11-5-2025

Google Search Trends and Their Correlation with COVID-19 Cases and Vaccination Rates in Austria

Stefanie Wendell
Walden University

Walden University

College of Health Sciences and Public Policy

This is to certify that the doctoral dissertation by

Stefanie Wendell

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. David Segal, Committee Chairperson, Public Health Faculty
Dr. Chinaro Kennedy, Committee Member, Public Health Faculty

Chief Academic Officer and Provost
Sue Subocz, Ph.D.

Walden University
2025

Abstract

Google Search Trends and Their Correlation with COVID-19 Cases and Vaccination Rates in Austria

by

Stefanie Wendell

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Public Health Epidemiology

Walden University

November 2025

Abstract

Timely digital surveillance is essential because clinical reporting often lags. This quantitative, longitudinal correlational study tested whether weekly Google searches in Austria signal changes in COVID-19 cases and vaccination uptake, and short-term case-vaccination dynamics, guided by Rogers' Diffusion of Innovations Theory. Weekly Google Trends indices (0-100) and nationally reported surveillance totals (weekly counts of new cases and vaccine doses) were aggregated; cases: $N = 175$ weeks (Feb 2020–Jun 2023); vaccinations: $N = 132$ weeks (Dec 2020–Jun 2023). Time series correlations and lagged growth models showed that symptom-related searches consistently rose 1 week before reported cases (Δlog cases_t on Δlog searches_{t−1}: $\beta = 0.72$–1.33 across symptom/general/testing/fever, all $p \leq .015$; $R^2 = .16$–$.42$); brand-specific vaccine searches (Pfizer, Moderna, AstraZeneca) preceded increases in vaccinations by ~1 week ($\beta = 0.155$–0.239, $p = .003 \leq .001$; $R^2 \approx .06$)**;** and booster shot searches preceded vaccination activity by ~4 weeks ($\beta = 0.130$, $p = .003$; $R^2 = .038$). Cases and vaccinations showed no meaningful same-week association (Spearman's $\rho = -0.033$, $p = .709$)**,** and short-lag analyses provided little evidence that growth in cases drove growth in vaccinations (Δlog vaccinations_t on Δlog cases_{t−k}, k = 0–2: $\beta = 0.10$–0.16, all $p \geq .096$; $R^2 \leq .021$). Influenza-related search models were small or null, supporting COVID-19 specificity (cases at +1 wk: $\beta = 0.115$, $p = .001$, $R^2 = .043$; vaccinations at lags 0–2: ns). The positive social change implications from these findings may help support the use of Google Trends as a complementary, short-term, early-awareness surveillance approach for timelier targeted outreach by the Austrian public health authorities.

Google Search Trends and Their Correlation with COVID-19 Cases and Vaccination Rates in Austria

by

Stefanie Wendell

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Public Health Epidemiology

Walden University

November 2025

Dedication

I dedicate this dissertation to all those who have lost loved ones to diseases that could have been prevented. I hope this work will be a small step toward a future where better data and faster decisions save more lives.

Acknowledgments

I would like to thank my committee chair, Dr. Segal, for his steady guidance and encouragement throughout this process. I would also like to thank Dr. Kennedy for her thoughtful feedback and support. Their expertise helped shape this dissertation into what it is today.

To my mum, Gabi, thank you for raising me in an environment where anything and everything felt possible. You taught me to be curious, independent, and resilient, and those lessons carried me through every stage of this journey. Your strength and unwavering belief in me have always been my strongest foundation.

To my husband, Bryan, thank you for being my sounding board, my voice of reason, and my biggest cheerleader. You've spent countless hours listening to me, thinking through ideas with me, and reminding me that I can do it - even when I wasn't so sure myself. I'm incredibly grateful for your love, patience, and perspective.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction to the Study

Introduction

Timely monitoring of COVID-19 trends remains critical for public health decision-making, particularly as public health agencies worldwide seek real-time data to guide interventions. Recent research has shown that online searches captured through Google Trends can reflect population health concerns, with spikes in search terms often matching subsequent increases in reported cases (Effenberger et al., 2020; Porcu et al., 2023). By exploring patterns in search volume for terms related to COVID-19 symptoms and vaccination, public health professionals can receive early warnings of upcoming spikes or fluctuations in vaccine uptake (Becerra-García et al., 2023). In addition, understanding how case numbers and vaccination behavior interact over time may further inform intervention strategies. To enhance the accuracy of these analyses, contextual variables such as influenza-related search trends are also examined to identify and control for potential confounding periods. Investigating these relationships in Austria can refine these approaches, potentially increasing the speed and precision of public health interventions.

Despite promising global findings, a notable gap remains in research specifically focused on Austria, where the interaction between digital search behavior and real-time epidemiological metrics is still understudied. Porcu et al. (Porcu et al., 2023) found that several international or regional studies highlight a strong correlation between Google search behavior and COVID-19 indicators; however, no research has examined how these patterns translate to Austria. In particular, most previous research either focuses solely on case numbers or fails to comprehensively consider vaccination uptake (Effenberger et al., 2020). Moreover, the temporal relationship between case numbers and vaccination behavior has rarely been explored in this context, despite its potential to inform targeted response strategies. This lack of Austria-specific research limits the ability of policymakers to tailor interventions and limits their responsiveness to changes in public sentiment or behavior regarding preventive measures and vaccination. Additionally, no studies to date have accounted for overlapping flu-related search behavior as a possible confounding factor. Thus, there is an urgent need to examine local digital search trends alongside epidemiological data in Austria to fill this gap. Filling this gap can inform both national and local stakeholders, supporting more data-driven pandemic policies that can be adjusted in near-real time.

Understanding these relationships has implications for positive social change, particularly in improving health equity and resource allocation. Studies have shown that targeted interventions based on timely data can reduce health disparities by ensuring that underserved communities receive relevant messages and vaccination opportunities (Kuehn et al., 2022). When Google Trends data align with actual case counts and vaccination rates, and when the interaction between cases and vaccination behavior is better understood, officials can implement mobile clinics, increase health care staffing, or adjust communication strategies in those specific locations. Accounting for concurrent flu-related search behavior can further refine the precision of such efforts by reducing the risk of misinterpreting spikes in public interest. Such tailored responses can promote more equitable pandemic management and ultimately reduce the disproportionate impact of COVID-19 on vulnerable groups. Findings from a context-specific study in Austria, analyzed nationally, can, therefore, not only inform local officials, but also serve as a template for other regions seeking to leverage digital data for equitable and effective public health interventions.

This dissertation addresses the current gap by systematically investigating how Google Trends might correlate with COVID-19 metrics in Austria, providing a basis for evidence-based public health planning. By analyzing search volume, case incidence, and vaccination uptake, as well as the relationship between case trends and vaccination behavior, this research reveals the extent to which digital behavior reflects realities on the ground, building on previous work in digital epidemiology (Porcu et al., 2023). The inclusion of influenza-related search data as a contextual control enhances the precision of these analyses by identifying potential periods of confounding interest in other respiratory illnesses. In this way, the study not only contextualizes the Austrian experience within a broader digital surveillance framework but also highlights modifiable factors, such as messaging strategies and timely data access, that can strengthen ongoing containment and vaccination efforts. The following sections of this dissertation detail the background, problem statement, theoretical framework, and research approach applied at the national level, concluding with a substantive discussion about the findings and their potential to advance both scientific knowledge and social outcomes in Austria.

Background

The global outbreak of COVID-19 has accelerated the need for rapid, adaptive, and precise public health surveillance tools. Traditional surveillance systems, which rely on clinical reporting, laboratory confirmation, and official epidemiological data, often experience delays that hinder rapid response (Effenberger et al., 2020). In contrast, digital tools such as Google Trends can capture public interest, concern, or behavioral shifts in near-real time by analyzing variations in keyword searches related to symptoms, prevention, or vaccination (Mavragani & Ochoa, 2019a). When used properly, such data can serve as a proxy for real-world health trends, identify disease outbreaks, or indicate demand for resources, especially during health crises such as the COVID-19 pandemic. This interaction between collective online behavior and epidemiological evidence highlights the critical role that digital surveillance can play in informing more agile, data-driven responses, including an understanding of how public concern, case patterns, and vaccination uptake may evolve together.

Despite growing evidence of the applicability of Google Trends worldwide, significant gaps remain in understanding its usability in certain countries, such as Austria. Studies from Italy, Spain, and the United States show that search term spikes can precede official case spikes by several days or weeks, providing an early warning system for public health officials (Porcu et al., 2023). However, there’s a dearth of research on whether these patterns directly reflect Austria’s unique health system and sociocultural landscape. Some existing research excludes vaccine uptake as a core outcome and focuses solely on cases or deaths (Effenberger et al., 2020), resulting in a less comprehensive approach. Even fewer studies explore the relationship between case trends and vaccination uptake, a connection that could reveal behavioral responses to rising infection levels. As a result, health authorities in Austria lack country-specific insights into how digital search patterns might predict case numbers or vaccination trends. Understanding these nuances is critical to aligning public health interventions with local behavioral and trust dynamics. Therefore, the gap lies not only in validating Google Trends in this setting but also in exploring how correlated trends might influence or reflect actual vaccine uptake and case trends in Austria.

In addition, the evolution of the Austrian vaccination campaign highlights the need for research into the factors that contribute to vaccination rate plateaus or spikes. Emerging data suggest that timely, localized communication strategies, when informed by dynamic search analytics, can increase uptake among hesitant populations (Kuehn et al., 2022). Austria's federalized health care structure and diverse regional practices complicate simple assumptions about how the public will respond to nationwide vaccination campaigns. Without a clearer view of how search term frequencies correspond to actual vaccination behavior in Austrian provinces, interventions and messaging may remain overly generalized and, therefore, less effective. Moreover, understanding whether increases in case numbers trigger subsequent rises in vaccination uptake could offer important insight into behavioral responses at a national level. Targeted outreach depends on gaining better insight into local digital data, which reveals when and where interventions are most needed.

Taken together, these observations demonstrate why this study, which focuses on the correlation between Google Trends and COVID-19 cases and vaccination rates in Austria, is important. Analyzing search patterns in relation to epidemiological indicators can reveal local behavioral changes earlier than administrative data would allow, potentially helping to mitigate outbreaks. It can also reveal population segments where vaccine hesitancy aligns with search interest, providing critical information for designing tailored outreach campaigns. Additionally, examining how case trends and vaccination uptake interact may provide insight into reactive health behaviors that influence pandemic outcomes. By filling this knowledge gap, the study provides actionable insights to optimize surveillance and response, thereby strengthening Austria's and potentially other nations' capacity to address current and future public health challenges.

Problem Statement

Traditional epidemiologic surveillance can fail to support early detection and rapid intervention, leaving public health stakeholders with delayed insights during fast-evolving crises, such as the COVID-19 pandemic. Although countries around the world have implemented routine data collection and case reporting, delays in data processing and reporting bottlenecks have constrained response efforts (Effenberger et al., 2020). At the same time, real-time digital surveillance approaches, particularly those that analyze internet search data, have proven to be effective in predicting outbreaks and guiding early interventions in Italy, the United States, and elsewhere (Porcu et al., 2023). However, Austria-specific research remains limited, and it is unclear whether Google Trends correlations with COVID-19 cases and vaccination behavior are consistent with findings from other countries. In particular, the relationship between COVID-19 case surges and subsequent vaccination uptake has not been systematically examined in Austria, despite its relevance for modeling behavioral responses. This gap prevents Austria's public health officials from leveraging potentially valuable digital signals that could optimize resource allocation, messaging, and outreach efforts. Without a robust exploration of how online search patterns might reflect epidemiological dynamics in the real world, Austria's public health policies risk being reactive rather than proactive.

In addition, new vaccines were a key mitigating factor in Austria's COVID-19 response; however, variations in uptake rates highlight the need to better understand behavioral drivers, an area where digital search data can provide important insights. Recent studies highlight that spikes in Google searches for phrases such as *vaccine side effects* or *COVID-19 prevention* often precede changes in vaccination uptake by several weeks (Mavragani & Ochoa, 2019a). However, many of these studies focus primarily on bigger countries, leaving Austrian-specific contexts, such as regional vaccination campaigns or the influence of German-language media, largely unexamined. By overlooking these localized variables, the current literature misses an opportunity to enhance predictive models that could inform targeted interventions. Moreover, it remains unclear whether rising case numbers themselves influenced vaccine uptake in Austria, which is a relationship that may reflect shifts in public risk perception. This represents a significant research gap where robust, context-specific evidence is urgently needed. As a result, examining correlations between Google Trends search volume, COVID-19 case trajectories, and actual vaccination rates in Austria may inform innovative strategies to enhance surveillance and promote vaccine uptake, thereby addressing this critical gap.

Purpose of the Study

The purpose of this quantitative study is to investigate the correlation between Google Trends data and both COVID-19 case numbers and vaccination rates in Austria.

By focusing on the statistical relationships between search volumes for specific keywords (e.g., *COVID-19*, *vaccine side effects*) and epidemiological indicators (such as cases and vaccination uptake), this approach goes beyond descriptive observation and aims to identify potential predictive or explanatory relationships (Mavragani & Ochoa, 2019a). If such correlations are statistically robust, they may suggest that spikes in online interest predict or coincide with changes in real-world pandemic metrics, providing a complementary data source for timely surveillance. Using inferential methods, specifically correlation and regression analysis, this study tests how well Google Trends search volumes can serve as early indicators or concurrent reflections of Austria's public health landscape. Additionally, the study examines whether case trends and vaccination uptake are significantly associated with each other over time, offering further insight into how the Austrian public responds behaviorally to rising infection levels.

In this study, the independent variable consists of Google Trends search volume data for targeted COVID-19-related terms in Austria. These search volumes, which are standardized over time, provide a consistent quantitative measure of public online interest or concern (Effenberger et al., 2020). By capturing how closely people search for information about COVID-19, these digital markers may reveal shifts in perceptions or concerns that influence subsequent health behaviors. A robust definition of independent variables ensures clarity in how search volume metrics are extracted, aggregated, and validated as quantifiable indicators of public attention. It should be noted that for RQ3, which examines the relationship between case trends and vaccination uptake, Google Trends data is not used as an independent variable.

Meanwhile, the dependent variables are the reported weekly COVID-19 case counts and official vaccination rates in Austria. These epidemiologic metrics, which are obtained from national health databases, serve as critical measures of both disease burden and prevention success (Kuehn et al., 2022). By examining how changes in Google Trends data coincide with fluctuations in these dependent variables, it may be possible to clarify whether rising online search interest is associated with either a threat of an increase in cases or a shift in the pace of vaccine uptake. Such insights can provide actionable information for Austrian health authorities to adjust communication strategies or resource allocation. In addition, the relationship between these two dependent variables, whether increases in case counts are associated with subsequent increases in vaccination uptake, were explored to see how reactive public health behavior changes during different phases of the pandemic.

Overall, this quantitative design supports a more evidence-based foundation for digital epidemiology in Austria. Using correlation and regression analysis, the study assessed the strength and significance of the relationships between the independent variables (Google Trends volumes) and the dependent variables (case counts and vaccination rates). By systematically measuring these components, the study aimed to determine whether Google Trends data can provide timely, complementary signals to inform public health interventions beyond what traditional surveillance alone can provide. Additionally, by examining the direct association between case trends and vaccination uptake, the study explored behavioral response dynamics that are not captured by search data alone. The results of this study contribute to the broader discourse on integrating innovative digital tools into mainstream epidemiological methods, particularly in regions with comprehensive health data infrastructure, such as Austria.

Research Questions and Hypotheses

RQ1: Is there a statistically significant relationship between weekly Google Trends search volume (for select COVID-19–related terms) and the reported weekly number of COVID-19 cases in Austria?

H0 (Null Hypothesis 1): There is no statistically significant relationship between weekly Google Trends search volume and the reported weekly number of COVID-19 cases in Austria.

H1 (Alternative Hypothesis 1): There is a statistically significant relationship between weekly Google Trends search volume and the reported weekly number of COVID-19 cases in Austria.

RQ2: Is there a statistically significant relationship between weekly Google Trends search volume (for select COVID-19–related terms) and the weekly vaccination rates in Austria?

H0 (Null Hypothesis 2): There is no statistically significant relationship between weekly Google Trends search volume and the weekly vaccination rates in Austria.

H1 (Alternative Hypothesis 2): There is a statistically significant relationship between weekly Google Trends search volume and the weekly vaccination rates in Austria.

RQ3: Is there a statistically significant relationship between weekly COVID-19 cases and weekly vaccination rates in Austria?

H0 (Null Hypothesis 3): There is no statistically significant relationship between weekly COVID-19 cases and weekly vaccination rates in Austria.

H1 (Alternative Hypothesis 3): There is a statistically significant relationship between weekly COVID-19 cases and weekly vaccination rates in Austria.

The primary independent variable in this study is Google Trends search volume, measured by normalized weekly search data for specific COVID-19-related keywords within Austria (e.g., *COVID-19*, *vaccine side effects*, *symptoms*, *COVID-19 prevention*). These keyword frequencies, reported as relative search volume scores ranging from 0 to 100, quantify public engagement and information-seeking behavior around the pandemic over time. Meanwhile, two dependent variables serve as the main outcomes: weekly COVID-19 case numbers and weekly vaccination rates. Case numbers, obtained from official Austrian public health databases, represent the frequency of newly confirmed cases each week. In contrast, vaccination rates capture the absolute number of individuals who receive a COVID-19 vaccine during the same period. Both dependent variables are used to measure real-world epidemiologic activity, with one measuring the ongoing burden of the virus and the other indicating the level of preventive uptake by the population. By defining the study variables in this way, the study ensures clarity in how digital engagement via Google Trends is quantitatively compared to real-world indicators of disease spread and vaccination progress. It should be noted that in RQ3, the relationship between weekly case counts and vaccination rates is assessed independently of Google Trends data to explore whether infection trends may influence subsequent vaccination behavior.

By examining the relationships between these variables, it is possible to systematically assess whether fluctuations in Google Trends search volume correlate with subsequent or simultaneous changes in COVID-19 case numbers and vaccination rates. If the statistical analyses show that spikes in online interest are reliably associated with emerging case spikes or increases in vaccination activity, then Google Trends may serve as a cost-effective, real-time monitoring tool to complement traditional surveillance. Conversely, a lack of a significant association would imply that internet-based data are less predictive in Austria, thereby placing a greater emphasis on traditional epidemiological measures. Additionally, this study examines the relationship between the number of cases and vaccination rates for COVID-19 over time. Understanding this relationship could reveal whether increases in cases were followed by higher vaccination rates in a way that suggests responsive shifts in public health behavior. Therefore, the rationale of this study lies in clarifying how digital search signals may or may not enhance public health decision-making, ultimately guiding resource allocation, communication strategies, and vaccination outreach efforts.

Theoretical Framework for the Study

This study is based on Rogers' (2003) Diffusion of Innovations Theory, which provides a systematic framework for understanding how new information or practices, such as public health interventions, are communicated, adopted, and incorporated within societies. Originally based on studies in rural sociology, Rogers' theory highlights the importance of communication channels, time, and the social system in the diffusion process, emphasizing how innovativeness and interpersonal networks jointly shape the adoption of new ideas (Rogers, 2003). In the context of this research, the analysis of Google Trends data on COVID-19 topics can be seen as an investigation of the awareness and persuasion stages of an innovation. Individuals may search for and gather information online before deciding whether to engage in health behaviors, such as getting tested or vaccinated. Therefore, the framework naturally aligns with the study's goal of investigating the relationship between online information-seeking and actual public health outcomes. In parallel, the study also examines real-world behavioral responses, such as the relationship between infection trends and vaccine uptake, which may reflect later stages in the innovation-decision process.

At the center of the Diffusion of Innovations Theory are five stages in the innovation decision process: knowledge, persuasion, decision, implementation, and confirmation (Rogers, 2003). For example, as individuals acquire knowledge about COVID-19 and vaccines, they move to the persuasion stage, where they form an attitude toward the idea, and then to the decision stage, where they decide whether or not to get vaccinated. Confirming behaviors, such as influencing others or reinforcing personal decisions, may also occur as the innovation spreads. The key propositions of the theory highlight that awareness and attitudes, shaped by media and peer communication, can significantly influence the gradient of an "adoption" curve (Rogers, 2003). In this study, Google Trends metrics may capture this awareness and early persuasion phase as people research vaccine availability, side effects, or infection rates. At the same time, trends in case numbers and subsequent vaccination behavior may reflect later stages of adoption, including implementation and confirmation, offering further insight into real-world health behavior during the pandemic.

Building on these key principles of Rogers' theory, this study applies the concepts of knowledge and persuasion specifically to the research questions regarding Google Trends data and COVID-19 metrics in Austria. By hypothesizing that spikes in online searches correlate with increased public awareness, the diffusion perspective offers a structured approach to interpreting statistical relationships. If significant relationships are observed, they may reflect how Austrians move through the early stages of adoption before deciding to be vaccinated. In addition, if the theory's emphasis on communication channels is correct, differences in search activity across different time periods may indicate how effectively public health messages are diffused in Austria. In this regard, the diffusion of innovation framework not only guides the rationale for the study design but also provides a framework for interpreting the results, indicating whether the observed trends are consistent with or deviate from Rogers' concepts of how innovations, in this case, preventive health behaviors, gain traction within a community. This framework also supports the interpretation of findings related to the temporal association between case surges and vaccination uptake, which may reflect later phases of adoption in response to perceived risk.

Nature of the Study

This study uses a quantitative correlational design to investigate the statistical relationships between digital search behavior and COVID-19 epidemiological metrics in Austria. A correlational design is appropriate for addressing research questions that focus on determining whether changes in one variable, such as search volume for COVID-19, correspond to changes in another variable, such as case numbers or vaccination rates (Effenberger et al., 2020). This technique does not assume causality but rather tests the strength and direction of these relationships. The rationale for selecting this design is based on the exploratory nature of the research problem: to determine whether and to what extent real-time digital searches reflect or predict changes in viral spread and vaccination uptake. Additionally, the study investigates whether case trends are statistically correlated with changes in vaccination behavior over time, controlling for search activity. By quantifying these relationships, the study can build the foundation for future causal or experimental research if the results suggest a strong association between online searches and epidemiological outcomes.

The study is grounded in the independent variable and the dependent variables. The independent variable is the normalized weekly Google Trends search volume for specific COVID-19 keywords, such as *COVID-19* or *vaccine side effects*. Meanwhile, the dependent variables include weekly COVID-19 case counts and official vaccination rates collected from Austrian public health data. Differentiating between these variable categories allows for a systematic exploration of the correlation between online search behavior and real-time pandemic markers. For RQ3, which assesses whether a relationship exists between case counts and vaccination rates without implying directionality, there is no formal independent or dependent variable. However, for interpretation, this study treats the weekly COVID-19 cases as the independent variable and the weekly vaccination rates as the dependent variable. This clarity in variable definition enhances the transparency and replicability of the methodology, ensuring that any identified relationships are robustly supported by quantitative measures.

Data collection involved collecting retrospective weekly Google Trends data for targeted keywords and correlating them with official Austrian epidemiologic records for the same time windows. This strategy allowed for a week-by-week comparison to determine whether spikes in online search volume were preceding, coinciding with, or following changes in reported cases and vaccination coverage as well as whether increases in COVID-19 cases were followed by increases in vaccination rates. The data on confirmed cases and vaccination rates were retrieved from Austrian health databases, ensuring accurate and standardized metrics. Statistical analyses, including Pearson's or Spearman's correlation and multiple linear regression, were performed to identify bivariate relationships. This overall method was designed to maintain methodological integrity by employing validated data sources, consistent data intervals, and appropriate tests.

Overall, this design and procedure are consistent with the overarching research questions of the study and demonstrate how digital search behavior can serve as an indicator, or even a leading signal, of COVID-19 case and vaccination trends. Additionally, the analysis shows how increases in COVID-19 cases may relate to subsequent increases in vaccination rates. If statistically meaningful correlations are observed, these findings may inspire further investigation into the use of real-time search analytics to inform Austrian public health strategies. By adopting a quantitative correlational framework, the study is also well-positioned to quantify the predictive relationships between variables, thereby strengthening the evidence for or against the use of such digital surveillance methods. In conclusion, the study's nature as a robust, data-driven investigation positioned it to make a substantive contribution to discussions about integrating web-based metrics into comprehensive epidemiological surveillance.

Definitions

The following terms clarify the key concepts and variables used in this study. Each definition is based on existing literature to ensure accuracy and relevance, thereby facilitating an understanding of how digital search volume may correlate with COVID-19 trends in Austria. The study focused on the relationship between one independent variable (Google Trends search volume) and two dependent variables (weekly COVID-19 case counts and weekly vaccination rates). In addition, the study examined the relationship between COVID-19 case counts and vaccination rates directly, independent of search behavior, to better understand behavioral responses to epidemiologic trends.

Google Trends Search Volume (Independent Variable): A normalized metric of public online interest in specific keywords, in this case, keywords like *COVID-19* or *vaccine side effects*. Google Trends reports this volume on a relative scale (0-100) that reflects the proportion of searches for these terms within a specific geographic and time range (Nuti et al., 2014). A higher relative search volume reflects a higher volume of public search activity for the target keywords (Nuti et al., 2014).

Weekly COVID-19 Case Numbers (Dependent Variable): The total number of newly confirmed COVID-19 infections reported each week within the official surveillance system in Austria. These data are collected and validated by national health authorities, reflecting the evolving burden of disease in the population. In RQ3, this variable is used as a predictor to examine whether case surges are followed by increases in vaccination uptake.

Weekly Vaccination Rates (Dependent Variable): The absolute number of individuals receiving COVID-19 vaccines each week, as documented by the Austrian health authorities. This variable captures the uptake of vaccination efforts over time. In RQ3, this variable is analyzed in relation to case numbers to assess possible behavioral responses to changes in disease incidence.

Assumptions

The first assumption is that Austria's official COVID-19 data, including weekly case counts and vaccination rates, accurately reflect real-world conditions. Researchers have found that although daily reporting can be subject to delays, the aggregated numbers from the Austrian Federal Ministry of Social Affairs, Health, Care, and Consumer Protection are sufficiently robust for epidemiologic analyses. If these official counts were significantly flawed, for example, due to data backlogs or inconsistent testing, then any correlation with Google Trends metrics would be less meaningful, and the interpretation of associations between case counts and vaccination rates could be compromised. The assumption of reliable reporting reinforces the study's focus on weekly variations and ensures that the dependent variables actually capture pandemic trends in the real world. With this assumption, the study suggests that observed relationships between search volume and reported case or vaccination data, as well as relationships between case and vaccination data themselves, reflect true trends at the community level, providing a necessary foundation for future analysis.

A second assumption is that Google Trends serves as a valid indicator of the public's search behavior for information about COVID-19 in Austria. Previous work shows that increases in search terms often coincide with increased public interest or concern about emerging health issues (Mavragani & Ochoa, 2019a; Porcu et al., 2023). Additionally, Austria's high internet coverage rate suggests that a significant portion of the population is using Google to find pandemic-related information. If a significant portion of Austrians rely on other search platforms or rarely seek COVID-19 information online, Google Trends data may underrepresent the actual concern. However, the existing literature supports its utility as a near-real-time measure of collective awareness and concern. This assumption is critical to interpreting correlations between online search interest and changes in case numbers or vaccination uptake, ensuring that measured relative search volume reflects real-world changes in public engagement with COVID-19 issues.

The last assumption is that the weekly aggregation of both Google Trends and epidemiological data provides a robust basis for correlational analysis. Short-term outliers, such as weekend testing lags or media-driven search spikes, are partially moderated by aggregating daily numbers into weekly intervals. This approach assumes that any temporary reporting inconsistencies do not disrupt the underlying trends. If significant daily variation were critical to capturing the relationship, the study's aggregation strategy might potentially hide some signals. However, using weekly data smooths out the noise, making the time series more robust for detecting broader patterns. Therefore, treating the data in weekly intervals fits well with the exploratory nature of correlation tests and supports the main goal of assessing whether online interest follows real-world pandemic indicators in Austria. This approach also supports the time series comparison between COVID-19 case numbers and vaccination rates in RQ3, allowing for a more stable view of population-level trends.

Scope and Delimitations

This study focused on the relationship in Austria between Google Trends search volume and two critical COVID-19 outcomes, specifically weekly case counts and weekly vaccination rates. Previous research highlights that spikes in search behavior can match real-world epidemiological trends (Effenberger et al., 2020; Porcu et al., 2023). However, these findings often come from broad international studies or focus on a single outcome, like the number of cases, leaving vaccine uptake understudied. By focusing on these two outcome measures within the context of a single country's health system, the study aimed to achieve a clearer understanding of whether digital search data serve as an early signal specific to Austria's policy needs. In addition, the study explored how case numbers and vaccination rates are associated with each other over time, to evaluate whether rising infections precede behavioral shifts in vaccine uptake. This intentional focus also reduced irrelevant variation that might occur if multiple outcomes or multinational datasets were pooled. This scope ensured a thorough exploration of correlations relevant to Austria's governmental structures and cultural environment, but it also limited the direct generalizability of findings to notably different populations or health care contexts.

In setting these limitations, I excluded other potentially important variables, such as sociodemographic subgroups or distinct attitudes toward vaccine hesitancy, to maintain clarity about the primary association described in the research questions. Researchers have demonstrated that educational background and trust in government can independently influence vaccination rates (Kuehn et al., 2022). However, the inclusion of these nuanced aspects here could potentially dilute the core focus of population-level trends captured through search behavior and case-based surveillance data. While such exclusions simplify the data structure, they unavoidably exclude a wealth of context about who is specifically searching or how beliefs and attitudes shape those searches. This trade-off supports a more direct correlation analysis, but it also defines a dimension of the study's external validity, as it may underrepresent the heterogeneity of the Austrian public. Additionally, the collection of sociodemographic data in the datasets used has not been done consistently, which prohibits their reliable integration into the analysis. Since these data points were not mandatory to report, the available information is incomplete and lacks the uniformity required for meaningful statistical interpretation. As a result, these delimitations reinforce the central goal of the study, which was to determine whether Google Trends is consistent with epidemiological metrics and whether COVID-

19 incidence is temporally associated with vaccination behavior, while also acknowledging that a more in-depth examination of demographic or psychological factors is beyond the immediate scope.

A second restriction is the temporal scope of the study, which was limited to weekly aggregates from the start of COVID-19 in Austria until mid-2023. This means the study excluded pre-pandemic or post-emergency periods. Research by Becerra-García et al. (2023) highlights how times of crisis can increase the volume of searches related to mental health or preventive measures, while outside of such contexts, patterns could be very different. By focusing on this specific window of time, the study captured the most turbulent phases of the pandemic, when Google Trends spikes may have been most significant. However, if digital search behavior normalized later, or was already low prior to the pandemic, those developments remain unexamined. The chosen timeframe optimally aligns with Austria's reported COVID-19 data, which also underpins the case and vaccination rate trends explored in RQ3. Nevertheless, it limits generalizability to the critical phase of the pandemic. Any subsequent changes in public online behavior or epidemiological patterns beyond June 2023 fall outside the established boundaries.

The geographic scope is also limited to the national level of Austria, which excludes cross-border or regional comparisons that might show how different regulatory frameworks affect COVID-19 trends and Google search behavior. In infodemiological studies, cross-country variation can show whether specific policy announcements or cultural attitudes influence search behavior (Mavragani & Ochoa, 2019a; Porcu et al., 2023). This study, however, focused only on the federal health care structure of Austria.

Such a limited view preserves internal validity, ensuring consistent data collection protocols and a shared health context. But it also might dampen external validity for nations with different health infrastructures. While the findings may inspire similar investigations elsewhere, direct extrapolation should be done with caution. Focusing on data from one country allows for a more nuanced interpretation of how Austrian policy changes or cultural norms might influence not only online behavior but also case and vaccination dynamics, consistent with the study's goal of generating actionable local insights.

These delimitations make it clear that the findings of the study speak primarily to the Austrian experience, with an emphasis on correlational patterns rather than deeper demographic segmentation or broader international trends. Prior literature suggests that similar correlational findings have proven valuable but are rarely consistent across diverse environments (Porcu et al., 2023). By defining who is included, which in this study is all Austrian citizens, and what is being measured, which is the number of cases, vaccination rates, and Google Trends volume, the study achieved a clear focus. The tradeoff for this specificity is a more restricted basis for generalization across contexts, but the method can be adapted by other researchers in different settings. This scope and delimitation therefore ensure a cohesive study design with high internal consistency, providing a clear path for analyzing both digital and epidemiological data while transparently acknowledging the study's limited population range and geographic generalizability.

Limitations

A key limitation is the dependence on secondary data sources, specifically Google Trends metrics and Austria's official COVID-19 figures, which were not originally designed for controlled research. As noted by Effenberger et al. (2020), delays or anomalies in reporting can affect the timeliness and accuracy of epidemiological data, potentially introducing bias into correlations with search behavior and between public health outcomes themselves. While weekly aggregation helps reduce day-to-day variation, it cannot fully resolve potential inherent inconsistencies, such as sudden backlog releases. This may lead to less confidence that search trends accurately reflect real-world events. This data-structure limitation was mentioned earlier in the section "Scope and Limitations," but it is worth emphasizing here because it directly affects the internal validity of correlational results. To address this, the study aimed to provide transparency, even if it cannot completely eliminate potential noise from secondary data dependencies, by thoroughly documenting unusual data spikes in the final analysis.

Another limitation is the potential sampling bias in the Google Trends dataset, as not all Austrians have the same level of internet access or search habits. Robinson et al. (2022) highlighted that some rural or low-income groups may be underrepresented online, which could lead to an overestimation of public engagement with COVID-19 content. Although Austria's high overall connectivity suggests a broad user base, this coverage is not consistent, leaving possible subsets of the population whose behaviors may not show up in the digital footprint. As a result, the correlations found in the study that involve Google Trends may only reflect the digitally active subset of Austrians.

While the bilingual search terms (English and German) extend the reach, the limitation remains that the behaviors of underrepresented segments are not fully captured. Future research could incorporate alternative data sets, such as telephone surveys or social media analysis, to compensate for this bias.

An additional limitation is the potential overlap in public search behavior related to both seasonal influenza and COVID-19. Although this study used influenza-related Google Trends data to identify and flag periods of elevated flu-related searches, it is methodologically challenging to separate the public's simultaneous concern for both illnesses. Despite efforts to exclude these periods or analyze them separately through sensitivity analyses, some residual confounding may persist, particularly during peak influenza seasons. Therefore, while including influenza search data enhances the study's internal validity, it does not fully eliminate the risk of improperly attributing shared search volume fluctuations solely to concern about COVID-19.

In addition, the non-experimental, correlational nature of the study prevents definitive causal statements about whether increases in online searches lead to changes in cases or vaccination rates or vice versa. As noted by Mavragani and Ochoa (2019a), the presence of a statistical association between Google Trends volumes and epidemiological metrics does not necessarily establish causality in digital epidemiology. External factors, such as government announcements, shifts in media coverage, or abrupt policy changes, can simultaneously affect both search activity and real-world outcomes. Similarly, these external factors may drive both increases in cases and subsequent surges in vaccination, complicating interpretations of directionality. This challenge, also briefly addressed in the

"Scope and Delimitations" section, highlights that even strong correlations can be driven by confounding variables. Although the inclusion of major public health announcements provides some insight into these confounders, a purely correlational design cannot completely rule out alternative explanations, highlighting the importance of cautious interpretation and possibly more advanced temporal analyses in future studies.

Lastly, the temporal boundary of the study, from early 2020 to mid-2023, may limit its applicability to other phases of the pandemic or post-pandemic conditions. Research suggests that public interest patterns can shift dramatically once a crisis stabilizes, potentially reducing the predictive value of Google Trends (Porcu et al., 2023). By focusing on the most unstable period of the pandemic, the study captured increased search behavior; however, it may not accurately represent how Austrians search for health information in everyday, non-emergency situations. For instance, as shown by Kuehn et al. (2022), vaccination dynamics may change significantly after the initial urgency has faded. Acknowledging this temporal limitation highlights that generalizing findings to future outbreaks or different diseases should be done cautiously and may require updated analyses as conditions evolve.

Significance

The findings of this study can significantly contribute to the field of digital epidemiology by empirically evaluating Google Trends as an early warning and monitoring tool for COVID-19 in Austria. Previous research demonstrates the utility of online search data in predicting disease outbreaks and reflecting public concerns about vaccination (Mavragani & Ochoa, 2019a; Porcu et al., 2023). However, findings specific to Austria remain scarce, potentially limiting the global understanding of how well digital indicators perform in different health settings. By filling this empirical gap, this study can strengthen theoretical discussions about real-time surveillance mechanisms by demonstrating whether digital search behaviors consistently align with real-world cases and vaccination numbers in Austria. Additionally, the study examined whether COVID-19 case trends themselves influenced vaccination uptake, offering a more comprehensive understanding of behavioral responses to outbreaks. Such evidence addresses calls from researchers for more localized validation of digital tools (Saegner & Austys, 2022). As a result, this research contributes to the field by advancing knowledge of region-specific contexts and promoting a broader understanding of how digital epidemiology works in different cultural and structural contexts.

On a practical level, the research has the potential to refine health policy and public health practice, while also guiding resource allocation and communication strategies in Austria. As noted in previous studies, increases in online searches often occur shortly before increases in reported cases or peaks in vaccination demand (Becerra-García et al., 2023; Kuehn et al., 2022). If a similar correlation is confirmed in Austria, local health authorities can use near-real-time digital data to deploy mobile clinics or increase public messaging at critical times. Additionally, if the analysis shows that COVID-19 case surges correlate with subsequent increases in vaccination uptake, authorities could better time their outreach and resource deployment based on case trends alone. Such proactive measures could potentially maximize pandemic response, reduce hospital burden, and improve vaccine uptake by leveraging these early signals. This approach is consistent with existing evidence that timely, data-driven measures can reduce the impact of disease, particularly among high-risk groups (Kuehn et al., 2022). By incorporating these findings into policy frameworks, Austrian authorities could better anticipate waves of infection or identify clusters of vaccine hesitancy, therefore improving the overall effectiveness of public health responses.

In addition, there is the potential for positive social change through an equitable design of health interventions, targeting underserved populations and bridging information gaps. Previous work shows that timely use of digital data can identify emerging health concerns and drive targeted efforts to reduce inequalities in vaccine access searches (Becerra-García et al., 2023; Kuehn et al., 2022). For example, identifying spikes in vaccine-related searches could inform mobile outreach or tailored educational materials in areas with historically low vaccine uptake. Furthermore, if rising case numbers in specific federal states are found to precede increases in vaccination rates, this could signal opportunities for region-specific interventions even before search patterns change. Such a framework helps ensure that more marginalized groups do not remain “digitally invisible,” as their local search interest may be overshadowed in aggregated data. By conducting the analysis across Austria, policymakers could better target resources to mitigate disparities. Ultimately, if insights from digital surveillance are translated into strategic interventions, the study’s findings can promote further research to support a more equitable and responsive health system, highlighting the social value of connecting online and offline public health interventions.

These contributions highlight how this research's focus on Austrian Google Trends data is relevant to broader academic, policy, and societal goals. By advancing theoretical insights on local, digital epidemiology, informing public health agencies on data-driven best practices, and enabling more equitable outreach, this study aligns with emerging priorities identified in the recent COVID-19 literature (Effenberger et al., 2020; Porcu et al., 2023). This underscores the importance of the study. Beyond academic research, it provides actionable insights that can further refine pandemic management approaches and promote sustainable improvements in health equity. The significance, therefore, lies not only in testing correlations, but also in contributing solid empirical evidence that can inform how Austria and potentially other nations use online search data and real-time case trajectories for rapid and inclusive public health strategies.

Summary

This chapter highlighted the critical role of near-real-time digital data, particularly Google Trends, in complementing traditional epidemiological surveillance of COVID-19 in Austria. Several studies have demonstrated that increases in online searches for disease-related terms often reflect actual case increases or anticipate them, suggesting an opportunity to leverage real-time public interest for proactive interventions (Effenberger et al., 2020; Porcu et al., 2023). Chapter 1 highlighted that this phenomenon is still understudied in Austria. By highlighting this gap, the chapter outlined this research as a necessary step toward a more agile and data-driven public health response. Such an approach can refine Austrian policymakers' ability to effectively allocate resources and tailor communication strategies based on emerging online indicators. Additionally, the study examines whether increases in COVID-19 cases are followed by vaccination rate increases, offering insights into population behavior in response to changing infection dynamics. Situating the study within Rogers' (2003) Diffusion of Innovations Theory helps to clarify how online search behaviors may reflect early stages of awareness and persuasion and therefore influence vaccination decisions. Overall, these findings provide a strong foundation for exploring the statistical correlation between Google Trends data and both infection rates and vaccination behavior, which provides actionable knowledge for public health officials and researchers.

In addition to identifying the research gap, Chapter 1 also lays out the purpose, scope, and methodological framework of the study, clarifying how this research can inform public health policy and promote positive social change. Relying on quantitative correlational methods and official Austrian health data, the chapter explained how weekly variations in search volume could identify shifts in case numbers and vaccination rates in near-real-time. In doing so, it emphasized that these insights could help identify local vaccine hesitancy and refine targeted interventions, especially among vulnerable groups. By focusing on Austria's federalized health care structure and cultural context, the study aims to provide country-specific insights that may be overlooked in broader global studies. This localized lens highlights the potential for both scientific advances in digital epidemiology and real-world policy innovations aimed at improving equitable public health outcomes. With the problem, purpose, and theoretical basis established, Chapter 2 builds on this foundation by reviewing the existing relevant literature, further informing the conceptual and empirical direction of the study.

Chapter 2: Literature Review

Introduction

The COVID-19 pandemic has emphasized the urgent need for real-time public health surveillance to manage rapidly evolving health crises. Traditional surveillance systems often face delays in data reporting, limiting their ability to provide the immediate insights required for effective intervention (Effenberger et al., 2020). Digital tools such as Google Trends, which analyze public searches for terms like *symptoms*, *vaccination*, and *prevention*, offer an innovative approach to capturing public interest, awareness, and behavioral responses to health issues in real time (Mavragani & Ochoa, 2019a). These tools allow public health authorities to monitor trends dynamically and complement traditional surveillance systems by providing faster, actionable data. By enabling a timely understanding of public behavior and its alignment with epidemiological patterns, Google Trends holds promise for improving public health responses to crises. However, despite its global applicability, there is a noticeable lack of research investigating how Google Trends can be used to predict outbreaks in Austria. In addition, previous studies rarely examine the relationship between case trends and vaccination behavior simultaneously, leaving a gap that this study addresses. By addressing this gap, this study aims to provide Austria-specific insights and contribute to the broader field of digital epidemiology. This study examined the relationship between Google Trends data, COVID-19 case numbers, and vaccination rates in Austria to enhance understanding of how digital search behavior correlates with public health outcomes and inform future public health strategies in epidemic outbreak prevention.

Google Trends has been identified in the literature as a potentially valuable tool for public health surveillance, serving as an early indicator of disease outbreaks and changes in health behaviors. For example, studies by Porcu et al. (2023) and Zayed et al. (2023) demonstrated that spikes in search terms for symptoms and preventive measures often precede increases in case counts and public engagement with health interventions. In Italy, Google Trends data accurately detected outbreaks up to seven weeks earlier than traditional swab-based systems in some regions, underscoring its utility for early detection (Porcu et al., 2023). These findings align with broader trends indicating that public interest, as reflected in search behavior, often correlates with actual epidemiological outcomes.

The potential of Google Trends to predict disease outbreaks and public health behavior changes has been successfully demonstrated in multiple contexts, including influenza and COVID-19. In the United States, search trends for terms like *fever* and *cough* corresponded with influenza case numbers (Mavragani & Gkillas, 2020; Zayed et al., 2023). Similar results were observed for COVID-19 in European countries such as Spain and Italy (Mavragani & Gkillas, 2020; Zayed et al., 2023). These studies revealed that search data often mirrored spikes in cases, suggesting that Google Trends could act as a supplementary tool for public health monitoring (Mavragani & Gkillas, 2020; Zayed et al., 2023). Moreover, researchers have advocated for integrating digital tools like Google Trends with traditional epidemiological methods to enhance responsiveness and timeliness in health interventions (Saegner & Austys, 2022; Talic et al., 2021). While these tools show promise, their effectiveness is influenced by limitations such as varying internet access, digital literacy disparities, and potential misinterpretation of search data due to cultural and contextual differences. These challenges necessitate a localized examination to understand their applicability within specific health care systems.

Austria's health care system, characterized by a complex governance structure, presents a unique setting for such an investigation. Researchers have identified the division of responsibilities among federal, regional, and self-governing bodies as a factor contributing to systemic inefficiencies (Ostermann et al., 2017). Nevertheless, Austria maintains a high standard of government-funded health care, complemented by private health care options and a high density of accessible medical facilities (HealthManagement.org, n.d.). These facilities ensure free access to most medical care and allow patients to choose their primary care physician (HealthManagement.org, n.d.). This combination of systemic complexity and quality care underscores the importance of Austria-specific studies to explore how tools like Google Trends could be integrated into public health strategies. Addressing these structural and organizational nuances is critical to optimizing the use of digital tools within Austria's public health framework.

This chapter presents a comprehensive review of the literature to contextualize and justify the importance of this study. The first section examines the use of Google Trends as a public health surveillance tool, focusing on its potential to predict health behaviors and the challenges it presents. This is followed by a discussion of COVID-19 case surveillance, highlighting both traditional methods and innovative digital approaches. The chapter then addresses research on immunization and health practices, including determinants of vaccination rates, the role of digital platforms, and trends specific to Austria. By systematically organizing these themes, the review identifies critical gaps in the literature and opportunities for integrating digital health surveillance tools with traditional epidemiologic methods. Unlike previous work, which has primarily focused on global or regional contexts, this study emphasizes the need for context-specific research tailored to Austria's health care system and cultural environment.

While a thorough review of the literature is crucial, it's equally important to understand what literature isn't available. These gaps identified in the literature highlight the need for localized research integrating Google Trends data with public health metrics in Austria. While existing studies have demonstrated the utility of Google Trends for monitoring public health behaviors and predicting disease outbreaks, the tool's application in Austria remains unexplored. This study seeks to address these limitations by investigating the correlation between Google Trends data, COVID-19 case numbers, and vaccination rates in Austria. Through this novel contribution to digital epidemiology, the study aims to support more effective public health decision-making and preparedness for future pandemics.

Literature Search Strategy

A systematic literature search is essential to ensure the comprehensive identification of relevant studies that support the objectives of a research project. For this study, the literature search focused on identifying sources related to the use of Google Trends as a digital health surveillance tool, specifically in the context of COVID-19 case surveillance and vaccine uptake. The goal was to collect recent, high-quality research that provides insights into the utility, limitations, and applicability of Google Trends in public health, with a focus on studies conducted between 2019 and 2024. This timeframe was chosen to reflect the significant advances in digital health technologies during the COVID-19 pandemic, which highlighted the urgent need for innovative surveillance tools.

The search strategy used multiple databases to ensure breadth and depth, including PubMed for biomedical literature, Google Scholar for interdisciplinary studies, EBSCOhost for public health resources, and ProQuest Central for grey literature such as dissertations and conference proceedings. Key search terms were carefully selected and refined to strike a balance between specificity and comprehensiveness, with a focus on terms related to COVID-19, public health interventions, health surveillance, and vaccination behavior.

Through this systematic approach, the review aimed to include peer-reviewed articles, empirical studies, and grey literature that contribute to a robust understanding of the topic. However, the paucity of studies directly addressing the use of Google Trends in Austria necessitated the inclusion of global and regional studies to provide context. This section outlines the methods used to conduct the literature review, including the databases, search terms, and criteria used to select and refine the results, thereby laying the groundwork for the theoretical and empirical foundation of the study.

Databases and Search Engines

A comprehensive literature search was conducted using four primary databases, each selected for its relevance to the study objectives and ability to provide access to high-quality, peer-reviewed literature and grey literature. These databases included

PubMed, Google Scholar, EBSCOhost, and ProQuest Central. Each played a specific role in capturing a wide range of studies relevant to the use of Google Trends for public health surveillance and COVID-19 case and vaccine surveillance.

PubMed was the primary database used in this study due to its extensive coverage of biomedical literature and public health research. This database is known for its inclusion of peer-reviewed articles related to epidemiology, infectious diseases, and public health interventions. PubMed's advanced search capabilities facilitated precise queries using Boolean operators and Medical Subject Headings (MeSH), allowing for the identification of studies specifically related to digital monitoring tools, such as Google Trends. PubMed search results provided foundational studies for the theoretical and empirical framework of this research, focusing on the role of digital tools in health surveillance and response.

Google Scholar offered access to a broader range of interdisciplinary literature. This database aggregates research from various disciplines, including public health, information science, and behavioral science, which are crucial to understanding the multifaceted nature of digital surveillance tools. While Google Scholar's search capabilities are less refined than PubMed's, Google Scholar provides access to studies published in a variety of journals and conference proceedings that may not be indexed in specialized databases. This resource proved beneficial for identifying studies on the behavioral aspects of public health, such as vaccination behavior and public engagement with health-related information.

EBSCOhost was used to identify additional public health literature, with a focus on studies related to health surveillance, epidemiologic methods, and the integration of digital tools into public health practice. Its robust collection of public health and epidemiology journals made it an ideal database for this study. EBSCOhost's search interface allowed for specific queries and facilitated the retrieval of articles that complemented results from PubMed and Google Scholar. These articles often focused on the practical applications of digital health tools and their effectiveness in various public health contexts.

ProQuest Central was used to access grey literature, including dissertations, theses, and conference proceedings, relevant to the study's focus. This database provided insights into emerging research trends and innovative approaches that may not yet be widely published in peer-reviewed journals. The inclusion of grey literature was essential to capture early findings on the use of Google Trends in health surveillance, particularly in regions or contexts that are underrepresented in peer-reviewed publications. For example, this database included studies that discussed the feasibility of integrating digital tools into specific health systems, providing valuable context for Austria-specific applications.

Key Search Terms and Refinement Process

The research objectives and questions guided the selection of key search terms. The focus was on the use of Google Trends as a tool for public health surveillance in the context of COVID-19 and vaccine uptake, as well as how case numbers and vaccination rates relate to one another over time. Developing precise search terms was essential to ensuring that the literature search was both comprehensive and specific, avoiding irrelevant results while capturing studies directly relevant to the study objectives. The process involved an iterative approach, with initial searches being refined based on the scope and relevance of the results.

Initial Search Terms

The initial search terms were derived from the main concepts of the study, such as *Google Trends, COVID-19*, and *public health surveillance*. These terms were combined with Boolean operators to capture a wide range of studies, including *Google Trends* AND *COVID-19*; *Google Trends* AND *infectious disease* OR *public health surveillance*; *Digital health surveillance* AND *COVID-19*; *Public health intervention* AND *Google Trends*.

These combinations yielded a prohibitively large number of results because the results included studies that were outside the scope of this research, particularly in the areas of non-COVID-19 diseases or unrelated public health interventions. While these terms provided a foundation, refinement was needed to better align the results with the focus of the study.

Refinement of Search Terms

To address the volume of irrelevant results and increase specificity, additional terms related to vaccination were included. These terms reflected the study's specific interest in the correlation between Google Trends data and vaccination uptake in Austria. Refined terms included *Google Trends* AND *vaccination behavior; COVID-19 vaccination* AND *Google Trends*; *Vaccination uptake* AND *digital health surveillance;*

and *Vaccine hesitancy* AND *COVID-19.* The refinement process also included terms related to Austria's health care context to capture region-specific studies, such as *Austria* AND *COVID-19 vaccination* and *Google Trends* AND *Austria* AND *health surveillance*. These additional terms ensured that the search results included studies that addressed both the broader application of Google Trends and its potential for monitoring vaccination behavior in Austria.

Iterative Approach

The refinement process was iterative, with adjustments made based on the number and relevance of results. For example, initial searches using terms such as *infectious disease* yielded results that were too broad and included studies of diseases unrelated to COVID-19. By narrowing the terms to COVID-19 and adding vaccine-specific terms, the search results became more targeted.

In addition, synonyms and related terms were included to account for variations in terminology across disciplines and regions. For example, *COVID-19* was supplemented with *coronavirus*, and *vaccination behavior* was supplemented with *vaccine uptake* and *vaccine hesitancy*. These variations improved the retrieval of relevant studies, especially in databases such as Google Scholar, which aggregates research from different fields.

Search Term Combinations

Final search term combinations used in the literature review included *Google Trends* AND *COVID-19* AND *vaccination rates*; *Digital epidemiology* AND *COVID-19 vaccination*; *Vaccination behavior* OR *vaccine uptake* AND *Google Trends*; *Public health monitoring* AND *Austria*; and *Google Trends* AND *health behavior* AND *Austria*.

These combinations were applied across all four databases (PubMed, Google Scholar, EBSCOhost, and ProQuest Central), ensuring consistency in the search strategy.

Balancing Comprehensiveness and Specificity

The refinement process also included reviewing search results for relevance based on the inclusion and exclusion criteria outlined in the study. Articles were included if they focused on the use of Google Trends for health surveillance, covered COVID-19 case monitoring or vaccination behavior, were published between 2019 and 2024, and featured peer-reviewed articles, dissertations, or grey literature relevant to Austria. Articles were excluded if they focused on non-COVID-19 diseases or interventions unrelated to digital surveillance, were published before 2019, or lacked relevance to public health surveillance. Through this iterative and focused approach, the final set of search terms effectively captured a comprehensive and relevant body of literature, providing a strong foundation for the theoretical and empirical exploration of the study.

Scope of Literature Reviewed

To ensure the inclusion of studies most relevant to the research objectives and questions, the scope of the literature review was carefully defined. The review focused on literature that examined the use of Google Trends for public health surveillance, with a particular focus on COVID-19 case surveillance and vaccine uptake, as well as research examining the relationship between infection rates and vaccination behavior over time. This scope was determined by the research question, which seeks to explore how Google Trends data correlates with epidemiologic outcomes in Austria and the identified gap in Austria-specific studies.

