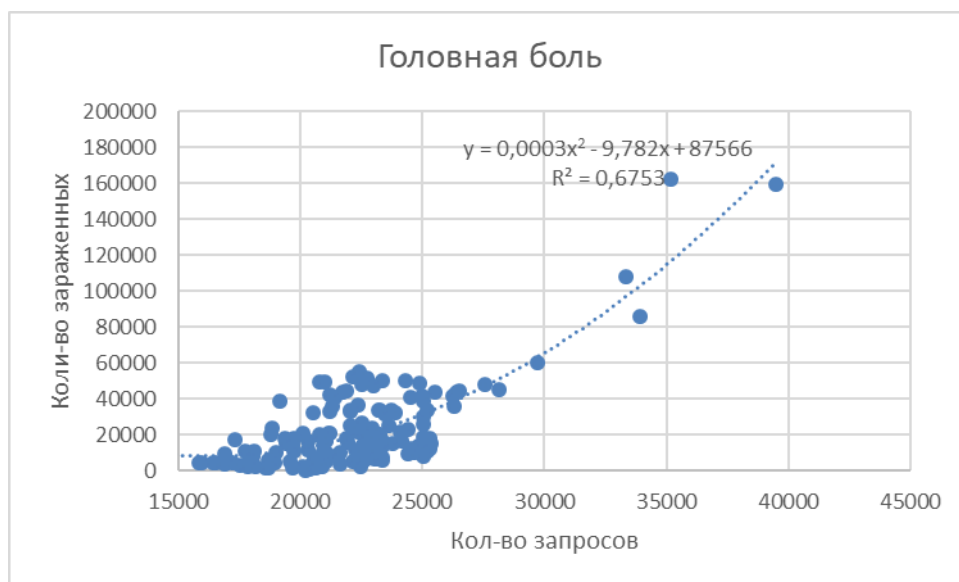


Рисунок 1. Сравнение динамики запросов «менингит» и симптомов по данным Google trends (GT) Yandex.Wordstat (YW). По оси ординат на графиках GT (слева) указано относительное количество запросов, по оси ординат на графиках YW (справа) указано абсолютное количество запросов.



Диаграммы рассеяния между количеством запросов «головная боль», «тошнота» (по данным Yandex.Wordstat) и реальными случаями заболеваемости COVID-19 в период 16.03.2020 – 02.01.2023



Диаграммы рассеяния между количеством запросов «обоняние», «головная боль», «тошнота», «температура» (по данным Yandex.Wordstat) и реальными случаями заболеваемости COVID-19 в период 16.03.2020 по 15.03.2021