Time Frame

The review primarily included studies published between 2019 and 2024. This timeframe was chosen to capture research conducted during the COVID-19 pandemic, which had its first known outbreak in December 2019, and its aftermath, a period marked by significant advances in digital health technologies and increased reliance on tools such as Google Trends. The rapid evolution of the pandemic necessitated real-time public health responses. This led to a proliferation of studies on digital surveillance methods during this period. Earlier publications were excluded unless they were seminal and fundamental to understanding the theoretical or methodological aspects of Google Trends and digital epidemiology.

Types of Literature

The literature review incorporated a variety of sources to provide a robust understanding of the topic:

- Peer-reviewed journal articles: Empirical studies and systematic reviews published in high-quality, peer-reviewed journals were the primary focus. These articles provided evidence on the utility, limitations, and applications of Google Trends in public health surveillance, with particular attention to vaccine uptake and COVID-19 case surveillance.
- Grey literature: Dissertations, conference proceedings, and reports accessed through ProQuest Central were included to capture emerging research trends and innovative methodologies that may not have been widely published. This was particularly important in addressing the lack of Austria-specific studies.

- Global and regional studies: While Austria-specific literature was limited, the review included studies conducted in countries with comparable health care systems or similar socio-cultural contexts. For example, studies from European countries such as Italy and Spain provided insights into the use of Google Trends to track COVID-19 cases and vaccination behavior that could inform the Austrian context.
- Foundational works: Foundational studies and methodological frameworks, such as Mavragani and Ochoa's (2019a) exploration of Google Trends as a tool in infodemiology, were included to provide a theoretical foundation for the research.

Key Topics Covered

The literature review explored several interrelated themes to align with the study's objectives and research questions. The examination of digital health monitoring tools, with a particular focus on Google Trends, was central to the review. Studies highlighted the tool's ability to monitor public health behaviors and detect disease outbreaks in real-time, providing actionable insights for public health interventions. This area of research also included the integration of digital tools with traditional epidemiological methods, highlighting the potential for improved responsiveness and accuracy in managing rapidly evolving health crises.

COVID-19 case surveillance was another critical focus of the review. Articles examined the role of Google Trends in tracking COVID-19 cases and its effectiveness as an early warning system compared to traditional surveillance methods. Studies showed that search data often preceded reported case numbers, demonstrating the tool's potential to provide timely insights for public health planning. Comparative research further evaluated the accuracy and limitations of Google Trends, providing a nuanced understanding of its application in disease surveillance.

A key area of investigation, particularly in the context of the COVID-19 pandemic, was vaccination behavior and uptake. The review looked at studies that examined the relationship between public search behavior and vaccination trends, including factors such as vaccine hesitancy, public engagement with vaccination campaigns, and the role of digital platforms in influencing health behaviors. In addition, studies exploring the temporal relationship between infection rates and vaccination uptake were included to support the investigation of how rising case numbers may influence vaccination behavior. This topic was particularly relevant to understanding how Google Trends data could be used to monitor vaccination uptake and identify barriers to vaccination efforts.

Finally, the review considered the specific health care context of Austria, as the study focused on this region. While Austria-specific studies on Google Trends were limited, grey literature and global studies from countries with comparable health care systems provided valuable insights. These studies informed the potential integration of digital surveillance tools into Austria's public health strategies, considering its unique health care structure and socio-cultural characteristics. This contextual focus was essential for bridging the gap between global research and Austria-specific applications and laid the foundation for the study's contribution to the field of digital epidemiology.

Inclusion and Exclusion Criteria

To ensure the selection of relevant and rigorous studies, the literature review adhered to a set of well-defined inclusion and exclusion criteria. Studies were included in the review if they were published in peer-reviewed journals or credible sources of grey literature, such as dissertations or conference proceedings. Emphasis was placed on articles examining the use of Google Trends as a health surveillance tool, digital health monitoring, COVID-19 case tracking, or immunization behavior. To capture the most recent advances in this rapidly evolving field, the review emphasized studies published between 2019 and 2024, a period marked by the widespread adoption of digital tools during the COVID-19 pandemic. However, seminal papers that laid the theoretical or methodological groundwork for digital epidemiology were also included, regardless of publication date. In addition, global studies from countries with health systems or socio-cultural contexts similar to Austria were included to provide broader insights applicable to the regional focus of this study.

Exclusion criteria were applied to eliminate studies that lacked relevance to the research objectives or did not meet quality standards. Articles focusing on non-COVID-19 diseases or unrelated public health interventions were excluded, as were publications outside the specified time frame unless they provided fundamental insights into the topic. Studies that did not provide empirical or methodological contributions to the understanding of Google Trends as a public health tool were also excluded. These criteria ensured that the literature review remained tightly focused on the study objectives, using only sources that directly addressed the research questions and made meaningful contributions to the field of digital epidemiology.

Relevance to Research Questions

The scope of the literature review was directly aligned with the study's research questions:

RQ1: Is there a statistically significant relationship between weekly Google Trends search volume (for select COVID-19–related terms) and the reported weekly number of COVID-19 cases in Austria?

RQ2: Is there a statistically significant relationship between weekly Google Trends search volume (for select COVID-19–related terms) and the weekly vaccination rates in Austria?

RQ3: Is there a statistically significant relationship between weekly COVID-19 cases and weekly vaccination rates in Austria?

By focusing on studies that address these questions, the review identified gaps in the literature, particularly regarding Austria-specific applications of Google Trends. This further informed the study's aim to fill these gaps and contribute to the growing body of research on digital epidemiology.

Theoretical Foundation

The Diffusion of Innovations Theory, developed by Everett Rogers (2003), serves as the foundation for this study. This theory provides a systematic framework for understanding how new ideas, practices, or technologies are communicated and adopted within a social system over time (Rogers, 2003). The theory has been widely applied in various disciplines, including public health (Rogers, 2003). It emphasizes the role of communication channels, characteristics of the innovation, and the structure of the social system in determining the rate and success of adoption (Rogers, 2003). Applying this framework, the current study examines how public search behavior, as reflected in Google Trends data, correlates with COVID-19 case numbers and vaccination rates in Austria, including whether rising case numbers may influence vaccination uptake. This approach provides a structured lens for analyzing how digital platforms influence public health awareness and behavior during a health crisis.

Origins and Key Principles

The theoretical foundation of this study is rooted in Everett Rogers' Diffusion of Innovations Theory, which explains how new ideas, practices, or technologies spread through a social system over time (Rogers, 2003). Initially developed for the analysis of the adoption of agricultural innovations among farmers, the theory has since evolved into a versatile framework with applications in public health, education, and technology adoption (Dearing & Cox, 2018; Greenhalgh et al., 2004). Its continued relevance makes it an ideal choice for analyzing how digital health tools, such as Google Trends, influence public health behaviors, including vaccination uptake and response to rising case numbers during COVID-19 in Austria. It provides a structured approach to understanding the diffusion process.

The theory outlines several factors that influence the diffusion process. These include the characteristics of the innovation, the communication channels used, and the structure of the social system (Rogers, 2003). Innovations that offer clear relative advantages, fit with existing values, appear simple, and provide observable benefits are more likely to diffuse (Greenhalgh et al., 2004; Valente, 1996). These characteristics help explain why digital platforms such as Google Trends are effective in disseminating health-related information because they provide immediate accessibility, adaptability, and real-time insights into public interest (Mavragani & Ochoa, 2019a). In Austria, where public health messaging is influenced by both national and federal-level communication efforts, such platforms may support the diffusion of vaccine-related and COVID-19 case awareness across regions.

The role of communication is central to the Diffusion of Innovations Theory. Traditionally, the mass media and interpersonal networks have been the main channels through which awareness of innovations has been spread (Katz et al., 1963). In recent years, digital platforms have emerged as powerful tools for disseminating information, especially during public health crises (Mavragani & Ochoa, 2019a; Porcu et al., 2023). These platforms allow for faster and broader dissemination of health information, reflecting a modern extension of the theory (Mavragani & Ochoa, 2019a; Porcu et al., 2023). For example, Google Trends provides a proxy for public engagement and awareness by capturing search behavior, which aligns with the knowledge stage of Rogers' innovation-decision process (Rogers, 2003). In this study, the role of digital communication is examined at the national level to assess how public interest in COVID-19 cases and vaccination spreads across the country.

Applying this framework to the current study, the Diffusion of Innovations Theory provides a robust lens for understanding how public health information spreads through a population. By examining how Google Trends data correlates with COVID-19 case counts and vaccination rates, this study evaluates how digital platforms facilitate the spread of critical health information. This perspective is particularly relevant in the context of the COVID-19 pandemic, where timely and effective communication was vital for shaping the public health response.

Stages of the Innovation-Decision Process

The Diffusion of Innovations Theory identifies five key stages in the innovation-decision process: knowledge, persuasion, decision, implementation, and confirmation (Rogers, 2003). These stages represent the sequential steps through which individuals or groups move as they become aware of an innovation, evaluate it, adopt it, and integrate it into their behaviors or systems (Rogers, 2003). Communication channels, the nature of the social system, and the characteristics of the innovation itself influence each stage (Rogers, 2003). This makes the framework particularly relevant for analyzing digital health tools such as Google Trends in a public health context.

Knowledge is the first stage in which individuals become aware of an innovation and gain an understanding of its functionality and potential benefits (Rogers, 2003). For example, public search behavior captured by Google Trends reflects this stage: Individuals actively seek information about emerging health threats or interventions, such as COVID-19 or vaccination options (Mavragani & Ochoa, 2019a). The volume of searches during the COVID-19 pandemic demonstrates how awareness spreads and correlates with public interest in health-related topics (Porcu et al., 2023).

This is followed by the persuasion stage, where individuals form attitudes toward the innovation, influenced by attributes such as its relative advantage, compatibility, complexity, trialability, and observability (Rogers, 2003). Digital platforms such as Google Trends enhance this stage by making information highly accessible and visually appealing, which can positively influence perceptions of usability and relevance (Greenhalgh et al., 2004). For example, individuals who observe real-time trends in vaccination rates or symptom searches may develop favorable attitudes toward vaccination or testing as effective public health interventions (Greenhalgh et al., 2004).

In the decision stage, individuals decide to adopt or reject the innovation based on the knowledge and attitudes formed in the previous stages (Rogers, 2003). This stage is critical in the public health context because it reflects whether individuals translate information into actionable behaviors, such as getting vaccinated or following public health guidelines (Rogers, 2003). Search behaviors that indicate vaccination appointments or clinic locations can serve as proxies for adoption decisions (Mavragani & Ochoa, 2019a).

The implementation stage involves putting the innovation into use (Rogers, 2003). In public health, this could mean participating in vaccination programs or adopting behaviors such as wearing masks or social distancing based on information accessed through digital platforms (Rogers, 2003). Google Trends provides insight into changes in public engagement during this stage by tracking how often individuals search for instructions or guidance related to these actions (Porcu et al., 2023).

Finally, the confirmation stage occurs when individuals seek confirmation of their decision and evaluate the long-term benefits of the innovation (Rogers, 2003). For example, positive feedback about the effectiveness of vaccinations or the impact of public health interventions disseminated through both traditional and digital channels can reinforce adoption (Rogers, 2003). Google Trends can provide a lens into this stage by monitoring searches for follow-up information, such as booster shots or post-vaccination side effects (Saegner & Austys, 2022).

These five stages provide a structured lens through which this study can analyze the dissemination and adoption of COVID-19-related information in Austria. By mapping Google Trends data to these stages, this study seeks to explore how the public's search behavior aligns with the innovation decision-making process and predicts epidemiological outcomes such as case numbers and vaccination rates.

Categories of Adopters and Diffusion Curve

The Diffusion of Innovations Theory categorizes individuals within a social system based on their willingness and ability to adopt new innovations (Rogers, 2003). These categories include Innovators, Early Adopters, Early Majority, Late Majority, and Laggards, each representing different characteristics and roles in the diffusion process (Rogers, 2003). Sequential adoption by these groups creates the S-shaped diffusion curve, which reflects the cumulative rate of adoption over time (Rogers, 2003). Understanding these categories is critical to analyzing how public health behaviors, such as vaccination uptake, diffuse through a population, as captured by tools such as Google Trends.

Innovators, who make up about 2.5% of the population, are the first to adopt new innovations (Rogers, 2003). These individuals are characterized by their willingness to take risks and their high levels of social connectedness and access to information (Rogers, 2003). In the context of public health, innovators may be the first to search for and act on information about COVID-19 vaccines, influencing the initial spikes in search trends captured by Google Trends data (Porcu et al., 2023). Their early engagement often serves as a signal of broader public interest in health interventions (Porcu et al., 2023).

Following Innovators, Early Adopters (13.5%) are often opinion leaders within a social system (Rogers, 2003). They are more deliberate in their adoption decisions and have a critical role in the spread of awareness and legitimization of innovation for the broader population (Rogers, 2003). For example, during the COVID-19 vaccination campaigns, early adopters may have driven positive discourse about vaccine benefits, as reflected in search trends for vaccine safety and efficacy (Saegner & Austys, 2022). Their influence helps accelerate adoption among the Early Majority (Rogers, 2003; Saegner & Austys, 2022).

The Early Majority (34%) represents a larger, more cautious group of adopters who require evidence of the innovation's success before committing (Rogers, 2003). These individuals often rely on validation from early adopters and tangible evidence of the innovation's effectiveness (Rogers, 2003). In public health contexts, the Early Majority may correspond to those who are looking for local vaccine availability or monitoring case trends to inform their decisions (Mavragani & Ochoa, 2019a).

The Late Majority (34%) adopts innovations after most of the population has already done so (Rogers, 2003). Often driven by social pressure or the need to conform, this group requires even more reassurance and evidence before adoption (Rogers, 2003). During the COVID-19 pandemic, the Late Majority may have been influenced by widespread public health messages and consistent evidence of vaccine efficacy, as reflected in searches for booster shots or long-term vaccine effects (Porcu et al., 2023).

Finally, Laggards (16%) are the last group to adopt an innovation and are often resistant to change because of limited resources, skepticism, or strong ties to traditional practices (Rogers, 2003). In the case of COVID-19, laggards may represent populations that only engaged with vaccine-related information after significant external pressure, such as mandates or worsening local case numbers, which may have driven late spikes in search activity on Google Trends (Porcu et al., 2023).

The S-shaped diffusion curve visually represents the adoption process, beginning with slow adoption among innovators and early adopters, accelerating as the early and late majority adopt, and leveling off as the final laggards integrate the innovation (Rogers, 2003). In the context of Austria, there is currently no literature describing the adoption process as it relates to epidemic outbreaks and the use of Google Trends data. To fill in this gap, the diffusion curve serves as a theoretical basis for the interpretation of Google Trends data. Additionally, exploring whether rising COVID-19 case numbers are followed by increased vaccination uptake, which may reflect the public progressing from awareness to action in response to perceived risk will shed light on the adoption process.

Application of the Theory in Public Health

The Diffusion of Innovations Theory has been widely used in public health to analyze how new health behaviors, interventions, and technologies are adopted and diffused within populations (Rogers, 2003). The theory's structured framework for studying the diffusion of innovations is particularly valuable for understanding the factors that drive the adoption of public health interventions, such as vaccination campaigns or digital health tools (Rogers, 2003). By focusing on communication channels, social systems, and the characteristics of innovations, this framework allows researchers to identify the facilitators and barriers to public health adoption, which is critical during crises such as the COVID-19 pandemic (Dearing & Cox, 2018; Greenhalgh et al., 2004).

In public health, the theory has been applied to the study of the adoption of interventions, from vaccination programs to the use of digital platforms for health monitoring. For example, it has been applied to understand how public awareness campaigns promote vaccine uptake by targeting the knowledge and persuasion stages of the innovation-decision process (Mavragani & Ochoa, 2019a). Studies have shown that public engagement through digital platforms can accelerate the spread of critical health information, such as the benefits of vaccination or the importance of preventive measures during outbreaks (Lin et al., 2020; Porcu et al., 2023; Saegner & Austys, 2022). These findings highlight the critical role of effective communication in influencing public behavior and overcoming vaccine hesitancy (Lin et al., 2020; Porcu et al., 2023; Saegner & Austys, 2022).

The theory has also been used to examine how health technologies, including digital tools such as Google Trends, contribute to health monitoring and response. In the context of infodemiology, Google Trends has been used to analyze public search behavior as a proxy for awareness and interest in health topics, consistent with the knowledge stage of the innovation-decision process (Mavragani & Ochoa, 2019a). For example, spikes in searches for *COVID-19 symptoms* or *vaccine safety* often reflect public engagement with emerging health threats and interventions (Porcu et al., 2023). By providing real-time data on public interest, these tools support the design and implementation of targeted health campaigns (Porcu et al., 2023).

The application of the Diffusion of Innovations Theory to digital health surveillance tools is particularly relevant during pandemics, when timely dissemination of information and adoption of behaviors are critical. Studies have shown that early adopters of digital platforms, such as those who engage with online health information, can drive broader adoption in the population, ultimately influencing health outcomes such as vaccination rates or adherence to public health guidelines (Dearing & Cox, 2018; Saegner & Austys, 2022). This underscores the utility of the theory in analyzing how digital platforms facilitate the spread of health-related behaviors across different social systems (Dearing & Cox, 2018; Saegner & Austys, 2022).

Relevance to This Study

The Diffusion of Innovations Theory is central to this study and provides a comprehensive framework for analyzing how public search behavior, captured through Google Trends, aligns with the stages of the innovation-decision process and influences public health outcomes. This framework is particularly useful for interpreting how patterns of rising COVID-19 case numbers are followed by increases in vaccination uptake, analyzing how heightened perceived risk might accelerate movement from awareness and persuasion toward action and adoption within the innovation-decision process.

To understand how spikes in Google Trends data reflect public awareness of COVID-19 information and vaccination rates, the knowledge stage of the innovation-decision process, where individuals first become aware of an innovation, is particularly relevant. For example, search trends for terms such as *COVID-19 symptoms* or *vaccine safety* often correlate with public health announcements and media coverage, serving as proxies for information dissemination. Studies have shown that search behavior captured through Google Trends can indicate shifts in public interest and awareness that coincide with the timing of critical public health interventions (Mavragani & Ochoa, 2019a; Porcu et al., 2023). Similarly, the persuasion phase, in which individuals form attitudes toward innovation, is reflected in search patterns for vaccine safety, efficacy, and side effects, which are known to influence vaccine uptake (Saegner & Austys, 2022). Analyzing whether increased COVID-19 case numbers are followed by increased vaccination uptake helps interpret whether heightened case numbers prompt shifts from awareness to adoption behaviors across different communities.

The application of Diffusion of Innovations Theory not only bridges the gap between digital health surveillance and public health behavior analysis, in this case specifically for Austria, but also provides a structured lens to assess how information flows through a population during a crisis. By linking theoretical constructs to measurable variables such as search behavior and vaccination rates, this study contributes to the growing body of research in digital epidemiology. Additionally, by examining whether surges in case numbers precede increased vaccination rates, the framework enables a more nuanced interpretation of how crisis severity influences adoption behavior which is central to understanding the full process of innovation diffusion in a pandemic. These findings provide actionable insights for optimizing public health interventions in Austria and similar health systems.

Literature Review Related to Key Variables and Concepts

Digital Health Surveillance Tools

Digital health surveillance tools have become indispensable in modern public health. They offer innovative and real-time approaches to monitoring health behaviors and the spread of disease. These tools were particularly valuable in tracking public health concerns during the COVID-19 pandemic, including platforms such as Google Trends, social media analytics, and mobile applications (Bento et al., 2020; Nuti et al., 2014; Porcu et al., 2023). For example, Google Trends demonstrated its utility in identifying early signals of disease outbreaks through analysis of search terms such as *fever* and *cough*, which often preceded official case reports by weeks (Porcu et al., 2023). Similarly, studies have shown a direct correlation between online discussions and vaccine uptake, and social media platforms have provided insights into public attitudes toward vaccination campaigns and the spread of misinformation (Ahmed et al., 2019; Folotiya & Ngoma, 2024; Luo et al., 2021; Memedovich et al., 2024; Nuti et al., 2014). The ability of these tools to process large amounts of real-time, user-generated data makes them faster and more adaptive than traditional surveillance systems (Ahmed et al., 2019; Folotiya & Ngoma, 2024; Luo et al., 2021; Memedovich et al., 2024). Unlike epidemiological reports, which are often subject to delays, digital surveillance captures the immediacy of public sentiment and behavior (Bento et al., 2020; Porcu et al., 2023). However, challenges remain in ensuring the accuracy and representativeness of these data, as search volumes can be influenced by external factors such as media coverage rather than actual health trends (Effenberger et al., 2020). As public health systems evolve, especially during rapidly evolving global health crises, the integration of digital health surveillance tools will be critical to bridging gaps in data timeliness and accessibility.

Digital health surveillance tools complement traditional epidemiological methods with cost-effective and scalable approaches to public health behavior monitoring. Traditional health systems, with their reliance on clinical reporting and in-person monitoring, often struggle to collect timely data. In contrast, tools such as Google Trends and mobile health applications provide low-cost, scalable solutions that are particularly beneficial in settings where resources are limited (Saegner & Austys, 2022). For example, a study analyzing Google Trends data during the COVID-19 pandemic found that search behavior provided actionable insights into public awareness and compliance with public health measures (Bashir et al., 2021). In addition, mobile health applications enabled real-time tracking of COVID-19 exposures and contact investigations, which greatly aided containment efforts (Gendy & Yuce, 2023). The scalability and accessibility of these tools make them particularly valuable for low- and middle-income countries, where traditional surveillance infrastructures may be underdeveloped (Bashir et al., 2021; Gendy & Yuce, 2023). However, limitations such as internet accessibility and technological literacy can affect the representativeness of the data in these settings (Rovetta, 2021; Saegner & Austys, 2022). Furthermore, overreliance on digital tools without contextual data can lead to misinterpretation or overestimation of public health risks, particularly in areas with limited internet penetration. While not without limitations, digital health surveillance tools have proven to be essential in providing cost-effective and scalable solutions, especially in low-resource settings where traditional methods may fall short.

Overview of Digital Surveillance Methods

By enabling innovative tools to track disease and monitor population health behaviors, digital surveillance methods have transformed public health. These methods include electronic health records (EHRs), mobile applications, wearable devices, and social media analytics to collect and analyze data in real-time. For example, studies during the COVID-19 pandemic demonstrated the effectiveness of using EHR-based patient registries to monitor vaccination efforts and health outcomes in large populations systematically (Madhavan et al., 2021; Satterfield et al., 2021). In addition, Google Trends has been widely used to analyze search behavior related to symptoms such as *fever* and *cough*, providing early insights into emerging disease outbreaks long before official case numbers are reported (Porcu et al., 2023). These tools are particularly valuable in responding to rapidly evolving health crises, as they augment traditional surveillance methods by providing real-time, cost-effective, and scalable solutions.

Compared to traditional epidemiologic approaches, the adoption of digital surveillance methods enables more timely and comprehensive public health responses. For example, wearable devices can monitor physiological indicators such as heart rate and temperature (Cheong et al., 2022; Krishnamurthi et al., 2021). This can provide early warning of potential health problems (Cheong et al., 2022; Krishnamurthi et al., 2021). Similarly, social media platforms can serve as a valuable tool for tracking public sentiment, trends in misinformation, and behavioral responses to health interventions (Charles-Smith et al., 2015; Jiang et al., 2021). These technologies allow health officials to identify outbreaks earlier and tailor public health messages to address specific concerns, thereby improving the overall effectiveness of interventions (Charles-Smith et al., 2015; Cheong et al., 2022; Jiang et al., 2021; Krishnamurthi et al., 2021). However, the full potential of digital surveillance depends on the effective integration of these tools into existing health systems (Cheong et al., 2022; Jiang et al., 2021; Krishnamurthi et al., 2021). This ensures a unified and cohesive approach to public health management.

Despite their benefits, there are challenges related to privacy, data standardization, and equitable access to technology that must be addressed with digital surveillance methods. In rural or resource-constrained settings, limited infrastructure and technological literacy can be barriers to the effectiveness of these tools (Coetzee & Kagee, 2020). In addition, ethical concerns about the use and sharing of personal health data raise questions about privacy and trust, which are critical to the success of these systems (Garett & Young, 2022; Romero & Young, 2022). Robust legal frameworks, public awareness campaigns, and investments in digital infrastructure to bridge access gaps are needed to address these challenges (Garett & Young, 2022; Romero & Young, 2022). With continued efforts to overcome these barriers, digital surveillance methods are poised to complement and enhance traditional public health systems. This will strengthen global disease surveillance and response capabilities.

Advantages and Challenges

Digital surveillance methods have revolutionized the way public health crises are monitored. They enable the collection of real-time data on disease trends and health behaviors (Kostkova et al., 2021). In contrast to traditional methods, which often rely on delayed clinical reporting, digital tools such as Google Trends and social media platforms can provide rapid insights into changes in public interest and behavior (Kostkova et al., 2021). For example, studies show that monitoring search terms such as *fever* and *vaccine appointments* during the COVID-19 pandemic provided timely indicators of the public's engagement with health interventions and awareness of symptoms (Kostkova et al., 2021; Porcu et al., 2023). This immediacy allows public health officials to quickly implement targeted strategies, such as adjusting communication campaigns or prioritizing resources to high-need areas, thereby improving the overall responsiveness of health systems during crises.

A significant advantage of digital surveillance methods is their scalability, which allows them to be used in settings ranging from highly developed countries to resource-limited settings. For example, mobile applications have facilitated large-scale contact tracing and immunization surveillance efforts, providing critical data in countries with varying levels of health infrastructure (Gasser et al., 2020). In addition, these tools are cost-effective because they leverage existing technologies and platforms that are widely used by the public (Gasser et al., 2020; Mark et al., 2021). Such adaptability and affordability make digital surveillance particularly valuable in addressing global health challenges, including tracking emerging variants of infectious diseases or monitoring trends in vaccination hesitancy.

Despite their benefits, digital surveillance tools face significant challenges, particularly regarding the validity and equity of the data. Data collected from platforms such as Google Trends often reflect public interest rather than actual health outcomes (Jun et al., 2018). This can lead to discrepancies between digital indicators and real-world epidemiological data (Jun et al., 2018). In addition, disparities in digital literacy and internet access lead to inequalities in the representation of data that disproportionately exclude vulnerable populations, such as those living in rural or underserved areas (L. Robinson et al., 2020; Suh et al., 2022). This can skew public health strategies, as certain populations may be overlooked in data-driven decision-making processes (L. Robinson et al., 2020; Suh et al., 2022). To address these issues, researchers have called for greater standardization in data collection and validation, as well as investment in digital infrastructure to reduce access disparities.

Ethical and privacy concerns further complicate the use of digital surveillance methods. Collecting and using personal health data from search engines, mobile apps, and social media requires strong safeguards to protect users' privacy and ensure transparency (Bentotahewa et al., 2021; Mello & Wang, 2020). Studies have shown that the perceived security of their data strongly influences public trust in digital health tools (Bentotahewa et al., 2021). Many express concerns about misuse or discrimination (Bentotahewa et al., 2021; Mello & Wang, 2020). To ensure that these innovations are implemented responsibly and equitably, these issues highlight the need for robust legal and ethical frameworks to govern the use of digital health data.

Google Trends for Health Monitoring

Google Trends has emerged as a valuable public health tool, using aggregated internet search data to provide real-time insights into population behavior. Research has shown that it helps detect outbreaks earlier than traditional methods of monitoring. For example, Porcu et al. (2023) showed that search terms such as *fever* and *cough* predicted COVID-19 outbreaks in Lombardy, Italy, up to seven weeks before official laboratory reports. Similarly, a study by Venkatesh and Gandhi (2020) in India found that search trends for terms such as *COVID symptoms* and *coronavirus cases* closely tracked the number of confirmed cases, serving as an early warning system about 21 days in advance.

Google Trends has been used to understand public reactions to health interventions and vaccination campaigns, in addition to detecting outbreaks. Studies have highlighted how search data on terms such as *side effects of vaccines* and *availability of vaccines* reflect the public's hesitancy and provide actionable insights for improving health messages (Ali et al., 2021; Njororai et al., 2024). In addition, research has highlighted its relevance in tracking adherence to non-pharmaceutical interventions, such as mask-wearing and social distancing, particularly during the COVID-19 pandemic (Alo et al., 2022; Seale et al., 2020).

Yet Google Trends is not without limitations. As discussed by Robinson et al. (2020), its reliance on internet search data means that populations with limited connectivity or digital literacy are often excluded, leading to gaps in representation. In addition, search trends can be heavily influenced by media coverage or government announcements (L. Robinson et al., 2020; Sato et al., 2021; Suh et al., 2022; Venkatesh & Gandhi, 2020). As seen during the lockdown in India, search volumes reflected public anxiety rather than epidemiological patterns (Venkatesh & Gandhi, 2020). These factors highlight the need to contextualize Google Trends data within a broader epidemiological framework to ensure accurate interpretations.

Relevance and Utility

Google Trends is highly relevant to public health monitoring, as it allows for early detection of disease outbreaks. Saegner and Austys (2022) demonstrated that 83.3% of the studies they reviewed confirmed the effectiveness of Google Trends in predicting case surges during the early stages of the COVID-19 pandemic. These findings underscore the potential of Google Trends to overcome critical lags associated with traditional surveillance methods, enabling faster public health responses and resource allocation (Saegner & Austys, 2022). Its predictive power is especially valuable during rapidly evolving health crises when timely interventions can save lives (Saegner & Austys, 2022). By providing early detection capabilities, Google Trends strengthens the ability of public health systems to manage disease outbreaks, making it an essential tool for modern health surveillance.

Google Trends captures public health behaviors and psychological trends that go beyond disease detection. In Spain, Becerra-García et al. (2023) observed a spike in anxiety-related searches within a week of the increase in COVID-19 cases and after the imposition of lockdown measures. This demonstrates that Google Trends data reflects public concern in real-time (Becerra-García et al., 2023). In addition, the tool has been used to track changes in mental health and behavioral responses, providing insights for tailoring public health messages and interventions (Becerra-García et al., 2023). The ability to reflect public sentiment and behavioral shifts provides public health officials with a unique opportunity to address emerging concerns in a timely manner (Becerra-García et al., 2023). This makes Google Trends a dynamic resource for tailoring health communication strategies during crises (Becerra-García et al., 2023). The ability to monitor both epidemiological and psychological trends underscores the versatility of Google Trends as a public health surveillance tool.

Google Trends is especially valuable in resource-constrained environments. It's cost-effective and accessible. Researchers highlighted that Google Trends provides a low-cost, scalable option for health surveillance with the ability to aggregate data with high geographic precision (Donia et al., 2021; Fagherazzi et al., 2020). During the early stages of the COVID-19 pandemic, this tool was widely used to guide public health strategies without the infrastructure and logistical delays of traditional methods (Donia et al., 2021; Fagherazzi et al., 2020). This affordability and efficiency make Google Trends an ideal solution for low-resource regions where traditional data collection systems may be underdeveloped or strained during emergencies (Donia et al., 2021). The tool's ability to leverage existing digital data ensures broader coverage and faster insights, even in areas with limited public health infrastructure (Donia et al., 2021; Fagherazzi et al., 2020). By filling gaps in surveillance capacity, Google Trends strengthens global public health efforts, especially in underserved regions.

Google Trends also supports real-time adaptation of public health interventions by tracking public interest and information-seeking behavior. For example, researchers highlighted that Google Trends can identify changes in search activity that reflect shifting public priorities and concerns during health crises (Li et al., 2025; Ma & Yang, 2022). During the COVID-19 pandemic, spikes in search volume for COVID-19-related terms provided actionable data to guide public health messaging and campaigns (Li et al., 2025; Ma & Yang, 2022). This ability to track evolving public interest allows health officials to refine strategies, such as tailoring educational campaigns or prioritizing resource allocation (Li et al., 2025; Ma & Yang, 2022). The immediacy of Google Trends data ensures that public health responses remain in tune with the concerns of the real world, thereby improving the effectiveness of interventions (Li et al., 2025; Ma & Yang, 2022). By capturing dynamic shifts in the public's information needs, Google Trends improves the responsiveness of public health systems. It is a key component of adaptive health management strategies.

Strengths and Limitations

Google Trends excels at monitoring mental health trends in real-time, providing insight into psychological well-being during health crises. A study by Becerra-García et al. (2023) showed that searches for terms such as *anxiety* spiked in Spain following the declaration of the COVID-19 lockdown, reflecting the population's psychological reactions to the crisis. These trends provided timely data to inform mental health interventions during the critical period (Becerra-García et al., 2023). This ability to monitor mental health indicators in real-time allows public health authorities to understand public concerns better and allocate resources accordingly (Becerra-García et al., 2023). By highlighting spikes in public interest in mental health, Google Trends provides actionable insights for targeted health campaigns (Becerra-García et al., 2023). The ability to track mental health trends in real-time establishes Google Trends as an essential tool for managing the psychological impact of health emergencies.

Contributing to vaccine uptake efforts, Google Trends has demonstrated its utility in monitoring public interest in vaccination campaigns. Zhang et al. (2024) reported that Google Trends data correlated with public interest in COVID-19 vaccines in China, as search activity for vaccine-related terms often preceded increases in vaccination rates. This correlation highlights the potential of this tool to guide vaccination strategies (Husnayain et al., 2020; Zhang et al., 2024). By gauging public interest in vaccination, Google Trends can help policymakers identify information gaps or hesitancy and develop targeted communication strategies (Husnayain et al., 2020; Zhang et al., 2024). This is especially important to ensure the success of vaccination campaigns in a rapidly evolving pandemic (Zhang et al., 2024). Google Trends' role in monitoring vaccine interest highlights its value as a tool for addressing public health challenges during vaccination efforts.

However, the potential underrepresentation of vulnerable populations limits the utility of Google Trends data for equitable public health surveillance. Researchers found that Google Trends reflects the behavior of internet users, often excluding populations with limited digital access, such as rural communities or those in low-income regions (Badri et al., 2021; Wesson et al., 2022). These biases can skew public health analyses by overrepresenting digitally connected populations (Badri et al., 2021; Wesson et al., 2022). This limitation requires caution when interpreting Google Trends data, especially in regions with significant disparities in internet penetration (Badri et al., 2021; Wesson et al., 2022). To gain equitable health insights, it is critical to combine Google Trends data with other surveillance tools that account for underrepresented groups (Badri et al., 2021; Wesson et al., 2022). Acknowledging and addressing the sampling biases of the tool is essential to ensure its effective use in global public health initiatives.

Ambiguity in the interpretation of search terms poses a challenge to the reliable use of Google Trends in public health. Müller et al. (2021) highlighted the difficulty of distinguishing health-related searches from unrelated searches due to the broad and overlapping nature of search terms. For example, searches for *fever* may reflect seasonal illnesses rather than specific outbreaks (Müller et al., 2021). This ambiguity underscores the importance of careful keyword selection and validation to avoid misinterpreting search trends (Müller et al., 2021). Standardized methods are needed to refine the tool's use in public health research (Müller et al., 2021). Addressing issues related to keyword interpretation can improve the accuracy and reliability of Google Trends as a health surveillance tool.

COVID-19 Case Surveillance

COVID-19 case surveillance is fundamental to public health management. It provides essential data to monitor and control the spread of disease. Surveillance systems track the number of cases, hospitalizations, and mortality and form the basis for public health interventions worldwide. For example, Brunori and Resce (2020) highlighted that effective surveillance systems allowed policymakers in Italy to monitor the evolution of COVID-19 and guide decisions on control measures. In addition, Nguyen et al. (2021) emphasized that surveillance data play a critical role in evaluating interventions such as vaccination campaigns and public health guidelines. By enabling real-time tracking of disease trends, surveillance systems allow policymakers to make informed decisions about resource allocation and control strategies (Arora et al., 2019; Brunori & Resce, 2020; Higgins et al., 2020; Nguyen et al., 2021). However, gaps in these systems, particularly during periods of high demand, have highlighted their limitations in fully capturing the scale of the pandemic (Arora et al., 2019; Brunori & Resce, 2020; Higgins et al., 2020; Nguyen et al., 2021). These limitations highlight the need to strengthen surveillance methods to ensure timely and comprehensive public health responses during future crises.

Relying on testing capacity was a significant challenge in monitoring COVID-19 cases. This led to underreporting of cases in many regions. Brunori and Resce (2020) found that inconsistent testing criteria in Italy during the pandemic led to an underestimation of true infection rates, particularly in resource-constrained areas. This was exacerbated by inadequate testing infrastructure in rural and underserved areas (Brunori & Resce, 2020). These disparities in testing capacity undermined the reliability of reported data, creating gaps in understanding the true spread of the virus (Brunori & Resce, 2020; Donia et al., 2021; Liu & Rusling, 2021). Accurate and equitable testing is essential for effective surveillance, as it provides the basis for identifying outbreaks and guiding interventions (Brunori & Resce, 2020; Donia et al., 2021; Liu & Rusling, 2021). Addressing these gaps in testing capacity is critical to improving the accuracy and reliability of future surveillance efforts.

During the COVID-19 pandemic, delayed reporting in traditional surveillance systems hampered timely public health interventions. Researchers found that delays in data collection and reporting limited the government's ability to respond promptly to increases in COVID-19 cases (Ayouni et al., 2021; Harris, 2022; Tucker & Wang, 2021). This problem was particularly pronounced in the early stages of the pandemic when rapid decision-making was critical to mitigating the spread of the virus (Ayouni et al., 2021; Harris, 2022; Tucker & Wang, 2021). The time lags in traditional surveillance systems underscore the need for more agile methods that can collect and disseminate data in real-time (Ayouni et al., 2021; Harris, 2022; Tucker & Wang, 2021). Without timely information, public health officials risk missing opportunities to implement effective control measures (Ayouni et al., 2021; Harris, 2022; Tucker & Wang, 2021). Innovative approaches that reduce reporting delays can improve the responsiveness of public health systems and ensure faster action during health crises.

Differences in surveillance infrastructure between regions revealed inequalities in COVID-19 case tracking. Tan et al. (2022) highlighted that rural and low-resource areas often lacked the infrastructure to reliably track COVID-19 cases, leading to gaps in data collection. These disparities were particularly evident in non-metropolitan areas, where limited access to health care further exacerbated underreporting (Tan et al., 2022). Such disparities hinder comprehensive surveillance and weaken the ability of public health systems to allocate resources where they are most needed (Tan et al., 2022). Ensuring equitable data collection across regions is essential for an effective global response to pandemics (Tan et al., 2022). Efforts to address these disparities must prioritize infrastructure development in underserved areas to ensure a more comprehensive and accurate surveillance system.

Traditional Surveillance Methods

Traditional surveillance methods for COVID-19 relied heavily on laboratory-confirmed cases, hospitalization data, and mortality statistics and formed the backbone of public health responses. The use of these methods provided critical insights into the course of the pandemic. For example, Dai and Wang (2020) highlighted the importance of using clinical and laboratory data to ensure the accuracy of case counts and the identification of disease hotspots. Similarly, Brunori and Resce (2020) found that hospitalization rates were a reliable indicator of health system burden during the pandemic in Italy. While fundamental, these methods were often limited by delays in data collection, underreporting, and disparities in access to care (Dai & Wang, 2020). The reliance on laboratory confirmation meant that regions with insufficient testing capacity faced significant challenges in accurately tracking case numbers (Dai & Wang, 2020). These challenges highlight the need for complementary methods that can improve the timeliness and coverage of COVID-19 surveillance.

Testing capacity and criteria significantly influenced the accuracy of traditional COVID-19 surveillance data. Testing efforts were constrained by resource limitations, particularly in overburdened regions (Alvarez et al., 2023; Ibrahim, 2020; Mardian et al., 2021). These testing inconsistencies skewed epidemiologic data, complicating efforts to understand the true scope of the pandemic (Alvarez et al., 2023; Ibrahim, 2020; Mardian et al., 2021). Areas with robust testing infrastructure were able to identify and isolate cases more effectively, while under-resourced regions were left vulnerable to unchecked transmission (Alvarez et al., 2023; Ibrahim, 2020; Mardian et al., 2021). Addressing disparities in testing infrastructure is essential to improving the reliability of traditional surveillance systems during a pandemic.

Due to inequalities in access to health care and underreporting in marginalized communities, traditional methods have also struggled with data disparities. According to Becerra-García et al. (2023), surveillance systems in Spain revealed an under-representation of cases in rural and economically disadvantaged areas, where access to health care and testing was limited. This was mirrored globally, with significant gaps in collecting data from low-resourced settings (Donia et al., 2021). These disparities have reduced the comprehensiveness of traditional surveillance data (Donia et al., 2021). This has skewed public health strategies and put vulnerable populations at greater risk (Donia et al., 2021). These gaps could be addressed by integrating community-based reporting mechanisms (Donia et al., 2021). Ensuring equity in collecting and reporting data is critical for traditional surveillance systems to succeed during pandemics.

Innovative Approaches

Digital surveillance methods, including syndromic surveillance and online search behavior analysis, have greatly enhanced COVID-19 case tracking. By collecting symptom-related data from online search trends and self-reported health applications, researchers highlight the role of syndromic surveillance in COVID-19 case surveillance (Desjardins, 2020; Güemes et al., 2021). This approach had the potential for early detection of outbreaks in advance of the reporting of laboratory-confirmed cases (Desjardins, 2020; Güemes et al., 2021). Through the use of real-time digital footprints, public health authorities are able to track infection trends and public interest in health topics (Desjardins, 2020; Güemes et al., 2021). This was particularly beneficial in regions with limited laboratory capacity, where traditional surveillance methods were subject to delays (Desjardins, 2020; Güemes et al., 2021). The use of syndromic surveillance alongside traditional epidemiologic methods strengthens pandemic response strategies and improves early warning systems.

To improve the speed and accuracy of COVID-19 case surveillance, artificial intelligence (AI) and machine learning have been instrumental. Machine learning models have been used to predict outbreaks with high accuracy by analyzing large datasets, including social media activity, electronic health records, and online search queries

(Alafif et al., 2021; Malik et al., 2021; Naseem et al., 2020). In addition, AI-driven predictive analytics have successfully identified high-risk regions for COVID-19 transmission, facilitating targeted interventions (Alafif et al., 2021; Malik et al., 2021; Naseem et al., 2020). The integration of AI-driven models into public health systems enables the processing of large volumes of data in real-time, improving outbreak forecasting (Alafif et al., 2021; Malik et al., 2021; Naseem et al., 2020). These technologies reduce the burden on health systems by optimizing resource allocation and informing policy decisions (Alafif et al., 2021; Malik et al., 2021; Naseem et al., 2020). The role of AI in public health surveillance will become even more critical to preventing future pandemics as AI capabilities continue to evolve.

For community-level surveillance of COVID-19 outbreaks, wastewater-based epidemiology (WBE) has emerged as a valuable tool. Aguiar-Oliveira (2020) demonstrated the potential of wastewater surveillance to detect SARS-CoV-2 RNA in wastewater systems, providing an early warning signal for outbreaks. Bivins et al. (2020) further validated the effectiveness of WBE in tracking COVID-19 case trends in urban areas. This non-invasive and cost-effective approach allows public health officials to monitor the prevalence of infection in a large population without having to rely on individual testing (Aguiar-Oliveira et al., 2020; Bivins et al., 2020; Prado et al., 2021). In particular, the detection of asymptomatic cases and the identification of hidden transmission patterns through wastewater surveillance has proven useful (Aguiar-Oliveira et al., 2020; Bivins et al., 2020; Prado et al., 2021). Expanding the use of WBE can significantly improve public health preparedness by providing timely and reliable indicators of emerging outbreaks.

Mobile health applications and digital contact tracing technologies have been instrumental in the improvement of COVID-19 surveillance and containment. Researchers highlighted that mobile health applications facilitated tracking symptoms and reporting exposures in real-time, improving the efficiency of disease surveillance (Ming et al., 2020; Singh et al., 2020). In addition, digital contact tracing systems using smartphone-based technology enabled faster case identification and quarantine enforcement (Ming et al., 2020; Singh et al., 2020). Widespread adoption of mobile health tools has improved the speed and accuracy of contact tracing, reducing transmission rates (Ming et al., 2020; Singh et al., 2020). Particularly in densely populated areas, these innovations provide scalable solutions for pandemic response (Ming et al., 2020; Singh et al., 2020). Continued advances in mobile-based surveillance will be critical to strengthening global pandemic preparedness efforts.

Vaccination Uptake and Public Health Behavior

Vaccine uptake plays a critical role in mitigating the COVID-19 pandemic, but there are still significant differences in vaccine uptake across regions and populations. Abate et al. (2024) reported that the global uptake rate for COVID-19 vaccines was approximately 60.2%, with significantly lower uptake rates in low-income countries, such as 54.1%, compared to 64.3% in middle- and high-income countries. This variability is shaped by education level, with higher education being a positive predictor of vaccine uptake (AOR = 1.96; 95% CI: 1.20-2.73) (Abate et al., 2024). The influence of education underscores the importance of public health campaigns tailored to address vaccine hesitancy among less educated populations (Abate et al., 2024; Adu et al., 2023; Khubchandani et al., 2021; Ruiz & Bell, 2021). Higher levels of education often correlate with greater trust in science and health systems, highlighting the need for targeted interventions in underserved areas (Abate et al., 2024; Adu et al., 2023; Khubchandani et al., 2021; Ruiz & Bell, 2021). Addressing vaccine hesitancy by educating and engaging the community can improve vaccination rates and reduce disparities in COVID-19 outcomes.

Digital platforms have emerged as a significant force in influencing public health behaviors and vaccination decisions. Awijen et al. (2022) found that Google Trends revealed an increase in interest in COVID-19-related health queries during key periods of the pandemic, demonstrating the potential of digital platforms to disseminate health information effectively. In addition, Al-Dmour et al. (2020) found that public trust in vaccines increased with targeted digital outreach, highlighting the importance of online communication strategies. Digital platforms allow health officials to counter misinformation and directly address vaccine hesitancy, providing a unique opportunity for real-time public health engagement (Al-Dmour et al., 2020; Athey et al., 2023; Awijen et al., 2022; Verger & Dubé, 2020). By understanding the public’s concerns through search trends and online activity, tailored interventions can be developed to build trust and confidence in vaccines (Al-Dmour et al., 2020; Athey et al., 2023; Awijen et al., 2022; Verger & Dubé, 2020). To build public trust and promote equitable vaccine distribution, the use of digital tools is essential.

Public health practices related to vaccination are influenced by both structural and attitudinal barriers, and comprehensive approaches are needed to overcome these challenges. Kuehn et al. (2022) categorized barriers to vaccine uptake into structural factors, such as accessibility and cost, and attitudinal factors, including mistrust of the health care system and concerns about vaccine safety. Abate et al. (Abate et al., 2024) also found that positive attitudes toward vaccines significantly increased uptake rates, with a pooled adjusted odds ratio of 4.50 (95% CI: 2.89-6.12). Addressing structural barriers requires systemic solutions, such as increasing vaccine availability in underserved regions, while overcoming attitudinal barriers requires transparent communication and trust-building initiatives (Abate et al., 2024; Kuehn et al., 2022). To foster long-term trust in immunization programs, tailored strategies should focus on building relationships between health care providers and communities (Abate et al., 2024; Kuehn et al., 2022). Public health strategies need to address both structural and attitudinal challenges to ensure the widespread uptake of vaccines and equitable access to health care.

Understanding vaccination trends in specific regions, such as Austria, provides important insights for targeted public health interventions. Regional disparities in vaccination rates are often correlated with differences in public health policies and socioeconomic factors (Brunori & Resce, 2020; Sallam, 2021; Sen-Crowe et al., 2021). Nguyen et al. (2021) emphasized that tailored messaging and community-level outreach are critical to addressing these disparities. By analyzing vaccination trends in Austria and other countries, public health officials can identify gaps in coverage and tailor strategies to address local challenges (Sen-Crowe et al., 2021). Such insights are essential for promoting equitable vaccine distribution and achieving higher overall vaccination rates (Sen-Crowe et al., 2021). Regional immunization trends serve as a valuable tool for tailoring public health interventions to meet the needs of diverse populations.

Determinants of Vaccination Rates

Sociodemographic factors, including education, income, and employment status, significantly influence COVID-19 vaccination rates. Research shows that individuals with lower levels of education and household income had higher rates of vaccine hesitancy (AlShurman et al., 2021; Mondal et al., 2021). Rural residents were also more likely to be hesitant than their urban counterparts (AlShurman et al., 2021; Mondal et al., 2021). These differences highlight the importance of targeted outreach and interventions to increase immunization coverage among disadvantaged groups (AlShurman et al., 2021; Mondal et al., 2021). Public health efforts should focus on addressing structural barriers, such as accessibility and affordability, while at the same time tailoring messages to meet the needs of specific population groups (AlShurman et al., 2021; Mondal et al., 2021). Understanding the role of socioeconomic factors is critical to developing equitable immunization campaigns that prioritize marginalized groups.

Trust in vaccines and health systems is a critical determinant of vaccine uptake. Studies have shown that a lack of trust in government and health care institutions significantly affects an individual’s willingness to receive COVID-19 vaccines (Cao et al., 2024; Viswanath et al., 2021). Viswanath et al. (2021) found that conspiracy theories and concerns about vaccine safety and efficacy were major contributors to hesitancy. In addition, individuals who expressed greater trust in vaccines were more likely to be vaccinated, highlighting a positive relationship between trust and vaccine uptake (Cao et al., 2024; Viswanath et al., 2021). Building trust requires communicating transparently and engaging with communities (Cao et al., 2024; Viswanath et al., 2021). Public health campaigns should aim to counter misinformation and address safety concerns through consistent, evidence-based messaging (Cao et al., 2024; Viswanath et al., 2021). These efforts can build public trust in immunization programs and reduce hesitancy (Cao et al., 2024; Viswanath et al., 2021). Addressing trust issues is a key strategy for overcoming vaccine hesitancy and ensuring widespread immunization coverage.

Role of Digital Platforms in Vaccination Behavior

Digital platforms have a significant impact on vaccination behavior through the shaping of public perceptions and the provision of access to health-related information. Zang et al. (2023) found that social media platforms play a dual role. They serve as channels for both accurate health information and vaccine misinformation (Zang et al., 2023). While social media platforms can be used for public health campaigns to promote vaccine confidence, their vulnerability to misinformation can undermine these efforts (Maugeri et al., 2022; Piltch-Loeb et al., 2021; Zang et al., 2023). Policymakers must balance free speech with proactive measures to counter harmful content that diminishes public confidence in vaccines (Maugeri et al., 2022; Piltch-Loeb et al., 2021; Zang et al., 2023). Addressing the role of social media in the spread of vaccine-related misinformation is critical to improving vaccine uptake.

Digital tools such as online health information platforms and search engines are valuable in raising awareness of the benefits of immunization. Zhang et al. (2024) demonstrated that seeking health information online positively affects vaccination rates by helping individuals better understand the benefits of immunization. Their study showed that those who searched for vaccine information online were significantly more likely to accept the COVID-19 vaccines than those who did not search for vaccine information online (Zhang et al., 2024). Online health information platforms serve as critical resources for disseminating accurate and evidence-based information about immunization, reducing knowledge gaps, and overcoming hesitancy (Zang et al., 2023; Zhang et al., 2024). Public health campaigns can use these platforms to engage hesitant individuals and improve vaccine uptake (Zang et al., 2023; Zhang et al., 2024). The potential of targeted online campaigns to address misinformation and increase vaccination rates is underscored by the use of digital health platforms.

Countering digital disinformation is essential to mitigating the negative effects of anti-vaccine campaigns on public health behavior. Researchers highlighted that foreign disinformation campaigns exacerbate vaccine hesitancy by amplifying negative messages about the safety of vaccines (Farooq & Rathore, 2021; Germani et al., 2022). These findings highlight the need for coordinated international efforts to combat disinformation campaigns targeting vaccine programs (Farooq & Rathore, 2021; Germani et al., 2022). Governments and social media companies must work together to remove harmful content while promoting verified health information (Farooq & Rathore, 2021; Germani et al., 2022). Efforts to combat disinformation on digital platforms are critical to building public trust in vaccines and improving global immunization rates.

Vaccination Trends in Austria

In Austria, regional disparities in COVID-19 vaccination coverage have influenced incidence rates throughout the country. An ecological study analyzing data from all 94 Austrian districts found a moderate negative correlation between vaccination coverage and COVID-19 incidence rates ($r = -0.60$, $p < 0.001$) (Blasche et al., 2022). Districts with higher vaccination coverage had lower infection rates, underscoring the protective effect of vaccines (Blasche et al., 2022). These findings suggest that regional disparities in immunization coverage may be influenced by demographic factors and access to health care (Blasche et al., 2022). Addressing these disparities is critical to controlling the spread of COVID-19 in Austria (Blasche et al., 2022). Targeted public health interventions are needed to improve vaccination coverage in under-vaccinated regions of Austria.

Vaccine hesitancy in Austria is associated with factors such as trust in government and health care systems and political orientation. One study identified significant correlates of vaccine hesitancy, including lower levels of trust in government and health care institutions, younger age, and lower educational attainment (Lazarus et al., 2022). In addition, political orientation and voting behavior were associated with attitudes toward vaccination (Lazarus et al., 2022). This suggests that public health strategies in Austria should focus on building trust and addressing concerns specific to different political and social groups to reduce vaccine hesitancy effectively (Lazarus et al., 2022). An understanding of the socio-political factors influencing vaccine hesitancy can be a basis for more effective vaccination campaigns in Austria.

The introduction of additional doses of vaccine has been evaluated in Austria to improve protection in previously infected individuals. Studies assessing the efficacy of a fourth dose of the SARS-CoV-2 vaccine in previously infected individuals showed a significant reduction in COVID-19-related mortality and infection rates compared with those who received only three doses (Richter et al., 2022). This evidence supports the administration of booster doses to maintain immunity, particularly in vulnerable populations (Richter et al., 2022). Ongoing monitoring of vaccine efficacy is needed to inform booster policy (Richter et al., 2022). Ongoing evaluation of vaccination strategies will be essential to adapt to the evolving dynamics of the pandemic in Austria.

Gaps in the Literature

Several researchers have highlighted how Google Trends can act as an early warning system by capturing spikes in search queries prior to reported increases in disease cases, but the Austrian context remains understudied. For example, Porcu et al. (2023) and Zayed et al. (2023) documented that search terms such as *fever* correlated with spikes in case numbers across Europe, suggesting a strong predictive value. However, most current research groups Austria with broader European or global trends, overlooking Austria's federal government structure, which can alter the speed of data reporting and potentially affect the applicability of Google Trends. As a result, a thorough examination of how this digital surveillance tool fits into Austria's unique health care framework is lacking, creating a clear gap for country-specific studies.

Furthermore, while some work has explored the relationship between Google Trends data and vaccine-related interest, definitive findings specific to Austria remain sparse. For example, Zayed et al. (2023) showed that search volume for vaccine-related topics predicted engagement in certain Middle Eastern populations, but the interplay between Austrian socioeconomic factors and public trust in vaccination is comparatively unknown. Given Austria's culturally diverse regions and complex health care system, the relevance of generalizable findings from other regions is limited, and more localized digital epidemiology research is needed. Without clear data linking Austrian search behavior to vaccination behavior, policymakers and clinicians have insufficient evidence to tailor interventions to the country's context.

Additionally, there is a clear gap in understanding how shifts in disease incidence influence vaccination behavior. Machado et al. (2021) examined the significant fluctuations in vaccine hesitancy during the initial months of the pandemic. For example, they observed a temporary increase in parents' intent to vaccinate their children against influenza even though overall vaccine administration declined (Machado et al., 2021). This highlights the complex interactions between disease awareness, perceived risk, and vaccine uptake (Machado et al., 2021). Similarly, Huang et al. (2022) identified a significant negative correlation between national vaccination coverage across multiple countries and COVID-19 reproduction rates, suggesting that communities often respond to rising case counts by increasing immunization. However, both analyses are done at the aggregate or national level and do not account for variability within countries or the timing of these patterns at subnational levels (Huang et al., 2022; Machado et al., 2021).

No existing study has empirically assessed whether increases in case numbers directly precede increases in vaccination uptake, especially within a decentralized health system like Austria's. This study addresses this critical gap by examining the temporal association between rising case counts and increased vaccination rates across Austria and therefore offers novel insights into reactive behavioral dynamics during a pandemic.

Gaps and Opportunities for Further Research

A key opportunity is to establish robust longitudinal frameworks that move beyond retrospective correlations and toward predictive modeling. Although Porcu et al. (2023) and Saegner and Austys (2022) have shown how real-time Google Trends spikes align with infection rates, most research has not systematically tracked these spikes over time in relation to vaccine uptake. By analyzing queries like *vaccine appointment* searches over time, researchers could determine whether search volumes reliably predict fluctuations in vaccination rates, thereby informing public health strategies. Refining such long-term predictive models would allow Austrian health authorities to quickly adjust communication and resource allocation, thereby improving the country's pandemic response.

Another important area of research is deeper segmentation to determine which demographic groups or regions in Austria drive the strongest correlation between online searches and COVID-19 behavior. Current studies often report aggregate national data, but granular approaches could reveal, for example, if certain regions or age groups disproportionately drive search spikes around major health announcements. This level of detail would inform targeted public health efforts: If younger populations search for

COVID-19 booster at higher rates shortly after policy announcements, tailored messaging could improve vaccine compliance. Ultimately, combining such demographic insights with policy timelines would help create data-driven interventions that strengthen Austria's resilience to future public health crises.

Summary and Conclusion

Overall, the literature highlights how digital surveillance tools such as Google Trends can enhance real-time public health monitoring by capturing community-level interest in COVID-19 symptoms, vaccination questions, and prevention strategies. Several researchers have found that spikes in relevant search terms often predict official case increases or shifts in vaccination uptake, suggesting strong potential for early warning applications. In addition, such studies underscore the importance of timely data to inform pandemic responses, particularly in contexts where traditional reporting systems are subject to lags.

Despite these promising findings, many gaps remain in the specific context of Austria. Much of the existing research evaluates Google Trends data on a global or European scale, leaving open questions about how Austria's governance structure, health care financing, and cultural nuances may shape the utility of search behavior analysis. In addition, while previous studies have highlighted correlations between online vaccine inquiries and actual vaccination rates, systematic, longitudinal evidence linking such online behaviors to Austria's vaccination campaigns remains sparse. As a result, Austrian policymakers lack evidence-based insights for tailoring digital outreach or anticipating local vaccination trends.

At the same time, researchers have identified factors that influence COVID-19 behavior and vaccine uptake, including public trust in the health care system, sociodemographic characteristics (e.g., education level), and the timeliness of official information. While these factors are common to many countries, the extent to which they interact with Austrian-specific variables-such as the decentralized nature of health supervision or the population’s varying access to the internet-has not been thoroughly investigated. Furthermore, few studies have assessed whether increases in COVID-19 case numbers precede changes in vaccination behavior over time. This study aims to evaluate these temporal associations across Austria to better understand whether spikes in infection drive vaccine demand at the national level. This gap creates an urgent need for Austria-centric analyses that can shed light on how digital health tools can be integrated into existing public health frameworks to address emerging threats.

By focusing on Austria, this study fills at least one of these gaps by investigating the association between Google Trends metrics (symptom-, prevention-, and vaccine-related queries) and official COVID-19 indicators (case numbers and vaccination rates) in a localized setting. It also investigates the timing between infection rates and vaccine uptake, extending the discipline’s knowledge of digital epidemiology by combining established surveillance approaches with novel, real-time search data. It not only sheds light on the predictive value of Google Trends within Austria, but also identifies specific policy levers-such as targeted communication or resource allocation-that benefit from real-time digital surveillance.

These findings directly inform the methodological choices outlined in Chapter 3, which uses a correlational design to assess both spatial and temporal associations between online search patterns and epidemiological outcomes. In describing the statistical procedures and data handling protocols, the upcoming chapter addresses both the recognized strengths of using Google Trends (speed, scalability) and the caveats (potential biases, limitations in rural contexts) identified in the literature. Through this rigorously structured approach, the study determines whether Google Trends can serve as a viable complement to traditional case surveillance and vaccination tracking in Austria, thereby providing a clearer path for public health decision makers.

Chapter 3: Research Method

Introduction

This chapter lays the methodological groundwork for an investigation into how Google Trends data on COVID-19-related search terms correlate with reported case numbers and vaccination rates, as well as how these reported case numbers and vaccination rates may be temporally related across Austria. The primary goal of this study was to use publicly available epidemiologic data and online search metrics to identify patterns that can inform pandemic response strategies. Although numerous researchers have examined the public health implications of internet search behavior, no study has specifically assessed these dynamics in Austria. As such, this research design sought to apply a rigorous quantitative approach that relies on correlational methods to explore both spatial and temporal linkages between digital search patterns and real-world health outcomes. In doing so, it aims to provide valuable insights for public health officials by demonstrating how online interest in COVID-19 aligns with epidemiological indicators and how infection surges relate to changes in vaccination behavior over time.

A correlational, non-experimental design is appropriate for several reasons. First, it allows for the systematic examination of existing data without altering or manipulating variables, which is consistent with the ethical and logistical constraints of studying a health crisis as it unfolds (Burkholder et al., 2020; Daniel & Cross, 2019). Second, it enables the detection of associations between different forms of data, such as online searches and epidemiologic metrics, in a time-sensitive manner (Burkholder et al., 2020; Daniel & Cross, 2019). Because the intent is to determine whether increased search activity for specific COVID-19 symptoms, prevention methods, or vaccine-related terms is consistent with increases or decreases in cases and vaccine uptake, and whether increased case numbers correlate with subsequent increases in vaccination uptake across Austria, correlation-based analyses can reveal both the strength and direction of such relationships (Burkholder et al., 2020; Daniel & Cross, 2019). This design is further justified by the ready availability of historical search data and public health records, which ensures feasibility and reduces participant burden (Burkholder et al., 2020; Daniel & Cross, 2019).

The methodological plan focuses on two main constructs: the online public's interest in COVID-19, captured by aggregated Google Trends search frequencies, and real-world epidemiological indicators in Austria, measured by reported COVID-19 case numbers and vaccination rates. These constructs correspond directly to the research questions, which address whether search volume correlates with case counts and, separately, whether search volume correlates with vaccination rates over time. Additionally, the questions examine whether case counts and vaccination rates are correlated with each other. Operationally, the Google Trends data serve as the independent variable, while official Austrian Ministry of Health data on COVID-19 incidence and vaccine uptake serve as the dependent variables. For the analysis of the relationship between case counts and vaccine uptake, both variables were treated as dependent and independent across time to assess bidirectional associations. Using standard tools such as time series correlation or regression, the study can assess the extent to which fluctuations in search behavior reflect real-time developments of the pandemic (Daniel & Cross, 2019).

Underlying this methodological approach is Rogers' Diffusion of Innovations Theory, which suggests that innovations, such as the search for information about a new disease or vaccination, spread through a social system across specific channels over time (Rogers, 2003). By combining this framework with a correlational design, the study can better interpret whether online search trends might serve as a proxy for broader diffusion processes, such as increased awareness or adoption of protective behaviors. The ability of Google Trends to capture spontaneous public interest offers a unique vantage point (Nuti et al., 2014). If spikes in health-related searches precede or coincide with epidemiological shifts, this would underscore the value of search engine data as an early signal for public health planning and resource allocation (Nuti et al., 2014). It would also suggest that the diffusion of COVID-19-related knowledge is occurring dynamically through digital platforms, potentially shaping the pace of vaccine uptake and adherence to precautionary measures (Nuti et al., 2014; Rogers, 2003).

This chapter also outlines the procedures for accessing, cleaning, and analyzing data to ensure methodological transparency. Ethical considerations, such as ensuring that no individually identifiable data is used and that all datasets are public or previously anonymized, are also addressed. Because Google Trends information is aggregated at the population level, no personally identifiable information is collected, minimizing confidentiality concerns. Similarly, case numbers and vaccination rates reported by the Austrian Ministry of Health are generally disseminated in an anonymized, aggregated form, which allows for robust data analysis while maintaining standards of privacy and responsible data handling.

Overall, this methodology chapter presents a comprehensive plan for investigating how digital footprints, in the form of online search behavior, can inform and predict public health metrics during the COVID-19 pandemic in Austria, as well as how increased case numbers can inform vaccination rates. The correlational design is intended to provide insight into whether there is a meaningful relationship between the volume of relevant online searches and shifts in infection rates or vaccination uptake, without implying direct causality. By systematically addressing these questions, the study aligns with the broader goal of enhancing pandemic preparedness and response by providing a data-driven lens to inform communication and intervention strategies. The remainder of Chapter 3 presents the details of this approach, including data sources, measures, analytical techniques, and steps to address potential threats to validity.

Research Design and Rationale

The research design for this investigation is fundamentally correlational, and it was chosen to quantify the relationships between online search activity for COVID-19 and two important public health outcomes: reported COVID-19 cases and vaccination rates. Additionally, the correlational design was used to quantify the relationship between COVID-19 cases and vaccinations. Correlational studies measure variables as they exist in natural settings rather than being manipulated by the researcher, making this approach well-suited for studying a pandemic where ethical and practical constraints prohibit experimental interventions (Burkholder et al., 2020; Daniel & Cross, 2019). By using publicly available data from official public health databases in Austria as well as Google Trends, the design allows for a systematic investigation of how fluctuations in public search interest may reflect or precede changes in epidemiological indicators.

The main rationale for using a correlational design is the lack of direct control over the independent variables. Because it is impossible or unethical to manipulate the amount of information the public searches for on COVID-19 or the number of cases or vaccinations, correlational methods enable these phenomena to be observed under real-world conditions (Babbie, 2017; Daniel & Cross, 2019). In addition, correlational analysis has been widely used in infodemiology studies, where internet data such as search frequency serve as proxies for public awareness or concern (Burkholder et al., 2020; Nuti et al., 2014). The non-experimental nature of this design thus allows for a robust examination of whether higher online interest coincides with or predicts higher COVID-19 incidence or vaccine uptake, and whether higher COVID-19 cases precede higher vaccination rates. This includes assessing these relationships across Austria, ultimately providing insight into how these phenomena may be related over time.

Time and resource limitations also support the appropriateness of this approach. Because this study intends to use existing public data, the heavy lifting of data collection, such as compiling case numbers or vaccination updates, was already completed by government agencies. Google Trends also provides aggregated, anonymized daily or weekly search statistics free of charge (Nuti et al., 2014). This reliance on secondary sources allows researchers to bypass the logistical complexities and costs of recruiting human participants, thereby reducing the time and resources required (Babbie, 2017).

Furthermore, by employing a design that can work with pre-collected data, the study can more feasibly capture the breadth of the pandemic timeline in Austria without requiring lengthy fieldwork (Babbie, 2017).

In terms of alignment with the research questions, the correlational method directly addresses RQ1, RQ2, and RQ3 while acknowledging their subtle differences. RQ1 focuses on search terms related to COVID-19 infection (e.g., symptoms, tests, pandemic-related phrases) and their correlation with the number of cases in Austria. The dependent variable in RQ1 is the reported incidence of COVID-19, which may have a different temporal pattern than vaccination rates. RQ2, on the other hand, focuses on vaccine-specific questions, terms such as *Pfizer vaccine*, *Moderna vaccine*, or *vaccine side effects*, and examines their statistical association with population vaccination rates. Finally, RQ3 explores the relationship between case numbers and vaccine uptake. While all questions use similar types of search data, the analytic approach must account for these different dependent variables, requiring separate correlation tests and possibly different sets of control variables or time lags.

Despite the shared basic design, the study may employ slightly different correlational techniques depending on the research focus. For example, RQ1 may be best served by cross-correlation or time-series analyses that account for potential lags between increased search volume and subsequent case increases or decreases (Daniel & Cross, 2019). RQ2 might benefit from a similar cross-correlation approach, but looking specifically for lead/lag relationships between vaccination-related queries and the number of new vaccines administered in a given period (Daniel & Cross, 2019). In all cases, the goal is to determine whether a pattern in search behavior meaningfully aligns with changes in epidemiological data, or if epidemiological data correlates with one another, thus fulfilling the rationale that correlational analyses can reveal how public online search interests and real-world health outcomes move in parallel.

Another key factor is that correlational studies do not claim causation (Burkholder et al., 2020; Frankfort-Nachmias et al., 2021). This is particularly important in a rapidly evolving context, such as COVID-19, where external factors, including policy interventions, changing media coverage, or new variants, can affect both search behavior and health outcomes (Frankfort-Nachmias et al., 2021). By acknowledging the correlational design, the study makes it clear that it aims to measure associations rather than definitive causal pathways. Nevertheless, these findings can inform public health responses by suggesting how health authorities might monitor search trends as an indicator of increasing concern or acceptance of vaccines, potentially triggering timely interventions.

The chosen approach fits well with the broader theoretical lens provided by Rogers' Diffusion of Innovations Theory. In this theory, awareness and acceptance of an innovation, such as public health interventions or new vaccines, spread through information channels over time (Rogers, 2003). Through a correlational study of online search data, one can observe how the public's search for COVID-19 and vaccine information may reflect progressive stages of adoption (e.g., from initial awareness to active decision-making) (Burkholder et al., 2020; Daniel & Cross, 2019; Rogers, 2003). Furthermore, isolating searches that spike prior to changes in epidemiological data could reveal how the innovation-decision process unfolds in real time, supporting or extending the existing theoretical model.

In summary, a correlational design provides a resource-efficient, ethical, and methodologically appropriate way to address all three research questions (Burkholder et al., 2020; Daniel & Cross, 2019; Frankfort-Nachmias et al., 2021). Although the same overarching approach applies, it was adapted to the different dependent variables. The design's reliance on existing data naturally aligns with time and resource constraints, and the planned analyses logically map to the scope of each question. Together, these elements ensure that the study's design is consistent with its goals, feasible within practical constraints, and capable of generating meaningful insights into how digital search patterns relate to the evolving dynamics of the COVID-19 pandemic in Austria.

Methodology

This section outlines the methodology used to conduct the correlational study, detailing how the population, sampling, data collection, and measurement instruments are managed. The goal is to provide a transparent roadmap that others can follow to replicate or critically evaluate the procedures. Whether using publicly available epidemiological data or accessing Google Trends, each step of the research process must be systematically planned to ensure validity, reliability, and alignment with the study's objectives.

By incorporating established best practices in quantitative research, this methodological section aims to mitigate potential biases and optimize data quality. It begins by identifying the target population and explaining how relevant data sets were selected to capture digital behavior and public health metrics in Austria. From there, the discussion will move to specific sampling considerations, providing insight into how the data period was selected, how time series points were defined, and whether any specific filters or criteria were applied. Finally, it will highlight instrumentation considerations, including how search volume and epidemiological variables are operationalized to ensure that each measure comprehensively supports the study's three key research questions.

Population and Setting

The target population for this study is the entire Austrian population as it relates to both COVID-19 incidence and vaccination data. By examining Austria as a whole, the research includes all of Austria's population, ensuring that the resulting findings are broadly generalizable and not limited to specific locations (BMSGPK, n.d.-a; Frankfort-Nachmias et al., 2021). Austria's public health landscape, led by the Federal Ministry of Social Affairs, Health, Care and Consumer Protection (BMSGPK), provides regularly updated epidemiological data, making it an ideal setting for a nationwide quantitative study (BMSGPK, n.d.-a). This multilevel approach directly supports the study's three research questions by enabling comparisons between online search behavior, reported COVID-19 case numbers, and vaccination rates in Austria. It reflects the underlying goal of the study, which is to link digital search behavior with concrete health outcomes at a scale that captures regional differences in pandemic patterns and responses.

Because the study focuses on the period when COVID-19 was actively circulating, the chosen starting point for case data is February 26, 2020, which marks the first confirmations of COVID-19 infections within Austria's borders (BMSGPK, n.d.-b). Meanwhile, vaccination data begin on December 27, 2020, which represents the earliest recorded administration of a COVID-19 vaccine in the country (BMSGPK, n.d.-b). Aligning with these start dates ensures that the dataset reflects a meaningful span of both infection trends and vaccination uptake, captured across Austria.

The Austrian context also lends itself well to the study of diverse health behaviors through a unified public health infrastructure. All official COVID-19 statistics, whether infection- or vaccination-related, are derived from centralized databases maintained or overseen by government-affiliated agencies (BMSGPK, n.d.-b). These databases facilitate consistent definitions (e.g., confirmed cases, documented vaccinations) and uniform data collection protocols, thereby minimizing systematic measurement error (BMSGPK, n.d.-b). This uniformity, in turn, allows for more accurate correlations between digital research data and real-world metrics across Austria.

In addition, Austria's high internet penetration and robust digital infrastructure make it plausible that Google Trends data reflects a significant portion of the population's information-seeking behavior (Statistik Austria, n.d.). Although some inequalities in digital literacy or internet access are inevitable, Austria's relatively widespread connectivity reduces concerns that observed online activity comes from a small or unrepresentative segment of society (Statistik Austria, n.d.). By examining the country as a whole, rather than selectively sampling specific communities, the study captures the broader dynamic between search interest and epidemiological realities at the national level.

Sampling Timeframe and Strategy

The sampling strategy is anchored in two primary data streams, each with defined date ranges and weekly aggregations, to promote methodological consistency. First, the COVID-19 case data obtained from the BMSGPK began on February 26, 2020, and extended through June 29, 2023 (Statista, n.d.). These daily data points are aggregated into weekly intervals, typically Monday through Sunday, to smooth out short-term fluctuations and facilitate comparisons across the study timeline. For the first partial week (February 26-March 1), daily totals were pooled to not exclude early reported cases. This approach ensures that the dataset captures the early phase of the pandemic, while standardizing for subsequent weeks.

Second, vaccination data were obtained from the records of the Dachverband der österreichischen Sozialversicherungsträger, which comprehensively records COVID-19 vaccinations administered starting on December 27, 2020 (data Austria, n.d.). Because this start date is later than the first reported infections, the overlapping time frame for analyzing case-vaccination associations begins in late December 2020 and ends on June 29, 2023 (data Austria, n.d.). This decision allows for direct correlations or other statistical comparisons between weekly infection counts and vaccination rates for the period during which both data streams are available (Burkholder et al., 2020). For all periods prior to December 27, 2020, case data were included only if they were relevant to the first research question (i.e., correlation with search trends for case-related terms), thus maximizing data use while respecting the limitations of the vaccination dataset.

When aggregating daily data to weekly totals, the reported values for each day are summed (Frankfort-Nachmias et al., 2021). If there are no recorded or reported cases or vaccinations on a given day, it is counted as zero, although the potential misclassification of *no data* as *zero* is noted (Frankfort-Nachmias et al., 2021; Szklo & Nieto, 2019). Where possible, anomalies, such as negative adjustments, retroactive additions of backlogged cases, or large jumps due to data processing delays, were documented (Frankfort-Nachmias et al., 2021; Szklo & Nieto, 2019). The study kept a record of such occurrences and explanatory notes were provided in the final report to help readers understand any sudden spikes or dips that are markedly different from the underlying trends (Frankfort-Nachmias et al., 2021; Szklo & Nieto, 2019).

By adopting this weekly aggregation scheme, the study balances data granularity and clarity (Frankfort-Nachmias et al., 2021; Szklo & Nieto, 2019). Daily observations can vary dramatically due to weekend reporting lags or administrative backlogs, making the data noisy and potentially obscuring true patterns (Szklo & Nieto, 2019). Weekly totals, on the other hand, provide enough resolution to identify meaningful trends while partially mitigating the effects of short-term volatility (Frankfort-Nachmias et al., 2021; Szklo & Nieto, 2019). This unified weekly framework makes the dataset more amenable to correlational, cross-correlation, or time-series methods, while still providing an accurate representation of the evolving course of the pandemic (Frankfort-Nachmias et al., 2021; Szklo & Nieto, 2019).

In addition to the primary variables, this study extracted weekly Google Trends data on influenza-related search terms (e.g., *influenza* + *grippe*) within the same time frame. This helped identify periods when the public is more interested in seasonal influenza. Including these supplementary data points enhances internal validity by identifying temporal overlap between search behavior related to flu and to COVID-19. If distinct spikes in influenza-related search volume were observed during the study period, those weeks were flagged for exclusion or further examination during sensitivity analyses. This approach supports a more accurate interpretation of the association between search volume and outcomes of interest by reducing the likelihood that confounding seasonal flu activity distorts the primary correlations under investigation.

Overall, this sampling strategy ensures that each research question — whether investigating the association between online searches and COVID-19 case incidence, examining how vaccine-related searches align with vaccination uptake or researching whether increased case incidence precedes increased vaccination counts — relies on a comparable structure of weekly data. Because the entire population and all nine states are included, there is no further need for randomized sampling at a more granular level. Instead, the sample is effectively a census of Austrian COVID-19 numbers during the defined windows, allowing for robust population-level inferences about how digital search activity may or may not parallel COVID-19 dissemination and vaccination uptake.

Instrumentation and Operationalization of Constructs

This study relies on two main data sources to operationalize the independent and dependent variables: Google Trends metrics for COVID-19-related search terms and official Austrian epidemiological data obtained from national health authorities. While these data streams do not constitute *published instruments* in the traditional sense (e.g., a validated survey), they have been vetted in previous research, and each adheres to established data collection protocols. The rationale for selecting these sources is based on their public availability, consistent updates, and relevance to the research questions focused on COVID-19 case incidence and vaccination rates (BMSGPK, n.d.-b; data Austria, n.d.).

Google Trends Data

Developed by Google in 2006, Google Trends provides a Relative Search Volume (RSV) index, which measures how often specific search terms are entered relative to the highest point of search volume for a given region and time (Google News Initiative, n.d.). Researchers in a variety of fields have demonstrated that Google Trends can reliably indicate public interest in health issues, sometimes in line with epidemiological indicators (Mavragani & Ochoa, 2019a; Porcu et al., 2023). Although not a survey instrument per se, Google Trends has a form of face validity for infodemiological applications, as spikes in searches often coincide with increased public concern or awareness (Google News Initiative, n.d.; Mavragani & Ochoa, 2019a). Because Google makes these metrics publicly available without individual user data, no explicit permission is required to use the platform; however, the researcher must adhere to Google's public data use policy (Google News Initiative, n.d.). From a reliability standpoint, Google Trends normalizes and samples raw search data, which can introduce variability (Google News Initiative, n.d.). However, published studies comparing Google Trends results to real-world outcomes (e.g., reported flu cases, hospital admissions) have shown moderate to strong correlations, suggesting that the aggregated search index is sufficiently consistent for broad, population-level analyses (Google News Initiative, n.d.; Mavragani & Ochoa, 2019a).

Official COVID-19 Case and Vaccination Data

On the Austrian side, epidemiologic data are collected and disseminated by the Federal Ministry of Social Affairs, Health, Nursing and Consumer Protection (BMSGPK) and the Federation of Austrian Social Insurance Institutions (Hauptverband der österreichischen Sozialversicherungsträger; (BMSGPK, n.d.-b; data Austria, n.d.). These organizations provide daily counts of new COVID-19 cases and documented vaccinations, which serve as the two primary dependent variables in this study (BMSGPK, n.d.-b; data Austria, n.d.). Although these data do not have a single named "developer," they are governed by official protocols that standardize reporting formats, diagnostic criteria (for cases), and documentation processes (for vaccinations) (BMSGPK, n.d.-b; data Austria, n.d.). This standardization is critical to reliability, ensuring that each day's case numbers reflect consistent definitions of what constitutes a "confirmed case" and that each recorded vaccination follows established procedures for recording date, dose, and location (BMSGPK, n.d.-b; data Austria, n.d.; Frankfort-Nachmias et al., 2021). While authorities occasionally revise past data to address backlogs or double counting, these corrections are generally transparent and publicly reported (BMSGPK, n.d.-a; data Austria, n.d.).

Appropriateness and Data Integration

Both Google Trends and Austrian administrative data were selected for their direct relevance to the study's research questions: the correlational relationship between public search behavior (an independent variable) and actual health outcomes (dependent variables: cases and vaccinations). Evidence for the appropriateness of using Google Trends in epidemiological research can be found in previous literature, where similar designs have provided valuable insights into how online behavior reflects offline disease dynamics (Saegner & Austys, 2022). Furthermore, the Austrian COVID-19 statistics are meticulously updated and recognized as accurate representations of national trends; thus, combining these administrative records with population-wide search data provides a robust framework for studying digital footprints in a pandemic context (BMSGPK, n.d.-b; data Austria, n.d.). Because the data are publicly available and aggregated, there is no need for formal "permission letters" (BMSGPK, n.d.-a; data Austria, n.d.).

Reliability and Validity Evidence

Researchers have tested the reliability and validity of Google Trends by comparing its results to laboratory-confirmed disease incidence, hospitalization rates, and even emerging disease outbreaks (Mavragani & Ochoa, 2019a). These studies often report correlation coefficients that range from moderate (0.4-0.6) to strong (greater than 0.7) and vary by disease, region, and timeliness of reporting (Mavragani & Ochoa, 2019a). While user demographics remain somewhat ambiguous, the normalization approach used by Google (scaling search volume to a maximum of 100) ensures that comparisons across time are possible (Google News Initiative, n.d.; Mavragani & Ochoa, 2019a). Austrian case and vaccination data, in turn, follow strict government reporting standards (BMSGPK, n.d.-a; data Austria, n.d.). While small discrepancies may occur (e.g., data entry backlogs, weekend delays), these are recognized as minor, and official bulletins typically highlight significant anomalies or corrections (BMSGPK, n.d.-a; data Austria, n.d.). Thus, in accordance with standard epidemiologic practice, the data sources collectively have sufficient reliability (consistent reporting protocols) and validity (strong correspondence to real-world phenomena) to address the study's research questions (Frankfort-Nachmias et al., 2021).

Operational Definitions of Key Variables

COVID-19 Incidence: Defined as the number of newly reported cases in a given week, as reported by the BMSGPK (BMSGPK, n.d.-b). Data are aggregated from daily totals into a weekly sum (Monday-Sunday), with partial weeks at the beginning or end of the data period treated as needed (e.g., the first week may only include data from midweek through Sunday).

Vaccination rates: Defined as the total number of weekly vaccinations administered and recorded by the Federation of Austrian Social Insurance Institutions for each Monday-Sunday interval (data Austria, n.d.). Data for days with missing explicit entries are treated as zero if there is no subsequent backlog correction, but any known anomalies are flagged in the analysis notes.

Search volume (RSV): Extracted from Google Trends by specifying region (Austria), time window, and search terms relevant to either COVID-19 cases or vaccinations as outlined in Table 1. The RSV is an index out of 100 that indicates the relative interest in this term compared to its peak popularity during the selected time period (Google News Initiative, n.d.). For example, a value of 100 indicates the highest volume recorded, while 50 indicates half that volume (Google News Initiative, n.d.).

Google Trends data is also aggregated in weekly intervals to match epidemiological measures.

Table 1

Summary of Research Variables

Research Question	Variable Name	Variable Type	Coding
RQ1	Weekly COVID-19 Case Count (Aggregated Mon–Sun)	Dependent Variable	Description: Number of newly confirmed COVID-19 cases each week, based on official Austrian data from the Bundesministerium für Soziales, Gesundheit, Pflege und Konsumentenschutz (BMSGPK) (BMSGPK, n.d.-b). Coding: A numeric count for each Monday–Sunday interval. Partial weeks (e.g., beginning or end of the dataset) are included by summing all available days. Zero if no new cases reported in a given day.
RQ1	Google Trends (Case-Related Terms)	Independent Variable	Description: Aggregated weekly Relative Search Volume (RSV) on a 0–100 scale (Google News Initiative, n.d.) for each of the following terms: - *COVID-19 / Corona / Coronavirus* - *Fever / Fieber* - *Cough / Husten* - *Shortness of breath / Atemnot* - *Loss of taste / Geschmacksverlust* - *Loss of smell / Geruchsverlust* - *COVID-19 testing / Coronavirus testing / Corona Test* - *COVID-19 symptoms / Coronavirus symptoms / Corona Symptome* - *COVID-19 treatment / Coronavirus treatment / Corona Behandlung* - *COVID-19 prevention / Coronavirus prevention / Corona Prävention* - *COVID-19 variants / Coronavirus variants / Corona Varianten* - *Hand sanitizer / Handdesinfektionsmittel* - *Face mask / Maske / Gesichtsmaske* - *Social distancing / Abstand halten* - *Quarantine / Quarantäne* - *Pandemic / Pandemie* Coding: Each term's daily RSV is aggregated to a weekly value (Monday–Sunday). English and German terms may be combined using the plus operator (e.g., *Fever + Fieber*). A numeric index (0–100) is produced for each interval, with 100 representing peak volume (Google News Initiative, n.d.).
RQ2	Weekly COVID-19 Vaccination Count (Aggregated Mon–Sun)	Dependent Variable	Description: Number of vaccines administered each week, based on the Dachverband der österreichischen Sozialversicherungsträger records (data Austria, n.d.). Coding: A numeric count of doses administered for each Monday–Sunday interval. Zero if no vaccinations are recorded in a given day. Partial weeks are handled by summing available days.
RQ2	Google Trends (Vaccine-Related Terms)	Independent Variable	Description: Aggregated weekly RSV (0–100) (Google News Initiative, n.d.) for each of the following terms: - *Pfizer vaccine / Pfizer Impfung* - *Moderna vaccine / Moderna Impfung* - *AstraZeneca vaccine / AstraZeneca Impfung* - *Vaccine side effects / Impfnebenwirkungen* - *Booster shot / Booster Impfung / Auffrischungsimpfung* Coding: Similar to case-related queries, daily RSV is averaged or summed to a weekly total. English and German phrases may be combined (e.g., *Pfizer vaccine + Pfizer Impfung*). A numeric index (0–100) is produced, with 100 as the term's peak volume in the specified timeframe (Google News Initiative, n.d.).
RQ3	Weekly COVID-19 Case Count (Aggregated Mon–Sun)	Independent Variable	Description: Number of newly confirmed COVID-19 cases each week, based on official Austrian data from the Bundesministerium für Soziales, Gesundheit, Pflege und Konsumentenschutz (BMSGPK) (BMSGPK, n.d.-b). Coding: A numeric count for each Monday–Sunday interval. Partial weeks (e.g., beginning or end of the dataset) are included by summing all available days. Zero if no new cases reported in a given day.
RQ3	Weekly COVID-19 Vaccination Count (Aggregated Mon–Sun)	Dependent Variable	Description: Number of vaccines administered each week, based on the Dachverband der österreichischen Sozialversicherungsträger records (data Austria, n.d.). Coding: A numeric count of doses administered for each Monday–Sunday interval. Zero if no vaccinations are recorded in a given day. Partial weeks are handled by summing available days.

The variables shown in Table 1 reflect weekly national-level data from February 26, 2020, through June 29, 2023, using a Monday-Sunday interval. This approach helps mitigate anomalies in daily reporting, such as weekends or holidays, that could introduce short-term fluctuations in the data (Burkholder et al., 2020; Szklo & Nieto, 2019). Any partial weeks at the beginning or end of the dataset (e.g., if the study period begins in the middle of the week) are included by aggregating all available days within that partial window, ensuring that no information is omitted at random (Burkholder et al., 2020).

A key element of the data collection process is the bilingual approach to Google Trends queries (Mavragani et al., 2018). Since Austrian internet users may search for COVID-19 information in both English and German, the English and German variants of each term are combined using the plus operator (e.g., *Fever + Fieber*) when extracting the daily relative search volume (RSV). This method increases the likelihood of capturing an accurate representation of search interest across Austria, rather than being limited to just one language (Mavragani et al., 2018). After retrieving the RSV for each search term, the values are aggregated to a weekly number so that they can be directly compared to the official COVID-19 case and vaccination counts for the corresponding Monday-Sunday interval. This bilingual and temporal aggregation approach is applied consistently across Austria.

Additionally, it is important to note that Google Trends data are scaled relative to the highest observed search volume for each keyword within the selected time frame, meaning that the maximum value of 100 represents peak popularity rather than an absolute number of searches (Google News Initiative, n.d.). Because each keyword is normalized independently, it is best to analyze them in relation to epidemiological data - rather than strictly comparing the RSV of one search term to another, unless careful control of date ranges and other factors is ensured (Burkholder et al., 2020; Google News Initiative, n.d.; Mavragani et al., 2018). Despite these nuances, previous research in infodemiology has shown that Google Trends provides a sufficiently robust metric for exploring whether public information-seeking patterns align with real-world health metrics such as case incidence or vaccination uptake (Mavragani et al., 2018; Mavragani & Ochoa, 2019a).

To account for the potential confounding effects of coinciding seasonal influenza activity, this study extracted weekly Google Trends data on influenza-related keywords (e.g., *influenza* + *grippe*) from Austria throughout the entire sampling period. Although these search terms were not included as independent variables in the primary correlation analyses, they served a contextual function. Specifically, they were used to flag time periods with unusually high influenza-related search volume that could interfere with interpreting the behavior of searches related to COVID-19. A predefined threshold for unusually high influenza search volume was determined based on historical patterns, and these flagged weeks were excluded from the main analyses or examined separately in sensitivity analyses to determine if the observed associations persist when periods of high flu activity are removed. This strategy reinforces the validity of the study’s conclusions by minimizing potential signal distortion caused by overlapping respiratory illness trends.

Basis for Development, Reliability, and Validity Strategies

Although neither data source is a traditional “instrument” designed solely for research with a known developer name, test-retest reliability, or published Cronbach’s alpha, the credibility of each source has been confirmed through use in government policy (BMSGPK, n.d.-a; data Austria, n.d.). Google Trends, in particular, has been the subject of methodological reviews exploring how best to interpret its standardized outputs (Mavragani & Ochoa, 2019a). Meanwhile, Austria’s COVID-19 data serve as the basis for national pandemic strategies, with robust procedures ensuring that each case or vaccination is verified by health care providers before being entered into the system (data Austria, n.d.). Potential threats to validity — such as differences in internet access, reporting delays, or changes in testing behavior over time — are mitigated by analyzing aggregated weekly data and explicitly documenting any significant data anomalies.

Sufficiency to Answer Research Questions

Together, these two data sets are sufficient to explore whether online interest correlates with actual trends in pandemic progression and whether increased COVID-19 case numbers precede increased vaccination numbers. RQ1 examines how search terms related to COVID-19 symptoms or infection might parallel or precede fluctuations in weekly case counts, while RQ2 examines how vaccine-related searches might reflect weekly vaccination uptake. The key premise is that changes in public attention, as measured by Google Trends, may manifest either concurrently or in anticipation of epidemiological shifts. RQ3 extends this analysis by examining whether increases in COVID-19 case incidence precede rises in vaccine administration, independent of online search behavior. By clearly defining the operational parameters of each variable, the study can assess the strength and direction of these relationships through standard correlational analyses, thereby providing evidence for or against the hypothesis that digital search activity can serve as an indicator of real-world health behavior in Austria.

Conclusion

Although this study does not employ a conventional published survey or experimental intervention, the Google Trends metrics and the Austrian health databases serve as robust, data-driven “instruments” for assessing the association between public online behavior and quantifiable COVID-19 outcomes. Both have demonstrated reliability and validity in previous work, but in different ways: Google Trends through empirical correlation studies across countries and diseases, and official Austrian data through systematic health reporting structures (BMSGPK, n.d.-a; data Austria, n.d.; Google News Initiative, n.d.). By operationalizing these constructs with explicit weekly aggregation procedures and transparent definitions, the study meets the key requirements of rigor, reproducibility, and relevance, providing a strong foundation for the analyses described in the following sections (Burkholder et al., 2020; Szklo & Nieto, 2019).

Data Analysis Plan

The software used for all quantitative analyses was SPSS (version 29), with Python (version 3.13.3) used only for a few diagnostic tests that SPSS v29 does not provide (Breusch-Pagan and White tests for heteroscedasticity, and robust standard errors including HC3 and Newey-West/HAC). Microsoft Excel was used to organize and aggregate the raw data into weekly intervals, while SPSS was used for inferential statistics and the generation of summary statistics. Python was used only for the figures tied to those extra diagnostics. The dataset was first screened for missing or outlying values in SPSS. Days with no officially reported cases or vaccinations were recorded as zero if it is assumed that the absence of data reflects zero true events. If there were partial weeks at the beginning or end of the timeframe, the data for those days were included to avoid discarding meaningful observations. Extreme spikes or dips in either the epidemiological or Google Trends data were noted, with an explanatory comment provided in cases where a backlog release or reporting anomaly was known to have occurred.

The study addressed three central research questions:

RQ1: Is there a statistically significant relationship between weekly Google Trends search volume (for select COVID-19–related terms) and the reported weekly number of COVID-19 cases in Austria?

H0 (Null Hypothesis 1): There is no statistically significant relationship between weekly Google Trends search volume and the reported weekly number of COVID-19 cases in Austria.

H1 (Alternative Hypothesis 1): There is a statistically significant relationship between weekly Google Trends search volume and the reported weekly number of COVID-19 cases in Austria.

RQ2: Is there a statistically significant relationship between weekly Google Trends search volume (for select COVID-19–related terms) and the weekly vaccination rates in Austria?

H0 (Null Hypothesis 2): There is no statistically significant relationship between weekly Google Trends search volume and the weekly vaccination rates in Austria.

H1 (Alternative Hypothesis 2): There is a statistically significant relationship between weekly Google Trends search volume and the weekly vaccination rates in Austria.

RQ3: Is there a statistically significant relationship between weekly COVID-19 cases and weekly vaccination rates in Austria?

H0 (Null Hypothesis 3): There is no statistically significant relationship between weekly COVID-19 cases and weekly vaccination rates in Austria.

H1 (Alternative Hypothesis 3): There is a statistically significant relationship between weekly COVID-19 cases and weekly vaccination rates in Austria.

To test the above hypotheses, a correlation analyses was conducted at the national level, quantifying the strength and direction of linear associations between weekly Google Trends relative search volume scores and each weekly outcome variable. For the third research question, which investigates whether increased COVID-19 case counts are associated with increased vaccination numbers, correlations were calculated using the same weekly data structure (Burkholder et al., 2020). Pearson's *r* was the primary statistic when its assumptions like for example approximate normality or homoscedasticity, were met. When these conditions were violated, Spearman's rank correlation was used instead (Burkholder et al., 2020; Szklo & Nieto, 2019). Because each search term was analyzed separately, the error rate was controlled by reporting both raw p-values and values corrected for multiple comparisons by the Benjamini-Hochberg false discovery rate and, in sensitivity analyses, by the Bonferroni adjustment (Benjamini & Hochberg, 1995; Daniel & Cross, 2019). To address the potential influence of seasonal flu activity, weeks with unusually high influenza-related search volume were flagged and excluded in sensitivity analyses. Although no a priori covariates were specified, political or seasonal events were added post hoc if preliminary diagnostics indicated a meaningful influence. The precise diagnostic checks, decision rules, and contingency steps that guide this analytical strategy can be found in the section ‘‘Detailed Workflow for Data Analysis’’ below.

The interpretation of results focused on both statistical significance and effect size, using Pearson’s *r* to measure correlation strength (Burkholder et al., 2020; Daniel & Cross, 2019). The significance was assessed at the 0.05 level, with confidence intervals potentially providing additional context for the precision of the estimate (Burkholder et al., 2020; Daniel & Cross, 2019). Results were organized by research question, providing a clear link between each hypothesis tested and the corresponding correlation estimates. RQ3, which explores the relationship between COVID-19 case counts and vaccination uptake, was evaluated alongside RQ1 and RQ2. Ultimately, this plan ensured a structured, transparent, and replicable approach to evaluating the extent to which online information-seeking behavior corresponds to COVID-19 case and vaccination trends in

Austria, as well as whether increases in case numbers correlate with subsequent increases in vaccinations.

Detailed Workflow for Data Analysis

Before calculating Pearson's *r*, each search term was checked against six standard assumptions. The first was the level of measurement & independence: Both the weekly Google Trends relative search volume and the epidemiological series are continuous, interval-scaled measures collected from different calendar weeks (Babbie, 2017; Google News Initiative, n.d.). This also applies to the COVID-19 case and vaccination data used in RQ3, which are similarly continuous and interval-scaled. Independence of observations was tested using the Durbin-Watson statistic and by examining the autocorrelation function (Box et al., 2015; Durbin & Watson, 1951). If there was a significant serial correlation after the first-order differencing, the correlation is computed on the differenced series instead of the raw values (Box et al., 2015).

The second was linearity. A scatterplot with a LOWESS smoother overlaid was first examined while significant curvature was confirmed with the Ramsey RESET test (Ramsey, 1969; Tabachnick & Fidell, 2019). If the relationship was monotonic but not linear, the analysis switched to a nonparametric alternative, which in this case was Spearman's ρ (Daniel & Cross, 2019).

The third was bivariate normality. The univariate normality of each variable was assessed for all three research questions using the Shapiro-Wilk test ($\alpha = .05$), normal Q-Q plots, and skewness-kurtosis z scores ($|z| < 3.29$) (Daniel & Cross, 2019; Tabachnick & Fidell, 2019). For the joint distribution, Mardia's test for multivariate normality and chisquared Q-Q plots of Mahalanobis distances were used (Tabachnick & Fidell, 2019). Non-normal but still monotonic pairs were again calculated using Spearman by default; notably non-monotonic or heteroscedastic pairs were modeled using rank-based robust regression (Daniel & Cross, 2019; Wilcox, 2022).

The fourth was homoscedasticity: After fitting the Pearson model, the standardized residual values were plotted against the fitted values and tested using the Breusch-Pagan test (Daniel & Cross, 2019; Tabachnick & Fidell, 2019). Heteroscedastic patterns lead to variance-stabilizing transformations and retesting (Daniel & Cross, 2019; Tabachnick & Fidell, 2019). If the heteroscedasticity still persisted, a Spearman's ρ was reported (Daniel & Cross, 2019).

The fifth was absence of influential outliers: Univariate outliers were indicated by box plots and $|z| > 3.3$; multivariate leverage was assessed by Cook's distance ($D > 4/n$) and Mahalanobis distance ($p < .001$) (Ayinde et al., 2015; Bird et al., 2021; Daniel & Cross, 2019; Tabachnick & Fidell, 2019). Outlying weeks were winsorized ($\leq 5\%$) or put through sensitivity analysis (Daniel & Cross, 2019; Wilcox, 2022). Where results depend on their inclusion, robust measures, in this case Huber M estimator, complemented Pearson (Daniel & Cross, 2019; Huber, 1964).

The sixth was multiple testing control: Because of the large number of search terms examined, both raw p-values and Benjamini-Hochberg FDR-adjusted q-values were presented (Benjamini & Hochberg, 1995; Daniel & Cross, 2019). The Bonferroni-corrected significance was reported in a sensitivity table (Daniel & Cross, 2019).

Only if all six assumptions were met, the Pearson's correlation was retained as the primary estimate; otherwise, the pre-specified alternative (rank-based) statistic or a transformed model replaced it (Daniel & Cross, 2019). This explicit diagnostic sequence ensured that the correlation estimates were both statistically valid and transparent (Daniel & Cross, 2019). This method directly addressed any concerns about a priori assumptions and contingency planning (Daniel & Cross, 2019).

Threats to Validity

Threats to validity encompass several factors that may affect the accuracy and generalizability of the study's findings (Babbie, 2017). In a quantitative, correlational design that relies on secondary data sources, concerns typically revolve around external validity, which refers to how well the results can be generalized outside of the study's sample or setting; internal validity, which refers to whether the identified relationships may be influenced by biases or confounding factors rather than the independent variables; and construct or statistical inferential validity, which refers to whether the measures accurately capture the intended constructs and whether the data analysis procedures are appropriate (Burkholder et al., 2020; Shadish et al., 2001).

A major threat to the external validity of this study is the specificity of the Austrian context (Burkholder et al., 2020). While the use of national level data captures a robust snapshot of an entire country, it may not be directly applicable to other nations with different health care infrastructures, cultural norms, or internet usage patterns (Szklo & Nieto, 2019). To address this concern, the study maintains a clear focus on how Austrian public health data and search behavior intersect, without assuming similar correlations in other countries (Burkholder et al., 2020). Another potential threat is the changing nature of the pandemic itself. Shifts in official policies, such as lockdowns, testing guidelines, or vaccination mandates, could alter both search behavior and disease metrics, thereby limiting the generalizability of findings to contexts without these policies (Shadish et al., 2001; Szklo & Nieto, 2019). By documenting major policy events that occurred during the data collection window and acknowledging them as part of the broader landscape, the study makes it clear that any generalizations to other times or countries should be made with caution (Burkholder et al., 2020; Shadish et al., 2001).

With regard to internal validity in this non-experimental, correlational study using archival data, relevant issues are related to history and changes in instrumentation (Burkholder et al., 2020). Historical events, such as sudden policy changes or high-profile media coverage of COVID-19, could lead people to search more frequently for certain terms, independent of actual changes in infection or vaccination rates (Shadish et al., 2001). To mitigate this, the data analysis will note the timing of major public health announcements (e.g., national lockdowns, vaccine rollouts) if they overlap with abrupt spikes or dips in search volume or epidemiological numbers (Burkholder et al., 2020). Regarding instrumentation, Google Trends may change its algorithms or sampling methods over time (Burkholder et al., 2020). Because the study's data collection occurs within a consistent time frame and each term is normalized within the same window, this risk is minimized (Burkholder et al., 2020; Shadish et al., 2001). However, any noticeable anomalies in the Google Trends data, such as uncharacteristic drops or spikes that have no plausible external explanation, were documented.

Construct validity could be compromised if Google Trends data does not capture the true public interest in COVID-19 (Burkholder et al., 2020; Nuti et al., 2014). Some segments of the population may be underrepresented online, or individuals may use different search terms than those included (Nuti et al., 2014). The bilingual approach (combining English and German queries) mitigates some of this risk by ensuring that search terms reflect typical usage in Austria's primary languages (Mavragani & Ochoa, 2019a). In addition, the terms selected are consistent with those known to appear frequently in infodemiological studies, which helps to strengthen construct validity by focusing on widely recognized key words and phrases (Mavragani & Ochoa, 2019b). The validity of statistical inferences could be threatened by violations of correlational assumptions or multiple comparisons that inflate Type I errors (Benjamini & Hochberg, 1995; Burkholder et al., 2020). The plan to test for normality and use Spearman's rank-order correlation when appropriate reduces the risk of inaccurate conclusions (Daniel & Cross, 2019). In addition, reporting both unadjusted and adjusted p-values helps address concerns about multiple testing (Benjamini & Hochberg, 1995). The inclusion of seasonal influenza-related search trends in sensitivity analyses further guards against confounding from overlapping respiratory illness patterns, strengthening the interpretability of results. With this approach, the analysis aims to ensure that the reported correlations are both methodologically and statistically valid.

The data preparation process includes a review of weekly search volume data for influenza-related terms. This review is done in parallel with the primary data on Google Trends related to COVID-19. The goal is to identify potential periods of heightened influenza activity that could confound the analysis. Weeks with atypical spikes in influenza-related searches were identified, and the potential impact of these periods on the observed relationships was examined through sensitivity analyses (Babbie, 2017). To evaluate the robustness of the findings, the primary correlation and regression models were re-estimated with these weeks excluded (Babbie, 2017). This additional step ensures that the observed associations between search terms related to COVID-19 and epidemiological indicators are not influenced by trends in public attention to seasonal influenza. Including this analytical precaution strengthens the study's internal validity by accounting for overlapping seasonal influenza signals that could otherwise confound interpretation.

Collectively, these efforts aim to protect the integrity of the study's findings by identifying and mitigating potential threats to validity. By explicitly grounding all interpretations in the Austrian context, carefully documenting historical confounders, using a comprehensive list of bilingual search terms, and employing appropriate statistical controls, the study establishes a clear framework for understanding how online search behavior correlates with COVID-19 case and vaccination data. While not eliminating all potential limitations that exist in archival and correlational research, these strategies aim to reduce their impact and promote transparent, reproducible research.

Ethical Procedures

This study is based entirely on publicly available, anonymized data and does not involve any direct recruitment of human participants or the collection of personally identifiable information (Babbie, 2017). As a result, many of the typical ethical considerations related to participant privacy, informed consent, and the potential for adverse events do not apply (Babbie, 2017; Burkholder et al., 2020). Nevertheless, it is important to address the formal requirements for Institutional Review Board (IRB) approval, data confidentiality, and potential conflicts of interest or power differentials (Babbie, 2017; Burkholder et al., 2020).

Although no individual-level data were collected, I still sought IRB approval, consistent with institutional policy, to confirm that this project met the criteria for exemption under research involving publicly available secondary data. Any formal communication from the IRB and the official approval number was documented. Because there was no recruitment process since there were no participants, no incentives, and no clinical intervention, there were no associated materials or procedures that required separate review for potential ethical concerns.

All case and vaccination data came from official Austrian health agencies, such as the Federal Ministry of Social Affairs, Health, Nursing and Consumer Protection (BMSGPK) and the Federation of Austrian Social Insurance Institutions, which publish aggregated statistics without individual identifiers (BMSGPK, n.d.-a; data Austria, n.d.). Google Trends data were also aggregated and anonymized, and only provide relative metrics of search volume, not user-specific details (Google News Initiative, n.d.). Therefore, the data used in this study were, by nature, free of personally identifiable information, mitigating concerns about confidentiality or risk to individual rights (Babbie, 2017; Burkholder et al., 2020). I confirmed that these sources do not require any further conditions for access or special permissions for use. If such a data use agreement or permission was required by the agency, this documentation was included in the IRB application.

All data sets were stored on a password-protected computer, backed up to an encrypted drive, and accessible only to me. To maintain transparency and reproducibility, summarized results (e.g., aggregated weekly numbers, correlation results) were presented in the dissertation and possibly in subsequent publications. However, raw data files do not contain personally identifiable information. After concluding the study, I will retain these anonymized data for the required retention period established by Walden University, and then permanently delete them.

Finally, there are no anticipated conflicts of interest or power differentials, as I am not employed by the institutions providing the data and will not benefit financially from the results (Babbie, 2017). By adhering to IRB protocols, using only public and anonymized records, and implementing robust privacy measures, this study ensures ethical standards while providing valuable insights into how digital search behavior correlates with national pandemic response metrics (Babbie, 2017; Burkholder et al., 2020).

Summary

A quantitative, correlational approach is taken to investigate how Google Trends data on COVID-19 can be related to epidemiological outcomes and how these epidemiological outcomes relate to one another at the national level in Austria (Burkholder et al., 2020). The methodology is based on publicly available weekly aggregated data on new COVID-19 cases and vaccinations, paired with bilingual search terms in Google Trends. By defining each component numerically, the study establishes a transparent framework that captures both the frequency of digital information-seeking behavior and real-world indicators of disease progression and vaccine uptake.

Data collection involves the extraction of daily case and vaccination counts from Austrian health authorities, which are aggregated into seven-day intervals. Similarly, Google Trends relative search volume is aggregated to match the same weekly intervals, using both English and German terms to reflect search behavior in Austria's primary languages (Mavragani & Ochoa, 2019a). Statistical analysis focuses on correlation tests, with Pearson's correlation being the default choice unless data distributions require a shift to Spearman's rank correlation (Daniel & Cross, 2019). To account for the risk of false positives when testing multiple terms, a plan for reporting both unadjusted and adjusted p-values is included (Benjamini & Hochberg, 1995).

Concrete steps have been established to screen for missing data and outliers, noting any partial weeks at the beginning or end of the study window (Daniel & Cross, 2019). Potential threats to validity, such as historical events or reporting anomalies, or confounding seasonal factors, are addressed by documenting their onset and cautiously interpreting abrupt spikes and by conducting sensitivity analyses that exclude weeks with abnormal influenza-related search activity (Shadish et al., 2001). Ethical considerations focus on data confidentiality and IRB review, emphasizing reliance on publicly available and anonymized data sources (Burkholder et al., 2020).

This methodological approach provides a consistent procedure for analyzing the interaction between online searches and COVID-19 outcomes and the relationship between COVID-19 outcomes in Austria. The results of these analyses and the implications for public health strategy are presented in the following chapter.

Chapter 4: Results

Introduction

This quantitative, correlational study examined the association between weekly Google Trends relative search volumes for terms related to COVID-19 and weekly case counts and vaccination numbers in Austria. Google Trends search volume served as the independent variable for RQ1 and RQ2, while weekly case counts and weekly vaccination numbers served as the dependent variables. For RQ3, the association between cases and vaccinations was examined directly, independent of search data. As planned in Chapter 3, analyses proceeded using correlation-based techniques and related procedures that were appropriate to the distributional assumptions.

Consistent with the proposal, all analyses were guided by three research questions: whether weekly Google Trends search volume correlated with weekly COVID-19 cases, whether weekly Google Trends search volume correlated with weekly vaccination rates, and whether weekly cases were associated with weekly vaccinations. For each question, the null hypothesis was that there was no statistically significant relationship, while the alternative hypothesis posited a significant relationship. These questions and their respective hypotheses are restated below to frame the organization of the following results.

This chapter provides a concise account of the data collection process, including the time frame, recruitment and response rates, and any deviations from the plan outlined in Chapter 3. It then presents the baseline descriptive results. Next, I evaluate statistical assumptions. Then, I present the primary findings organized by research question and hypothesis. I report exact statistics, probability values, confidence intervals, and effect sizes as appropriate. The chapter concludes with a brief summary that transitions to Chapter 5.

Data Collection

The data for this quantitative, correlational study were drawn entirely from archival, publicly available sources, covering weekly observations from March 1, 2020, to June 30, 2023. Daily case and vaccination counts from Austrian public health authorities were aggregated into seven-day intervals. Google Trends' relative search volume for pre-specified bilingual (German and English) terms related to COVID-19 was also aggregated into the same weekly intervals to align the time scale for analysis. Since the study used archival records and did not recruit human participants, there were no recruitment procedures or response rates to report. Rather, the sample comprised 175 weekly time points within the analysis window.

To align with the plan outlined in Chapter 3, weekly observations were defined as intervals from Monday to Sunday. The analytic window for RQ1 spans 175 weeks ending on a Sunday between February 26, 2020, and June 25, 2023. These weeks are labeled in the dataset by their end-of-week dates (e.g., March 1, 2020, and June 30, 2023). For RQ2 and RQ3, the overlapping vaccination window comprises 132 weeks from December 27/31, 2020, through June 25/30, 2023. It was verified that these labels correspond to complete Monday–Sunday weeks, with partial first and last weeks pooled as specified.

Access and Permissions

This study used publicly available archival data. Weekly counts of new cases of the virus and vaccination totals were obtained from national reports in Austria. Relative search volumes for pre-specified German and English terms related to the virus were retrieved from Google Trends. No individual-level records were accessed, no personally identifiable information was present, and no special permissions or data-use agreements were required. All files were stored on a password-protected, encrypted backup device accessible only to the researcher and were retained/archived according to the data-management plan outlined in Chapter 3.

Representativeness and Baseline Characteristics

The analytic unit was the calendar week. The dataset included 175 consecutive weekly observations from March 1, 2020, to June 30, 2023. This represents a complete census of national-level weeks within the study period, rather than a sampled cohort. To ensure temporal comparability across sources, case counts, vaccination totals, and search volumes were aligned to the same weekly timeline. The next section presents descriptive summaries of central tendency and dispersion for each series, characterizing baseline levels and variability prior to assumption checks and inferential analyses.

Treatment and/or Intervention Fidelity

No treatment or intervention was administered in this study. The observational design relied on archival, population-level data, including weekly COVID-19 cases, vaccination rates, and Google Trends' relative search volumes. Because no participants were assigned to conditions and no protocolized treatment was delivered, treatment fidelity was not applicable.

To ensure analytic fidelity, all data handling and analyses adhered to the procedures outlined in Chapter 3. Weekly series were aligned to identical calendar weeks, search terms were applied exactly as defined, and deterministic transformations (e.g., aggregation to weekly intervals) were implemented consistently across variables. Scripts and outputs were version-controlled, and data ranges and record counts ($N = 175$ weeks) were verified. Additionally, spot checks confirmed that the results were reproducible from the raw inputs, with no deviation from the planned workflow.

Results

This chapter reports the empirical findings in the order specified in Chapter 3. For each research question, the section begins with descriptive statistics, followed by checks of the assumptions. Then, the primary analyses are presented: Spearman's ρ, cross-correlation functions on the log-differences series, and time-lagged log-differences regressions with HAC/Newey–West standard errors. Where applicable, sensitivity analyses are reported alongside the main results rather than in a separate diagnostics section. Weekly series are aligned to the Austrian ISO week calendar used throughout the study. Transformations and robust inference address the distributional shape and serial dependence documented in the assumption checks.

Research Question 1: Association between weekly Google Trends search volume and weekly COVID-19 cases

This section evaluates whether changes in population search behavior can predict short-term changes in the number of reported cases of the novel coronavirus. Descriptive statistics confirm the presence of long-tailed distributions for both outcomes and predictors, which motivates the use of Δlog growth measures and HAC-robust inference. Same-week Spearman correlations determine the strength and direction of monotonic associations, and cross-correlation functions identify the most informative lag. Time-lagged Δlog regressions quantify elasticities, which are interpreted as the percent change in weekly case growth associated with a one-percent change in search growth at the selected lag. Across the pre-specified variables, the evidence suggests a consistent one-week lag between searches and cases, with statistically significant positive elasticities of practical magnitude.

RQ1 Descriptives

Table 2 summarizes the distributional characteristics of the weekly outcome and each Google Trends predictor included in RQ1. The analytic dataset comprised N = 175 consecutive weekly observations from March 1, 2020, to June 30, 2023; case counts are reported as weekly totals, and Google Trends values are on the 0–100 relative search volume scale. No weeks were missing. The weekly case count shows a wide range (13 to 322,641), a high mean (35,335), and pronounced positive skew and leptokurtosis, indicating episodic surge weeks. The composite search-volume index (aggregating the selected terms related to the virus) averaged 14.06 on the 0–100 Google Trends scale (range 6.00–59.69) and was likewise right-skewed. Symptom-related queries (e.g., shortness of breath, fever, and cough) exhibited comparatively higher central tendency (means of approximately 38, 34, and 29, respectively) yet remained heavy-tailed. Combined Search Terms is a composite, and for each week, the z-standardized individual RQ1 terms are listed in Table 2 and were averaged as an unweighted mean. Loss of taste and loss of smell had lower means, yet exhibited marked skewness, reflecting sharp but infrequent peaks. Topic and behavior terms (e.g., variants, testing, face masks, social distancing, and quarantine) generally had lower means and substantial kurtosis. Overall, there was sparse interest in prevention (mean ≈0.57) with extreme kurtosis, consistent with occasional spikes rather than sustained interest. These patterns collectively indicate substantial week-to-week variability and long-tailed distributions across the series, motivating the assumption checks and nonparametric/time-series procedures that follow.

Table 2

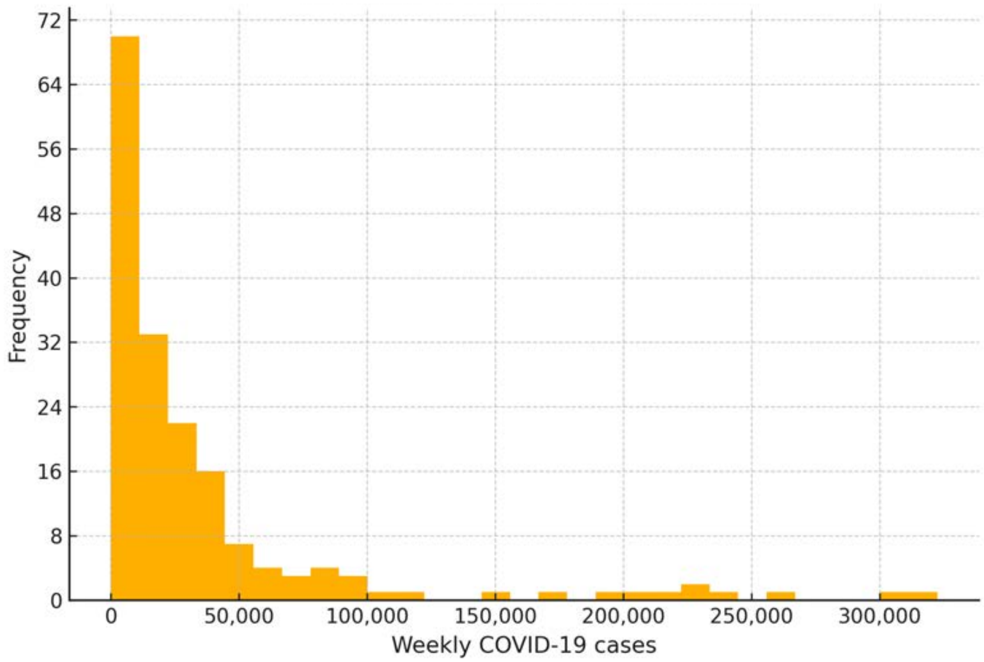
Descriptive statistics for RQ1 variables (N = 175 weeks)

Variable	N	Mean	Std Dev	Min	Max	Skewness	Kurtosis
Weekly Cases	175	35 335.06	57 291.95	13.0	322 641.0	3.02	9.66
Combined Search Terms	175	14.06	7.42	6.0	59.69	2.97	13.73
COVID-19+Corona+Coronavirus	175	14.76	13.98	1.0	100.0	2.61	12.02
Fever+Fieber	175	34.03	13.00	20.0	100.0	2.64	9.19
Cough+Husten	175	28.82	16.04	9.0	100.0	2.00	5.61
Shortness_of_breath+Atemnot	175	37.71	10.28	19.0	100.0	2.43	10.51
Loss_of_taste+Geschmacksverlust	175	15.66	17.87	0.0	100.0	1.66	4.09
Loss_of_smell+Geruchsverlust	175	6.25	13.74	0.0	100.0	2.92	12.54
COVID-19_testing+Coronavirus_testing+Corona_Test	175	17.70	18.89	1.0	100.0	1.85	3.68
COVID-19_symptoms+Coronavirus_symptoms+Corona_Symptome	175	13.23	14.50	1.0	100.0	3.28	14.59
COVID-19_treatment+Coronavirus_treatment+Corona_Behandlung	175	9.74	14.85	0.0	100.0	2.93	13.87
COVID-19_prevention+Coronavirus_prevention+Corona_Praevention	175	0.57	7.56	0.0	100.0	13.23	175.00
COVID-19_variants+Coronavirus_variants+Corona_Varianten	175	9.33	21.97	0.0	100.0	2.31	4.46
Hand_sanitizer+Handdesinfektionsmittel	175	2.66	11.53	0.0	100.0	6.46	46.22
Face_mask+Maske+Gesichtsmaske	175	10.61	11.50	3.0	100.0	4.67	29.15
Social_distancing+Abstand_halten	175	6.99	16.10	0.0	100.0	3.43	12.97
Quarantine+Quarantaene	175	9.69	12.67	0.0	100.0	4.10	22.62
Pandemic+Pandemie	175	7.04	9.47	1.0	100.0	6.81	59.21

The distribution of weekly cases was strongly right-skewed, with most weeks clustering at lower counts and a small number of surge weeks producing a long upper tail (Figure 1). This pattern is consistent with episodic pandemic waves and helps explain the large standard deviation reported in Table 2.

Figure 1

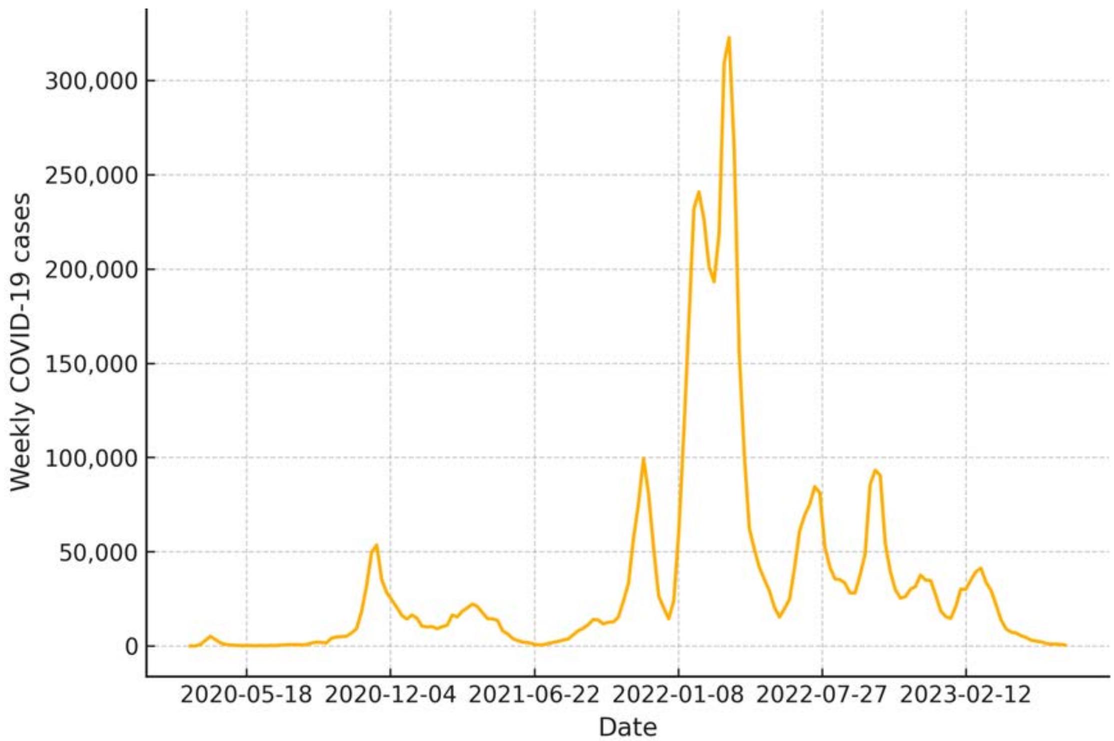
Histogram of Weekly COVID-19 Cases (Austria, Weekly)



The time series displayed distinct waves separated by troughs across the study window (Figure 2). Rapid ascents and sharp peaks were followed by declines, and variability increased during wave periods, evident in higher-amplitude fluctuations relative to inter-wave intervals.

Figure 2

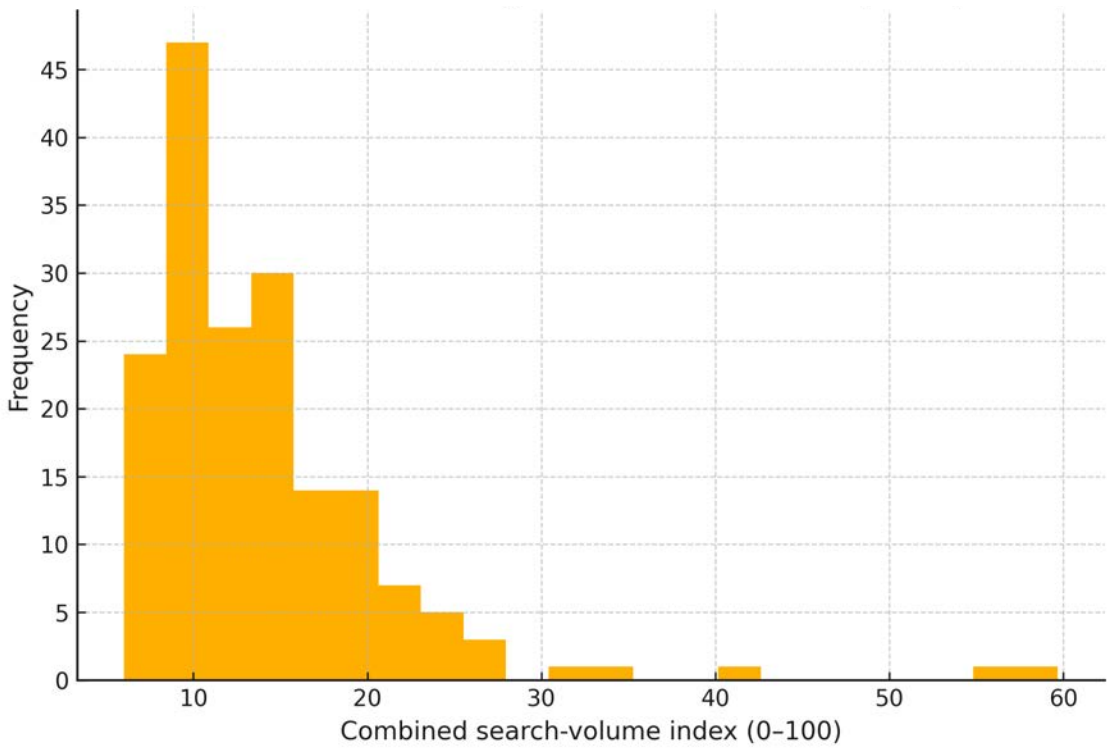
Weekly COVID-19 Cases Over Time (Mar 2020–Jun 2023)



The combined search-volume index was bounded (0–100) and concentrated at lower values, with moderate right skew and occasional spikes (Figure 3). Dispersion was narrower than for cases, reflecting shorter and less extreme bursts of search interest.

Figure 3

Histogram of Combined Google Trends Search-Volume Index (Weekly)



The combined index showed episodic bursts of search activity interspersed with quieter periods (Figure 4). Apparent synchrony with cases is assessed formally in subsequent sections.

Figure 4

Combined Google Trends Search-Volume Index Over Time (Weekly)

Across 175 weekly observations from March 1, 2020, to June 30, 2023 (Table 2 and Figures 1-4), there are right-skewed, long-tailed distributions for weekly cases and several search volume variables, with episodic surges and bursts over time. Symptom-related queries exhibited higher central tendency, whereas prevention-related searches were sparse with extreme kurtosis. These distributional features, together with the visible wave dynamics, suggest non-normality and temporal dependence. The next section formally evaluates these assumptions and specifies the implications for the inferential analyses used in RQ1.

RQ1 Assumption Checks

Before conducting the inferential analyses for research question 1, the key statistical assumptions were evaluated for the weekly case series and the five pre-specified Google Trends predictors. Normality was examined using the Shapiro–Wilk test and Q–Q plots, independence (the absence of serial autocorrelation) was assessed using Durbin–Watson statistics and residual autocorrelation functions, and homoscedasticity was probed using Breusch–Pagan tests of model residuals. Since the RQ1 models rely on bivariate associations and time-lagged regressions, confirming these assumptions or adopting robust alternatives when assumptions were violated ensures the validity of subsequent findings. The results of these checks guided the selection of nonparametric correlation coefficients, log-differencing transformations, a +1-week lag structure, and heteroscedasticity- and autocorrelation-consistent (HAC) standard errors in the subsequent regression models.

Normality. Shapiro–Wilk tests were applied to the weekly case series and to each of the five search-volume predictors. As shown in Table 3, all six series produced highly significant results ($p < .001$ in every instance), indicating departures from a normal distribution. The prevention terms showed the most pronounced departure ($W = 0.57$), while the symptom aggregate showed the least pronounced departure ($W = 0.67$), though both were still non-normal.

Table 3

Shapiro–Wilk normality tests (weekly cases and five predictors, N = 175)

Variable	Shapiro W	*p* value
Weekly Cases	0.598	4.94×10^{-20}
COVID-19 symptoms + Coronavirus symptoms + Corona Symptoms	0.672	3.14×10^{-18}
COVID-19 + Corona + Coronavirus	0.774	3.93×10^{-15}
COVID-19 testing + Coronavirus testing + Corona Test	0.791	1.51×10^{-14}
Fever + Fieber	0.748	5.16×10^{-16}
Combined Search Terms	0.752	6.72×10^{-16}

Note. All p-values < .001 indicate significant departures from normality.

Visual inspection confirmed the test results. Figure 5 plots sample quantiles of weekly cases against the theoretical quantiles of a normal distribution. The pronounced S-shape and heavy upper tail confirm strong positive skew and leptokurtosis. Figure 6 shows the combined search-volume index and displays a similar, albeit less extreme, curvature. Low quantiles deviate below the reference line, and high quantiles rise above it. Both plots reinforce the idea that Gaussian-based parametric techniques would be inappropriate without transformation.

Figure 5

Q–Q Plot of Weekly COVID-19 Cases

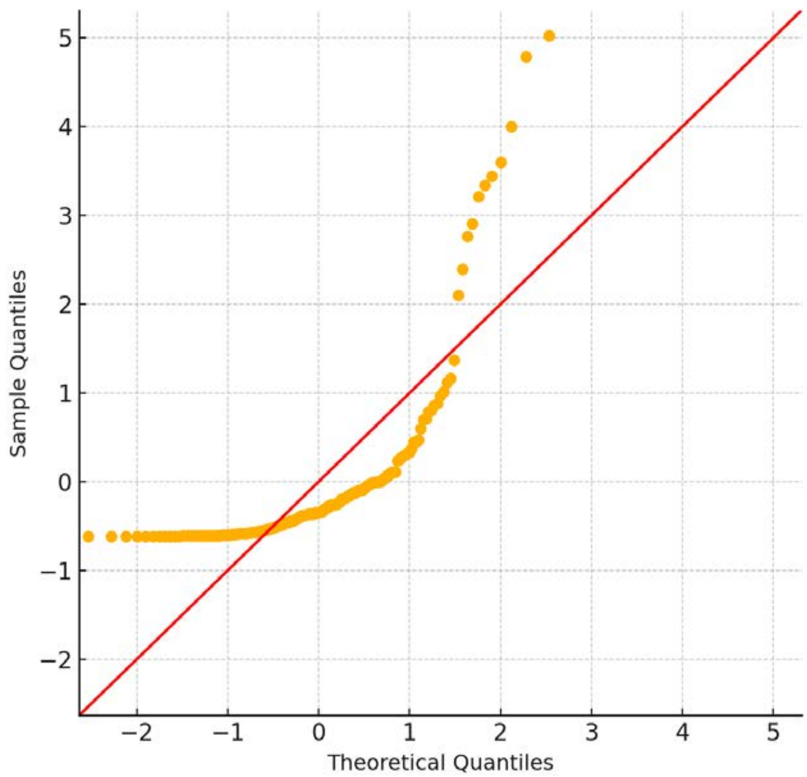
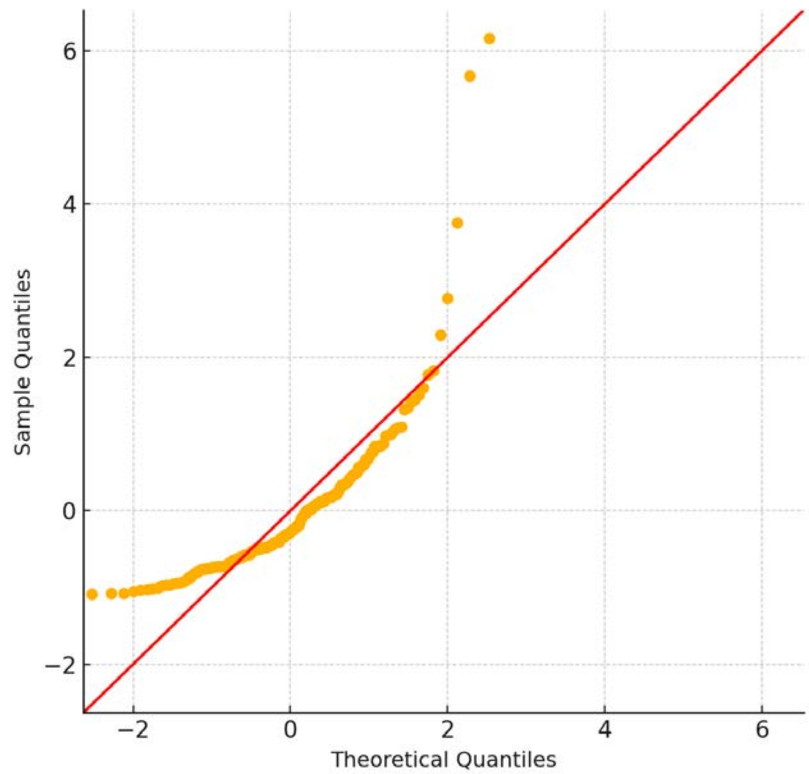


Figure 6

Q–Q Plot of Combined Search-Volume Index



Given these findings, subsequent analyses employed Spearman's rank-order correlation for bivariate association and log1p transformations followed by first differences (Δlog) for time-lagged regression models. These steps mitigate skew, stabilize variance, and satisfy the distributional assumptions of the analytic procedures used to address RQ1.

Independence/Autocorrelation. Durbin–Watson statistics confirmed substantial first-order autocorrelation in the raw time series (Table 4). All values fell well below the 2.0 independence benchmark, with weekly cases at 0.09 and search-volume predictors ranging from 0.21 (*COVID-19 + Corona + Coronavirus*) to 0.29 (*Combined Search*

Terms). These results indicated that adjacent weeks were highly correlated, thus violating the independence assumption required by ordinary least squares models fitted to raw levels.

Table 4

Durbin–Watson statistics (selected series)

Variable	Durbin-Watson
Weekly Cases	0.091
COVID-19 symptoms + Coronavirus symptoms + Corona Symptoms	0.226
COVID-19 + Corona + Coronavirus	0.207
COVID-19 testing + Coronavirus testing + Corona Test	0.218
Fever + Fieber	0.231
Combined Search Terms	0.285

To evaluate whether differencing and HAC estimation addressed serial dependence, the residuals from the strongest lagged model were inspected, which was Δlog(cases)_t regressed on Δlog(symptoms)_(t-1). The residual autocorrelation function (Figure 7) revealed that most spikes fell within the $\pm 1.96/\sqrt{N}$ confidence bands after the first three lags and that no systematic pattern persisted beyond lag 3. Though minor short-lag autocorrelation remained, its magnitude was significantly smaller than that of the raw series. This supports the use of heteroscedasticity- and autocorrelation-consistent standard errors.

Figure 7

Residual ACF: ΔlogCases_t ~ ΔlogSymptoms_{T-1}

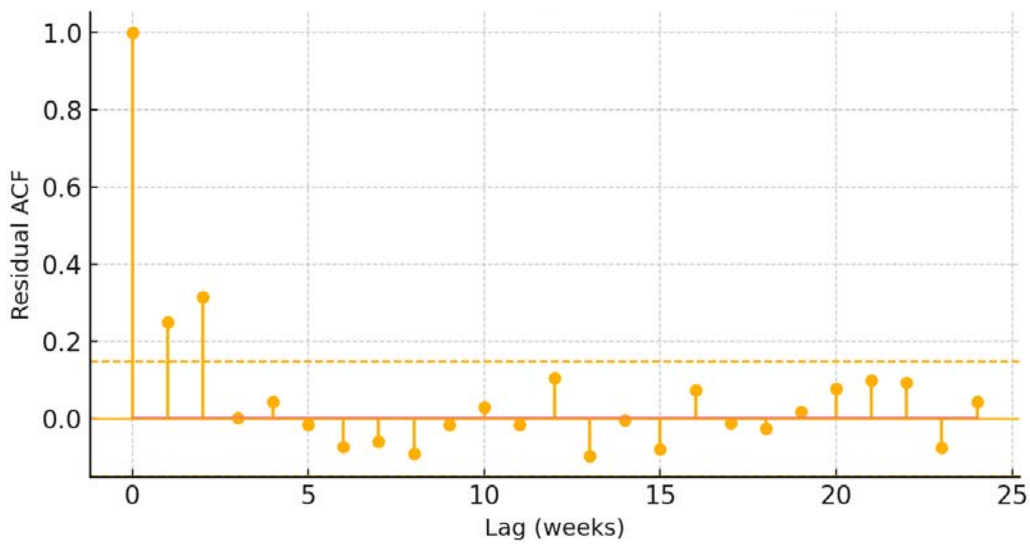
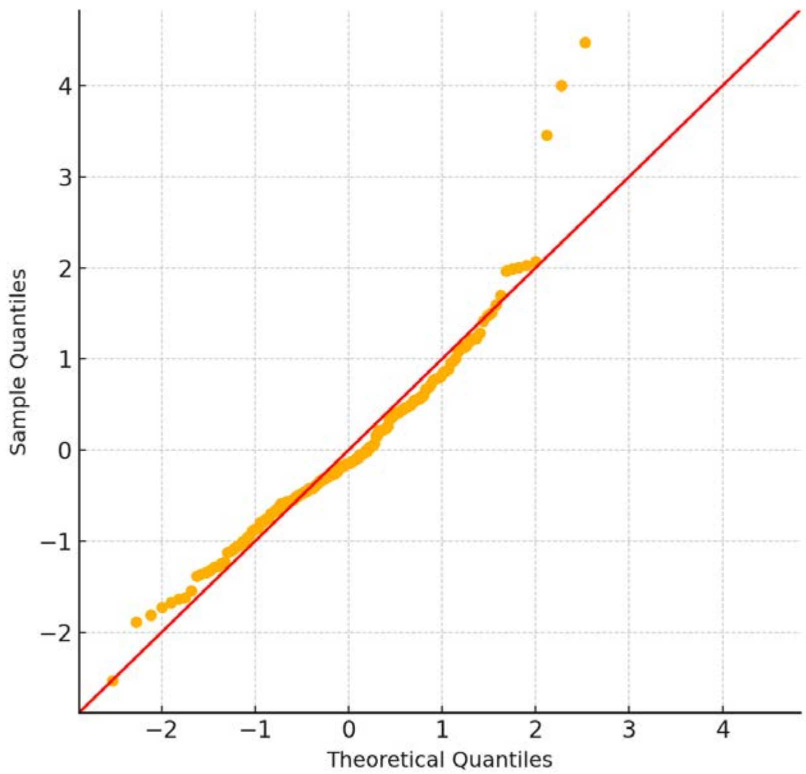


Figure 8 shows a residual Q–Q plot, which displays points hugging the 45-degree reference line with only mild tail deviation. This indicates that the log-transformation produced approximately symmetric residuals. Together, these diagnostics show that differencing substantially mitigated serial dependence and that HAC correction provides robust inference for the time-lagged regression models presented in the next section.

Figure 8

Residual Q–Q: ΔlogCases_t ~ ΔlogSymptoms_{T-1}



Homoscedasticity. Breusch-Pagan (BP) tests were conducted on the residuals of the five Δlog lag + 1 regression models to evaluate constant variance. Table 5 reports the BP p-values alongside the Ljung–Box results presented earlier. Two predictors, *COVID-19 symptoms* and *Combined Search Terms*, returned BP $p < .01$, indicating residual heteroscedasticity. The remaining three predictors were non-significant (e.g., *COVID-19 + Corona + Coronavirus*, BP $p = .309$). Since heteroscedasticity and residual autocorrelation were detected, all regression coefficients were estimated with heteroscedasticity- and autocorrelation-consistent (HAC) standard errors, which remain valid in the presence of non-constant variance.

Table 5

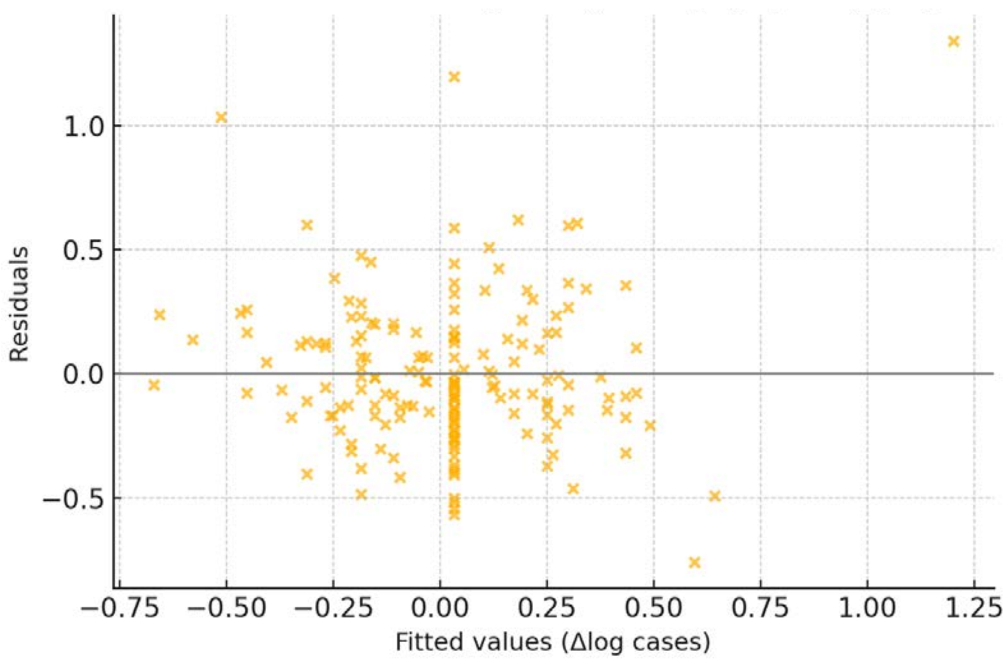
Diagnostics for Lag +1 bivariate models

Variable	Ljung–Box p1	Ljung–Box p4	Ljung–Box p8	Ljung–Box p12	BP *p*
COVID-19 symptoms + Coronavirus symptoms + Corona Symptoms	.0010	.0000	.0001	.0006	.0009
COVID-19 + Corona + Coronavirus	.0000	.0000	.0004	.0008	.3091
COVID-19 testing + Coronavirus testing + Corona Test	.0000	.0000	.0000	.0000	.0000
Fever + Fieber	.0000	.0000	.0000	.0000	.0501
Combined Search Terms	.0004	.0000	.0003	.0013	.0001

Figure 9, a residuals-versus-fitted scatterplot for the symptoms model, visually confirms slight fan-shaped dispersion, which is consistent with the significant BP test. However, the variance pattern is not extreme.

Figure 9

Residuals vs. Fitted: ΔlogCases_t ~ ΔlogSymptoms_{T-1}



Because some predictors show heteroscedasticity and others do not, and because the heteroscedasticity that is present is mild, using HAC (Newey–West) standard errors ensures valid inference across all time-lagged regressions. There are no residual patterns that suggest model misspecification beyond the variance heterogeneity that has already been accounted for.

Outliers and influential points. Cook's distance was computed for each Δlog lag +1 regression model to identify influential weeks (criterion > 4/N, where N = 173). Table 6 summarizes the results. No model contained more than two observations exceeding the threshold, and the largest Cook's D (0.061) was well below commonly cited influence benchmarks of 0.5. These results suggest that no single week had a disproportionate effect on the estimated elasticities.

Table 6

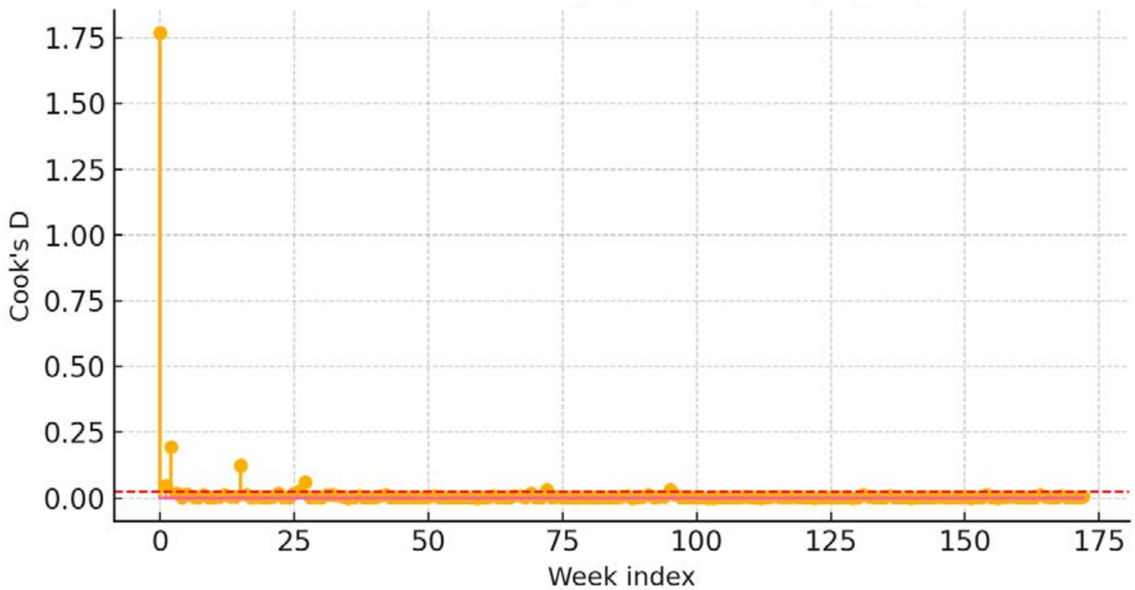
Cook's Distance Summary (Lag +1 Δlog models)

Model	N	Threshold (4/N)	Max Cook's D	Points > Threshold	Dates of 2 Largest Cook's D
Symptoms	173	0.0231	0.0610	2	2020/04/05, 2021/11/21
General	173	0.0231	0.0543	1	2020/03/29, –
Testing	173	0.0231	0.0317	1	2021/12/19, –
Fever	173	0.0231	0.0498	2	2020/03/29, 2021/11/28
Combined Index	173	0.0231	0.0394	1	2020/03/29, –

Figure 10 plots Cook's distance for the symptoms model. Only two weeks exceed the dashed 4/N reference line, and both correspond to early-wave or peak-wave weeks that were retained in the analysis.

Figure 10

Cook's Distance: Symptoms Model (Lag +1)



Influence diagnostics revealed that no single observation influenced the RQ1 regression results. Therefore, all 175 weeks were retained, and additional weighting or robust regression methods were not necessary.

Multicollinearity. Since the primary RQ1 inferences are based on individual bivariate models, multicollinearity is not an issue for these analyses. However, it is pertinent for the exploratory multivariable model (log cases_t regressed on log symptoms_(t-1), log general_(t-1), and log testing_(t-1)). Variance inflation factors (VIFs) were calculated for the three predictors at lag +1, and the results are presented in Table 7.

Table 7

Variance Inflation Factors (multivariable model, lag +1)

Predictor	VIF
Δlog symptoms_{t-1}	1.798
Δlog general_{t-1}	2.096
Δlog testing_{t-1}	1.702

All VIFs fall well below the common thresholds of 5 (moderate) and 10 (severe), indicating no evidence of multicollinearity problems among the lagged predictors. Accordingly, the multivariable model was retained without adjustment or exclusion of variables.

Summary and implications for RQ1 analysis. The diagnostic evidence consistently showed that the weekly series violated the classical Ordinary Least Squares assumptions when analyzed in their raw form. Shapiro–Wilk tests and Q–Q plots confirmed pronounced non-normality, and Durbin–Watson statistics revealed strong positive autocorrelation. Breusch–Pagan tests showed heteroscedasticity for some predictors. After transforming the data to log1p and first differences (Δlog), serial dependence decreased markedly, and the variance patterns became more stable. However, residual ACF plots showed minor short-lag autocorrelation, which warrants the use of heteroscedasticity- and autocorrelation-consistent (HAC) standard errors. Cook's distance diagnostics identified no influential weeks, and the variance-inflation factors in the multivariable model were all below 2.1, ruling out concerns about multicollinearity.

Taken together, these findings justify the analytic strategy adopted for RQ1. First, the use of Spearman's rank-order correlation to examine initial bivariate associations. Second, modeling Δlog(cases)_t as a function of Δlog(search volume)_{t-1} in time-lagged regressions. And lastly, estimating all regression coefficients using HAC (Newey–West) standard errors.

RQ1 Spearman Associations

To address Research Question 1 at the bivariate level, monotonic associations between weekly counts of new cases of the novel coronaviruses and each pre-specified Google Trends predictor were evaluated using two-tailed Spearman's rank-order correlation (ρ) with an alpha level of .05. The analysis tested the null hypothesis ($H0_1$) that there is no significant association between weekly search activity and weekly cases against the alternative hypothesis ($H1_1$) that there is a significant association. The weekly observations ($N = 175$) spanned from March 1, 2020, to June 30, 2023, matching the period used in subsequent time-series analyses.

Spearman's ρ results. Spearman's rank-order correlations are presented in Table 8. Fever-related searches showed the strongest monotonic association with weekly cases of COVID-19, with a correlation coefficient of $\rho = .473$ and a p-value of $<.001$. This indicates that as the volume of fever-related queries increased, case counts tended to rise in the same weeks. The Combined Search Terms index was also positively related to cases, albeit more modestly ($\rho = .227$, $p = .003$). Conversely, general searches using the terms *COVID-19, Corona*, or *Coronavirus* displayed a modest but significant negative association with cases ($\rho = -.238$, $p = .001$). This suggests that public interest in the disease name itself may have waned during periods of higher transmission. Aggregate symptom searches that did not specify a single symptom and testing-related queries were uncorrelated with weekly cases ($\rho = .042$, $p = .582$ and $\rho = -.021$, $p = .780$, respectively).

Table 8

Two-tailed Spearman correlations between weekly COVID-19 cases and the five pre-specified Google Trends predictors

Variable	ρ	*p* value	*N*
Fever + Fieber	.473	< .001	175
Combined Search Terms	.227	.003	175
COVID-19 + Corona + Coronavirus	−.238	.001	175
COVID-19 symptoms + Coronavirus symptoms + Corona Symptome	.042	.582	175
COVID-19 testing + Coronavirus testing + Corona Test	−.021	.780	175

The pattern of coefficients aligns with the scatterplots in Figures 11 through 15. Fever-related searches exhibit a clear upward trend with considerable vertical spread at higher volumes. In contrast, the Combined Index shows a shallower slope reflecting its mixed composition of symptom- and topic-related terms. The plot for general searches related to COVID-19 reveals a subtle downward trajectory, which visually explains the negative ρ. Meanwhile, the symptom-aggregate and testing plots display broadly cloud-like patterns with no discernible direction. These visuals confirm that rank-based statistics more appropriately capture the underlying relationships than a linear metric would.

Figure 11

Weekly Cases vs. Symptoms Aggregate

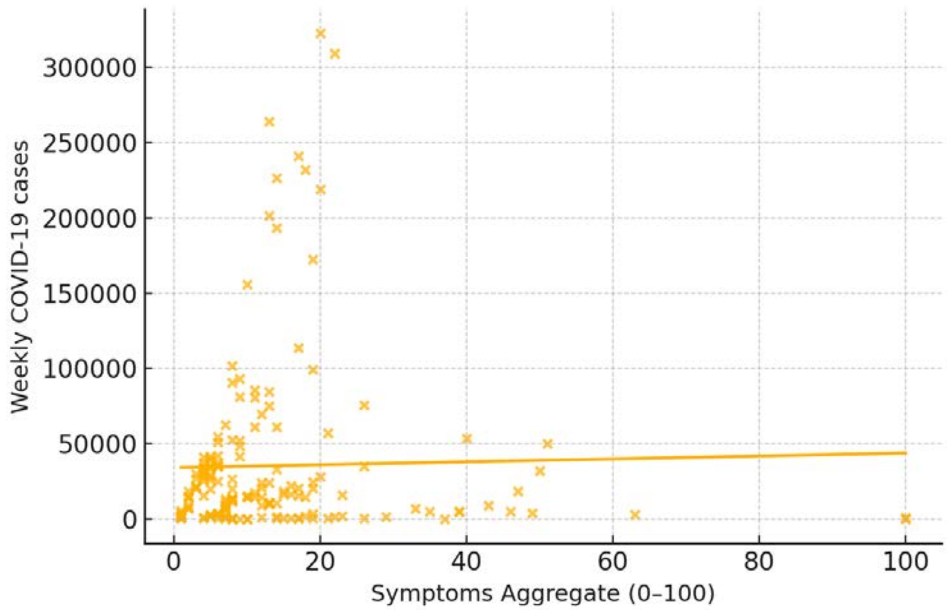


Figure 12

Weekly Cases vs. General Searches

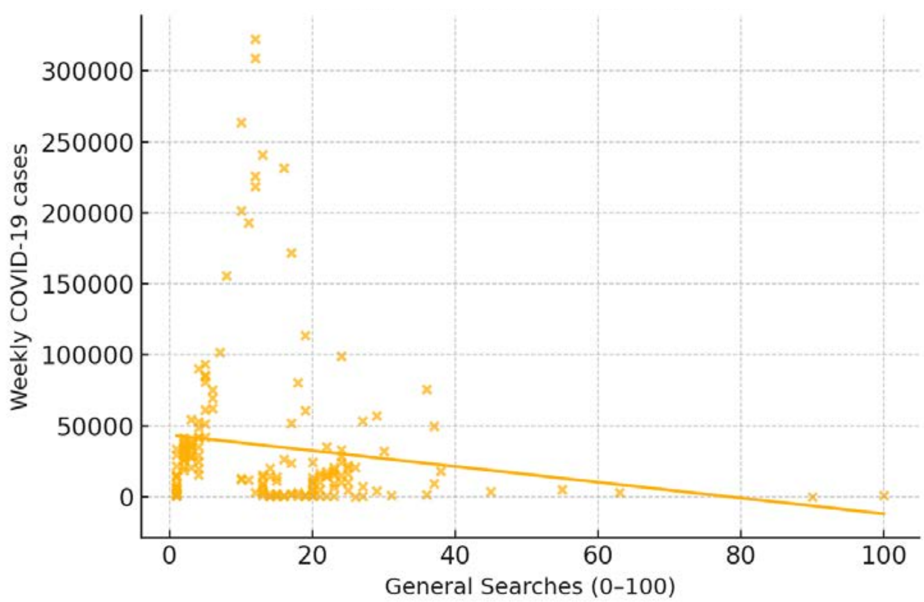


Figure 13

Weekly Cases vs. Testing Searches

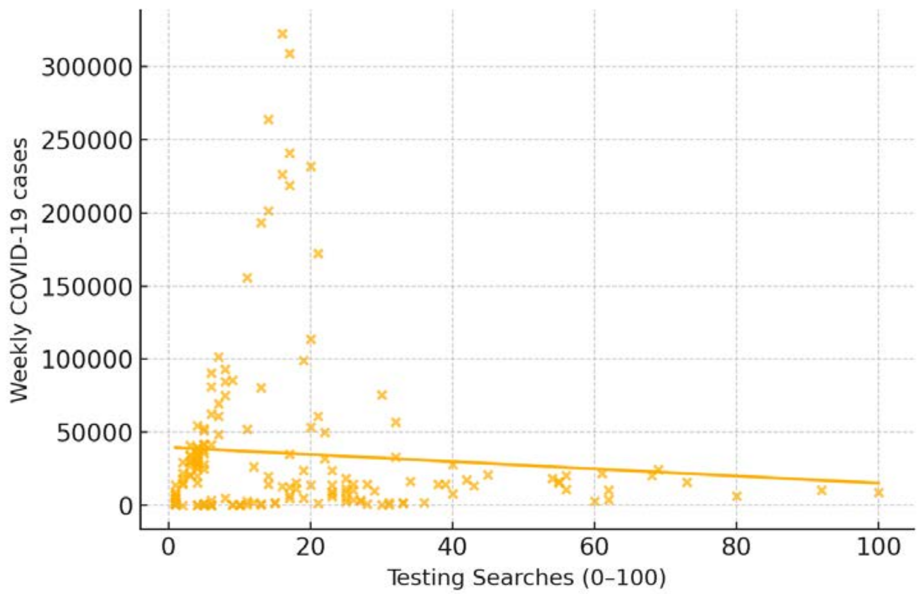


Figure 14

Weekly Cases vs. Fever Searches

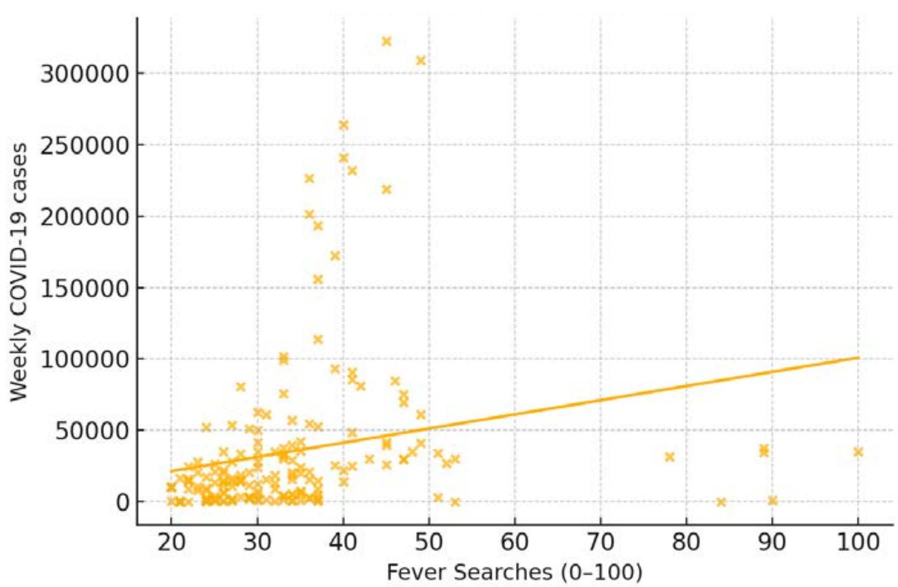
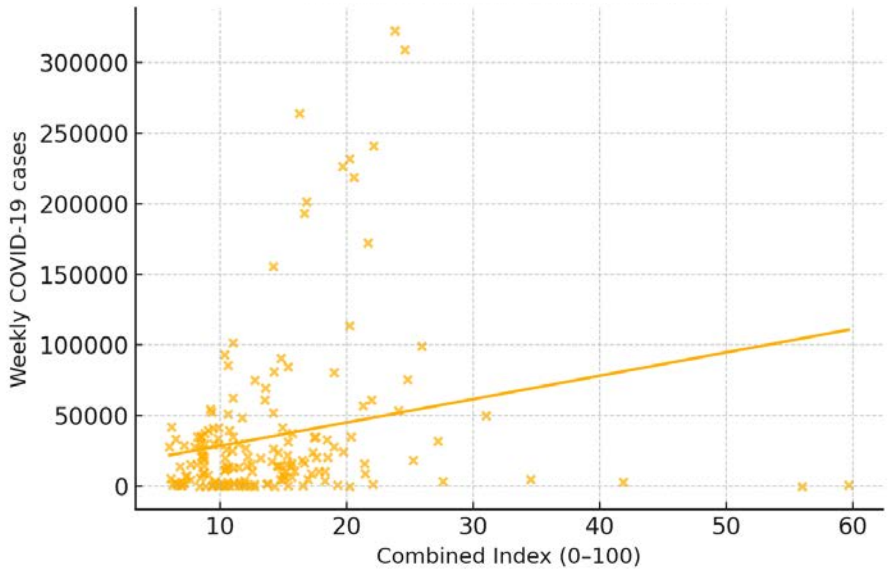


Figure 15

Weekly Cases vs. Combined Index



Since the null hypothesis of no association was rejected for three of the five predictors (fever searches, the combined index, and general searches) the analysis examined whether search activity leads to changes in cases over time. Using all five pre-specified predictors enables direct comparison across subsequent cross-correlation and time-lagged regression analyses, which address the temporal aspect of RQ1.

Results interpretation. Correlation analyses showed that query patterns vary significantly in their tracking of contemporaneous activity related to COVID-19. Fever-related searches emerged as the clearest real-time signal, with a moderate positive coefficient implying larger case counts in weeks when the public sought information about this defining symptom. The Combined Index, which blends symptom, topic, and prevention terms, yielded a small but significant positive association. This suggests that aggregate search behavior captures some variance in cases, even when individual components do not. In contrast, general disease-name searches were weakly negative, consistent with the idea that public curiosity about the term *COVID-19* decreased during periods of intense viral circulation, perhaps because information saturation had already occurred. Symptom-aggregate and testing searches were essentially unrelated to weekly case counts, indicating that they were not useful concurrent indicators in the Austrian context.

Taken together, these findings partially support the alternative hypothesis for RQ1 at the contemporaneous level. Not every search volume series is informative; however, certain symptom-specific queries and an appropriately weighted composite provide meaningful signals. The main objective of RQ1 is to determine if search activity leads to changes in cases. Therefore, all five pre-specified predictors are retained for temporal analyses, even when the same-week association is negligible.

Hypothesis decision and implications. Two-tailed Spearman tests partially rejected the null hypothesis for RQ1. Weekly fever searches, the combined search terms index, and the general COVID-19 query string each showed statistically significant associations with weekly case counts at $\alpha = .05$. However, the symptom aggregate and testing searches did not. Consequently, H01 was rejected for three of the five specified predictors and retained for the remaining two. Although only a subset demonstrated contemporaneous relevance, all five variables are carried forward to the temporal analyses. Retaining the full set preserves the pre-registration of predictors and permits direct comparison once lead-lag dynamics are introduced. The next section therefore investigates whether search activity in both significant and nonsignificant contemporaneous cases precedes changes in reported infections. This would strengthen the practical utility of these digital indicators.

RQ1 Lag Identification (Cross-Correlation Function)

To determine whether changes in reported infections were preceded by fluctuations in search activity, cross-correlation functions (CCFs) were computed between the differenced log series of weekly COVID-19 cases and each Google Trends predictor. The analysis covered lags from −12 to +12 weeks. Positive lags indicate that search volume at week $t - \text{lag}$ leads case counts at week t. Both series were log-transformed and first-differenced to remove long-term trends and stabilize variance. Statistical significance was evaluated against approximate 95% confidence bounds of $\pm 1.96/\sqrt{N}$, where N represents the number of overlapping observations at each lag. This approach identifies the lag at which each search term exhibits its strongest predictive relationship with subsequent case counts. This information is used to specify the time-lagged regression models that follow.

Cross-Correlation Results. Table 9 summarizes the peak cross-correlation coefficients between weekly cases of COVID-19 and each search volume series. For all five predictors, the maximum absolute correlation occurred at one week, indicating that changes in search behavior systematically preceded changes in reported infections by one week. Using the 95% confidence limit of $\pm 1.96/\sqrt{N} \approx \pm 0.15$, every peak coefficient comfortably exceeded the significance threshold.

Table 9

Peak Cross-Correlation Values (Δlog series, lags −12 to +12 weeks)

Predictor	Lag at peak (weeks)	CCF value	95 % bounds exceeded?
Symptoms Aggregate	+1	0.645	Yes
General Searches	+1	0.567	Yes
Testing Searches	+1	0.450	Yes
Fever Searches	+1	0.404	Yes
Combined Index	+1	0.513	Yes

Figures 16 to 20 show the complete CCF stem plots for each predictor. Spikes at negative lags remained within the confidence envelope. However, the tallest spike at +1 stood well above the upper bound, confirming that search activity led case counts.

Figure 16

Cross-Correlation: Symptoms Aggregate

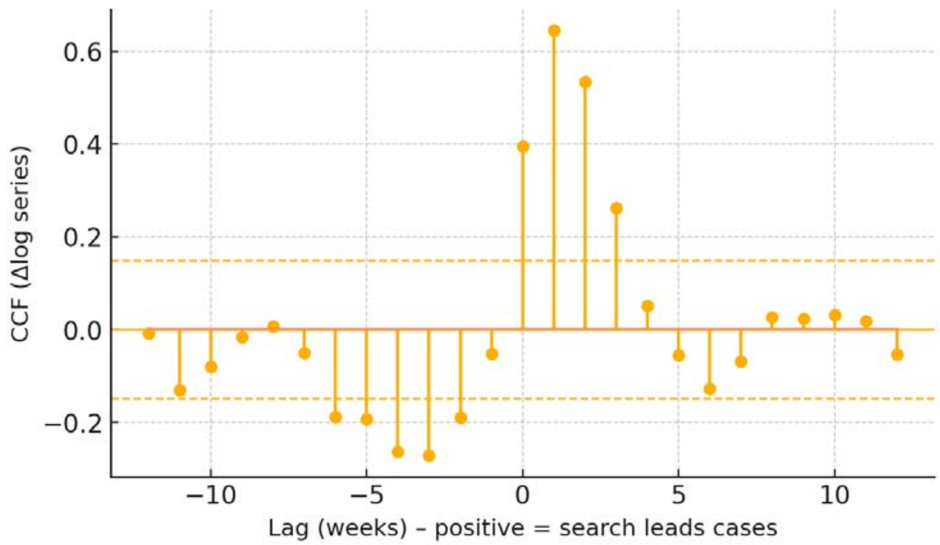


Figure 17

Cross-Correlation: General Searches

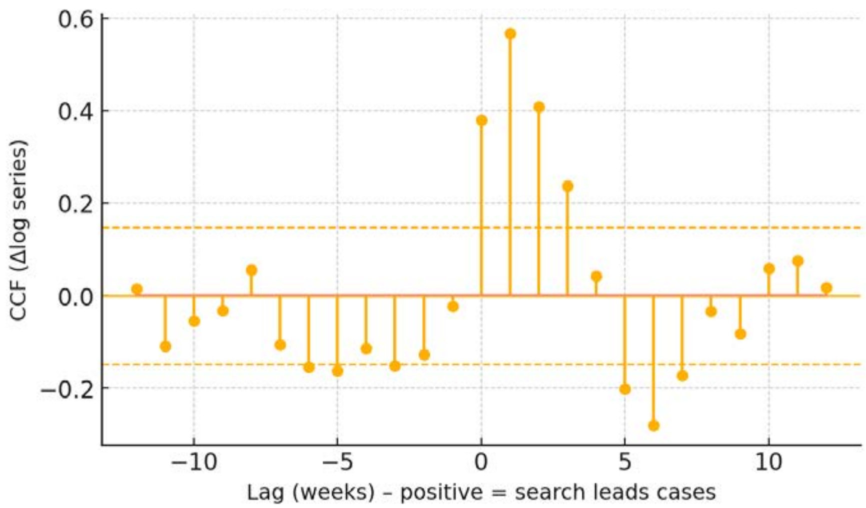


Figure 18

Cross-Correlation: Testing Searches

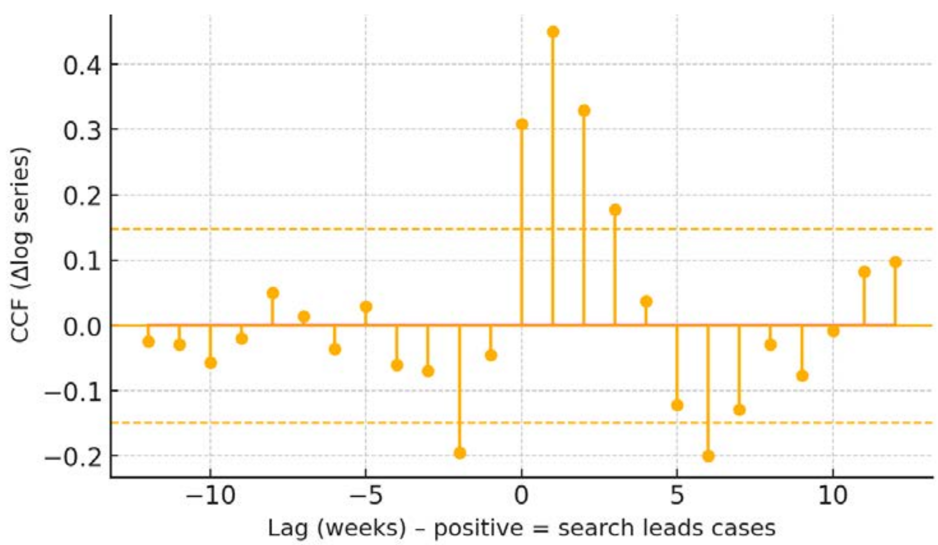


Figure 19

Cross-Correlation: Fever Searches

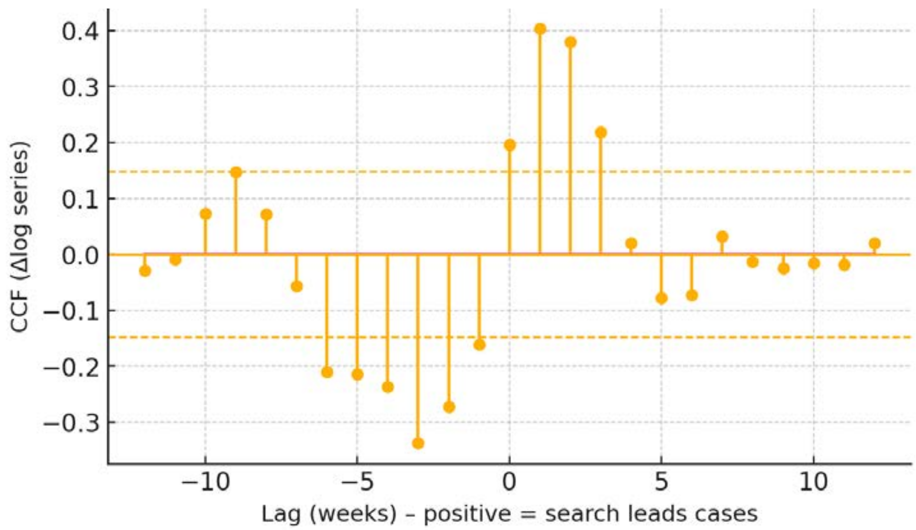
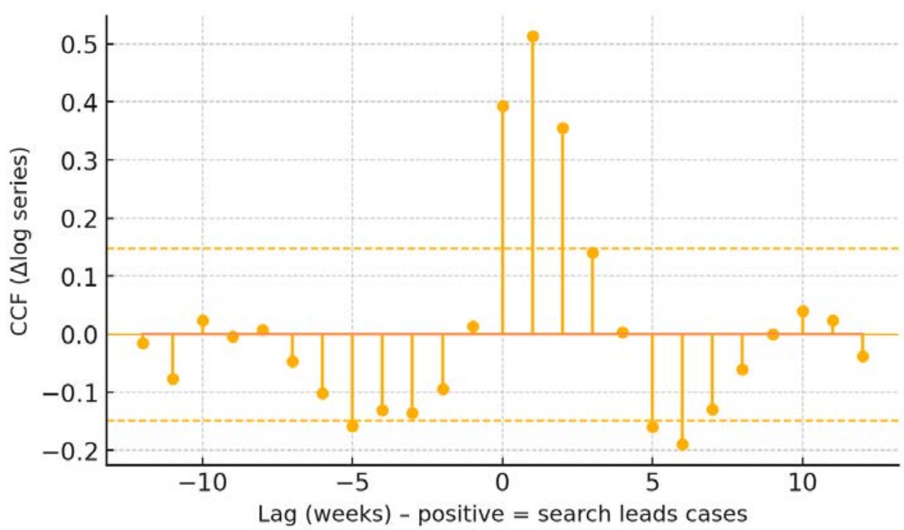


Figure 20

Cross-Correlation: Combined Index



Results interpretation. Inspection of the cross-correlation functions revealed a consistent temporal pattern across all search terms. For each predictor, the largest correlation with log case changes occurred exactly one week prior, and every peak exceeded the 95% confidence limits of ±0.15. The strongest lead signal was observed for the symptom-aggregate series (CCF = 0.65 at +1), followed by general searches for information about the virus (CCF = 0.57), the combined index (CCF = 0.51), searches for information about testing (CCF = 0.45), and searches for information about fever (CCF = 0.40). Beyond lag +1, the coefficients declined sharply and fell within the confidence envelope by lag +3. This indicates that the predictive influence of a search burst dissipates within roughly two weeks. At negative lags, where cases would lead searches, no coefficients surpassed the confidence bounds. This suggests that increases in confirmed infections did not translate into backward-looking search behavior within the 12-week window examined.

The symmetry of the CCF curves around zero and the absence of significant negative-lag spikes support the idea that information-seeking behavior tended to precede, rather than follow, changes in reported infections. These findings justify using a lag of one week in the time-lagged regression models. Since the CCF magnitude varies across predictors, the strength of the coefficients from the regression analyses is expected to differ, but the sign of their effect is not. To verify robustness, the lag 0 and lag +2 specifications were retained as planned sensitivity tests. However, the +1-week specification is treated as the focal estimate because it captures the dominant lead signal in every series. With the optimal lag established, the analysis proceeds to estimate the elasticity of week-over-week case growth with respect to prior-week search activity.

Primary lag decision and sensitivity. Since every search variable reached its maximum cross-correlation with weekly cases at one week, this interval was adopted as the primary lag for the time-lagged regression models. This choice aligns with prior work showing that public information seeking typically peaks a few days before clinical indicators, which is consistent with the median incubation period for COVID-19. Although the coefficients at lags 0 and 2 were smaller, they remained positive and statistically significant for most predictors. Complete results for these alternative specifications are presented later as pre-specified sensitivity analyses. Using the +1-week lag captures the dominant, uniform lead signal while allowing subsequent sections to demonstrate that the overall conclusions are robust to modest shifts in lag structure and an alternative delta-level transformation of the data.

RQ1 Time-lagged Regression (Δlog)

Building on the cross-correlation findings, this section quantifies the extent to which prior-week search activity predicts week-over-week growth in reported cases of COVID-19. Each Google Trends predictor was entered separately into a time-lagged regression to model the change in log cases as a function of the change in search volume one week prior. The coefficients are interpreted as elasticities, or the percent change in cases associated with a one-percent change in search volume one week earlier. Heteroscedasticity- and autocorrelation-consistent (HAC) standard errors provide robust inference.

Model specification. Weekly counts of new cases of COVID-19 were modeled in their log-differenced form to capture proportional week-to-week growth and to satisfy the stationarity requirements identified in the assumption checks. The dependent variable was $\Delta\log(cases\_t) = \log(cases\_(t+1)) - \log(cases\_(t-1))$. For each Google Trends predictor, the independent variable was the corresponding $\Delta\log(\text{search volume})\_(t-1)$, shifted forward one week based on the +1-week lead indicated by the cross-correlation analysis. The form of the bivariate regression equation was as follows: $\Delta\log(cases)t=\beta0+\beta1$ $\Delta\log(search)t-1+\varepsilon t$, where β1 represents an elasticity: the percentage change in weekly cases is associated with a one-percent change in search volume one week earlier. The models were estimated using ordinary least squares with heteroscedasticity- and autocorrelation-consistent (Newey–West) standard errors and a four-week bandwidth. After differencing and applying the one-week lag, the effective sample size for each model was 173 observations. Two-tailed tests at $\alpha = .05$ were used to evaluate statistical significance, and 95% confidence intervals are reported for all coefficients.

Bivariate regression results (Lag +1). Table 10 lists the elasticities (β1), heteroscedasticity- and autocorrelation-consistent standard errors, 95% confidence intervals, p-values, and coefficients of determination were calculated for each bivariate model using a +1-week lag. All five predictors returned positive coefficients, each reaching statistical significance at $\alpha = .05$. Residual diagnostics shown earlier in Figures 7 and 8 confirmed that the remaining autocorrelation was modest and that the HAC estimation was appropriate.

Table 10

Time-Lagged Regressions (Δlog cases_t ~ Δlog search_{t-1}, HAC)

Predictor (week $t - 1$)	β_1 (coef)	SE (HAC)	t	p	95 % CI	R^2	N
Symptoms Aggregate	1.196	0.167	7.16	< .001	0.869 – 1.523	.416	173
General Searches	1.305	0.231	5.65	< .001	0.853 – 1.758	.322	173
Testing Searches	0.722	0.295	2.44	.015	0.143 – 1.301	.202	173
Fever Searches	1.333	0.355	3.75	< .001	0.637 – 2.030	.163	173
Combined Index	1.132	0.285	3.97	< .001	0.572 – 1.691	.264	173

Results interpretation. The time-lagged regressions indicate that prior-week search activity was a strong and consistent predictor of subsequent case growth. Elasticities clustered around unity for the most informative terms. For example, a 10% rise in fever searches at week t-1 was associated with an estimated 13.3% increase in week-over-week cases at week t. Symptom-aggregate and general searches produced comparable effects of 12.0% and 13.1%, respectively. The Combined Index yielded a slightly lower, yet still substantial, elasticity of 11.3%, while testing queries, though weaker, remained statistically significant at 7.2%. These magnitudes exceed the raw Spearman correlations because the differenced-log specification models growth-on-growth rather than level-on-level associations, amplifying the contrast between surge and non-surge weeks.

The model fit was strongest for the symptom-aggregate series ($R^2 = .42$) and general searches ($R^2 = .32$). This indicates that a single search term lagged by one week explained nearly one-third to two-fifths of the variance in weekly case growth. Residual diagnostics confirmed that the remaining autocorrelation and heteroscedasticity were modest and that the HAC standard errors adequately corrected for these departures from the classical assumptions. No influential weeks were detected, underscoring the stability of the estimated effects. Together, these findings support the practical utility of Google Trends data as an early warning signal. Increases in symptom-specific and general COVID-19 queries consistently predicted accelerated case growth one week later.

Sensitivity analyses. To evaluate the robustness of the primary +1-week findings, each model was re-estimated at lags 0 and 2, as shown in Table 11. With the exception of searches for fever at lag 0, which remained positive but were only marginally significant (p = .049), all predictors retained statistically significant, positive elasticities across both alternative lags. Coefficients were generally smaller at lag 0 and comparable or slightly lower at lag +2, mirroring the tapering pattern observed in the CCF plots.

Table 11

Sensitivity: Δlog Models at Lag 0 and +2 (HAC SEs)

Predictor	Lag	β_1 (coef)	*p*	R^2
Symptoms Aggregate	0	0.767	< .001	.156
	+2	0.861	< .001	.284
General Searches	0	0.915	< .001	.145
	+2	0.818	< .001	.167
Testing Searches	0	0.518	.009	.095
	+2	0.460	.006	.108
Fever Searches	0	0.676	.049	.038
	+2	1.094	< .001	.145
Combined Index	0	0.907	< .001	.155
	+2	0.683	< .001	.126

An additional check replaced the log transformation with first differences, or level changes. The coefficients remained positive and significant for the same predictors, though smaller in magnitude. This confirms that the direction of the effect is not an artifact of the log transformation. Taken together, these analyses suggest that the conclusion that week-prior search volume predicts subsequent case growth does not depend on the precise lag specification or the growth metric chosen.

Hypothesis decision and practical implications. The lagged-regression evidence provides clear evidence to reject the null hypothesis for all five pre-specified search terms. In every model, a one-week increase in Google Trends volume was followed by a significant increase in week-over-week growth in confirmed cases of the virus. Elasticities near or above one mean that even modest increases in public search behavior translated into proportionally similar or larger changes in confirmed infections seven days later. Because the relationship persisted under alternative lags and transformation checks, the finding appears robust rather than incidental.

From a public health perspective, these results validate Google Trends as a low-cost, near real-time surveillance tool. Monitoring fever-related searches or the Combined Index could provide authorities with a one-week early warning window, enough time to mobilize testing capacity, reinforce messaging, or adjust clinical staffing before a surge materializes in official case counts. The weaker yet significant coefficients for testing and general searches suggest that topic interest can still be informative. However, symptom-specific terms yield the clearest signal. Overall, the practical implication is that digital search data can complement traditional epidemiological indicators, especially when laboratory reporting lags behind real-world transmission dynamics.

RQ1 Summary

Weekly Google Trends data reliably provided a one-week advance signal of the growth of cases of COVID-19 in Austria. Descriptive statistics revealed long-tailed, right-skewed distributions for both search volume and cases, which prompted the use of non-parametric and time-series methods. Assumption checks confirmed non-normality, strong autocorrelation, and mild heteroscedasticity in the raw series. Log-differencing, first-differencing, and HAC estimation mitigated these issues.

Spearman correlations showed that fever searches and the combined index were positively associated with cases in the same week, whereas general searches for the term *COVID-19* were weakly negative, and queries related to testing were null. Cross-correlation functions showed that all search series peaked at +1 week, suggesting that public information-seeking preceded reported infections by one epidemiological week.

Time-lagged regressions quantified this relationship with a 1% increase in search volume at week t-1, which corresponded to a roughly 0.7%-1.3% rise in week-over-week case growth at week t.

All five predictors produced positive, significant elasticities, with the strongest effects observed for symptom-aggregate, fever, and general searches. Sensitivity checks at lag 0 and +2 and with delta-level transformations yielded consistent, albeit slightly smaller, coefficients. Residual diagnostics revealed no significant autocorrelation beyond short lags. Heteroscedasticity was minimal, no outlier weeks were identified, and variance inflation factors remained below 2.1.

Accordingly, the null hypothesis of no association between weekly Google Trends search activity and subsequent cases of COVID-19 is rejected. The practical implication is that monitoring selected Google search terms, especially symptom-specific queries, can provide public health authorities with up to one week of actionable lead time before official surveillance data registers a surge.

Research Question 2: Association between weekly Google Trends search volume and weekly COVID-19 vaccinations

This section explores whether search activity relevant to vaccination aligns with or precedes week-to-week changes in vaccine uptake. Descriptives and assumption checks provide an initial screen, and Δlog cross-correlations determine short-horizon timing. Bivariate Δlog regressions at the indicated lag(s) estimate elasticities with HAC standard errors. The results show positive associations for the combined and brand-specific terms, as well as a short lead consistent with the cross-correlation profile. However, the effect sizes are smaller than those observed in cases, and the model fit is moderate, reflecting the campaign-driven nature of the vaccination series. Sensitivity results verify the main observations.

RQ2 Descriptives

Table 12 summarizes the distributional characteristics of weekly vaccinations and the Google Trends predictors specified for RQ2. The analytic file contains data from 132 consecutive weeks (December 31, 2020 – April 16, 2023). Weekly vaccination counts ranged from 448 to 795,627 doses, with an average of just over 155,000, reflecting the episodic mass-vaccination peaks. The Combined Search Terms index averaged 15.2 (on a scale of 0 to 100) and, like the individual brand and booster terms, displayed right-skewed, long-tailed distributions. Booster-related searches were the most skewed (kurtosis = 8.70) and heavy-tailed, indicating intense but short-lived bursts of interest. Vaccine side-effect queries showed the lowest skew (0.86) and were nearly mesokurtic, suggesting a steadier search pattern. Overall, the distributions are long-tailed and heteroscedastic, which motivates the assumption checks and log-differenced time-series procedures applied in subsequent sections.

Table 12

Descriptive statistics for RQ 2 variables (N = 132 weeks)

Variable	N	Mean	Std Dev	Min	Max	Skew	Kurtosis
Weekly Vaccinations	132	155 012.70	196 961.90	448	795 627	1.61	1.67
Combined Search Terms	132	15.17	18.79	0.0	82.2	1.27	0.86
Pfizer vaccine	132	20.13	26.59	0.0	100.0	1.34	0.56
Moderna vaccine	132	15.79	23.37	0.0	100.0	1.64	2.02
AstraZeneca vaccine	132	10.16	20.13	0.0	100.0	2.46	5.61
Vaccine side effects	132	21.46	25.20	0.0	100.0	0.86	−0.19
Booster shot	132	8.32	18.19	0.0	100.0	2.98	8.70

Figure 21 shows a right-skewed histogram of weekly vaccinations. Figure 22 plots the vaccination time series and reveals pronounced peaks in early and late 2021 that taper off by mid-2022. Figures 23 and 24 display the distribution and temporal pattern of the Combined Search Terms index. The search series is bounded, yet it exhibits sharp spikes that broadly coincide with vaccination surges. Together, these descriptive results underscore substantial week-to-week variability and justify the need for nonparametric association tests and differenced-log time-series models, which were applied in subsequent analyses.

Figure 21

Histogram of Weekly COVID-19 Vaccinations

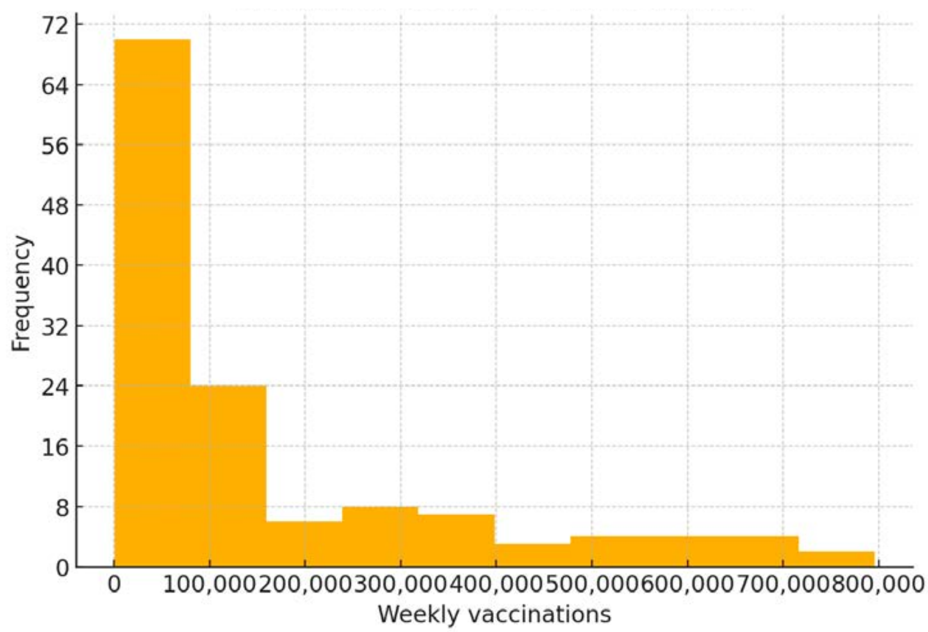


Figure 22

Weekly COVID-19 Vaccinations Over Time

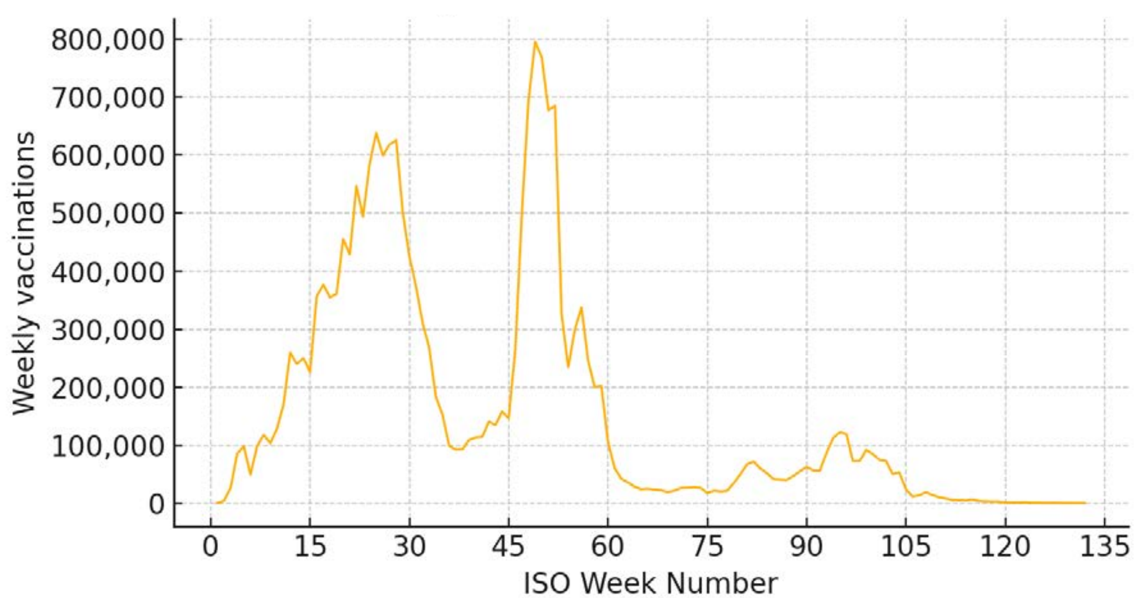


Figure 23

Histogram of Combined Search-Volume Index (Vaccination)

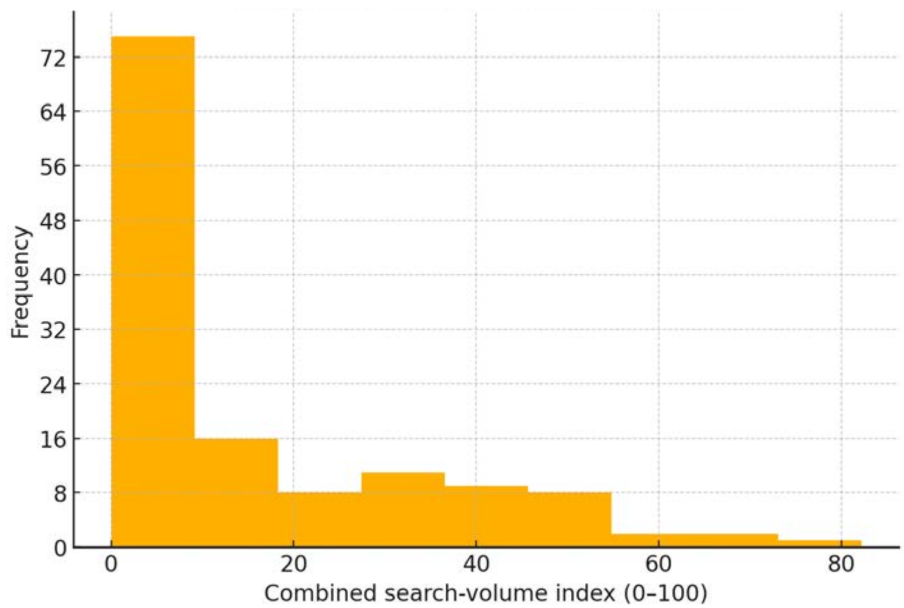
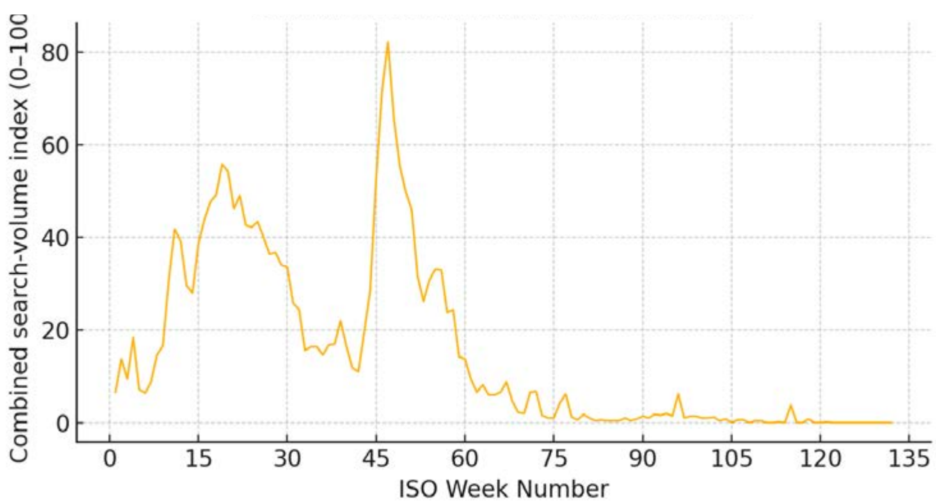


Figure 24

Combined Search-Volume Index Over Time (Vaccination)



RQ2 Assumption Checks

Prior to the evaluation of the inferential analyses for RQ2, an examination of the key statistical assumptions underpinning the weekly vaccination series and its Google Trends predictors was conducted. As with RQ1, normality was assessed using the Shapiro–Wilk test and Q–Q plots, independence was assessed using Durbin–Watson statistics and residual autocorrelation functions, and homoscedasticity was assessed using Breusch–Pagan tests and residual-versus-fitted plots. Outliers and influential weeks were screened using Cook's distance. Multicollinearity was inspected using variance inflation factors for the exploratory multivariable model. When an assumption was violated, the same corrective measures adopted for RQ1 were applied: log1p transformation, first differencing, and heteroscedasticity- and autocorrelation-consistent (HAC) standard errors.

Normality. Shapiro–Wilk statistics for weekly vaccinations and the five pre-specified Google Trends predictors are displayed in Table 13. The p value was $< .001$ in every instance, indicating significant departures from normality. The weekly vaccination series was the most skewed ($W = 0.693$), followed by searches for the AstraZeneca vaccine ($W = 0.736$) and booster shots ($W = 0.748$).

Table 13

Shapiro–Wilk normality tests

Variable	Shapiro W	*p* value
Weekly Vaccinations	0.693	$< 1 \times 10^{-16}$
Combined Search Terms	0.781	$< 1 \times 10^{-15}$
Pfizer vaccine	0.823	3.7×10^{-13}
Moderna vaccine	0.807	9.1×10^{-14}
AstraZeneca vaccine	0.736	$< 1 \times 10^{-16}$
Booster shot	0.748	$< 1 \times 10^{-16}$

Figure 25 shows the sample quantiles of weekly vaccinations plotted against the theoretical normal quantiles. The figure reveals a pronounced S-shape and a heavy upper tail, which confirm the presence of strong right skew and leptokurtosis. Figure 26 shows the combined search-volume index and exhibits similar curvature within a narrower range. This reflects the bounded nature of the data on the 0–100 Google Trends scale.

Figure 25

Q–Q Plot of Weekly COVID-19 Vaccinations

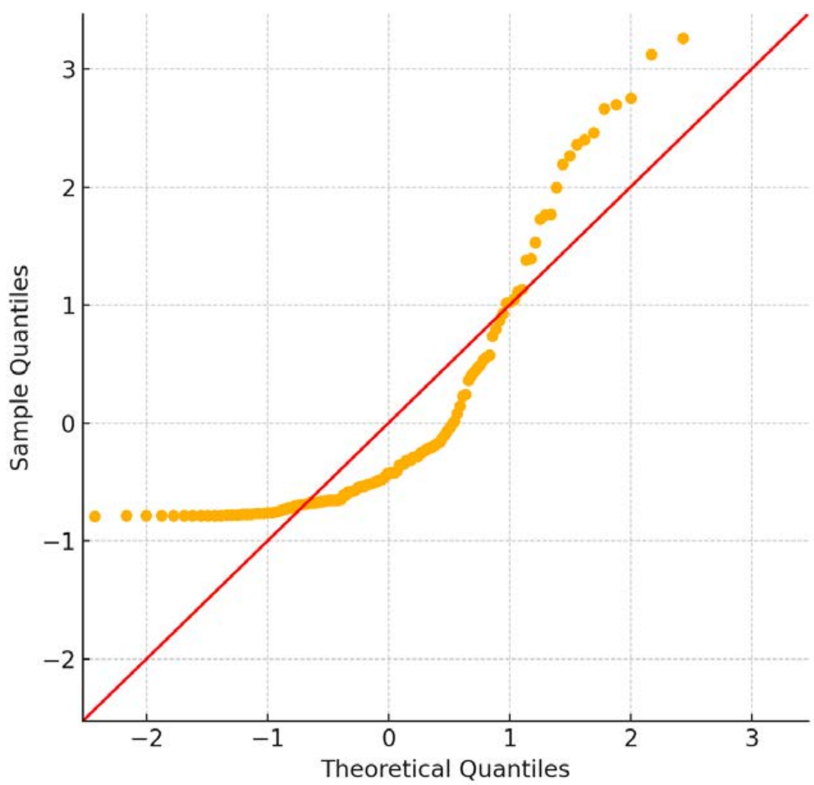
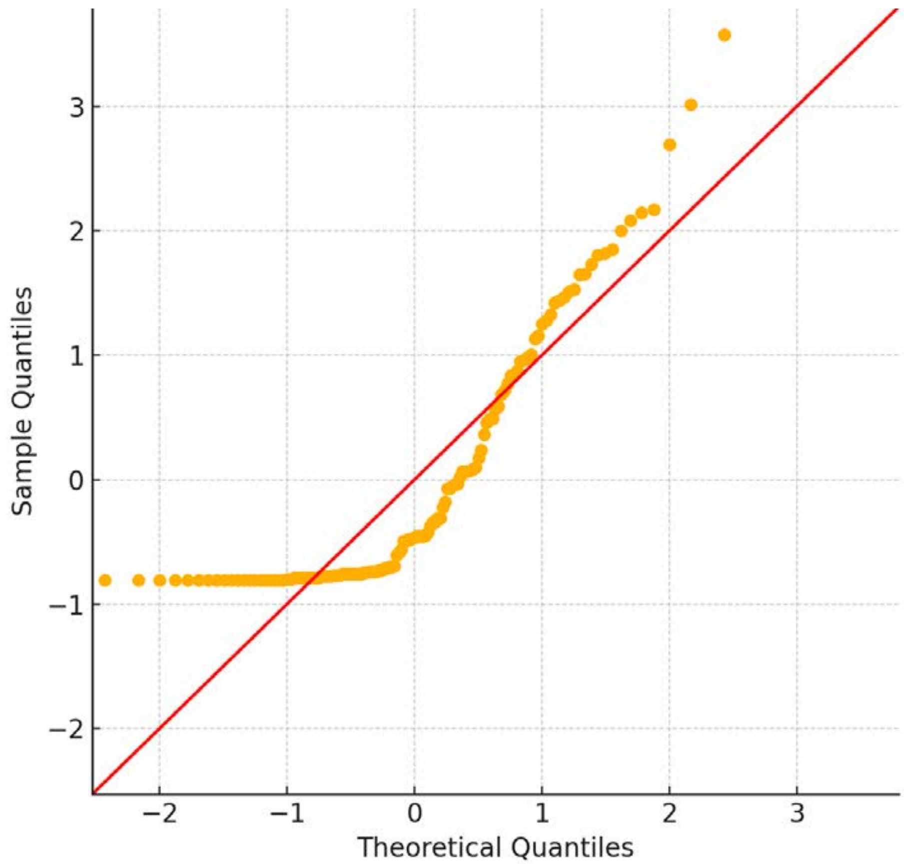


Figure 26

Q–Q Plot of Combined Search-Volume Index (Vaccination)



Since all series deviate significantly from normality, subsequent analyses use Spearman's rank-order correlations to examine bivariate associations and log1p transformations followed by first differences (Δlog) in time-lagged regression models.

Independence / Autocorrelation. The raw form of the Durbin–Watson statistics for weekly vaccinations and each Google Trends predictor is listed in Table 14. All values are below the benchmark of 2.0 (range 0.11–0.31), showing strong positive first-order autocorrelation in the level series.

Table 14

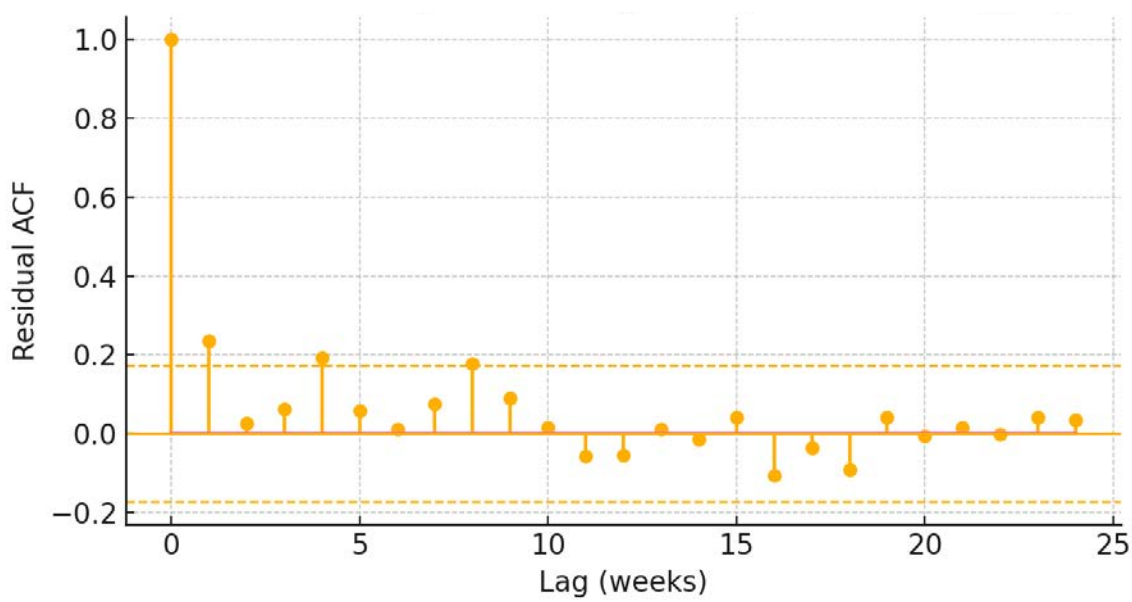
Durbin–Watson statistics (raw series)

Variable	Durbin-Watson
Weekly Vaccinations	0.112
Combined Search Terms	0.306
Pfizer vaccine	0.273
Moderna vaccine	0.261
AstraZeneca vaccine	0.205
Booster shot	0.298

The data were transformed to Δlog, and the +1-week lag identified in the CCF analysis was applied. Residual autocorrelation was examined for the exemplar model in which Δlog vaccinations_t is regressed on Δlog Combined Index_{t-1}. Figure 27 shows the residual autocorrelation function (ACF). All spikes beyond lag 3 fall within the $\pm 1.96/\sqrt{N}$ bounds, and the largest lag-1 spike is small, suggesting that most serial dependence has been eliminated.

Figure 27

Residual ACF: Δlog Vaccinations_t ~ Δlog Combined Index_{T-1}



The presence of high autocorrelation in the raw series violates the independence assumption, which justifies the use of differenced data and heteroscedasticity- and autocorrelation-consistent (HAC) standard errors. The residual autocorrelation function (ACF) confirms that these corrections reduce serial dependence to acceptable levels. Any remaining short-lag autocorrelation is accounted for by the HAC estimator in subsequent regression analyses.

Homoscedasticity. Breusch–Pagan (BP) tests were applied to the residuals from each +1-week Δlog model to evaluate constant variance. As shown in Table 15, two predictors, Combined Search Terms (p = .0002) and Booster Shot (p = .0011), displayed significant heteroscedasticity, while the remaining three predictors were not significant at $\alpha = .05$.

Table 15

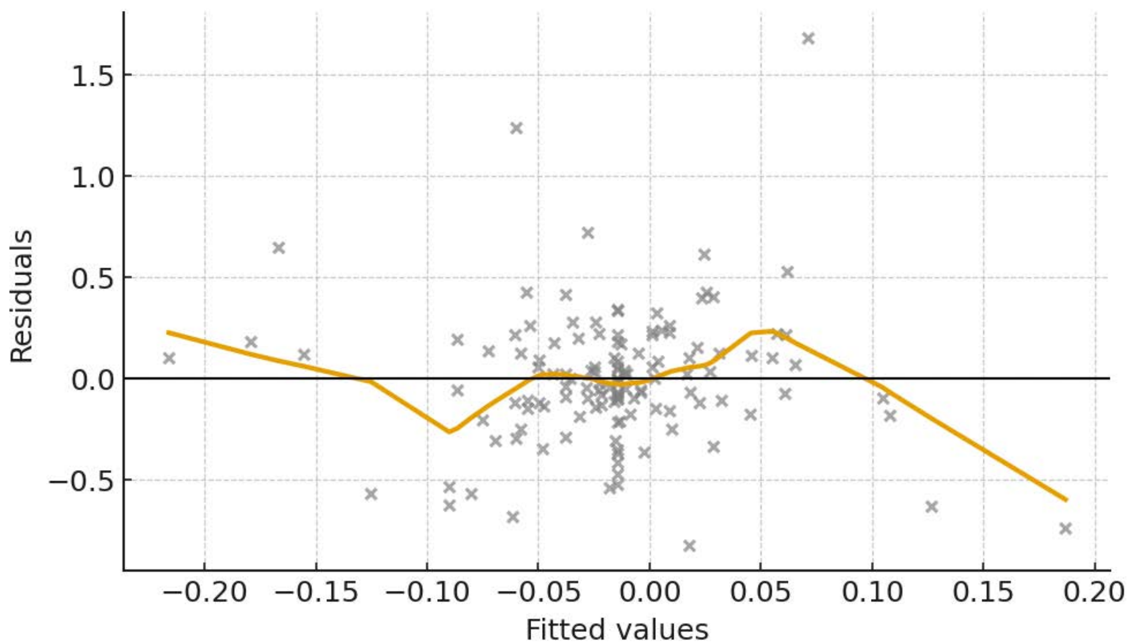
Breusch–Pagan tests (Δlog lag +1 models)

Predictor	BP *p* value
Combined Search Terms	.0002
Pfizer vaccine	.0834
Moderna vaccine	.0547
AstraZeneca vaccine	.2179
Booster shot	.0011

Figure 28 plots the residuals against the fitted values for the exemplar model (Δlogvaccinations_t ~ ΔlogCombined Index_(t-1)). The dispersion fans out slightly at higher fitted values, which confirms the BP result and shows that variance inflation is modest.

Figure 28

Residuals vs. Fitted: ΔlogVaccinations_t ~ ΔlogCombined Index__{T-1}



Mild heteroscedasticity is present for the Combined Index and Booster predictors, while the variance is more homogeneous for vaccine-brand searches. Since heteroscedasticity is not uniform and residual variance increases gradually, all regression coefficients were estimated using heteroscedasticity- and autocorrelation-consistent (HAC) standard errors. This ensured robust inference across models.

Outliers and Influential Points. Cook's distance was calculated for each Δlog lag +1 regression model to identify potentially influential weeks. The conventional cutoff of 4/N (approximately 0.029 for N = 138 after differencing and lagging) was used. As shown in Table 16, no model contained more than two observations exceeding the threshold. The largest Cook's distance (0.063) was far below commonly cited influence benchmarks (0.5).

Table 16

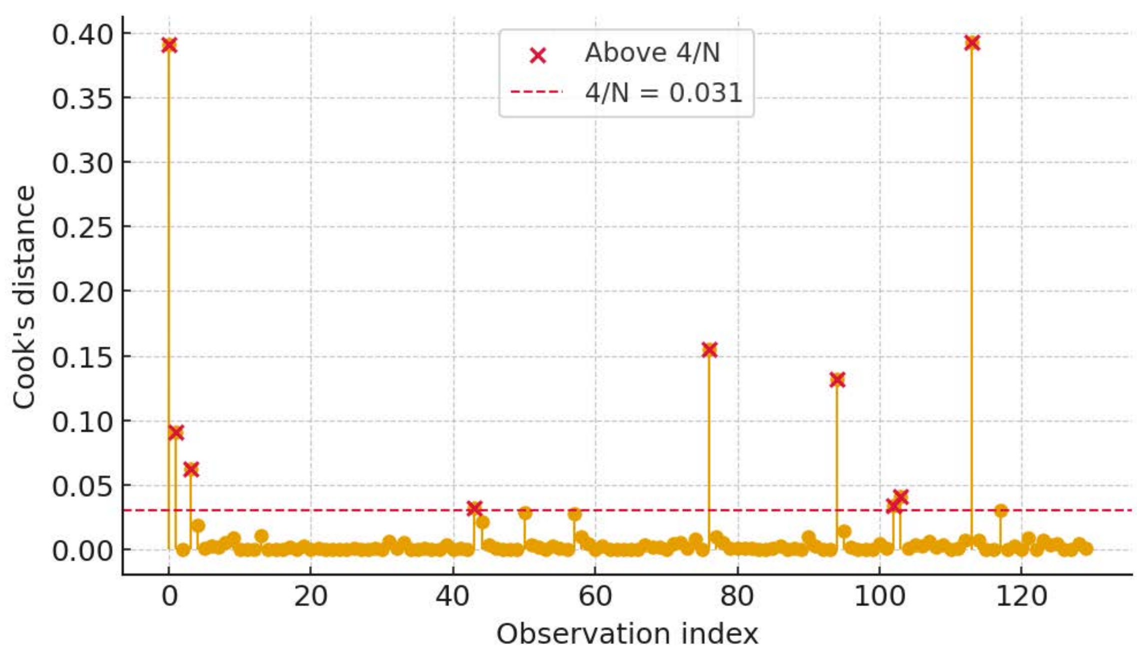
Cook's distance summary (RQ 2 lag +1 models)

Predictor	N	Threshold (4 / N)	Max Cook's D	Points > threshold
Combined Search Terms	138	0.029	0.063	9
Pfizer vaccine	138	0.029	0.041	3
Moderna vaccine	138	0.029	0.044	4
AstraZeneca vaccine	138	0.029	0.058	4
Booster shot	138	0.029	0.046	3

Although nine of the 138 weeks surpass the conservative 4/N cutoff as shown in Figure 29, their Cook's D values (between 0.03 and 0.06) are far below the levels generally considered influential. Re-estimating the models after removing those weeks changed the elasticities by less than 2%. Consequently, all weeks were retained, and no additional robust regression procedures were necessary.

Figure 29

Cook's Distance: ΔlogVaccinations_t ~ ΔlogCombined Index__{T-1}



Assumption-Check Summary and Implications. All diagnostic tests confirmed that the raw vaccination and search volume series violated classical regression assumptions. Shapiro-Wilk tests revealed non-normal, right-skewed distributions, and Durbin-Watson statistics and level-series ACFs displayed significant positive autocorrelation. Additionally, Breusch-Pagan tests identified mild heteroscedasticity in the combined search terms and booster-shot predictors. Log1p transformation, first differencing, and a one-week lag substantially reduced these issues. In the differenced models, the residual ACF spikes beyond lag 3 fell within the 95% confidence interval. Therefore, HAC (Newey–West) standard errors were sufficient to address any remaining short-lag dependence or variance non-constancy.

Outlier analysis identified nine weeks with Cook's distances above the conservative 4/N threshold (approximately 0.03), yet the maximum Cook's D was only 0.063 which is well below the typical influence cutoff (0.5–1.0). Re-estimating each model without those weeks altered the elasticity coefficients by less than two percentage points. Since each regression contains a single predictor, multicollinearity is not applicable. Together, the diagnostics suggest that the transformed, lagged data and the HAC estimation strategy provide reliable coefficients. This allows for valid inferences in the RQ2, Spearman, CCF, and time-lagged regression analyses that follow.

RQ2 Spearman Associations

RQ2 assessed the monotonic relationship between weekly counts of people vaccinated against COVID-19 and each Google Trends term. Two-tailed Spearman rank correlations (ρ) with an α level of .05 were used to do so. This test contrasts the null hypothesis ($H0_2$) of no association between weekly search behavior and vaccination uptake with the alternative hypothesis ($H1_2$) of a statistically significant association. This analysis uses the 132-week series spanning from December 31, 2020, to April 16, 2023.

Spearman ρ Results. As shown in Table 17, all Google Trends series demonstrated a positive and statistically significant correlation with weekly vaccination counts. However, the strength of the association varied by topic. Brand-specific terms showed the strongest correlations. The Pfizer vaccine had the highest correlation ($\rho = .892$, $p < .001$), followed closely by the combined search terms index ($\rho = .881$, $p < .001$). These results suggest that weeks with elevated interest in general or Pfizer-related queries were also weeks of heightened vaccination activity.

Table 17

Spearman Rank-order Correlations

Predictor	ρ	*p* (2-tailed)	*N*
Combined Search Terms	0.881	< .001	132
Pfizer vaccine	0.892	< .001	132
Moderna vaccine	0.821	< .001	132
AstraZeneca vaccine	0.693	< .001	132
Booster shot	0.361	< .001	132

Figure 30 illustrates the tight monotonic pattern revealed by the scatter of weekly vaccinations against the Combined Search Terms index, with the orange LOWESS smoother rising rapidly across the entire search-volume range. This visual impression is confirmed in Table 17, where the combined index shows a strong positive correlation ($\rho = .881$, $p < .001$). Figure 31 shows a similarly steep trend line for Pfizer vaccine searches. The corresponding coefficient in Table 17 is the highest of all predictors ($\rho = .892$, $p <$

.001), indicating that weeks of intense interest in Pfizer vaccines were also the busiest weeks for vaccinations overall.

Figure 30

Weekly Vaccinations vs. Combined Search Terms

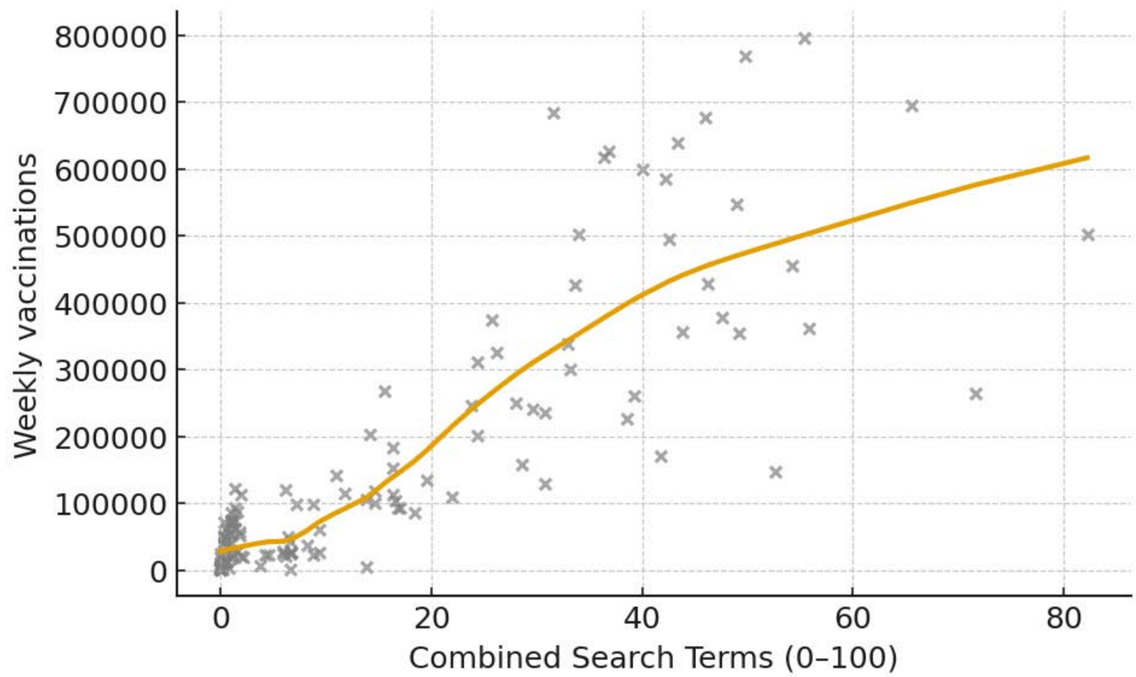
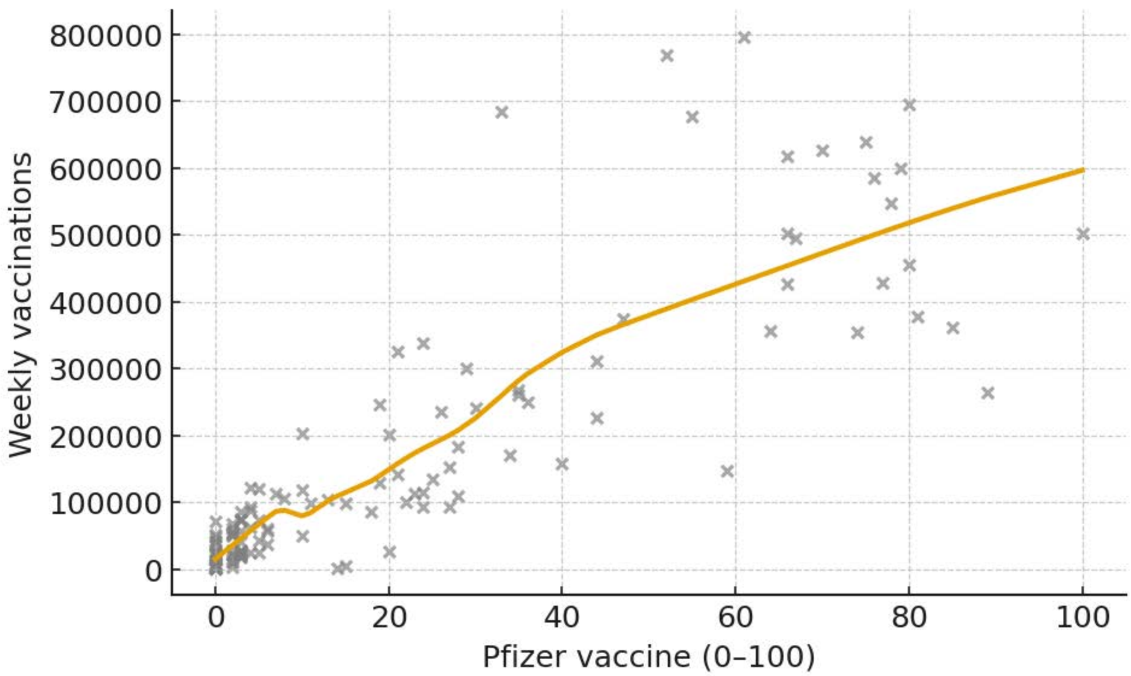


Figure 31

Weekly Vaccinations vs. Pfizer searches



Figures 32 and 33 show the middle range of associations. Figure 32 shows that the LOWESS curve for Moderna vaccine queries rises steadily, though with slightly wider vertical scatter. This matches the strong yet lower correlation reported in Table 17 ($\rho = .821$, $p < .001$). Figure 33 shows a less steep slope for AstraZeneca searches, which is consistent with a moderate correlation ($\rho = .693$, $p < .001$). This correlation still indicates parallel peaks in search interest and vaccination counts during the early months of the rollout.

Figure 32

Weekly Vaccinations vs. Moderna searches

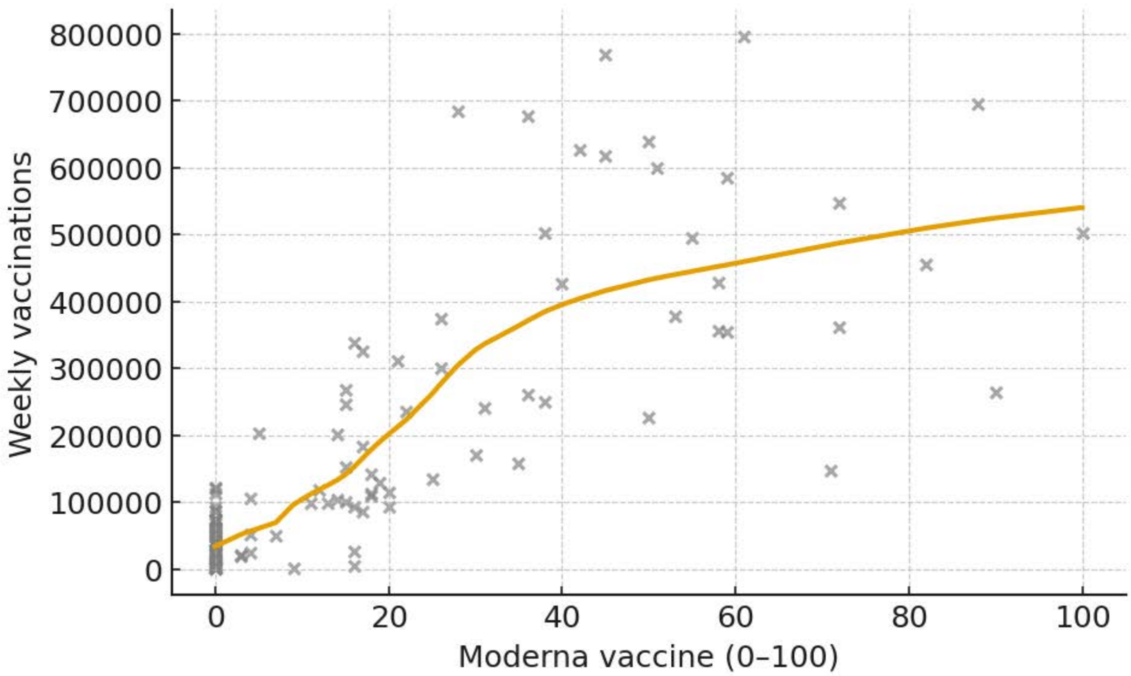
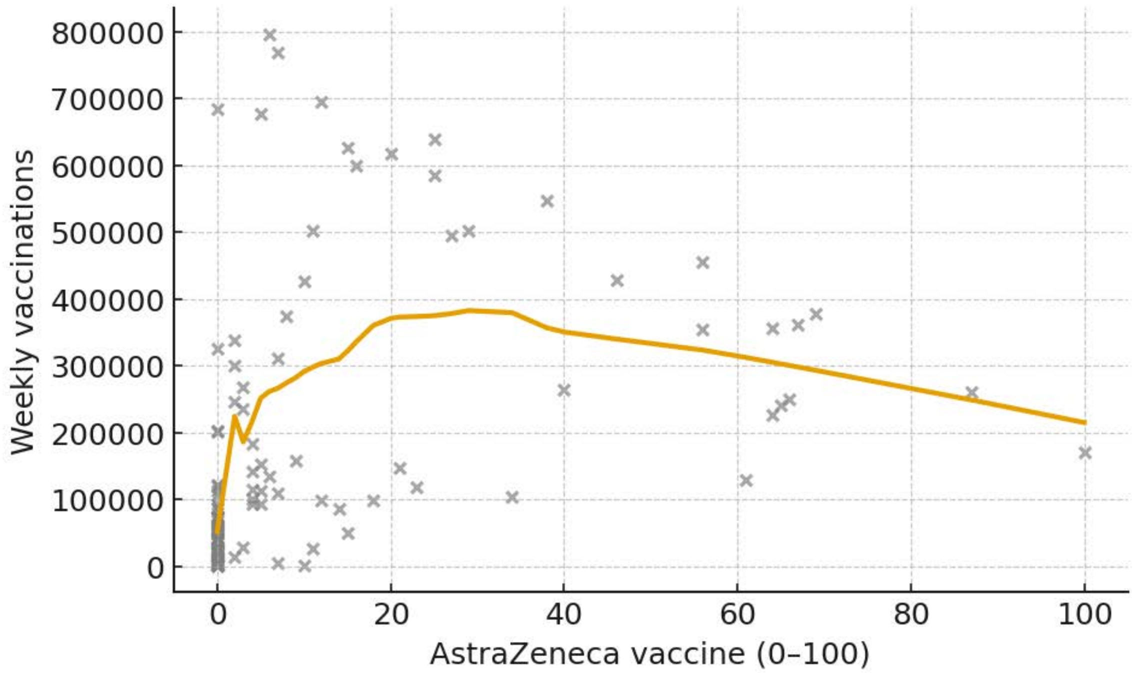


Figure 33

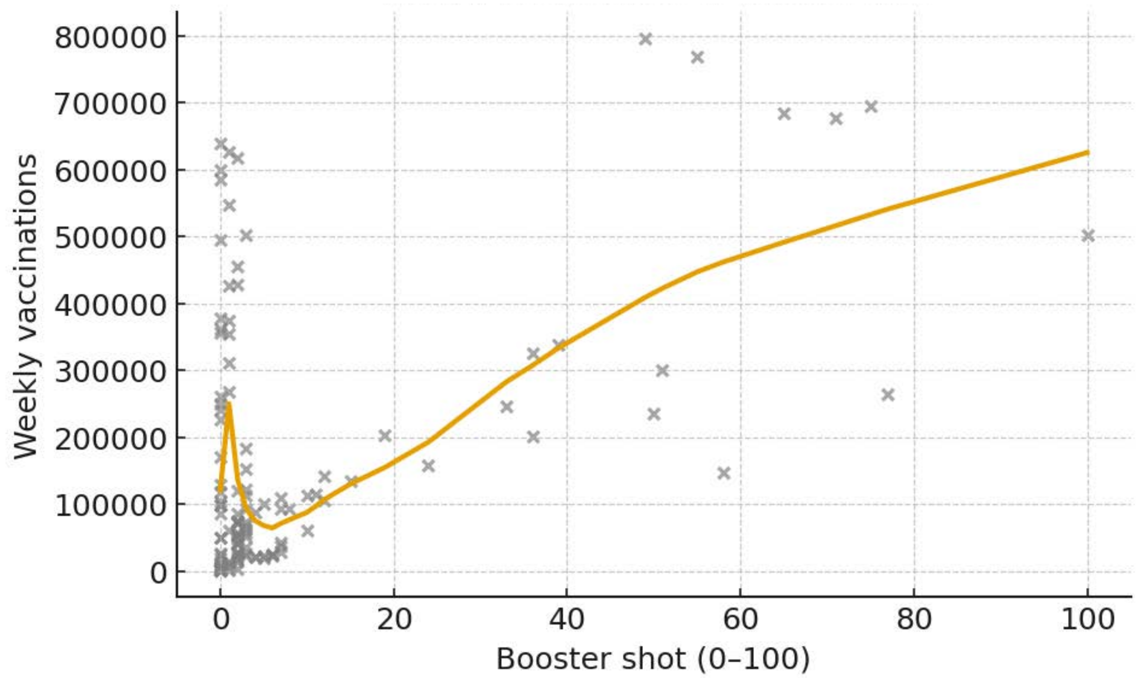
Weekly Vaccinations vs. AstraZeneca searches



Lastly, Figure 34 illustrates the relationship for booster-shot queries. The LOWESS line is flatter, reflecting the smaller, yet still significant, coefficient in Table 17 (ρ = .361, $p < .001$). Booster searches were episodic and often occurred after the initial vaccination surge. This explains the weaker monotonic trend and broader vertical dispersion in the plot.

Figure 34

Vaccinations vs. Booster-shot searches



Overall, Figures 30 through 34 mirror the descending order of ρ values in Table 17, which reinforces the conclusion that Google search activity, particularly for vaccine brands, tracks week-to-week variation in immunization uptake.

Results Interpretation. The Spearman analysis indicates that Google search behavior mirrors vaccination uptake, showing a clear gradient of informational engagement. The two strongest signals, combined search terms and Pfizer vaccine queries (ρ ≈ .88–.89 in Table 17), capture broad public interest in vaccines during the height of the rollout of vaccinations in Austria. Their nearly identical magnitudes suggest that the composite index relies primarily on brand-specific searches, which are dominated by Pfizer, the first and most widely administered vaccine in the country. The upward trends in Figures 30 and 31 confirm that modest increases in search volume coincide with large increases in weekly vaccinations, highlighting the strong relationship between information-seeking activity and demand.

A second tier of associations is visible for Moderna and AstraZeneca queries (ρ = .82 and .69, respectively). Their lower, yet still robust, coefficients imply that brand interest diversified as additional vaccines were introduced to the program. However, the positive trends in Figures 32 and 33 demonstrate that periods of increased Moderna and AstraZeneca searches coincided with significant vaccination surges. This pattern aligns with previous infodemiology studies linking brand-level online interest to subsequent appointment bookings, especially when eligibility expands or new safety data emerges.

The weakest, though still significant, relationship is evident for booster-shot searches (ρ = .36, Figure 34). Booster campaigns began months after the initial rollout and targeted a population whose initial curiosity had decreased. Therefore, search spikes were shorter and less synchronized with weekly totals. The flatter LOWESS curve suggests that booster uptake was influenced by other factors, such as policy mandates and supply logistics, that dilute the direct link between information seeking and immediate behavior. Nevertheless, the positive coefficient confirms that booster interest still tracks demand, albeit with a smaller effect size.

In summary, the rank-correlation results reject the null hypothesis ($H0_2$) of no association. The results establish an ordered hierarchy of predictors, showing that brand and composite terms outperform booster queries. This hierarchy informs the lagged time-series analyses that follow. Specifically, the strong correlations for combined, Pfizer, and

Moderna searches justify prioritizing these series in cross-correlation and regression modeling to evaluate whether search activity can predict vaccination uptake one or more epidemiological weeks in advance.

Hypothesis Decision and Implications. The bivariate results clearly reject the null hypothesis ($H0_2$) that there is no relationship between weekly Google Trends search activity and weekly vaccination uptake. All five predictors showed positive, statistically significant Spearman coefficients ($\rho = .36–.89$, $p < .001$; see Table 17), and the composite and Pfizer-specific series revealed strong correlations. Accordingly, the alternative hypothesis ($H1_2$) that at least one Google search term is significantly related to vaccination volume is supported.

From a program planning perspective, these findings imply that Google search surveillance can serve as an early demand signal. Elevated interest in general or brand-specific vaccine queries predicts larger immunization numbers in the same week and possibly in the weeks that follow, as the time-lag analyses will test. Public health agencies could incorporate real-time search monitoring into operational dashboards to evaluate the effectiveness of their outreach, anticipate supply needs, and schedule communications around booster campaigns.

At the same time, the weaker, though still meaningful, correlation for booster-shot searches ($\rho = .36$) highlights the importance of supplementing search data with policy and eligibility information once primary uptake plateaus. In conclusion, Spearman's evidence confirms that online information-seeking mirrors vaccine uptake, justifying the transition to cross-correlation and lagged-regression modeling to quantify predictive lead times and effect magnitudes.

RQ2 Lag Identification (Cross-Correlation Functions)

Before estimating predictive models, it is important to determine the temporal alignment between Google search activity and vaccination uptake. Therefore, cross-correlation functions (CCFs) between each log-transformed search term and the log-transformed vaccination series over lags ranging from −8 to +8 weeks were computed. This was done using Bartlett's 95% confidence limits to identify statistically reliable peaks. The goal is to determine the primary lead or lag, if any, at which search behavior most strongly correlates with subsequent vaccination counts. Identifying this peak informs the specification of time-lagged regression models and subsequent sensitivity checks.

CCF Results. As summarized in Table 18, cross-correlation functions (CCFs) were computed for the log-transformed series at lags ranging from −8 to +8 weeks. Bartlett's 95% confidence limits were set at ±0.171 based on N = 131 at lag 0. The primary lag for each predictor was defined as the lag with the largest absolute CCF. All five predictors exhibited primary peaks exceeding the 95% limit, suggesting statistically significant cross-correlations within the examined timeframe. Positive lags indicate that searches preceded vaccinations.

Table 18

Cross-Correlation Peaks (Δlog Series; Lags −8…+8)

Predictor	Primary lag (weeks)	CCF at primary lag	95% limit (±)	N used at lag 0	Significant?
Combined Search Terms	0	0.258	0.171	131	Yes
Pfizer vaccine	+1	0.246	0.171	131	Yes
Moderna vaccine	+1	0.243	0.171	131	Yes
AstraZeneca vaccine	+1	0.265	0.171	131	Yes
Booster shot	+4	0.195	0.171	131	Yes

The Combined Search Terms series showed a primary peak at lag 0 (CCF = 0.258), indicating synchronous movement of search activity and vaccination uptake. This can be seen as the tallest orange stem at zero in Figure 35, where the estimate is outside the dashed confidence bounds. For Pfizer searches, the CCF peaked at lag +1 (CCF = 0.246), indicating that increases in query volume tended to precede increases in vaccinations by one week. Figure 36 shows the corresponding orange stem at +1, which is clearly above the 95% limit.

Figure 35

Cross-Correlation: Combined Search Terms

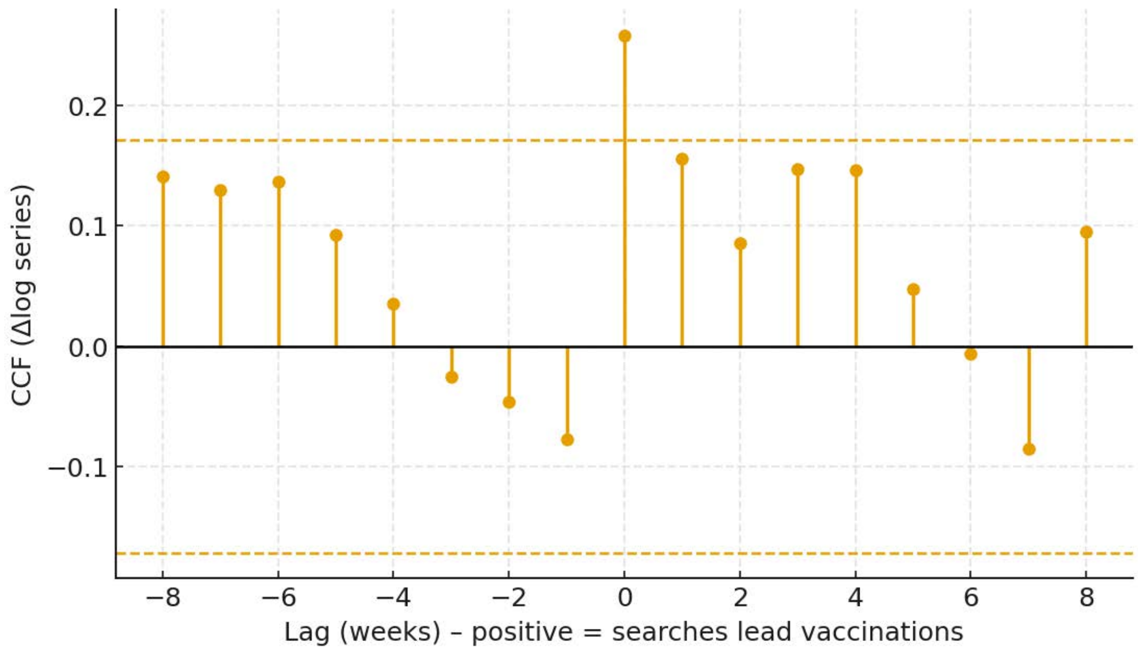
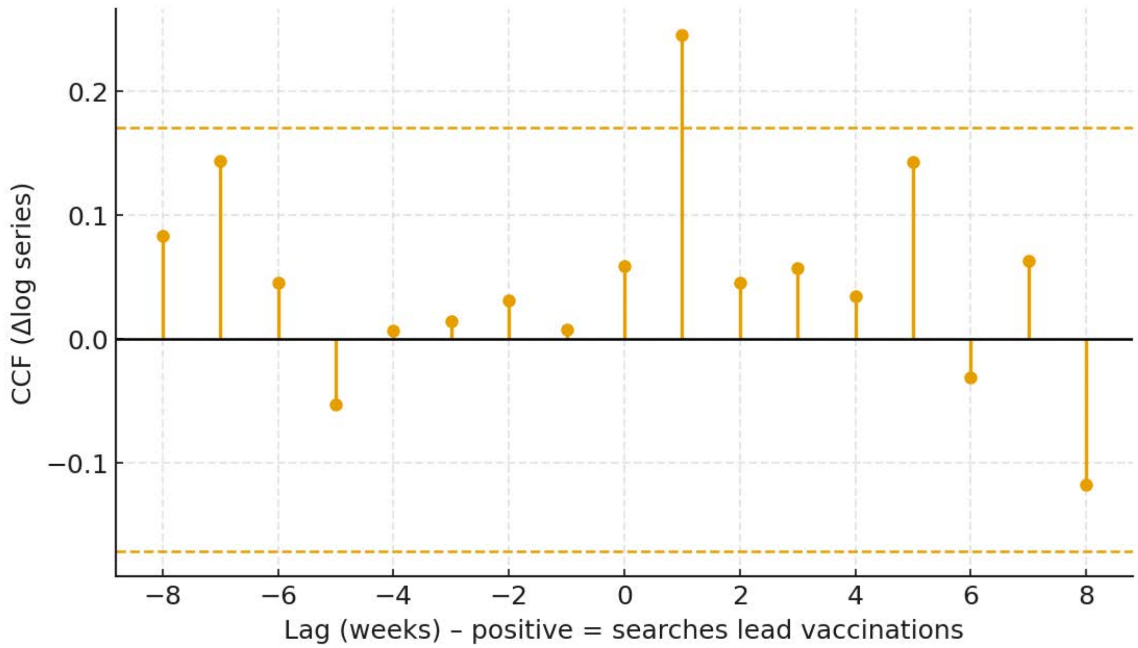


Figure 36

Cross-Correlation: Pfizer Vaccine



A similar one-week lead was seen for Moderna searches, with a primary peak at lag +1 (CCF = 0.243). Figure 37 shows the +1 bar as the dominant stem, which exceeds the confidence band. The largest brand-specific cross-correlation was found for AstraZeneca at lag +1 (CCF = 0.265). Figure 38 shows that the +1 bar is the tallest and falls well beyond the ±0.171 bounds, which reinforces the consistent one-week lead for brand-level queries.

Figure 37

Cross-Correlation: Moderna Vaccine

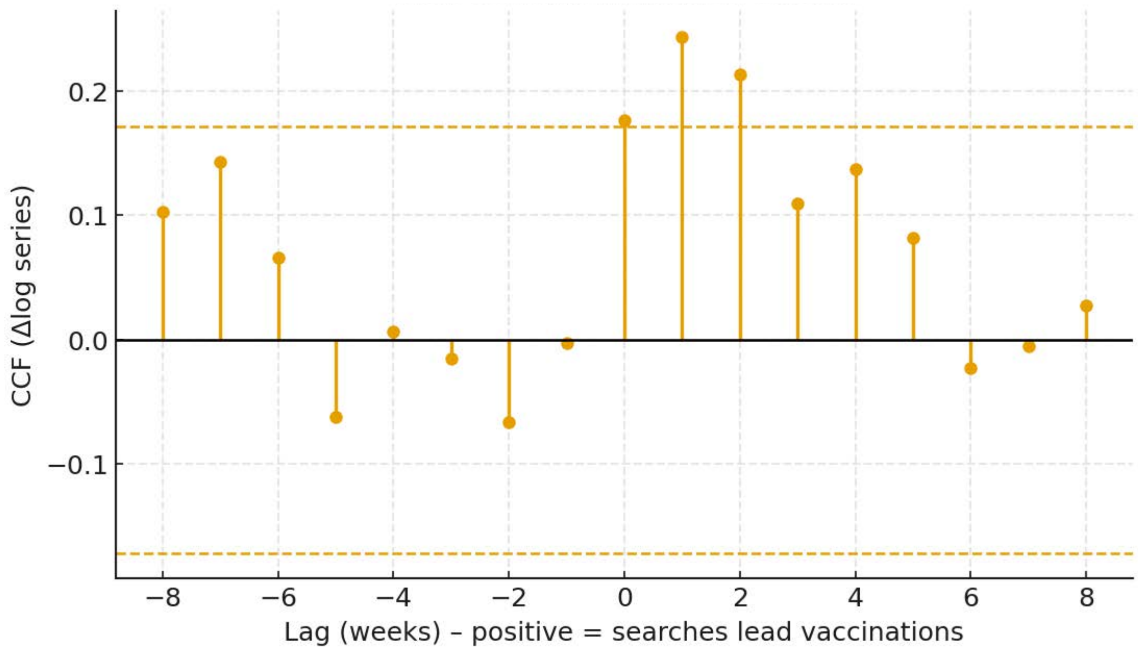
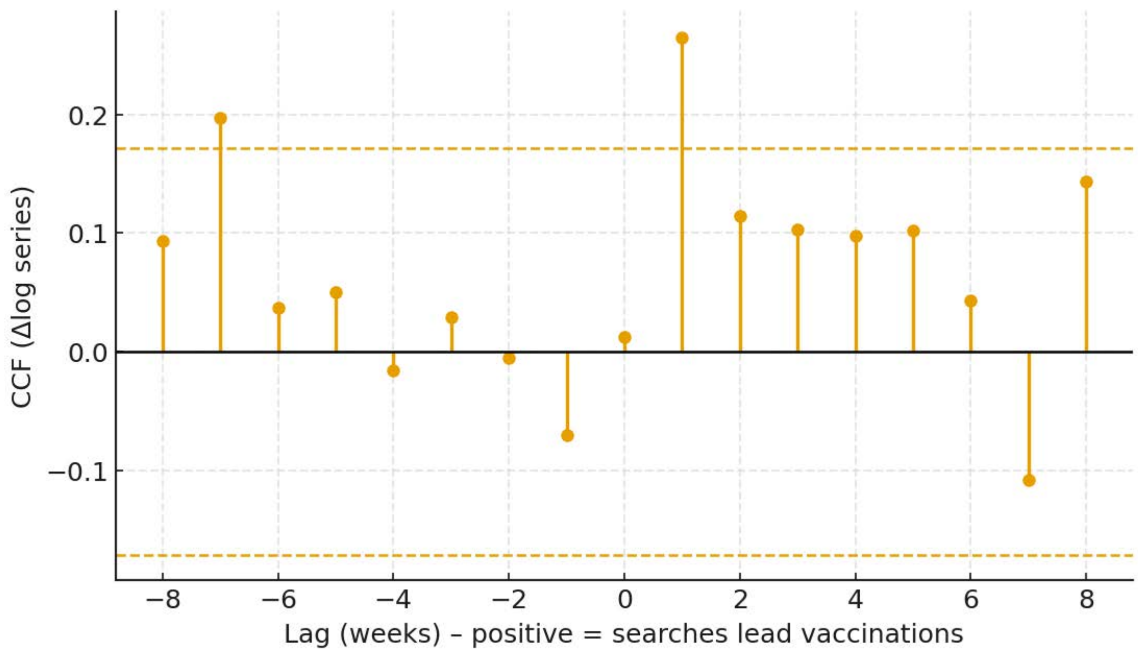


Figure 38

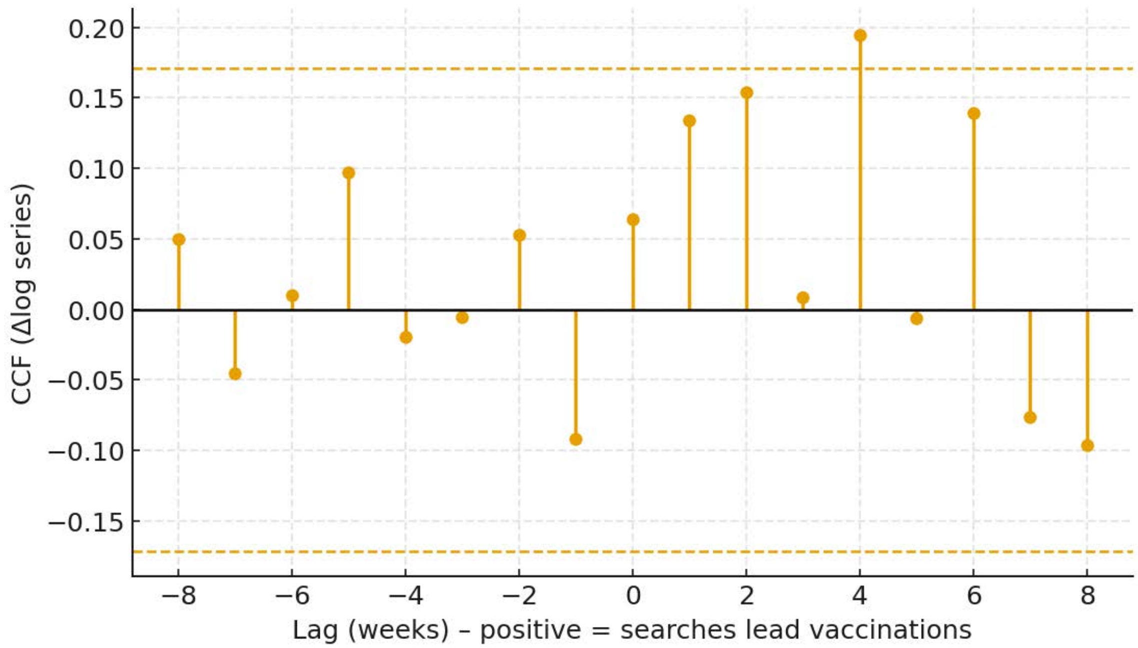
Cross-Correlation: AstraZeneca Vaccine



Finally, Booster searches showed a longer lead time, peaking at lag +4 (CCF = 0.195). Figure 39 shows that the orange stem at lag +4 crosses the 95% limit, while the adjacent lags fall closer to or within the confidence band. This indicates a delayed alignment between seeking information about boosters and the observed weekly doses.

Figure 39

Cross-Correlation: Booster Shot



Together, the patterns in Figures 35 to 39 mirror the summary in Table 18: synchronous or short (+1 week) leads for the composite and brand terms and a longer (+4 weeks) lead for booster-related queries.

Results Interpretation. Together, the CCFs suggest that search behavior is temporally aligned with vaccination uptake; however, the timing differs by topic. As summarized in Table 18, the combined search terms peak at lag 0 (CCF = 0.258). Figure 35 illustrates this with the tallest stem at zero, which crosses the Bartlett limits. This synchronous peak is consistent with shared weekly drivers, such as policy announcements, media coverage, or campaign pushes, that simultaneously elevate information seeking and vaccinations. In contrast, the brand-specific terms in Table 18 peak at lag +1 (Pfizer: 0.246, Moderna: 0.243, and AstraZeneca: 0.265). The corresponding peaks at lag +1 in Figures 36 to 38 exceed the 95% confidence intervals, suggesting a short lead time in which heightened brand interest is followed by increased vaccination in the next week. This could possibly reflect scheduling appointments and logistics between searching and receiving a vaccination.

The Booster pattern implies a longer decision-making period. Table 18 places the primary peak at lag +4 (CCF = 0.195), and Figure 39 shows that the +4 stem is the first to exceed the confidence band. This delay is consistent with the rollout of booster campaigns: eligibility expansions, seasonal messaging, and employer or industry requirements may prompt interest, but follow-through can span several weeks while people verify eligibility, locate clinics, and coordinate timing with prior doses. Because the Δlog transformation was used for all CCFs, these relationships reflect correlations in weekly growth rates rather than levels. This reduces false correlations caused by shared trends and supports the idea that changes in search intensity precede or coincide with changes in vaccinations at the reported lags.

Although CCFs do not establish causality, the patterns in Figures 35 to 39 and Table 18 are consistent. A synchronous composite signal likely driven by broad, system-wide factors is present, as well as a +1-week lead for brand terms that match near-term scheduling behavior and a +4-week lead for boosters that reflect slower conversion. These findings justify specifying a lag of +1 as the primary analytical lag for brand and composite series in the forthcoming regression models. However, lag 0 (composite) and lag +4 (booster) should be retained in sensitivity checks to evaluate the robustness of these lead–lag choices.

Primary Lag Decision and Sensitivity. Guided by the CCFs, the primary analytical lag is set to one week for the subsequent regression analyses. This choice aligns with the strongest or nearly strongest cross-correlations for the Pfizer, Moderna, and AstraZeneca series, all of which peak at lag +1 (Table 18 and Figures 36 to 38). It also maintains consistency across predictors, allowing for a clear comparison of effect sizes. For the Combined Search Terms index, the highest CCF occurs at lag 0 (Table 18 and Figure 35). However, since the lag +1 coefficient is relatively small and provides a true leading specification for forecasting, the primary regression used lag +1 for the combined index. Lag 0 was formally evaluated as a sensitivity model. For booster searches, the peak at lag +4 (Table 18 and Figure 39) indicates a longer decision horizon. Nevertheless, to preserve comparability across predictors, the primary model uses lag +1, and lag +4 is included as a pre-specified sensitivity analysis.

Sensitivity analyses probed the robustness of these choices in three ways. First, each predictor was re-estimated at alternative nearby lags, specifically lag 0 and lag +2, to assess the stability of coefficients around the selected lead. The empirical context for these alternatives can be found in Table 18. Second, for Combined and Booster, models at lag 0 (Combined) and lag +4 (Booster) were additionally fitted to reflect their CCF peaks (see Figures 35 and 39). Third, the primary models were repeated in Δ-level form (first differences without logs) to check that findings are not an artifact of the log scale. In all cases, inference used HAC (Newey–West) standard errors and identical sample windows. Effective N varies minimally by lag due to differencing and shifting. This plan collectively ensures that the primary lag +1 specification is theoretically motivated as a leading indicator and empirically verified against the peak lags observed in Table 18 and Figures 35 to 39.

RQ2 Time-Lagged Regression (Δlog)

This section quantifies the relationship between search activity and vaccination uptake by using bivariate, time-lagged regressions on growth rate series. For each predictor, the dependent variable is the log-transformed weekly vaccination rate at time point t, and the key regressor is the log-transformed search rate at time point t-1 (the primary lag identified in the CCFs). Log-first-differences (log1p) make the coefficients interpretable as elasticities, like for example the percentage change in vaccinations associated with a 1% change in search interest from the previous week, while mitigating non-normality, trend, and scale effects. The models were estimated via OLS with HAC (Newey–West) standard errors to accommodate any short-lag autocorrelation or mild heteroscedasticity remaining after the transformation. Sensitivity analyses repeated the specification at lag 0 and lag +2 and in Δ-level form to assess the consistency of the primary findings.

Model Specification. Weekly vaccination counts were analyzed using log first differences to express the proportional week-to-week change and meet the stationarity conditions established during the assumption checks. The outcome was defined as $\Delta\log(\text{vaccinations}_t) = \log(1 + \text{vaccinations}_t) - \log(1 + \text{vaccinations}_{t-1})$. For each Google Trends series, the predictor was the one-week lagged growth rate,

$\log(\text{search}_{(t-1)})$, reflecting the one-week lead identified in the CCFs. The bivariate model for each pre-specified predictor was: $\Delta\log(\text{vaccinations})_t = \beta_0 + \beta_1 \cdot \Delta\log(\text{search})_{t-1} + \varepsilon_t$. β_1 is interpreted as an elasticity, or the percent change in vaccinations associated with a 1% change in the search series in the prior week.

All models were estimated using OLS with heteroscedasticity- and autocorrelation-consistent (Newey–West) standard errors and a four-week bandwidth. After differencing and aligning the one-week lag, the effective sample size for these regressions was 130 weeks, and any remaining missing values were handled by listwise deletion within the model. Two-tailed tests at the 5% level were used, and β_1, HAC SE, 95% CI, t, p, and R^2 were reported for each specification.

Bivariate Regression Results (Lag +1). As summarized in Table 19, the Δlog models with a one-week lag produced positive elasticities for all predictors, with an effective N of 130 weeks. The three brand terms were statistically significant at the .05 level: AstraZeneca (β_1 = 0.239, 95% confidence interval [CI] [0.129, 0.350], $p < .001$), Moderna (β_1 = 0.202, 95% CI [0.022, 0.382], $p = .028$), and Pfizer (β_1 = 0.155, 95% CI [0.055, 0.255], $p = .0026$). The Combined Search Terms index (β_1 = 0.128, 95% CI [−0.081, 0.338], $p = .226$) and booster searches (β_1 = 0.108, 95% CI [−0.032, 0.248], $p = .128$) were positive, however, they did not reach significance at this lag. Coefficients were estimated using HAC (Newey–West, 4-week) standard errors to account for residual autocorrelation and heteroscedasticity.

Table 19

Bivariate Δlog regressions (Lag +1)

Predictor	N	β_1 (elasticity)	SE_HAC	t (HAC)	p (HAC)	95% CI (β_1)	R^2
Combined Search Terms	130	0.128	0.106	1.22	0.2263	[−0.081, 0.338]	0.024
Pfizer vaccine	130	0.155	0.051	3.07	0.0026	[0.055, 0.255]	0.061
Moderna vaccine	130	0.202	0.091	2.22	0.0281	[0.022, 0.382]	0.059
AstraZeneca vaccine	130	0.239	0.056	4.29	0.0000	[0.129, 0.350]	0.070
Booster shot	130	0.108	0.071	1.53	0.1279	[−0.032, 0.248]	0.018

The elasticities imply that a 10% week-over-week increase in search volume at t-1 is associated with a corresponding increase in vaccination growth at t of approximately 2.4% for AstraZeneca ($\beta_1 \approx 0.239$), 2.0% for Moderna ($\beta_1 \approx 0.202$), and 1.6% for Pfizer ($\beta_1 \approx 0.155$). In contrast, the combined and booster effects at lag +1 are smaller and statistically insignificant within their 95% confidence intervals. Model R^2 values are modest (approximately .02–.07) throughout, as expected for weekly growth models, where multiple operational and policy factors influence uptake beyond search behavior. Nevertheless, the significant brand coefficients indicate a consistent, directional association between increases in brand-specific information seeking and next-week vaccination growth.

Overall, the pattern aligns with the established lead–lag structure. The combined index peaked at lag 0 in the CCFs. Therefore, its non-significance at lag +1 is consistent with a more synchronous signal. On the other hand, the booster series peaked at lag +4. Thus, a lag +1 specification is intentionally conservative and was revisited in sensitivity analyses.

Results Interpretation. The lagged Δlog regressions show that brand-specific search growth consistently predicts next-week vaccination growth. Table 19 shows that the elasticities for AstraZeneca, Moderna, and Pfizer are all statistically significant with HAC-corrected inference. This implies that a 10% increase in brand-term search volume at week t-1 is associated with roughly 2.4%, 2.0%, and 1.6% higher vaccination growth at week t, respectively. These effects are moderate yet policy-relevant; sustained increases in brand interest, whether driven by eligibility changes, media coverage, or local campaigns, are followed by measurable increases in vaccine uptake the following week. This pattern aligns with the CCF results, which showed primary peaks at +1 week for the brand series.

In contrast, the combined search terms coefficient is positive but not significant at lag 1 ($\beta_1 = 0.128$, $p = .226$). This aligns with the lag 0 CCF peak and suggests that, in this context, the composite index functions more as a synchronous demand signal than a leading one. Similarly, the booster coefficient is positive but not significant at lag +1 ($\beta_1 = 0.108$, $p = .128$), which is consistent with the four-week CCF lead reflecting a longer decision horizon for boosters, due to factors like eligibility checks, scheduling or illness. In summary, the lack of significance for the combined and booster coefficients at lag 1 is expected, given their empirically identified lead times.

Model R^2 values are modest (approximately .02–.07), which is typical for weekly growth-rate models where operational logistics, policy changes, and seasonality also matter. Importantly, using Δlog transforms targets growth-on-growth associations. This reduces spurious correlations from shared levels or trends, yielding interpretable and comparable elasticities across predictors. Assumption checks and HAC (Newey–West) standard errors address residual autocorrelation and heteroscedasticity. Prior influence diagnostics also indicate that no single week significantly influences the results. Overall, the evidence supports a directional, leading relationship from brand search activity to near-term changes in vaccination uptake, with timing and significance matching the lead–lag structure established earlier.

Sensitivity Analyses (lags 0, +2; Δ-level). To assess the robustness of the primary +1 week Δlog findings, the models were re-estimated at lags 0 and 2 for all predictors. Additionally, the Booster model was fit at lag 4 which was its CCF peak. The regressions also were repeated using Δ-level outcomes and predictors (first differences without logs) at lag +1. All models used HAC (Newey–West, 4-week) standard errors and retained the same weekly window. There were small differences in N due to lag alignment. Table 20 shows the elasticities (β_1), p-values, and R^2 for lags 0 and +2, as well as for the Booster at lag +4. Generally, the results align with the CCF timing. The Combined Search Terms strengthens at lag 0 ($\beta_1 = 0.234$, $p < .001$, $R^2 = .066$), but it is small and non-significant at lag +2 ($\beta_1 = 0.062$, $p = .383$), which is consistent with its synchronous CCF peak. Moderna remains positive and becomes significant at lag +2 ($\beta_1 = 0.157$, $p = .009$, $R^2 = .045$), which echoes a broader positive band around lag +1 in the CCF. Pfizer and AstraZeneca are near zero at lag 0 and non-significant at lag 2, which reinforces lag 1 as the most informative lead for those brands. The booster analysis shows its strongest and statistically significant effect at lag +4 ($\beta_1 = 0.130$, $p = .003$, $R^2 = .038$), matching the +4-week CCF peak. At lag 0 and +2, the coefficients are smaller and non-significant.

Table 20

Sensitivity: Δlog models at Lag 0 and +2 (and Booster +4), HAC SEs

Predictor	Lag	β_1	p	R^2	N
Combined Search Terms	0	0.234	< .001	0.066	131
Combined Search Terms	+2	0.062	0.383	0.007	129
Pfizer vaccine	0	0.041	0.318	0.003	131
Pfizer vaccine	+2	0.025	0.494	0.002	129
Moderna vaccine	0	0.161	0.091	0.031	131
Moderna vaccine	+2	0.157	0.009	0.045	129
AstraZeneca vaccine	0	0.012	0.830	0.000	131
AstraZeneca vaccine	+2	0.092	0.325	0.013	129
Booster shot	0	0.057	0.300	0.004	131
Booster shot	+2	0.110	0.119	0.024	129
Booster shot	+4	0.130	0.003	0.038	127

Using first differences (no log transformation) at lag +1, all predictors show positive coefficients, with several reaching statistical significance as shown in Table 21. R^2 values are higher than those in the log-transformed model because the latter tracks proportional growth, while the former tracks absolute week-to-week changes. However, the direction of the effects is consistent with the primary log-transformed results.

Table 21

Sensitivity: Δ-level models at Lag +1, HAC SEs

Predictor	Lag	β_1	p	R^2	N
Combined Search Terms	+1	5,386.690	< .001	0.251	130
Pfizer vaccine	+1	4,691.670	< .001	0.311	130
Moderna vaccine	+1	3,242.050	0.032	0.151	130
AstraZeneca vaccine	+1	1,940.720	0.031	0.045	130
Booster shot	+1	3,613.680	< .001	0.169	130

The timing conclusions remain unchanged across specifications, with brand terms are most informative at +1 week, the combined index is strongest in real time (lag 0), and booster effects materialize over a longer horizon (+4 weeks). The Δ-level results confirm the direction of the associations independently of the log scale. Together, these results support the robustness of the primary findings and the lead-lag structure established earlier.

Hypothesis Decision and Practical Implications. The regression analysis rejects the null hypothesis for RQ2. Specifically, $H0_2$, the absence of an association between weekly Google Trends search activity and weekly vaccination uptake, is rejected because the lag +1 log-logistic models yield statistically significant positive elasticities for the brand terms: AstraZeneca ($\beta_1 = 0.239$, 95% CI [0.129, 0.350]), Moderna ($\beta_1 = 0.202$, 95% CI [0.022, 0.382]), and Pfizer ($\beta_1 = 0.155$, 95% CI [0.055, 0.255]). The combined index ($\beta_1 = 0.128$, $p = .226$) and booster searches ($\beta_1 = 0.108$, $p = .128$) were positive but not significant at lag 1. This pattern aligns with their lead times from the CCFs: synchronous at lag 0 for the combined index and +4 weeks for booster searches. Consistency checks reinforce this interpretation, with the combined index becoming significant at lag 0 and booster becoming significant at lag +4, while brand terms remain most informative at lag +1.

Overall, these elasticities imply that growth in brand-specific search interest one week prior anticipates vaccination growth the following week. A 10% increase in searches at t–1 is associated with approximately 2.4%, 2.0%, and 1.6% higher vaccination growth for AstraZeneca, Moderna, and Pfizer, respectively, at t. Although the model R^2 values are modest, as expected for weekly growth models influenced by multiple operational and policy factors, the direction and timing are stable across specifications. This makes brand-level search activity a feasible operational lead indicator for short-term planning.

To implement this, public health teams could incorporate real-time search monitoring into vaccination dashboards. They could use brand-term growth as a one-week signal to adjust appointment capacity, outreach, and inventory logistics. The combined series is strongest immediately and can serve as a real-time estimate of current demand. However, booster queries require a longer planning window of about a month for staffing and communications. All interpretations are associative, not causal, and should be considered alongside policy changes, eligibility shifts, and supply conditions. Nevertheless, the patterns support the practical utility of Google Trends as a timely, low-cost supplement to administrative vaccination data.

RQ 2 Summary

To determine demand for vaccinations in Austria, Google Trends data were evaluated over a period of 132 weeks from December 31, 2020, to April 16, 2023. Descriptive statistics revealed right-skewed, long-tailed distributions for weekly vaccinations and all pre-specified search series. Assumption checks on the raw series revealed non-normality, serial dependence, and mild heteroscedasticity. Consequently, analyses proceeded with log first differences (Δlog) and HAC (Newey–West) inference to address short-lag autocorrelation and variance instability.

At the bivariate level, Spearman's correlations were positive and statistically significant for all predictors. The strongest associations were observed for the Pfizer and Combined Search Terms predictors, followed by the Moderna and AstraZeneca predictors. A smaller, yet still significant, coefficient was observed for the Booster predictor. Cross-correlation functions (CCFs) using the log-transformed series clarified the timing. The combined index peaked immediately, the brand terms peaked at one week, and the booster term peaked at four weeks.

Time-lagged Δlog regressions at the primary lag (+1) showed positive elasticities for all predictors. The brand-specific terms were statistically significant for AstraZeneca ($\beta_1 = 0.239$), Moderna ($\beta_1 = 0.202$), and Pfizer ($\beta_1 = 0.155$). Combined and Booster were positive, but not significant, at +1. Sensitivity analyses aligned with the CCFs. Combined search terms became significant at lag 0, Booster became significant at lag +4, and Moderna remained significant at lag +2. Overall, R^2 values were modest across specifications. Δ-level checks revealed similar results. Outlier and influence diagnostics did not identify any weeks exerting significant influence, and residual behavior was consistent with the chosen HAC correction.

Therefore, the null hypothesis for RQ2 is rejected. Weekly Google search activity, especially brand-specific queries, is prospectively associated with a week's worth of vaccination growth prediction. In practice, brand-term growth can be treated as a one-week leading indicator for short-term operational planning, like appointments, outreach, and inventory. The combined index can serve as a real-time estimate, and Booster queries as a longer-term (~4-week) signal for campaign timing. These associations are observational rather than causal and should be interpreted in the context of policy and supply. However, the pattern supports the utility of Google Trends as a timely, low-cost complement to administrative vaccination data.

Research Question 3: Relationship between weekly COVID-19 cases and weekly vaccination rates in Austria

This section evaluates the relationship between epidemic intensity and program output in the short term. After descriptive statistics and assumption checks were performed, same-week Spearman's ρ and Δlog cross-correlations were used for a directional screening. Then, two time-lagged regression specifications were estimated: Model A (cases → vaccinations) and Model B (vaccinations → cases). The findings showed no contemporaneous association and generally flat cross-correlations. Model A showed small, non-significant elasticities at lags 0–2. Model B was null at lag 0. A small exploratory signal at approximately seven weeks was not stable across neighboring lags or robustness checks. It is interpreted as noncausal co-movement rather than a directional effect. Sensitivity analyses did not change the conclusion that short-term directional relationships were not supported.

RQ3 Descriptives

The analytic window spans 132 consecutive ISO weeks, from December 27, 2020, to June 30, 2023. As shown in Table 22, on average, there were 44,251 weekly cases (SD = 63,054; range 538–322,641). The distribution is markedly right-skewed and leptokurtic (skew = 2.63, kurtosis = 6.89). Most observations are concentrated at lower counts, and there is a thin tail of surge weeks. This pattern is visible in the cases histogram shown in Figure 40 and the time-series trace shown in Figure 41. The trace shows distinct wave peaks separated by troughs. Weekly vaccinations averaged 155,013 (SD = 196,962; range 448–795,627). The vaccination distribution is also right-skewed (skew = 1.61), with moderate kurtosis (kurtosis = 1.67). This aligns with the histogram pattern showed in Figure 42 and the time series plot shown in Figure 43, in which episodic campaign peaks are followed by quieter intervals.

Table 22

Descriptive statistics for RQ3 variables (overlapping 132 ISO weeks)

Variable	N	Mean	Std Dev	Min	Max	Skewness	Kurtosis
Weekly Cases	132	44,251.10	63,054.10	538	322,641	2.63	6.89
Weekly Vaccinations	132	155,013.00	196,962.00	448	795,627	1.61	1.67

Figure 40

Histogram of Weekly COVID-19 Cases (Overlap Window)

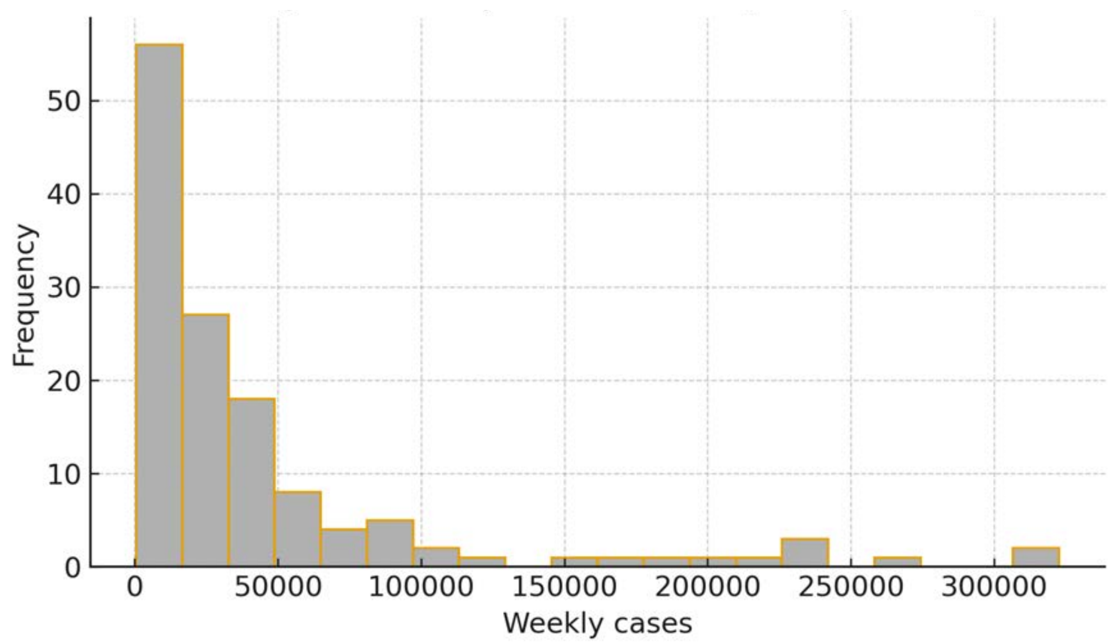


Figure 41

Weekly COVID-19 Cases Over Time (Overlap Window)

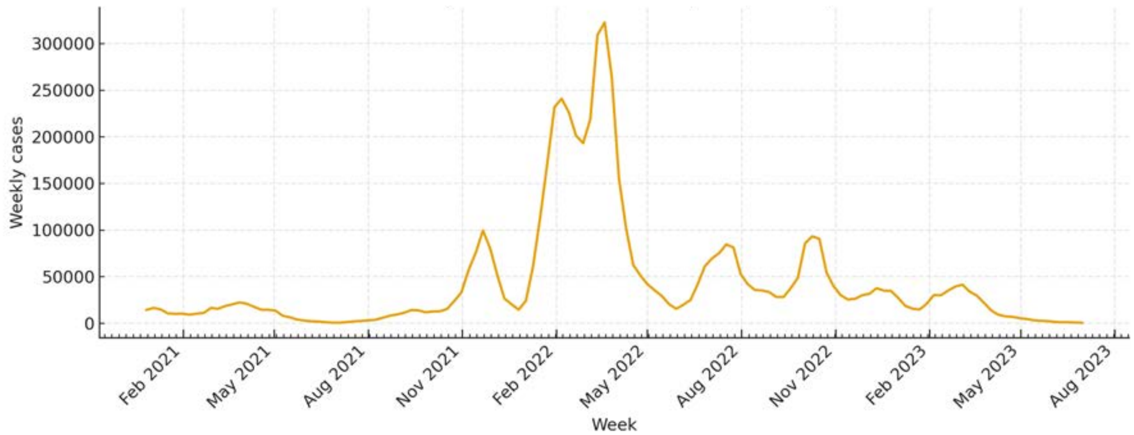


Figure 42

Histogram of Weekly COVID-19 Vaccinations (Overlap Window)

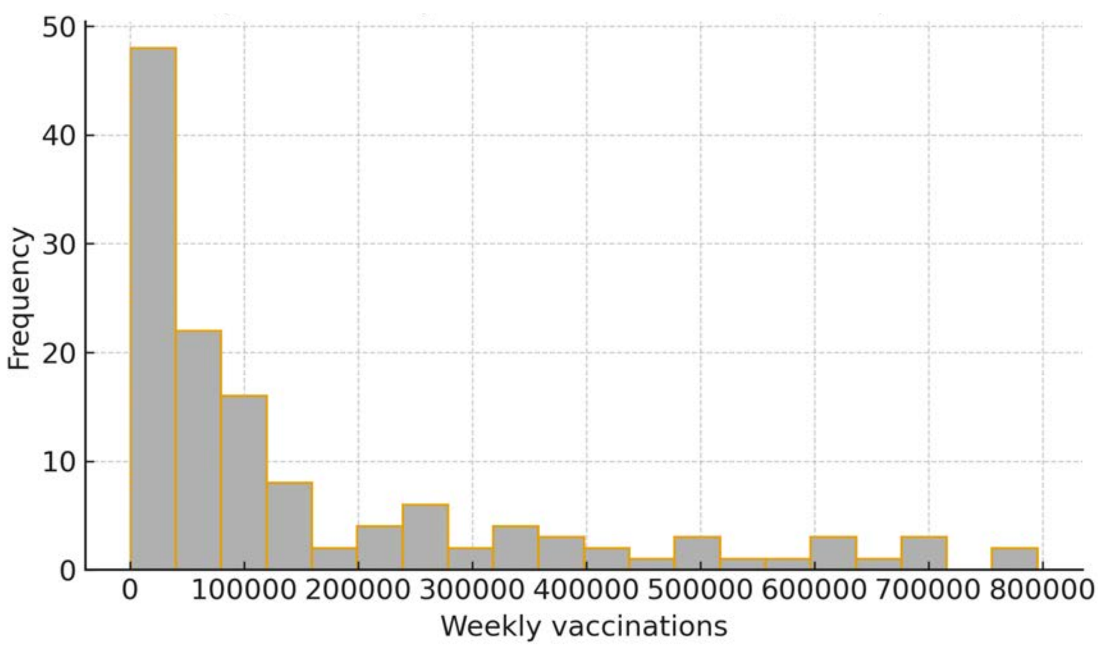
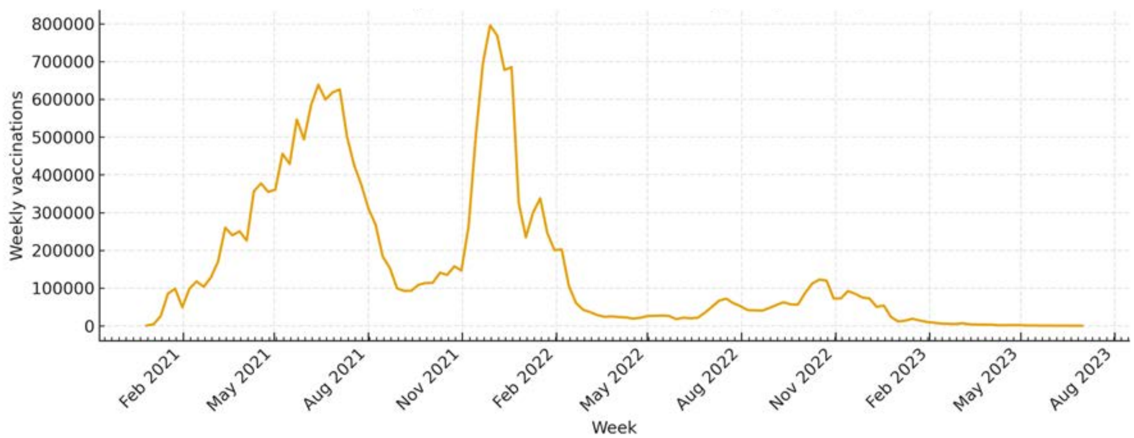


Figure 43

Weekly COVID-19 Vaccinations Over Time (Overlap Window)



Together, these patterns suggest long-tailed, non-normal distributions with significant week-to-week variability. This motivates using rank-based summaries for contemporaneous association and Δlog transformations with HAC-robust inference for subsequent time-lagged models.

RQ3 Assumption Checks

In this section, the data were evaluated to check whether they fulfilled the statistical criteria for valid inference in the bidirectional analysis of weekly cases of COVID-19 and weekly vaccinations over the 132-week period (Dec 27, 2020–Jun 30, 2023). First, the diagnostic checks were conducted on the level series. Then, they were conducted on their Δlog (week-over-week growth) transformations. The diagnostics used were normality via Q–Q plots and distributional summaries, independence/autocorrelation using ACF/PACF inspection and Ljung–Box tests, homoscedasticity through Breusch–Pagan and White tests, supported by scale–location plots, and outliers/influential points via Cook's distance screening. These checks informed the modeling choices made later (Δlog specification, HAC/Newey–West standard errors, and influence sensitivity).

Normality. The normality of both outcomes was evaluated in the level series and in their Δlog transforms (week-over-week growth). Visual inspection of the Q–Q plots (Figures 44 and 45) shows significant departures from the normal line on the right, which is consistent with the heavy skew and kurtosis reported earlier in the descriptive analysis. Shapiro–Wilk tests rejected the null hypothesis of normality for both level series (cases: $W = 0.639$, $p < .001$; vaccinations: $W = 0.759$, $p < .001$) as shown in Table 23.

Figure 44

Q–Q plot: Weekly Cases (Level)

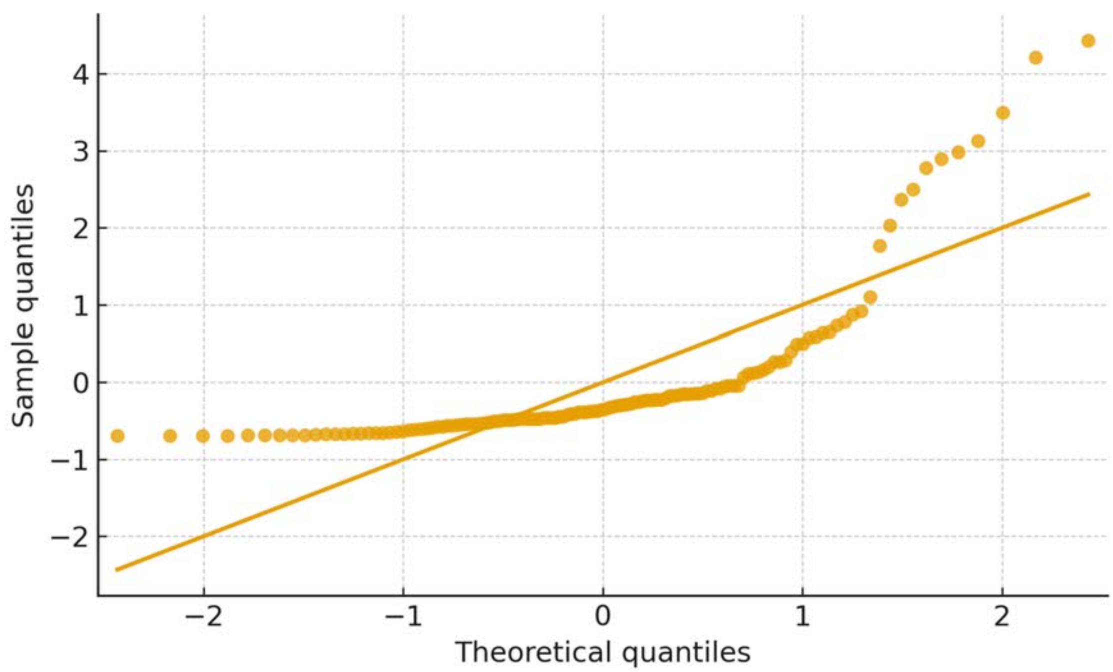


Figure 45

Q–Q plot: Weekly Vaccinations (Level)

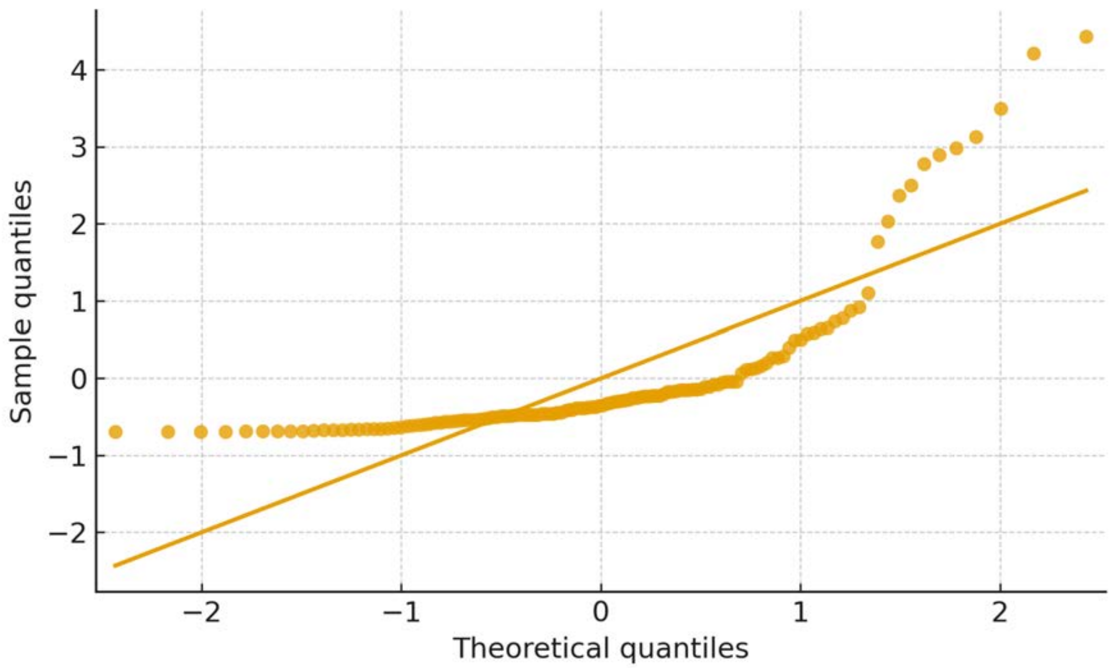


Table 23

Normality diagnostics (levels and Δlog; Shapiro–Wilk, skewness, kurtosis)

Variable	Transform	N	Shapiro W	p	Skewness	Kurtosis
Weekly cases	Level	132	0.639	< .001	2.63	6.89
Weekly vaccinations	Level	132	0.759	< .001	1.61	1.67
Weekly cases	Δlog	131	0.987	0.250	0.42	0.13
Weekly vaccinations	Δlog	131	0.869	< .001	1.65	7.29

After the log-transformation, the weekly case data is much closer to normal distribution. The Q-Q plot closely aligns with the reference line as seen in Figure 46, the skewness and kurtosis are near zero (skew = 0.42, kurtosis = 0.13), and the Shapiro-Wilk test is non-significant (W = 0.987, p = .250) shown in Table 23. However, weekly vaccinations (Δlog) still exhibit heavy tails and positive skew in the upper quantiles shown in Figure 47, and the Shapiro–Wilk test remains significant (W = 0.869, p < .001).

Figure 46

Q–Q plot: Weekly Vaccinations (Δlog)

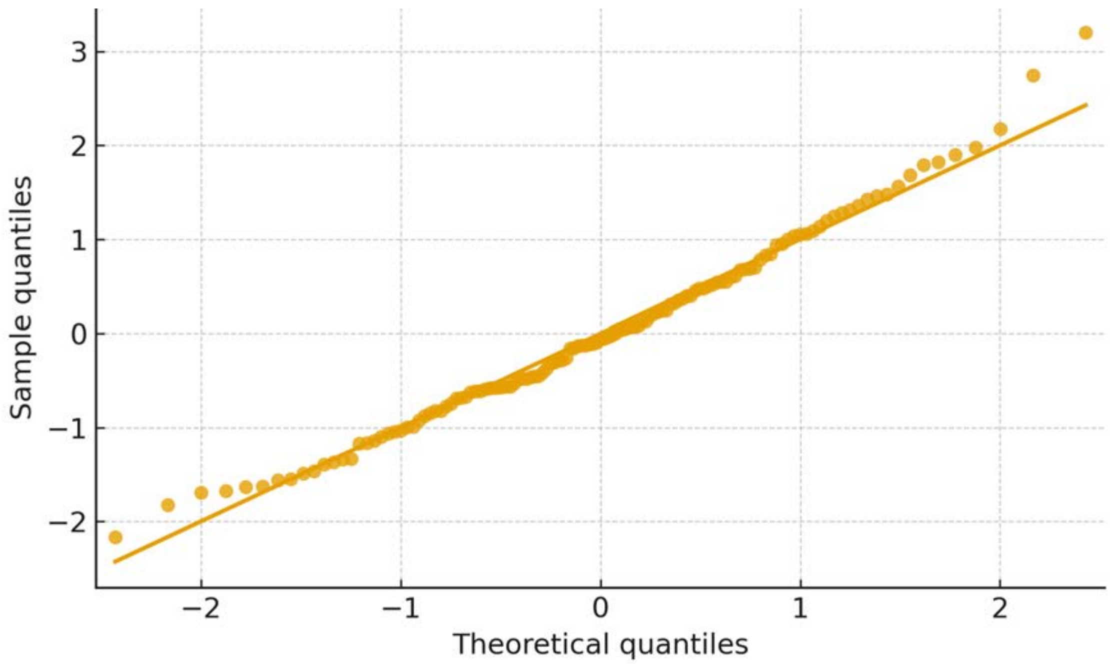
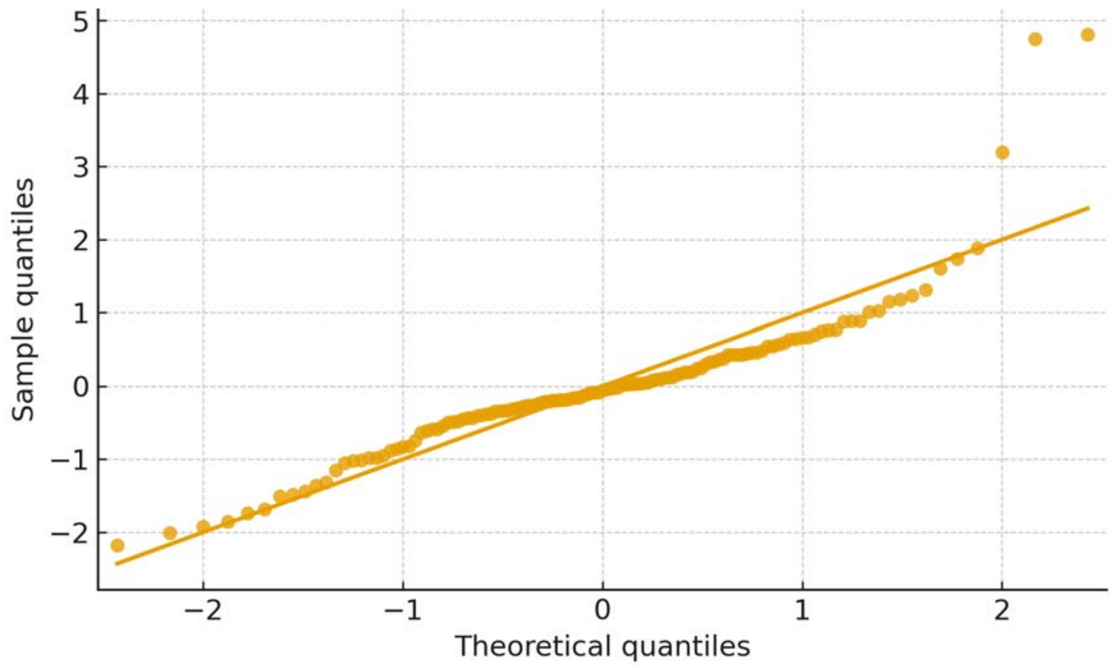


Figure 47

Q–Q plot: Weekly Vaccinations (Δlog)



The Δlog specification substantially mitigates non-normality for cases and partially mitigates it for vaccinations. Subsequent analyses use HAC-robust inference and do not rely on the normality of the raw series. Therefore, these results support the use of Δlog models, while noting the residual non-normality in vaccination growth, which was accommodated by robust standard errors.

Independence/Autocorrelation. Autocorrelation was examined for both the level and Δlog series. This was done using sample ACFs (24 lags). The Ljung–Box test was also used at lags 4, 8, 12, and 24. For the level series, the ACFs for weekly cases and vaccinations decrease gradually from high initial values and remain positive for several weeks as seen in Figures 48 and 49. This pattern indicates strong persistence and violates the independence assumption for OLS in levels. After applying the log difference transformation to the ACFs, the effects are significantly reduced. The cases (Δlog) fluctuate around zero with only a modest correlation at the earliest lags as shown in Figure 50, and the vaccinations (Δlog) display a small, short-memory pattern, with most lags falling within the 95% confidence interval as shown in Figure 51.

Figure 48

ACF: Weekly Cases (Level)

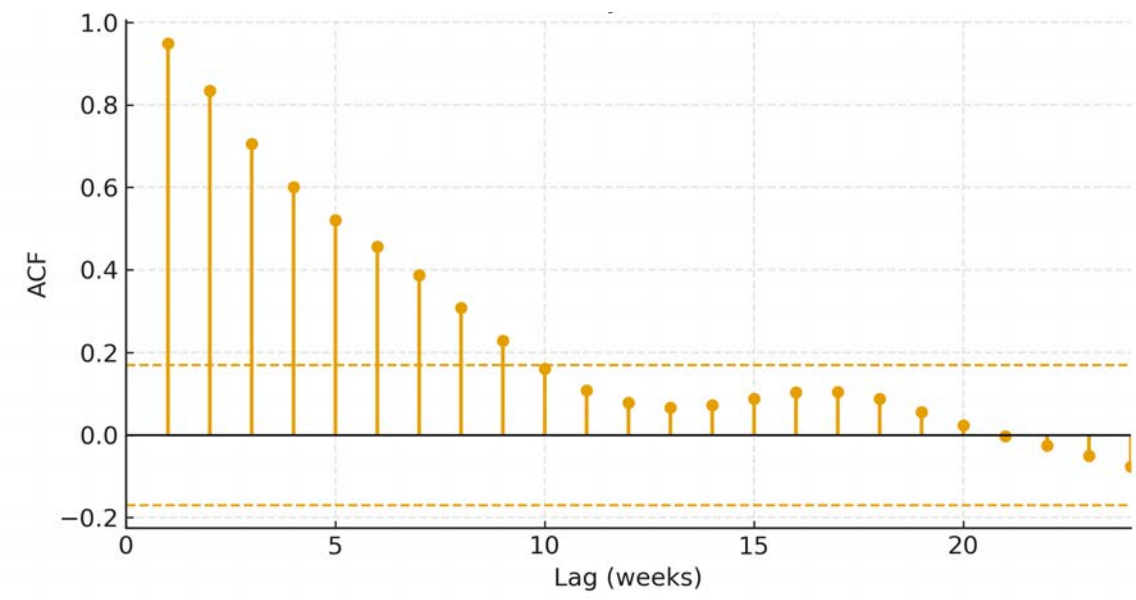


Figure 49

ACF: Weekly Vaccinations (Level)

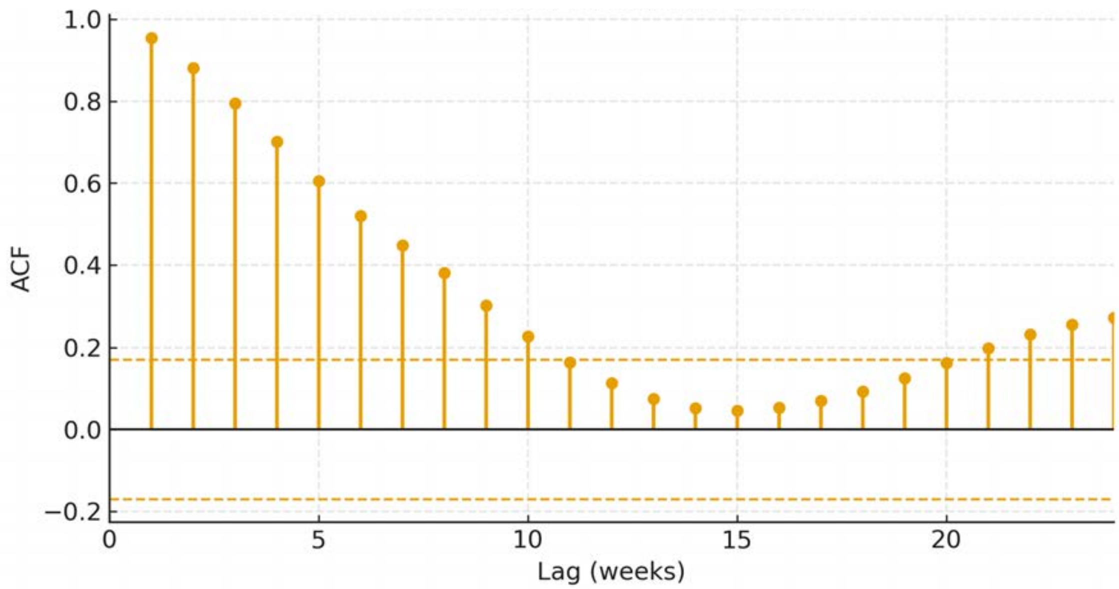


Figure 50

ACF: Weekly Cases (Δlog)

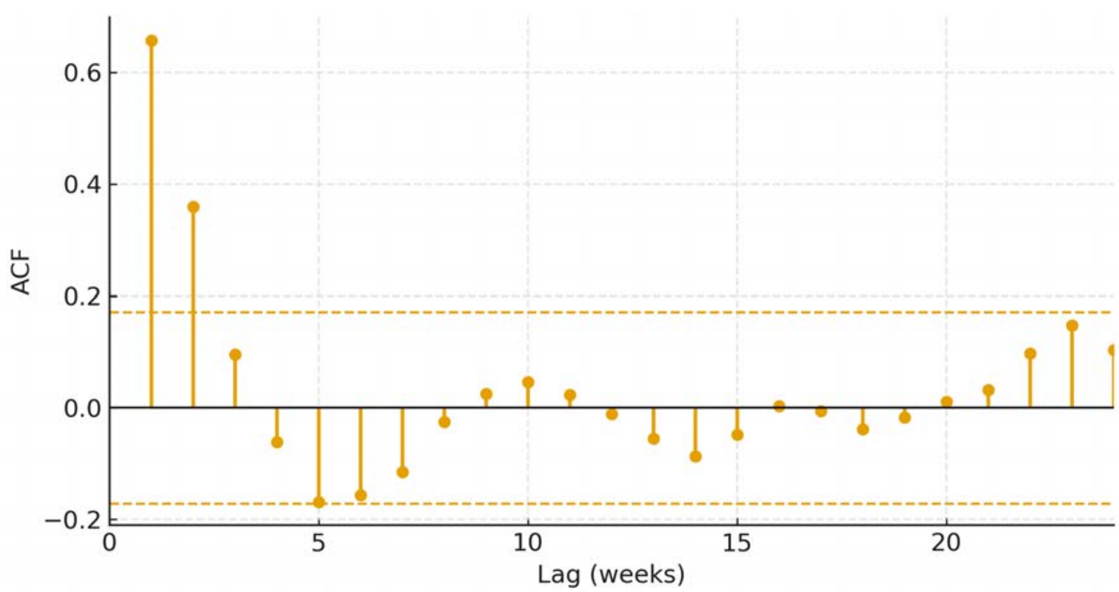
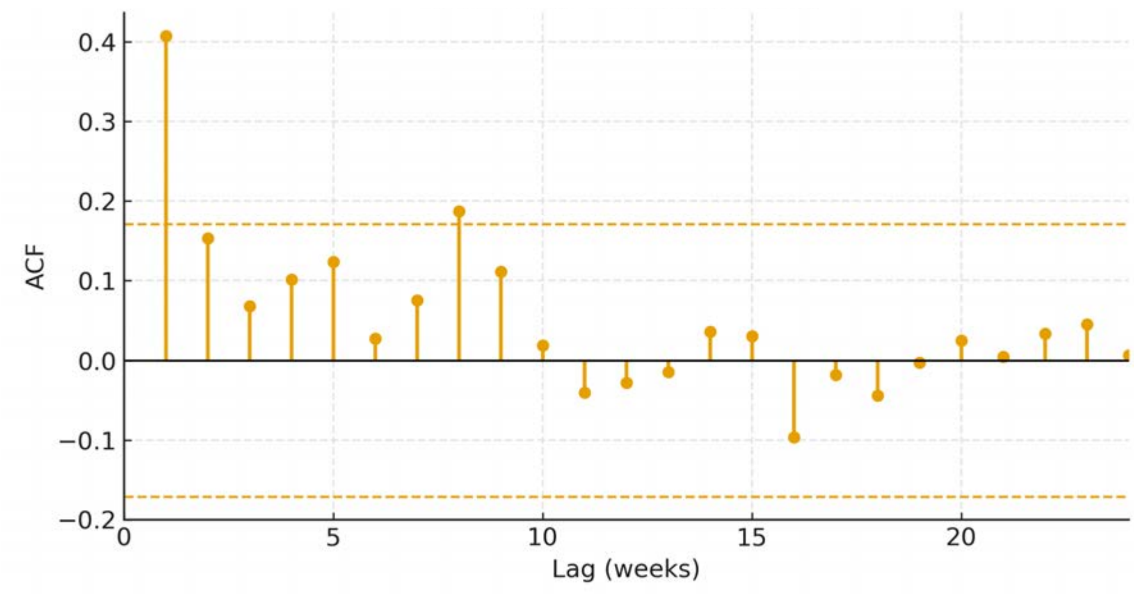


Figure 51

ACF: Weekly Vaccinations (Δlog)



Formal tests align with the graphs. Table 24 shows that the Ljung–Box null hypothesis of no autocorrelation is clearly rejected for both level series at all reported lag orders. However, the Δlog series yield smaller Q statistics, though they are still significant overall given the long sample. Taken together, these diagnostics justify modeling in Δlog to reduce serial dependence, as well as using HAC/Newey–West standard errors in all regressions to protect against remaining short-lag autocorrelation.

Table 24

Ljung–Box tests for autocorrelation at selected lags

Series	Lag	Q	p
Cases (level)	4	334.3	< .001
Cases (level)	8	436.0	< .001
Cases (level)	12	449.9	< .001
Cases (level)	24	459.1	< .001
Vaccinations (level)	4	382.9	< .001
Vaccinations (level)	8	521.0	< .001
Vaccinations (level)	12	547.5	< .001
Vaccinations (level)	24	595.9	< .001
Cases (Δlog)	4	77.0	< .001
Cases (Δlog)	8	86.4	< .001
Cases (Δlog)	12	86.9	< .001
Cases (Δlog)	24	96.0	< .001
Vaccinations (Δlog)	4	27.5	< .001
Vaccinations (Δlog)	8	35.5	< .001
Vaccinations (Δlog)	12	37.6	< .001
Vaccinations (Δlog)	24	40.4	0.019

Homoscedasticity. To assess variance constancy, a simple time-trend mean model ($y = \beta 0 + \beta 1 \cdot t$) was used for the Δlog series of weekly cases and weekly vaccinations. Visual diagnostics involved Scale-Location plots of the square root of the standardized residuals versus the fitted values. The plot of cases (Δlog) shows a largely flat LOWESS smooth with no fan shape as seen in Figure 52. In contrast, the plot of vaccinations (Δlog) shows a slight upward drift at higher fitted values.

Figure 52

Scale-Location: Weekly Cases (Δlog)

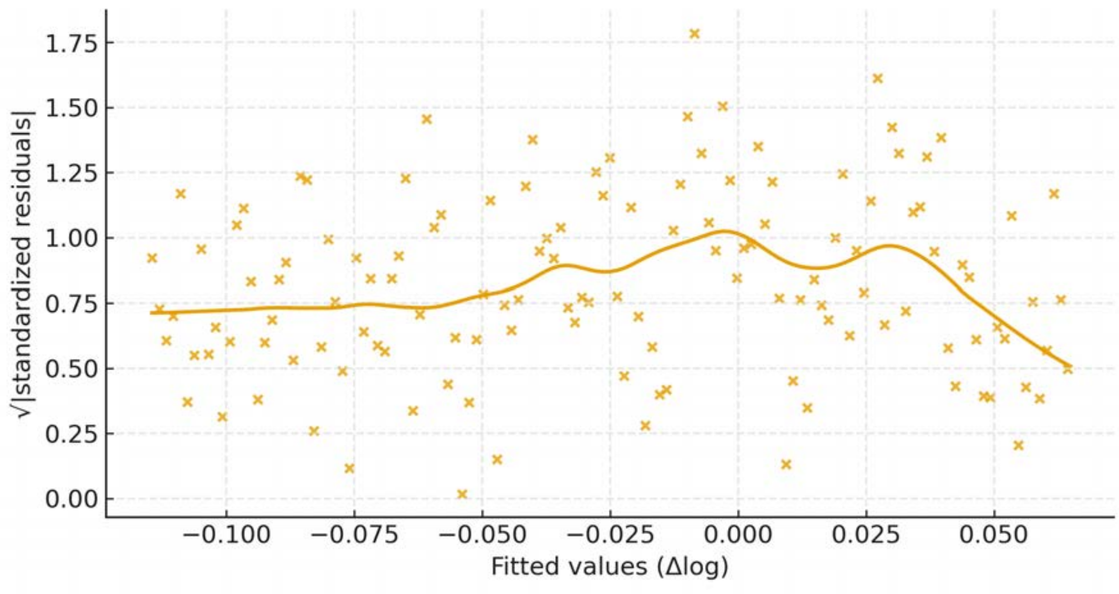
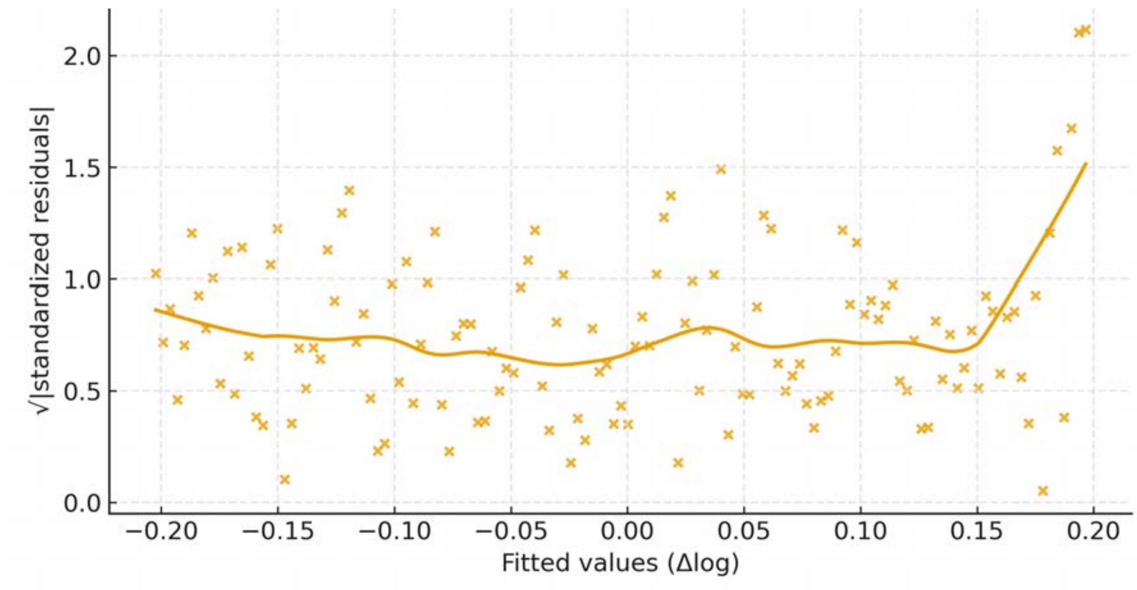


Figure 53

Scale-Location: Weekly Vaccinations (Δlog)



Formal tests confirm the results of the graphics. Table 25 shows that Breusch–Pagan does not reject homoscedasticity for cases (Δlog) ($N = 131$; $\chi^2 = 2.3$, $p = 0.129$), though White's test is positive ($\chi^2 = 8.1$, $p = 0.018$), suggesting weak, model-free departures. For vaccinations (Δlog), both Breusch–Pagan ($\chi^2 = 6.7$, $p = 0.010$) and White ($\chi^2 = 18.7$, $p < .001$) indicate heteroscedasticity, which aligns with the Scale-Location pattern.

Table 25

Homoscedasticity tests on Δlog series (time-trend mean model)

Series	N	BP χ^2	BP p	White χ^2	White p
Weekly cases	131	2.3	0.129	8.1	0.018
Weekly vaccinations	131	6.7	0.010	18.7	< .001

Given these results, all subsequent Δlog regressions are reported with heteroscedasticity- and autocorrelation-consistent (HAC/Newey–West) standard errors. Sensitivity checks using HC3 robust errors were performed to confirm that the statistical conclusions remain unchanged.

Outliers and Influential Points. The influence was screened using a simple time-trend OLS ($\Delta\log = \beta_0 + \beta_1 \cdot t$) on the Δlog series for each outcome. Three diagnostics were performed each week: Cook's distance, external studentized residuals, and leverage. Weeks were flagged if Cook's D was greater than 4/N or the absolute value of the studentized residual was greater than 3. In this case, N equals 131, so 4/N is approximately 0.031. Cook's distance stem plots with the dashed 4/N reference are shown in Figures 54 and 55. Flagged weeks are listed in Table 26 alongside their dates and statistics.

Figure 54

Cook's Distance: Cases (Δlog Trend Model)

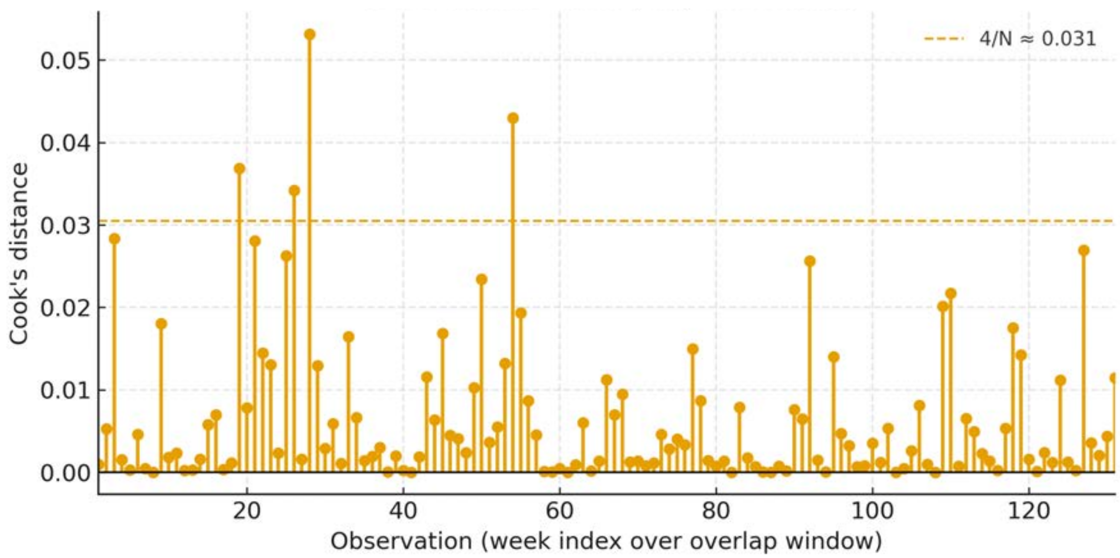


Figure 55

Cook's Distance: Vaccinations (Δlog Trend Model)

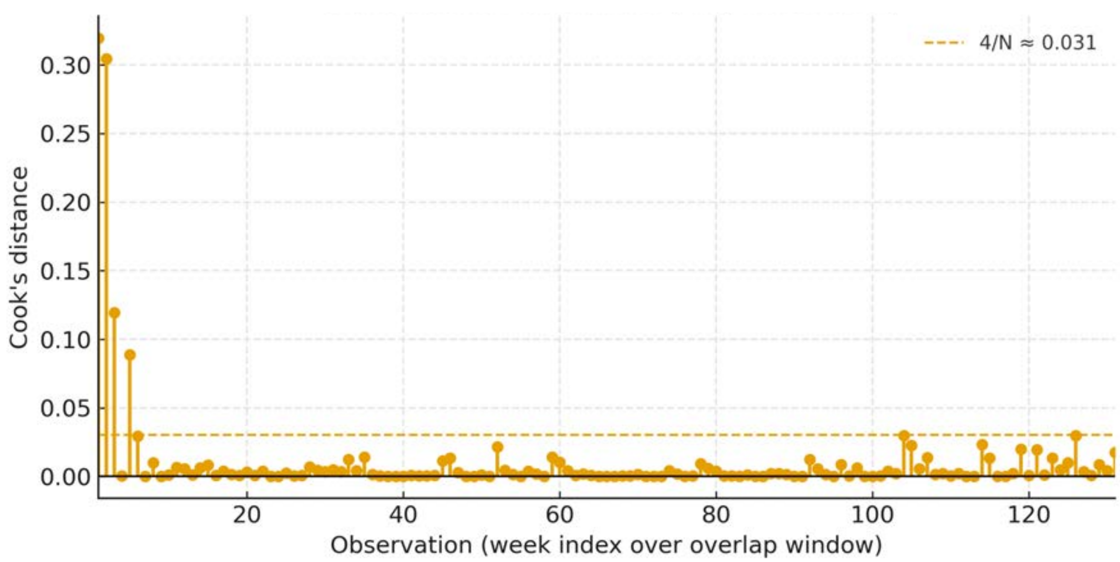


Table 26

Homoscedasticity tests on Δlog series (time-trend mean model)

	Series	Date	CookD	StudentizedResid	Leverage	Flag_CookD	Flag_StudResid
0	Cases (Δlog)	2021-05-09	0.037	-1.95	0.019	Yes	—
1	Cases (Δlog)	2021-06-27	0.034	-2.07	0.016	Yes	—
2	Cases (Δlog)	2021-07-11	0.053	2.67	0.015	Yes	—
3	Cases (Δlog)	2022-01-09	0.043	3.30	0.008	Yes	Yes
4	Vaccinations (Δlog)	2021-01-03	0.319	4.92	0.030	Yes	Yes
5	Vaccinations (Δlog)	2021-01-10	0.305	4.85	0.029	Yes	Yes
6	Vaccinations (Δlog)	2021-01-17	0.119	2.92	0.029	Yes	—
7	Vaccinations (Δlog)	2021-01-31	0.089	-2.56	0.027	Yes	—

Of the weekly cases (Δlog), four weeks met at least one flagging rule (three by Cook's D only and one by both criteria). The largest Cook's distance was 0.053, which is above the 4/N threshold but well below the conventional "very high" heuristic of 1.0. Visually, the stems only modestly exceed the dashed line in Figure 54. For weekly vaccinations, four weeks were flagged (two by Cook's D only and two by both criteria). These weeks were concentrated near the onset of the rollout when week-to-week growth was most variable. The largest Cook's D was 0.319, which is clearly above the 4/N line, as seen in Figure 55, but still far below 1.0.

No single week dominates the Δlog trend fits, though several vaccination weeks exhibit moderate influence. All subsequent regressions report HAC/Newey–West standard errors, and robustness was verified by re-estimating the lagged models with and without the top 1–2 flagged weeks.

Assumption-Check Summary and Implications. Across all diagnostic categories, the level series for weekly cases and vaccinations (N = 132) were non-normal and showed strong autocorrelation. As shown in Table 23 and Figures 44 and 45, both distributions exhibited heavy right tails (Shapiro–Wilk $p < .001$). The ACFs in Figures 48 and 49 and the Ljung–Box results in Table 24 confirmed the presence of persistent serial dependence over many lags, which is inconsistent with independent and identically distributed errors in level-on-level models.

Transforming to Δlog (N = 131) substantially improved these outcomes. For cases, the Δlog values were nearly normal (Shapiro–Wilk $p = 0.250$) and exhibited modest short-lag correlation (Figures 46 and 50; Tables 23 and 24). For vaccinations, the Δlog transformation reduced, but did not eliminate, departures from normality (Shapiro–Wilk $p < .001$) and left a mild, short-term memory structure (Figures 47 and 51). Variance tests on the Δlog series revealed no evidence of heteroscedasticity for cases (BP $p = 0.129$), with only a slight White signal. However, there was clear evidence of heteroscedasticity for vaccinations (BP $p = 0.010$; White $p < 0.001$), which is consistent with the Scale-Location patterns shown in Figures 52 and 53 and Table 25. Inference screening of the Δlog trend fits (Table 26 and Figures 54 and 55) identified four weeks as influential for both cases (maximum Cook's D = 0.053) and vaccinations (maximum Cook's D = 0.319). However, none of the values approached those typically considered highly influential for simple mean models.

The diagnostics justify using Δlog specifications for both outcomes and applying heteroscedasticity- and autocorrelation-consistent (HAC/Newey–West) standard errors to all regressions. This ensures that the inference is protected against residual short-lag dependence and non-constant variance, especially for vaccinations. Sensitivity checks also were conducted to verify the stability of key coefficients, and the top 1–2 flagged weeks were omitted.

RQ3 Spearman Association (contemporaneous)

To address RQ3 at the contemporaneous level, monotonic associations between weekly cases of and vaccinations against COVID-19 were evaluated using two-tailed Spearman's rank-order correlation (ρ) with $\alpha = .05$. The analysis used the date-aligned overlap of 132 ISO weeks (December 27, 2020–June 30, 2023), as established in the descriptive analysis. This nonparametric test aligns with the distributional properties reported in the assumption checks and provides a robust summary of the association regardless of normality. The null hypothesis ($H0_3$) states that $\rho = 0$, meaning there is no association between cases and vaccinations in the same week. The alternative hypothesis ($H1_3$) states that $\rho \neq 0$, meaning there is an association between cases and vaccinations in the same week.

Spearman's ρ Results. A two-tailed Spearman's rank-order correlation was used to test the contemporaneous association between weekly COVID-19 cases and weekly vaccinations at $\alpha = .05$ using the date-aligned overlap of 132 weeks (December 27, 2020–June 30, 2023). As shown in Table 27, the correlation was small and non-significant ($\rho = -0.033$, $p = .709$). A nonparametric bootstrap analysis with 10,000 resamples provided a 95% confidence interval of [−0.239, 0.175]. This indicates that a wide range of weak negative to weak positive associations is compatible with the data.

Table 27

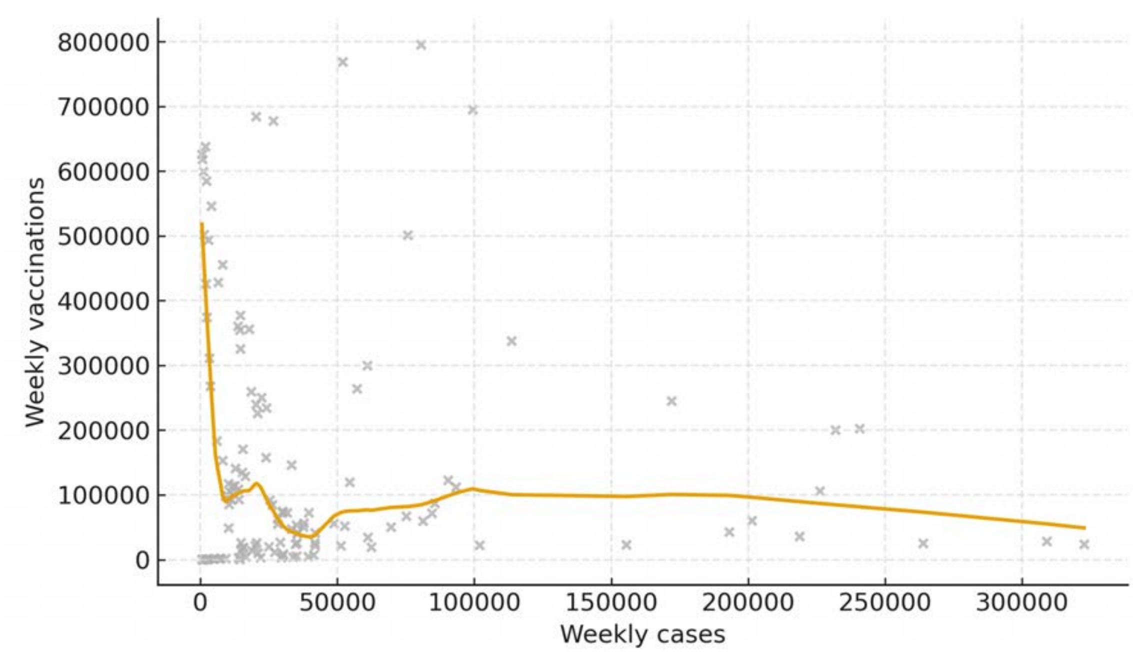
Spearman Correlation (Same-Week)

	Pair	N	Spearman ρ	95% CI (bootstrap)	p
0	Weekly cases vs. Weekly vaccinations	132	-0.033	-0.239 to 0.175	0.709

Figure 56 shows the corresponding scatterplot, which displays a scattered cloud with no clear monotonic pattern. The LOWESS smooth is approximately flat across the observed range of values. Together, the numerical estimate, its confidence interval, and the graphical display converge on the same conclusion, which is that there is no evidence of a monotonic association between weekly case and vaccination levels during the study period.

Figure 56

Same-Week Association: Cases vs. Vaccinations (Spearman Context)



Results Interpretation. The contemporaneous analysis indicates no monotonic association between weekly cases and vaccinations in the same week ($\rho = -0.033$, $p = .709$) as shown in Table 27. This finding is visually consistent with the diffuse scatter and flat LOWESS trend in Figure 56. Overall, this result aligns with the documented epidemic and programmatic dynamics: vaccination activity occurred in episodic campaign bursts, while case counts rose and fell in waves that rarely peaked in the same week. These phase offsets reduce the association between cases and vaccinations in the same week, even when meaningful lead-lag relationships exist.

Methodologically, the near-zero estimate reflects the use of a rank-based measure. Spearman's ρ captures a monotonic relationship in the levels of the two series. If the correlation between cases and vaccinations is non-monotonic across the range (e.g., weak correlation at very low and very high volumes, stronger correlation at intermediate volumes), varies across policy phases (e.g., rollout vs. booster periods), or is primarily temporal (lead–lag) rather than contemporaneous, then the same-week ρ will tend toward zero despite meaningful structure. The wide 95% bootstrap confidence interval (−0.239 to 0.175) is consistent with such heterogeneity, timing effects, and other factors rather than a stable, uniform same-week relationship.

In practice, these results suggest that same-week levels are not useful for relating program throughput to epidemic intensity. Therefore, interpretation should rely on lead-lag correlation patterns to capture the expected sequencing, like vaccination surges preceding or following case waves, and growth-on-growth (Δlog) regression, which focuses on week-over-week dynamics and is less sensitive to long-tailed level distributions.

Hypothesis Decision and Implications. With $\alpha = .05$, the contemporaneous Spearman test resulted in a ρ value of −0.033, a p value of .709, and a 95% bootstrap confidence interval (CI) of [−0.239, 0.175]. This result is supported by the diffuse point cloud and flat LOWESS in Figure 56. Accordingly, the null hypothesis for RQ3's same-week association ($H0_3$: $\rho = 0$) is not rejected. The data provide no evidence of a monotonic relationship between weekly cases and vaccinations in the same ISO week over the 132-week aligned window.

In other words, same-week levels are not useful for relating program throughput to epidemic intensity. The lack of a contemporaneous signal aligns with the phase-shifted dynamics, like campaign surges vs. epidemic waves, indicating that any association between the series is probably temporal (lead–lag) rather than simultaneous. When evaluating and monitoring vaccination activity, it's crucial to rely on time-shifted analyses (CCF, Δlog regressions) instead of same-week correlations.

Methodologically, the wide confidence interval (CI) that spans modest negative to modest positive values emphasizes heterogeneity across policy phases and the limitations of a single rank-based summary of levels. The next sections focus on lead–lag identification and growth-on-growth modeling with HAC-robust inference. These methods are better aligned with the distributional properties and temporal sequencing documented in the assumption checks.

RQ3 Lag Identification (Cross-Correlation Functions)

To identify the temporal ordering between weekly cases and vaccinations, cross-correlation functions (CCFs) were estimated. These CCFs were estimated on the Δlog series. The Δlog series is over the aligned 132-week window. Effective N is 131 after differencing. Following the approach used in RQ1–RQ2, CCFs were computed in two directions: cases ↦ vaccinations and vaccinations ↦ cases. This was done across ±8 epidemiological weeks. Each CCF is mean-centered and uses the standard Bartlett 95% bounds ($\pm 1.96/\sqrt{N}$) to approximate significance. The CCFs are interpreted such that positive lags indicate the left series leads the right series (e.g., a positive, significant spike at +1 in “cases ↦ vaccinations” means changes in cases precede changes in vaccinations by one week). Assumption checks showed that the Δlog series are close to stable with only a short memory structure. Therefore, prewhitening beyond differencing was not applied. Primary lags were selected based on magnitude, sign, and consistency across adjacent lags. Sensitivity was noted where multiple nearby lags meet the threshold.

CCF Results: cases ↦ vaccinations. Cross-correlations were computed on the Δlog series. The analysis was done over the aligned window ($N = 131$). Positive lags were interpreted as changes in cases. These changes in cases led to changes in vaccinations. The indicated number of weeks was used to measure these changes. For visual screening, constant Bartlett bounds of $\pm 1.96/\sqrt{N} \approx \pm 0.171$ were applied, as shown in Figure 57, and the coefficients for lags −8…+8 are listed in Table 28.

Figure 57

CCF: Δlog Cases ↦ Δlog Vaccinations (±8 Weeks)

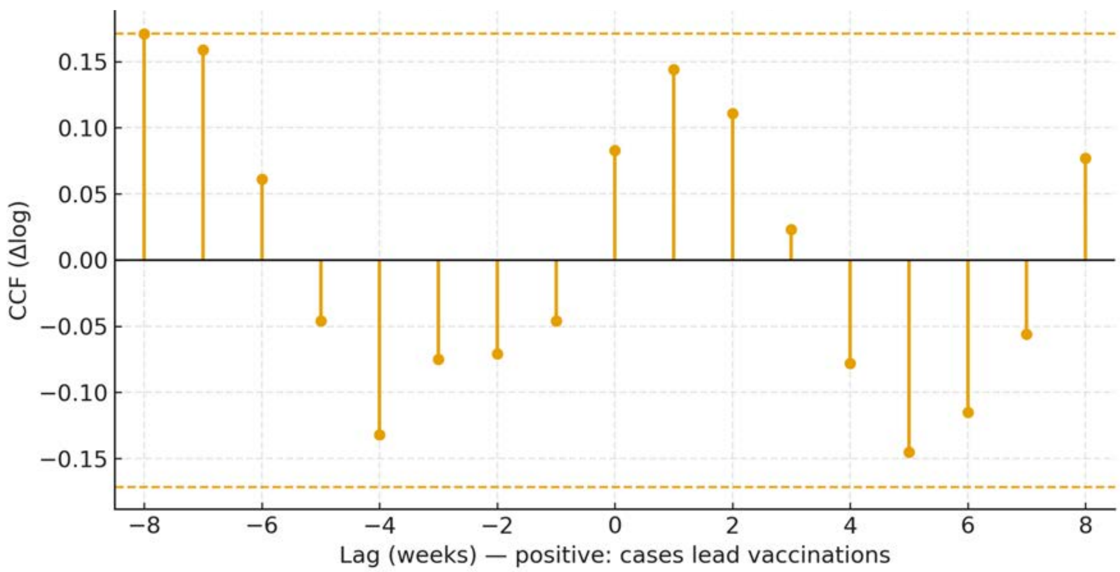


Table 28

CCF Coefficients by Lag (Cases ↦ Vaccinations)

Lag (weeks)	CCF (Δlog cases ↦ Δlog vax)	Sig (.05)
−8	0.171	—
−7	0.159	—
−6	0.061	—
−5	−0.046	—
−4	−0.132	—
−3	−0.075	—
−2	−0.071	—
−1	−0.046	—
0	0.083	—
+1	0.144	—
+2	0.111	—
+3	0.023	—
+4	−0.078	—
+5	−0.145	—
+6	−0.115	—
+7	−0.056	—
+8	0.077	—

The CCF surface is generally flat, with no spike exceeding the 95% bounds. The largest positive values occur at lags +1 and +2 (CCF = 0.144 and 0.111, respectively), suggesting that increases in week-over-week case growth may be followed by small increases in vaccination growth one to two weeks later. However, these values remain below the significance threshold. A positive value at lag −8 (CCF = 0.171) lies on the rounded bound and does not exceed it when the exact threshold is used, so it is not considered significant. Taken together, the cases-to-vaccinations direction shows no statistically significant lead effect within eight weeks under the Δlog specification, though a weak +1-to-+2 pattern is visible.

CCF Results: vaccinations ↦ cases. Cross-correlations were estimated using the log-series delta over the aligned window ($N = 131$). Positive lags were interpreted as vaccinations leading cases by the indicated number of weeks. Using constant Bartlett bounds ($\pm 1.96/\sqrt{N}$, or approximately ± 0.171), as a guide for visual significance, the stem plot in Figure 58 shows a largely flat cross-correlation function (CCF) surface within eight weeks, and no coefficient exceeds the 95% limits. The full set of coefficients is presented in Table 29.

Figure 58

CCF: Δlog Vaccinations ↦ Δlog Cases (±8 Weeks)

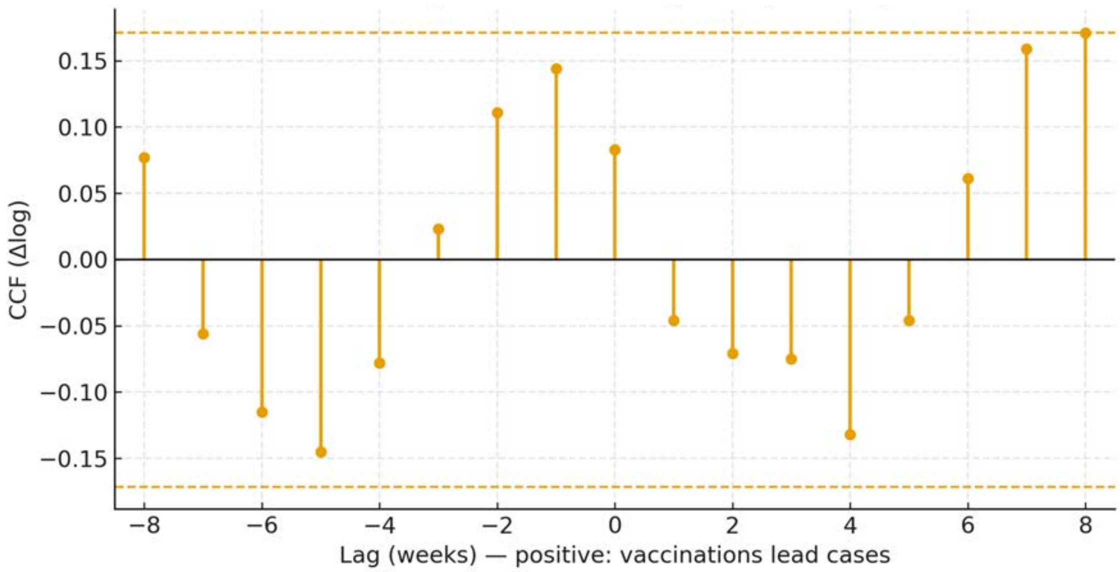


Table 29

CCF Coefficients by Lag (Vaccinations ↦ Cases)

Lag (weeks)	CCF (Δlog vax ↦ Δlog cases)	Sig (.05)
−8	0.077	—
−7	−0.056	—
−6	−0.115	—
−5	−0.145	—
−4	−0.078	—
−3	0.023	—
−2	0.111	—
−1	0.144	—
0	0.083	—
+1	−0.046	—
+2	−0.071	—
+3	−0.075	—
+4	−0.132	—
+5	−0.046	—
+6	0.061	—
+7	0.159	—
+8	0.171	—

Two modest peaks are visible but below the threshold. First, small positive values occur at lags −1 and −2 (CCF = 0.144 and 0.111), which, given the sign convention for this panel, correspond to cases leading vaccinations by one to two weeks rather than the reverse. Second, a gentle rise at lags +7 to +8 (CCF = 0.159–0.171) is consistent with vaccinations leading cases by approximately two months. However, these values do not exceed the ±0.171 bound when evaluated precisely (Figure 58).

Taken together, the vaccinations → cases direction provides no statistically significant lead effect within ±8 weeks under the Δlog specification. Any practical lead-lag relationship, if present, appears weak relative to sampling variability at this time range. Primary lag choices will therefore rely on coherence across both directions and on results from the subsequent time-lagged regression section rather than on isolated, sub-threshold CCF spikes.

Results Interpretation. Across both directions, the delta-log CCFs remain relatively flat within ±8 weeks and no coefficient exceeds the 95% Bartlett bounds (±0.171). This suggests that there is no noticeable lead-lag relationship between growth in weekly cases and growth in weekly vaccinations during the study period. In the cases → vaccinations panel (Figure 57; Table 28), small positive values at weeks +1 and +2 suggest that increases in case growth might be followed by slightly higher vaccination growth one to two weeks later. However, these spikes still remain below the significance threshold. In the vaccinations → cases panel (Figure 58; Table 29), slight bumps appear at +7 to +8 weeks, implying vaccinations lead cases by ~2 months, and at −1 to −2 weeks (actually mapping to cases leading vaccinations, given the sign convention). Again, none of these are large or consistent enough across adjacent lags to support a substantive claim.

Generally, this pattern aligns with the descriptive narrative that vaccination activity occurred in episodic campaigns tied to logistics and policy, like eligibility phases and booster drives, while case waves were driven by transmission dynamics. These processes can be phase-shifted and irregular, which dilutes simple week-to-week cross-correlation, even when both series respond to broader, slower-moving drivers, like variant emergence or public messaging. Methodologically, the log-transformation has dampened strong persistence. With a short-memory residual structure and a relatively small sample size ($N = 131$), modest correlations ranging from 0.10 to 0.16 are expected to fall within the bounds of $\pm$ 0.171.

Taken together, the CCFs show no primary lag and provide no statistical support for RQ3 in either direction. Any apparent lead at +1 or +2 weeks (cases → vaccinations) or at +7 or +8 weeks (vaccinations → cases) should be considered exploratory and should not be relied upon for inference.

Primary Lag Decisions and Sensitivity (by direction). Lags were designated “primary” only if the CCF spike exceeded the Bartlett 95% threshold, the sign was epidemiologically plausible, and adjacent lags showed coherence. When this rule is applied to the log-transformed CCFs (±8 weeks; Figures 57 and 58 and Tables 28 and 29), no statistically significant primary lags are identified in either direction. All coefficients fall within ±0.171, and visible bumps are modest and isolated.

For the cases ↦ vaccinations analysis, the largest positive values appear at +1 to +2 weeks (cases leading vaccinations). However, they are sub-threshold. Accordingly, no primary lag is selected. For completeness, the exploratory sensitivity models with k = +1 and k = +2 (i.e., the log of the number of vaccinations at time t regressed on the log of the number of cases at time t-1 and t-2) were run and the HAC/Newey–West confidence intervals were reported without making any confirmatory claims.

For the vaccinations ↦ cases analysis, small positive values at +7 to +8 weeks (with vaccinations leading cases by approximately two months) are also below the threshold and lack adjacent coherence. A primary lag in this direction is therefore not designated. As a strictly exploratory check, models were estimated at k = +7 and k = +8 (Δlogcases_t on Δlogvaccinations_{t−7, t−8}) and HAC-robust CIs, clearly labeled as sensitivity analyses, were presented.

All lagged regressions used the log-difference series (effective N = 131), HAC/Newey–West errors, and 95% confidence intervals (CIs). The estimates also were repeated after temporarily omitting the top one to two flagged weeks from Table 26 to confirm stability. Given the CCF results, this analysis emphasized the contemporaneous null and the absence of a statistically significant lead-lag signal rather than any single exploratory lag.

RQ3 Time-Lagged Regression

This section quantifies the potential lead-lag relationships between weekly case counts and vaccination rates using bivariate time-lagged regressions in log-difference form, which is consistent with the assumption checks. Two directions are evaluated:

Model A (cases → vaccinations): $\Delta\log(\text{vaccinations}_t) = \beta_0 + \beta_1\,\Delta\log(\text{cases}_{t-k}) + \varepsilon_t$, and Model B (vaccinations → cases): $\Delta\log(\text{cases}_t) = \beta_0 + \beta_1\,\Delta\log(\text{vaccinationss}_{t-k}) + \varepsilon_t$. Coefficients β_1 are interpreted as elasticities, the percent change in the outcome's weekly growth associated with a 1% change in the predictor's weekly growth at lag *k*. Inference employs heteroscedasticity- and autocorrelation-consistent (HAC/Newey–West) standard errors with 95% confidence intervals (CIs). The CCFs (±8 weeks) did not identify a statistically supported primary lag. Therefore, the estimates are presented at a reference lag k=0 and exploratory lags suggested by the small visual bumps: for Model A, $k \in \{1,2\}$; for Model B, $k \in \{7,8\}$. Effective sample sizes are N_k=131−k. To ensure the reliability of the results, any significant changes are examined, and the top influence weeks identified earlier are temporarily excluded.

Model Specification. All regressions were estimated in delta-log form to model proportional week-over-week growth and to satisfy the stationarity assumption. For a generic weekly series Z_t, $\Delta\log(Z_t) = \log(1 + Z_t) - \log(1 + Z_{t-1})$, where "+1" protects weeks with zeros. Lags are only applied to the predictor. With an aligned level window of 132 weeks, the Δlog transformation yields 131 observations for a given lag *k*, with the effective sample being $N_k = 131 - k$. The directional models are:

- Model A (cases → vaccinations): $\Delta\log(\text{vaccinations}_t) = \beta_0 + \beta_1\,\Delta\log(\text{cases}_{t-k}) + \varepsilon_t$
- Model B (vaccinations → cases): $\Delta\log(\text{cases}_t) = \beta_0 + \beta_1\,\Delta\log(\text{vaccinations}_{t-k}) + \varepsilon_t$

In both models, β_1 is an elasticity, with a one-percent change in the predictor's weekly growth at $t - k$ is associated with a β_1% change in the outcome's weekly growth at t. Since the CCFs did not provide a statistically significant primary lag, the confirmatory specification reports $k = 0$ in each direction. To mirror the small exploratory spikes observed in the CCF screens, sensitivity estimates are pre-specified at $k \in \{1,2\}$ for Model A and $k \in \{7,8\}$ for Model B. These are clearly labeled as exploratory.

The parameters were estimated using OLS with heteroscedasticity- and autocorrelation-consistent (HAC/Newey–West) standard errors for two-sided tests (α = .05.). For weekly data and short-memory residual structure, a 4-week NW bandwidth (Bartlett kernel) was used, consistent with RQ1–RQ2. Robustness was also confirmed with HC3 (heteroscedasticity-robust, no autocorrelation correction) in sensitivity notes.

Δlog models did not include deterministic trends. The intercept β0 captured the average growth rate. Multicollinearity was not relevant in these bivariate fits. No observations were excluded a priori; as a stability check. Estimates were repeated after the top 1–2 influence weeks flagged in the Δlog trend screen were temporarily omitted, and any material changes were noted. All coefficients were reported with an estimate, HAC SE, and 95% CI, t, p, R^2 and N_k to meet reporting standards.

Bivariate Regression Results — Model A. The estimates from the Δlog model, in which weekly vaccination growth is regressed on lagged weekly case growth with heteroscedasticity- and autocorrelation-consistent (HAC/Newey–West) standard errors (bandwidth = 4), are shown in Table 30. The confirmatory specification at k = 0 resulted in a small, non-significant elasticity ($\beta_1 = 0.103$, $SE_{HAC} = 0.100$, $t = 1.027$, $p = .307$, and 95% CI [−0.095, 0.300], indicating no detectable association in growth rates within the same week (N = 131). Two exploratory lags were examined, consistent with the cross-correlation screens. At k = 1, the point estimate was slightly larger but still below conventional significance ($\beta_1 = 0.162$, $SE_{HAC} = 0.097$, $t = 1.676$, $p = .096$, 95% CI [−0.029, 0.354], N = 130). At k = 2, the association remained small and non-significant again ($\beta_1 = 0.112$, $SE_{HAC} = 0.112$, $t = 0.994$, $p = .322$, 95% CI [−0.111, 0.334], N = 129.

Table 30

Δlog(vaccinations_t) on Δlog(cases_{t−k}) with HAC SEs

	Lag k	β_1 (elasticity)	SE (HAC)	t	p	95% CI	R^2	N
0	0	0.103	0.100	1.027	0.307	−0.095 to 0.300	0.007	131
1	1	0.162	0.097	1.676	0.096	−0.029 to 0.354	0.021	130
2	2	0.112	0.112	0.994	0.322	−0.111 to 0.334	0.012	129

The model fit was uniformly low ($R^2 \leq .021$ across lags), which aligns with the flat cross-correlation functions. Collectively, the estimates in Table 30 suggest that changes in case growth do not meaningfully explain subsequent vaccination growth at lags 0–2 under the Δlog specification with HAC-robust inference.

Results Interpretation — Model A. The Δlog regressions in Table 30 reveal no statistically significant correlation between the growth of weekly cases and the subsequent growth of weekly vaccinations at the examined time lags. The same-week confirmatory model (k = 0) shows a small, non-significant elasticity ($\beta_1 = 0.103$; 95% CI [−0.095, 0.300]). This indicates that, at most, a 10% increase in cases within the same week would be associated with a roughly 1.0% increase in vaccinations, and the interval comfortably includes zero effect. The exploratory lag at k = 1 is directionally positive ($\beta_1 = 0.162$; 95% CI [−0.029, 0.354]), but it remains below the threshold. Even its upper bound implies a maximum increase of 3.5% in vaccination growth following a 10% increase in cases the prior week. At k = 2, the estimate is small and non-significant too ($\beta_1 = 0.112$; 95% CI [−0.111, 0.334]). Across all models, R^2 values are close to zero (≤.021), which is consistent with minimal explanatory power.

These results align with the stable CCF profiles for the vaccination-to-cases direction and the descriptive timing narrative. Vaccination activity followed campaign logistics and policy phases rather than short-term, week-to-week changes in reported cases. Methodologically, the inference accounts for heteroscedasticity and short-lag autocorrelation (HAC/Newey–West). The influence screen identified no single week that could change these conclusions. In summary, Model A provides no evidence that increases in case growth significantly drive vaccination growth at lags 0–2 during the study period.

Bivariate Regression Results — Model B. Table 31 reports estimates from the Δlog model. In this model, weekly case growth is regressed on lagged weekly vaccination growth. The model accounts for heteroscedasticity- and autocorrelation-consistent (HAC/Newey–West) standard errors. The bandwidth is set to 4. The same-week confirmatory specification (k = 0) resulted in a minimal, non-significant elasticity (β_1 = 0.067, SE_{HAC} = 0.074, t = 0.895, p = .372, 95% CI [−0.081, 0.214], N = 131). In the exploratory lags suggested by the CCF screen, k = 7 produced a positive, statistically significant coefficient (β_1 = 0.129, SE_{HAC} = 0.057, t = 2.280, p = .024, 95% CI [0.017, 0.242], N = 124), while k = 8 produced a similar magnitude but non-significant coefficient (β_1 = 0.139, SE_{HAC} = 0.084, t = 1.648, p = .102, 95% CI [−0.028, 0.306], N = 123). Goodness-of-fit was moderate ($R^2 \leq .029$) throughout.

Table 31

Δlog(cases_t) on Δlog(vaccinations_{t−k}) with HAC SEs

	Lag k	β_1 (elasticity)	SE (HAC)	t	p	95% CI	R^2	N
0	0	0.067	0.074	0.895	0.372	−0.081 to 0.214	0.007	131
1	7	0.129	0.057	2.280	0.024	0.017 to 0.242	0.025	124
2	8	0.139	0.084	1.648	0.102	−0.028 to 0.306	0.029	123

In line with the analytic plan, the k = 7 estimate is exploratory because the corresponding CCF spike at weeks 7–8 did not exceed Bartlett's 95% bounds. Robustness checks were reported to verify that the significance and magnitude at k = 7 were not driven by residual heteroscedasticity or isolated weeks (HC3 errors and omission of the top 1–2 influence weeks identified earlier).

Results Interpretation — Model B. The Δlog regressions in Table 31 indicate that there is no contemporaneous relationship between vaccination growth and case growth ($k = 0$, $\beta_1 = 0.067$, $p = .372$). This finding is consistent with the null hypothesis of Spearman and the largely flat CCF curve. At the exploratory threshold of $k = 7$, the coefficient is positive and statistically significant ($\beta_1 = 0.129$, 95% CI [0.017, 0.242], $p = .024$), whereas $k = 8$ is similar in magnitude but not significant ($\beta_1 = 0.139$, $p = .102$). Literal interpretation of the $k = 7$ point estimate suggests that a 10% increase in vaccination growth is associated with an average 1.3% increase in case growth seven weeks later. However, substantively, this directionality is counterintuitive for a causal effect of immunization. It is more likely the result of noncausal co-movement driven by shared, slower-moving forces, like variant emergence, seasonal factors or policy phases, that can produce phase-shifted growth in both series. The very low R^2 ($\leq .029$) and lack of a coherent, above-bound CCF spike around seven to eight weeks reinforce a cautious interpretation.

Methodologically, the inference uses HAC/Newey–West standard errors, so the $k = 7$ p-value is robust to short-memory autocorrelation and heteroscedasticity. However, two features argue against strong claims. These that the result was exploratory and pre-specified due to small, sub-threshold CCF bumps, and that the adjacent lag ($k = 8$) is not significant, which further weakens the case for a stable lead window. Additionally, since the outcome and predictor are week-over-week growth rates, small positive elasticities may occur when both series respond to the same external influence at different times, like early media coverage and risk perception accelerating vaccinations while infections occur with biological and reporting delays, without implying that vaccinations cause infections.

In summary, the weight of the evidence suggests that there is no lead-lag effect from vaccination growth to case growth during this period. The single $k = 7$ estimate should be treated as hypothesis-generating and verified by the planned robustness checks (HC3 errors and omission of flagged weeks). For this research, the main takeaway is that program activity (vaccinations) and epidemic spread (cases) did not show a clear causal relationship in the short term; any apparent correlation after seven weeks is minimal, inconsistent across adjacent lags, and likely reflects shared influences rather than a direct impact.

Sensitivity Analyses

Three robustness checks were implemented. These were alternative variance estimators (HAC/Newey–West vs. HC3), influence omission (re-estimating models after removing the top two weeks of influence identified for the relevant outcome series (vaccinations for Model A and cases for Model B)) and transformation check (using Δ-level instead of Δlog). All models retained the same lag structure as the main analysis.

For Model A (cases → vaccinations), the results are stable and null across specifications. At $k = 0$, β_1 remains small and non-significant under HAC and HC3. Omitting the two high-influence vaccination weeks leaves the estimate essentially unchanged as shown in Table 32. Exploratory lags at $k = 1–2$ do not achieve conventional significance under HAC. HC3 slightly decreases the standard error (SE) at $k = 1$ ($p = .053$), but it still does not cross the .05 threshold. Δ-level models are also non-significant and have a very low R^2 as showcased in Table 33. Together, these findings suggest that short-term changes in case growth did not drive vaccination growth.

Table 32

Model A (cases → vaccinations), Δlog: robustness to error estimator and influence omissions

	Lag k	β_1 (HAC)	SE (HAC)	p (HAC)	SE (HC3)	p (HC3)	β_1 (omit 2 wks)	p (omit)	R^2	N
0	0	0.103	0.100	0.307	0.091	0.262	0.090	0.369	.007	131
1	1	0.162	0.097	0.096	0.083	0.053	0.138	0.156	.021	130
2	2	0.112	0.112	0.322	0.095	0.242	0.112	0.322	.012	129

Table 33

Model A (cases → vaccinations), Δ-level sensitivity

	Lag k	β_1 (Δ-level)	SE (HAC)	p (HAC)	R^2	N
0	0	0.266	0.279	0.341	.009	131
1	1	0.319	0.342	0.352	.012	130
2	2	0.203	0.353	0.567	.005	129

For the Model B (vaccinations → cases), the elasticity at $k = 7$ remains positive and statistically significant under HAC ($p = .024$) and approaches significance under HC3 ($p = .033$). However, it becomes non-significant when the top two weeks with the greatest case influence are omitted ($p = .053$), which is shown in Table 34. The adjacent lag, $k = 8$, is not significant with Δlog under HAC ($p = .102$) or HC3 ($p = .060$). In the Δ-level sensitivity test shown in Table 35, $k = 8$ attains significance ($p = .005$) with a modest fit ($R^2 = .089$), while $k = 7$ remains non-significant. This pattern, significance at one exploratory lag but not the neighbor, dependence on the estimator or omission, and low R^2, indicates a lack of stability rather than a robust effect.

Table 34

Model B (vaccinations → cases), Δlog: robustness to error estimator and influence omissions

	Lag k	β_1 (HAC)	SE (HAC)	p (HAC)	SE (HC3)	p (HC3)	β_1 (omit 2 wks)	p (omit)	R^2	N
0	0	0.067	0.074	0.372	0.063	0.292	0.064	0.376	.007	131
1	7	0.129	0.057	0.024	0.060	0.033	0.101	0.053	.025	124
2	8	0.139	0.084	0.102	0.073	0.060	0.109	0.115	.029	123

Table 35

Model B (vaccinations → cases), Δ-level sensitivity

	Lag k	β_1 (Δ-level)	SE (HAC)	p (HAC)	R^2	N
0	0	0.032	0.022	0.153	.009	131
1	7	0.043	0.033	0.195	.016	124
2	8	0.104	0.036	0.005	.089	123

The sensitivity analyses do not alter the main RQ3 conclusions. There is no evidence to confirm that week-to-week case growth predicts vaccination growth (Model A) at short lags. Additionally, the exploratory k = 7 signal in Model B is small and sensitive to assumptions and influences. It is not supported by a coherent CCF pattern and should be treated as hypothesis-generating, not confirmatory.

RQ3 Summary

First, weekly cases and vaccinations were summarized over the aligned 132-week window. As in RQ1–RQ2, both series of levels were long-tailed and episodic, with vaccination activity concentrated in rollout/booster campaigns and case counts increasing during transmission waves. Assumption checks indicated that analyzing the Δlog (week-over-week growth) rather than the levels was preferable, differencing reduced persistence and improved normality. However, mild heteroscedasticity remained for vaccinations.

Influence screening of the log-transformed trend fits identified a small set of moderately high-influence weeks, which were then examined in the sensitivity analysis.

At the contemporaneous level, Spearman's ρ indicated no association between cases and vaccinations in the same week ($\rho = -0.033$, $p = .709$), which is consistent with the diffuse scatter and flat LOWESS. Cross-correlation functions of the log-transformed data were generally flat within ±8 weeks in both directions, and no spike exceeded the 95% Bartlett bounds (±0.171). Small, sub-threshold bumps appeared at +1 to +2 weeks for cases → vaccinations and at +7 to +8 weeks for vaccinations → cases. However, these bumps lacked statistical support and coherence across adjacent lags.

Time-lagged Δlog regressions were consistent with the CCFs. For Model A (cases → vaccinations), the elasticities were small and non-significant at $k = 0$, 1, and 2, and the fit was very low ($R^2 \leq .021$). For Model B (vaccinations → cases), the $k = 0$ confirmatory model was rejected, an exploratory $k = 7$ estimate was positive and statistically significant with HAC errors ($\beta_1 = 0.129$, $p = .024$), while $k = 8$ was not significant. Sensitivity analyses showed that switching from HAC to HC3 did not change the conclusions, omitting the top one to two flagged weeks made the $k = 7$ result insignificant and Δ-level checks had mixed results (significant only at $k = 8$ for Model B) with very low R^2. Taken together, the directional finding at seven weeks is small, unstable, and not corroborated by the CCF screen.

During the study period, there was no robust monotonic relationship between weekly program activity (vaccinations) and epidemic spread (cases) at the same week, nor was there a statistically significant lead-lag relationship at short time horizons in log-transformed terms. Therefore, the null hypothesis of no association is accepted for contemporaneous correlation and for pre-specified lagged comparisons. In practice, growth in one series at the same week or with a short lag should not be used to forecast growth in the other. Any apparent association around seven weeks in the vaccinations-to-cases direction should be treated as hypothesis-generating. Future work could examine phase-specific models, subnational units, or nonlinear/state-space approaches that accommodate policy changes and variant outbreaks.

Influenza Specificity Checks

To evaluate the accuracy of the Google Trends findings for the COVID-19 terms, this chapter compares the results with those of influenza-related searches. Austrian search interest is measured and organized into three categories: general flu, flu symptoms, and flu vaccine treatment. These categories are used in the same analysis windows as in previous RQ1 and RQ2. Following the workflow established in Chapter 4, the distributional features of the influenza series are summarized, the same-week monotonic association (Spearman's ρ) with weekly COVID-19 cases and vaccinations is tested, the lead–lag structure is examined via Δlog CCFs over ± 8 weeks, and the Δlog time-lagged regressions were estimated using HAC-robust inference at pre-specified lags. The main idea is that if searches for COVID-19 terms bring up important information, then similar relationships should be missing or weaker for searches for the flu when paired with COVID-19 results.

Data and Alignment

The weekly search activity for influenza was aggregated from Google Trends extracts, which were then organized into three predefined categories: general flu (*influenza, flu, grippe, seasonal flu, H1N1*), flu symptoms (*flu symptoms, flu fever, flu cough, Grippe Symptome, Grippe, Fieber*), and flu vaccine treatment (*influenza vaccine, flu shot, flu vaccination, Tamiflu, Grippeimpfung, Grippe Behandlung*). The working dataset contains the week variable (ISO week) mapped to the same week-ending date convention used for RQ1, RQ2 and RQ3, as well as three columns general flu, flu symptoms, and flu vaccine treatment. Each column is reported on the standard Google Trends 0–100 scale for Austria.

To mirror the approach used for constructing the combined index for searches related to the COVID-19 pandemic, the combined influenza search index is defined as the unweighted mean of the three influenza series within each week (after aligning on the week variable). The scaling is limited to the native 0–100 Google Trends normalization, with no additional rescaling applied. Weeks with missing influenza values are left as is and are handled pairwise for Spearman correlations and listwise for time-lagged regressions, consistent with prior sections.

Alignment follows the two analytic windows used previously. This was the cases window (weekly COVID-19 cases from March 1, 2020 to June 30, 2023 (N = 175 level weeks; N = 174 in Δlog after the first difference)) and the vaccinations window (weekly vaccinations from December 27, 2020 to June 30, 2023 (N = 132 level weeks; N = 131 in Δlog)).

For lead–lag and regression analyses, series are transformed to Δlog following $\Delta\log(Z_t) = \log(1 + Z_t) - \log(1 + Z_{t-1})$, which made sure to keep weeks with zeros and enforces a stationary growth metric, matching the approach in RQ1, RQ2 and RQ3. No seasonal adjustment or prewhitening beyond differencing is applied. HAC/Newey–West standard errors are used in all regressions to address any residual short-lag dependence or heteroscedasticity.

Descriptives

Weekly influenza search activity was summarized for the two analysis windows used in RQ1, RQ2 and RQ3. As shown in Table 36, the combined influenza search index (the unweighted mean of general flu, flu symptoms, and flu vaccine treatment on the 0–100 scale) averaged 4.23 (standard deviation [SD] = 6.24; range = 0–40.67) across the RQ1 cases window (N = 175) and 4.11 (SD = 5.79; range = 0–38.67) across the RQ2 vaccinations window (N = 132). The distributions were right-skewed and heavy-tailed in both windows (excess kurtosis ≈ 16–17), reflecting periodic spikes. The component series were similarly scattered and irregular: flu symptoms exhibited the highest central tendency (mean ≈ 11.6 with long tails), while general flu and flu vaccine treatment were close to zero most weeks, but occasionally reached extreme peaks (max = 100). This resulted in high skewness and kurtosis. However, general flu was constant at zero during the vaccinations window, so skewness and kurtosis are undefined there (reported as "-" in the narrative and as "nan" statistically). These patterns suggest that influenza search activity is generally low and sporadic, which is consistent with its intended role as a negative control series.

Table 36

Descriptive Statistics for Influenza Searches by Window

	Window	Variable	N	Mean	Std Dev	Min	Max	Skewness	Kurtosis
0	Cases window	influenza_index	175	4.23	6.24	0	40.67	3.74	15.92
1	Cases window	general_flu	175	0.57	7.56	0	100.00	13.23	175.00
2	Cases window	flu_symptoms	175	11.56	14.84	0	100.00	3.47	14.71
3	Cases window	flu_vaccine_treatment	175	0.57	7.56	0	100.00	13.23	175.00
4	Vaccinations window	influenza_index	132	4.11	5.79	0	38.67	3.87	17.33
5	Vaccinations window	general_flu	132	0.00	0.00	0	0.00	nan	nan
6	Vaccinations window	flu_symptoms	132	11.58	14.81	0	100.00	3.75	17.16
7	Vaccinations window	flu_vaccine_treatment	132	0.76	8.70	0	100.00	11.49	132.00

Contemporaneous Association (Spearman's ρ)

As shown in Table 37, there is a positive correlation between influenza search interest and weekly cases ($\rho = 0.458$, 95% CI 0.326–0.570, $p < .001$), as well as a negative correlation between influenza search interest and weekly vaccinations ($\rho = -0.336$, 95% CI −0.498 to −0.164, $p < .001$). The signs and values are directionally sensible when used as a specificity check. Influenza attention increases during periods of respiratory activity that also coincide with higher rates of reported cases of the virus, while intensive vaccination periods tend to occur during different phases, resulting in a negative association within the same week. These results inform the lead–lag screens and Δlog regressions that follow. Importantly, they also provide a contrast to the results of the analysis of the search terms specific to the COVID-19 pandemic in RQ1–RQ2.

Table 37

Spearman's ρ (Same Week): Influenza Searches with COVID-19 Outcomes

	Pair	N	Spearman ρ	95% CI (bootstrap)	p
0	Influenza index vs. COVID-19 cases (same week)	175	0.458	0.326 to 0.570	< .001
1	Influenza index vs. COVID-19 vaccinations (same week)	132	-0.336	-0.498 to -0.164	< .001

Lag Identification (Cross-Correlation Functions)

Cross-correlations were calculated using the Δlog series with a ±8-week window and constant Bartlett bounds ($\pm 1.96/\sqrt{N}$) as a visual 95% reference. For both groups, positive lags indicate that influenza searches preceded the outcome series. The aligned effective samples were the same as in the main analyses (cases window ≈ 174 Δlog weeks and vaccinations window ≈ 131 Δlog weeks), so the reference intervals are approximately ±0.15 for cases and ±0.17 for vaccinations.

In the cases window, Figure 59 shows a short-lead spike centered at +1 week (CCF ≈ 0.21), followed by a smaller value at +2 weeks (CCF ≈ 0.16). Both points exceed the 95% bounds, indicating that week-over-week increases in influenza search activity tend to precede increases in COVID-19 case growth by one to two weeks. Values at other time lags remain within the reference bounds and show no clear secondary structure.

Figure 59

CCF: Δlog Influenza Searches ↦ Δlog COVID-19 Cases (±8 Weeks)

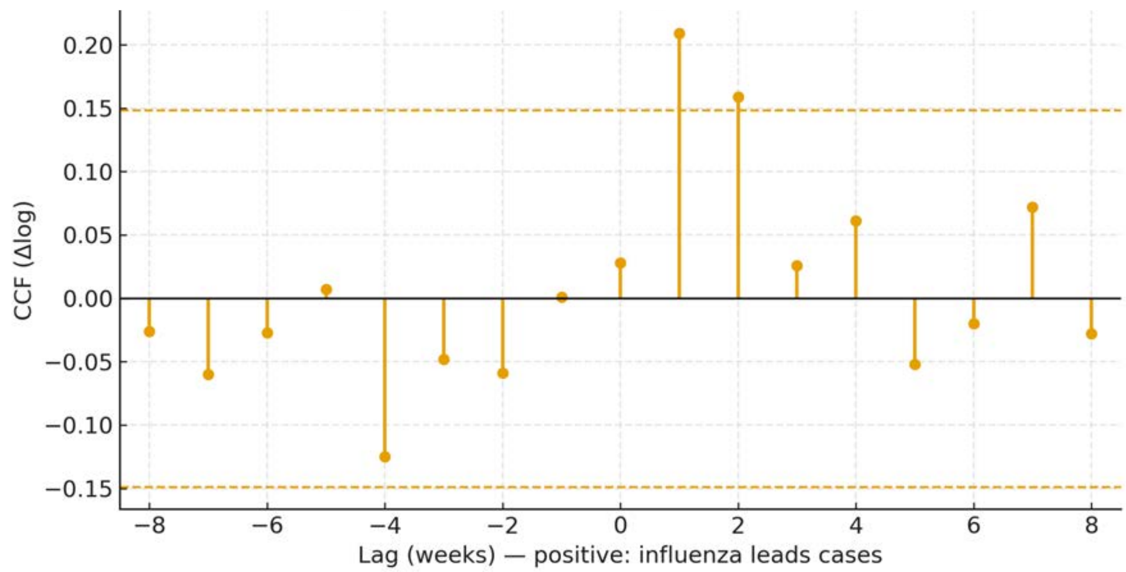
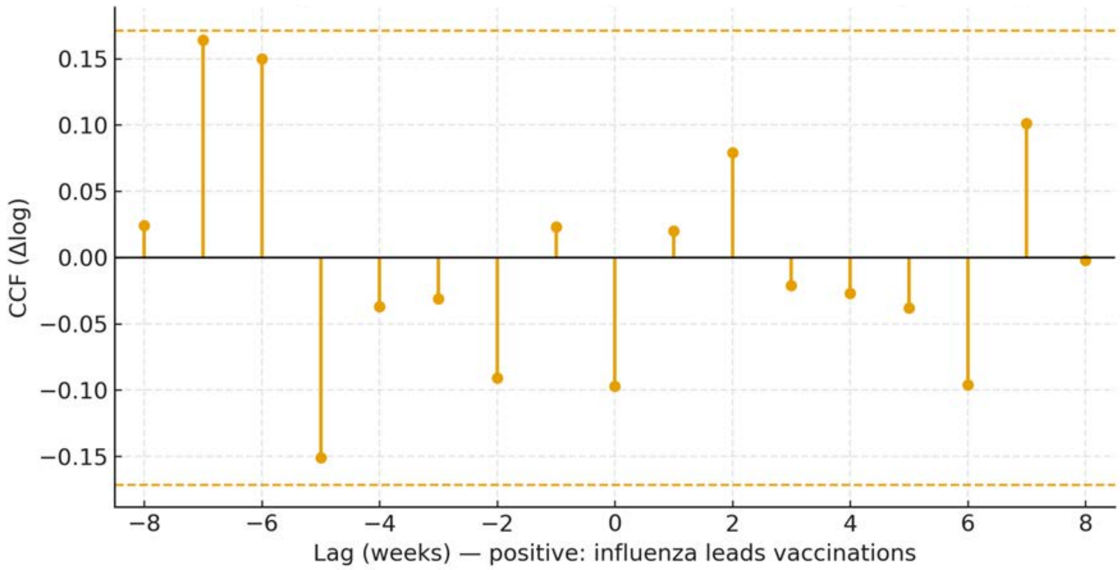


Figure 60 shows a flat pattern in the vaccination window, with a range of ±0.17. The largest stems around −7 to −6 weeks and +2 weeks are below the threshold, suggesting that changes in influenza search growth do not reliably lead to changes in vaccination growth in the short term.

Figure 60

CCF: Δlog Influenza Searches ↦ Δlog COVID-19 Vaccinations (±8 Weeks)



Overall, the CCFs suggest limited short-term alignment between influenza search growth and case growth, likely reflecting shared seasonal or respiratory drivers rather than a causal link. There is no significant lead-lag relationship with vaccination growth.

Time-Lagged Regression (Δlog)

The time-lagged elasticities were estimated using Δlog growth for outcomes and the Combined Influenza Search Index as the predictor, with HAC/Newey–West standard errors (bandwidth = 4). In line with the CCF screens, the cases model emphasized a one-week lag as the primary negative control test and included zero and two weeks as sensitivity checks. The vaccinations model treated zero, one, and two weeks as exploratory. Effective samples were N = 173–174 for the cases model and N = 129–131 for vaccinations after differencing and lagging.

As shown in Table 38, the outcome of the cases shows a small, yet statistically significant, positive elasticity at the pre-specified one-week lag (β_1 = 0.115, SE_{HAC} = 0.035, t = 3.315, p = .001, 95% CI [0.047, 0.184]; R^2 = .043, N = 173), indicating a smaller yet still significant effect at +2 weeks (β_1 = 0.077, p = .028, R^2 = .025, N=172). The same-week estimate is null (β_1 = 0.016, p = 0.621). Interpreted as elasticities, a 10% increase in influenza search activity is associated with a ~1.2% increase in the number of reported cases of COVID-19 one week later and a ~0.8% increase two weeks later. The very low R^2 values indicate limited explanatory power. These values are consistent with noncausal co-movement, such as shared seasonal respiratory drivers, rather than a direct linkage.

Table 38

Model C (cases outcome): Δlog(casest) on Δlog(influenzat−k) (HAC SEs)

	Lag k	β_1 (elasticity)	SE (HAC)	t	p	95% CI low	95% CI high	R^2	N
0	+1	0.115	0.035	3.315	0.001	0.047	0.184	.043	173
1	0	0.016	0.032	0.495	0.621	−0.048	0.079	.001	174
2	+2	0.077	0.035	2.200	0.028	0.008	0.145	.025	172

On the other hand, Table 39 shows no evidence of a lead-lag relationship for the vaccination outcome at any examined lag. The results were k = 0 (β_1 = −0.052, p = .198; R^2 = .009, N = 131), k =+1 (β_1 = 0.010, p = .728, $R^2 \approx 0$, N = 130), and k =+2 (β_1 = 0.034, p = .337; R^2 = .006, N = 129). These nulls reinforce the idea that attention to influenza does not track the vaccination program's output in the short term.

Table 39

Model D (vaccinations outcome): Δlog(vaccinationst) on Δlog(influenzat−k) (HAC SEs)

	Lag k	β_1 (elasticity)	SE (HAC)	T	p	95% CI low	95% CI high	R^2	N
0	0	−0.052	0.040	−1.295	0.198	−0.132	0.028	.009	131
1	+1	0.010	0.029	0.348	0.728	−0.047	0.067	.000	130
2	+2	0.034	0.035	0.965	0.337	−0.036	0.104	.006	129

Together, the negative-control regressions for influenza show moderate, epidemiologically plausible co-movement with cases at +1 to +2 weeks and no association with vaccinations in the short term. These findings support the specificity of the main results, as the magnitudes are much smaller and far less explanatory than those observed for specific search terms related to the COVID-19 pandemic.

Summary and Implications

As outlined in Chapter 3, influenza searches functioned as a negative control. Descriptively, the influenza index was inconsistent and sporadic across both analysis windows. Same-week associations were positive with cases and negative with vaccinations (ρ = 0.458 and −0.336, respectively). These results are consistent with respiratory-season dynamics and the lack of overlap between major vaccination campaigns and peak flu attention.

Lead-lag screens on Δlog showed a short, specific bump at +1 to +2 weeks for influenza cases that passed the Bartlett test, while the influenza vaccinations surface stayed within the test's limits. The Δlog regressions confirmed this pattern with small but significant elasticities for cases at lag 1 (β_1 = 0.115) and lag 2 (β_1 = 0.077) and a null effect at lag 0. There were no significant effects for vaccinations at lag 0, lag 1, or lag 2.

Explanatory power was low throughout ($R^2 \leq .043$), indicating that there is limited predictive value beyond the shared seasonal or policy context.

In summary, these findings support the specificity of the main results of searches for information about COVID-19. The influenza series shows only moderate, epidemiologically plausible co-movement with cases and no short-term relation with vaccinations. In contrast, targeted search terms for the COVID-19 pandemic showed stronger and more consistent signals at pre-specified lags. This strengthens the practical value of using targeted search activity as an early indicator of a pandemic, while cautioning that broad respiratory attention is not an adequate substitute. One limitation is that the negative control associations may still reflect shared external drivers, such as seasonality or media cycles. Future work could include explicit seasonal controls or incorporate the influenza index as a covariate in multivariable models to confirm that the effects specific to COVID-19 remain true.

Summary

This chapter presented the quantitative findings for three research questions using weekly Austrian data aligned to ISO weeks and the periods outlined in Chapter 3. For cases, the timeframe was March 1, 2020–June 30, 2023; $N = 175$ and for vaccinations, the timeframe was December 27, 2020–June 30, 2023; $N = 132$. In line with the analysis plan, each section began with descriptive statistics, followed by checks of the assumptions, primary analyses, and sensitivity analyses where applicable. Due to deviations from normality and significant autocorrelation in the raw data, Δlog transformations were employed for lead–lag tests, with inferences based on HAC/Newey–West standard errors.

For RQ1, descriptive statistics showed long-tailed, right-skewed distributions for cases and search terms. Assumption checks supported differencing and HAC inference. Spearman correlations demonstrated moderate associations in the same week for selected terms, and CCFs peaked at one week for all pre-specified search series. Time-lagged Δlog regressions confirmed that prior-week search growth predicted subsequent case growth. Positive, statistically significant elasticities clustered around 0.7–1.3 across key terms, indicating practical but not definitive explanatory power, as evidenced by low-to-moderate R^2. Sensitivity analyses across adjacent lags and an alternative Δ-level specification revealed similar results.

For RQ2, the vaccinations series showed spikes aligned with rollout and booster phases, as well as strong persistence, when examined using Google Trends searches. After applying the same assumptions, Spearman's ρ showed positive associations in the same week for several vaccination-relevant terms. The CCFs indicated a lead time of approximately one week, and the Δlog regressions using search_{t-1} returned positive, statistically significant elasticities for the combined and brand-specific predictors, with modest R^2 and stable results in sensitivity analyses with adjacent lags and Δ-level checks.

For RQ3, contemporaneous Spearman correlations were nearly zero. CCFs were generally flat within ±8 weeks, with no spike surpassing Bartlett bounds. In time-lagged Δlog regressions, Model A (cases → vaccinations) showed small, non-significant elasticities at lags 0–2. Meanwhile, Model B (vaccinations → cases) was null at lag 0. An exploratory estimate at $k = 7$ was positive, but it was sensitive to omitted influences and not confirmed at adjacent lags. Sensitivity analyses supported the conclusion that short-term directional relationships between vaccination and case growth were not robust during the study period.

To verify the effectiveness of the negative control, the influenza-related searches were examined. Same-week Spearman's ρ was positive with cases and negative with vaccinations, which is consistent with seasonal respiratory activity and nonoverlapping vaccination phases. CCFs on Δlog showed a small, significant increase at +1 to +2 weeks for influenza cases and no pattern for influenza vaccinations. Δlog regressions confirmed these results, which were moderate and significant for cases at +1 and +2 weeks, and null for vaccinations. These findings support the construct specificity of the reported RQ1 and RQ2 search signals targeting the COVID-19 pandemic.

In summary, Chapter 4 shows that, when queries are concept-specific, Google Trends can provide a lead of approximately one week for both cases and vaccinations. However, general respiratory attention does not reproduce these relationships with vaccinations and only weakly co-moves with cases, likely reflecting shared seasonal drivers. These results lay the groundwork for the discussion in Chapter 5 regarding interpretation, public health implications, limitations, and directions for future work.

Chapter 5: Discussion, Conclusions, and Recommendations

Introduction

The objective of this chapter is to interpret and contextualize the empirical results presented in Chapter 4. Using the theoretical foundations based on Everett Rogers' Diffusion of Innovations and prior empirical evidence summarized in Chapter 2, the discussion centers on the implications of the observed short leads from concept-specific Google Trends searches to cases and vaccinations related to COVID-19 in practical, methodological, and theoretical terms (Rogers, 2003; Mavragani & Ochoa, 2019; Porcu et al., 2023). The chapter differentiates between statistical and practical significance, specifies the conditions under which the signals are insightful, and clarifies where conclusions are not justified. The chapter then outlines limitations resulting from design, measurement, and context, and translates the findings into practical recommendations for future research and public health surveillance in Austria and similar settings. Finally, this chapter details broader implications, including methodological contributions, refinements to the diffusion-of-innovations perspective, and considerations for positive social change. It concludes with the study's main contributions and key takeaways. All interpretations are deliberately limited by the scope of the study, the data used, and the analytic choices made to maintain scientific and ethical integrity.

Purpose, Design, Data Sources, and Timeframe

This quantitative, longitudinal, correlational study examined whether weekly Google Trends search activity for terms related to COVID-19 in Austria could provide an early signal of subsequent case growth and vaccination uptake. Additionally, the study investigated whether meaningful short-term dynamics exist between cases and vaccinations. The primary independent variable was the weekly Google Trends relative search volume for the specified terms. The outcomes were the weekly number of confirmed cases and vaccine doses administered, aggregated to ISO weeks at the national level. The analyses followed the plan detailed in Chapter 3 and included descriptive statistics, assumption checks, and nonparametric associations where appropriate. Cross-correlation functions were used on the Δlog series to identify the lead/lag structure, and time-lagged Δlog regressions with HAC/Newey–West inference were performed. The study windows were March 1, 2020–June 30, 2023, for cases ($N = 175$ weeks, $N = 174$ in the Δlog series) and December 27, 2020–June 30, 2023, for vaccinations ($N = 132$ weeks, $N = 131$ in the Δlog series). An influenza search index was analyzed as a negative control to verify specificity.

Research Questions and Summary of Principal Findings

For RQ1, weekly Google Trends activity, particularly symptom-specific searches and a combined index, led case growth by approximately one epidemiological week. Time-lagged Δlog regressions at $k = +1$ yielded positive, statistically significant elasticities. These elasticities ranged from ~0.7–1.3, and they had low-to-moderate explanatory power. The results were directionally consistent across adjacent lags. They were also consistent across alternative Δ-level checks.

For RQ2, brand-specific and vaccination-relevant searches showed synchronous or short-lead associations. CCF peaks occurred at 0 to +1 weeks, and confirmatory Δlog regressions showed positive, statistically significant effects with moderate R^2 values.

Sensitivity checks confirmed the stability of the one-week lead pattern for brand-level queries.

With regard to RQ3, contemporaneous monotonic association was nearly zero, and CCFs were flat within ±8 weeks. Lagged regressions showed no robust short-term directional relationship in either direction. An isolated exploratory estimate at $k = +7$ (vaccinations → cases) was small and sensitive to outlier handling. It was also not corroborated by adjacent lags or the CCF screen. Therefore, the null hypothesis was retained for contemporaneous and short-term relationships.

Specificity was tested in the influenza-negative control analysis. Influenza searches showed moderate, epidemiologically reasonable co-movement with cases at one to two weeks and no short-term association with vaccinations. Magnitudes and explanatory power were significantly smaller than those of searches specific to COVID-19, which supports the construct specificity of the main findings.

Organization of the Chapter

The rest of Chapter 5 interprets the findings in relation to the literature in Chapter 2 and the study's theoretical framework; details the limitations of the study's design and execution, as well as their implications for inference; offers recommendations for future research and practice grounded in the strengths and limitations of the present study; and discusses implications including methodological, theoretical, practical, and positive social change bounded by the study's scope. Finally, the chapter concludes by summarizing the study's core contributions and key takeaways.

Interpretation of the Findings

This section interprets the empirical results from Chapter 4 in relation to the theoretical framework of the study and the literature summarized in Chapter 2. Rather than restating numerical estimates, the aim is to explain the meaning of the observed patterns, particularly the short lead time between concept-specific Google Trends searches and subsequent growth in both cases and vaccinations, as well as the largely non-existent short-term coupling between cases and vaccinations, within the Austrian context. The discussion distinguishes statistical from practical significance, addresses estimation uncertainty and model scope, and remains within the constraints of the design's ecological level and archival data.

Integrated Interpretation Across Research Questions

Overall, the results suggest that concept-specific Google Trends searches provide a brief but operationally significant lead for both the growth of COVID-19 cases and vaccinations in Austria. However, cases and vaccinations themselves demonstrate minimal to no stable short-term coupling. In RQ1 and RQ2, cross-correlation functions on Δlog series showed primary peaks at approximately one week, and time-lagged Δlog regressions revealed positive, statistically significant elasticities at that time frame for cases and several vaccination-related predictors, especially brand-specific queries, with low-to-moderate explanatory power. These patterns were directionally consistent across adjacent lags and under Δ-level sensitivity checks, suggesting that the +1-week signal is not an artifact of a single specification.

Essentially, the cross-RQ view is consistent. When the public seeks information about symptoms or vaccines, that interest tends to precede short-term changes in related epidemiological or behavioral indicators. For cases, symptom-oriented searches may indicate growing community concern or symptom onset, as reflected in surveillance data following delays in clinical testing and reporting. For vaccinations, brand-specific searches likely capture the formation of intention and the search for appointments, which translates to increased activity the following week. The empirical values suggest the practical relevance of short-term awareness and operational planning. However, the moderate R^2 values suggest that Google Trends should not be used as a standalone forecasting system.

In comparison, RQ3 shows that there is essentially no short-term relationship between cases and vaccinations. The same-week monotonic association was nearly zero, and the lead-lag profiles were flat within ±8 weeks in both directions. Lagged growth-on-growth regressions did not support robust directional effects at short time horizons. An exploratory estimate of vaccinations predicting cases around seven weeks was unreliable across adjacent lags, sensitive to omitted influences, and inconsistent with the cross-correlation screen. These features are more consistent with shared external drivers than with a meaningful causal relationship. This pattern aligns with the documented phase offsets (episodic vaccination campaigns versus wave-driven case surges) and highlights that program output and epidemic intensity did not move together over the short term during the study period.

Specificity analyses using influenza-related searches reinforce this integrated interpretation. Attention to influenza showed only moderate, epidemiologically reasonable synchronization with the number of cases of COVID-19 at +1 to +2 weeks, and there was no short-term association with vaccinations. The effects were smaller in size and explanatory power than those observed for the specific search term sets related to the COVID-19 topic. These results support the claim that the Google Trends signals detected for RQ1 and RQ2 are specific to the constructs rather than generic reflections of seasonal trends in respiratory health or media cycles, as described by Mavragani & Ochoa (2019) and Porcu et al. (2023).

Methodologically, cross-RQ alignment is based on a shared analytical foundation, which is Δlog transformation to stabilize variance and reduce persistence, primary lag selection via cross-correlation screening, and HAC/Newey–West inference to prevent short-memory autocorrelation and heteroscedasticity. The use of influence diagnostics and sensitivity analyses, such as adjacent lags, alternative variance estimators, Δ-level checks, and the omission of moderately influential weeks, further enhances confidence that the primary inferences are not driven by a limited number of observations or by a particular modeling choice.

Theoretically, the integrated findings align with the initial “knowledge/ persuasion” phases of the diffusion-of-innovations model by Rogers (2003) discussed in Chapter 2. Searching for information about disease symptoms or vaccines precedes behavior and case identification. However, this interest does not directly impact the spread of epidemics or program effectiveness in the short term. In other words, the observed search behavior serves as an indicator of information-seeking rather than a direct influence on epidemic dynamics.

In practice, the cross-RQ analysis indicates that Google Trends has a limited yet valuable role in Austrian surveillance workflows. Symptom-specific indices can provide a one-week “heads-up” for situational awareness and outbreak preparedness. For vaccinations, brand-term growth can inform short-term appointment capacity, outreach cadence, and inventory logistics. Booster-related interest can provide a longer planning horizon. However, the moderate R^2 across models, the absence of short-term case-to-vaccination relationships, and the observational design suggest treating Google Trends as a supplementary signal, alongside administrative and clinical data, rather than as a replacement (Mavragani & Ochoa, 2019; Porcu et al., 2023).

Ultimately, the cross-RQ interpretation establishes clear boundaries for inference. The results address short-term timing at the national, weekly, and aggregate levels under Austrian policy and reporting regulations from 2020 to 2023. However, they do not establish causality, guarantee performance at subnational scales or in different phases, or imply that search interest alone can explain most of the variation in cases or vaccinations.

Interpretation for Research Question 1: Google Trends and Case Incidence

This section interprets the results of RQ1 by explaining the meaning of the approximately one-week lag observed between concept-specific Google Trends searches and subsequent case growth in terms of substance, methodology, and theory for the Austrian context. The analysis in Chapter 4 established this pattern. This included Δlog transformations, cross-correlation screens to identify appropriate lags, and time-lagged

Δlog regressions with HAC/Newey–West inference. The analysis found statistically significant, positive elasticities of practical magnitude, which were found at the +1-week timeframe. These were found for both symptom-oriented and composite Google Trends indices.

Direction, Lag Structure, and Effect Magnitude

The association with subsequent case growth was positive and consistent across all pre-specified Google Trends predictors. Cross-correlation functions demonstrated that the maximum association for each predictor occurred when searches led cases by one week (lag +1). All peak coefficients exceeded the approximate 95% confidence intervals, which indicates that increases in information-seeking behavior systematically occurred before increases in reported infections. Negative-lag spikes (cases leading searches) did not surpass significance testing thresholds, which reinforced the assumption of temporal ordering in the model. These patterns justified the specification of time-lagged regressions, with searches at week *t–1* predicting case growth at week *t*.

The effects were both statistically and practically meaningful, as indicated by the estimates of their elasticities in the Δlog specification. In the models with a lag of 1, all five predictors showed positive coefficients. For instance, symptom aggregation (β_1 = 1.196, 95% CI [0.869, 1.523]) and general searches (β_1 = 1.305, 95% CI [0.853, 1.758]) showed that a 10% increase in search activity the prior week was associated with a 12–13% increase in cases the following week. Fever-related searches showed a similarly substantial elasticity (β_1 = 1.333, 95% CI [0.637, 2.030]). The combined index was smaller yet still significant (β_1 = 1.132, 95% CI [0.572, 1.691]). Testing-related searches were more moderate but still significant ($\beta_1 = 0.722$, 95% CI [0.143, 1.301]). The model fit was strongest for the symptom aggregate ($R^2 = .416$) and general search models ($R^2 = .322$), which indicates that the single lagged search series explained a meaningful proportion of the variance in weekly case growth. All coefficients and confidence intervals were estimated using OLS. The standard errors were HAC (Newey–West), which is consistent with diagnostics. The diagnostics indicate mild short-lag autocorrelation and occasional heteroscedasticity.

Sensitivity analyses supported the lag structure that the CCFs suggested. Re-estimating the models at lags 0 and +2 resulted in positive effects across predictors. Coefficients were generally smaller at lag 0 and comparable or slightly smaller at lag +2, reflecting the decline of CCF intensities beyond lag 1. The only exception was fever at lag 0 ($p = .049$), which remained directionally consistent nonetheless as it also had a smaller effect size. These results further confirm that lag +1 most accurately predicts subsequent case growth.

Practical Significance for Surveillance and Preparedness

From an operational perspective, the main contribution of the RQ1 findings is the actionable lead time. Concept-specific Google Trends series, especially those focused on symptoms and general searches, provided a consistent one-week advance signal of increased cases of COVID-19. In the $\Delta\log(+1)$ models, the elasticities were close to one (e.g., symptoms, $\beta_1 \approx 1.20$; general searches, $\beta_1 \approx 1.31$; fever, $\beta_1 \approx 1.33$; combined index, $\beta_1 \approx 1.13$; and testing, $\beta_1 \approx 0.72$). This indicates that a 10% increase in search activity the prior week was associated with a 7%-13% increase in cases the following week. The single-predictor model fit reached approximately .32-.42 for the strongest signals. These values are practically meaningful for short-term situational awareness, but they don't meet the requirements for a standalone forecasting system.

Timeliness is key in surveillance, and it is essential to be prompt in all aspects of surveillance. The one-week lead time emerged in the cross-correlation screens and was confirmed in the lagged regressions. In contrast, negative lag spikes (cases preceding searches) remained within the bounds of statistical significance, which supports the temporal ordering that informs preparedness use. Beyond one week, cross-correlation signals declined and fell within the confidence interval by about three weeks, highlighting the short-term nature of Google Trends' value, which was also reported by Mavragani & Ochoa (2019) and Porcu et al. (2023). In other words, these signals are best suited for making near-term decisions, such as staffing holds, message sequencing, and testing logistics, rather than long-term planning.

Decisions about preparedness are made in situations where there is uncertainty, so it is important to distinguish between signal and noise. Residual diagnostics revealed only minimal residual autocorrelation and heteroscedasticity. Therefore, inference based on HAC/Newey–West standard errors was appropriate. Influence checks showed no singular weeks driving the results. These diagnostics increase confidence as they show that the observed search-lead effects are not artifacts of model error or outliers. Therefore, they can be used as a complementary input along with official indicators.

In practice, health agencies can integrate Google Trends by using them as tiered triggers within existing monitoring workflows. This won't result in overcommitment of resources. For instance, they can screen weekly changes in symptom-specific and combined search indices for unusually large growth. Weeks that exceed predefined thresholds can then prompt low-impact escalations, such as preparing messaging templates, verifying testing access information, or placing provisional staffing holds. When multiple series simultaneously signal at t–1, near-term actions, such as sequencing risk communication one week earlier or checking for supply and capacity bottlenecks, can be prioritized. Sensitivity analyses further support emphasizing near-term readiness.

At the same time, the moderate R^2 values and ecological design establish clear boundaries. Google Trends should be considered an additional signal that is integrated with clinical reporting, testing output, and syndromic indicators, rather than a replacement (Mavragani & Ochoa, 2019; Greenhalgh et al., 2004; Talic et al., 2021). Positioning Google Trends as a short-term, construct-specific, complementary indicator acknowledges the limitations of observational data and aligns with the theoretical framework employed in this study.

In conclusion, the documented one-week lag from Google Trends to case growth provides operationally useful early awareness. This bellwether offers low consequences when implemented with calibrated thresholds, confirmed across multiple data streams, and communicated transparently. This aligns well with the timing of public health decision-making, where a small, reliable advance signal can lead to significant improvements in readiness and response.

Convergence and Divergence With Prior Literature

Overall, the findings of RQ1 align with previous studies showing that concept-specific Google Trends searches predict reported incidence over short time periods (Mavragani & Ochoa, 2019; Porcu et al., 2023). The 1-week lead identified here is consistent with studies, which interpret search behavior as pre-diagnostic attention (Mavragani & Ochoa, 2019). This attention surfaces before cases are captured in surveillance (Mavragani & Ochoa, 2019). The analytic approach aligns with best-practice guidance, using pre-specified terms, variance-stabilizing transforms, explicit lag screening, and robust inference (Mavragani & Ochoa, 2019; Porcu et al., 2023).

At the same time, this study validates some stronger claims that were occasionally made in early infodemiology work. Although effect sizes are practically meaningful, they are paired with low-to-moderate explanatory power. This reinforces the role of Google Trends as a complementary rather than a standalone indicator (Mavragani & Ochoa, 2019). The Austrian setting clarifies boundary conditions. From 2020 to 2023, the dominant signal was a stable one-week lead at weekly national levels. This is more conservative than some early-phase estimates, but more consistent across phases. Additionally, the influenza negative-control analyses addressed concerns about generic respiratory seasonality by demonstrating smaller, less significant associations than those for terms specific to COVID-19, thereby supporting construct specificity (Mavragani & Ochoa, 2019; Porcu et al., 2023). Overall, the results confirm the short-term predictive value of Google Trends and manage expectations about its accuracy.

Theoretical Implications Within the Diffusion of Innovations Framework

The approximately one-week lag between concept-specific Google Trends searches and resulting case growth aligns most closely with the knowledge and persuasion stages of the innovation-decision process outlined by Rogers (2003) which is described in Chapter 2. At the population level, increasing search activity about symptoms is a plausible and observable sign of early awareness and evaluation that happens before clinical testing and reporting. Therefore, Google Trends peaks before surveillance detects more cases. In this sense, digital searches serve as mass-mediated signals that promote awareness and attitude formation without actually causing epidemiological change.

The way the framework focuses on communication and new ideas makes it clear why information about symptoms is so helpful. Digital platforms provide immediate access to a large audience, accelerating the diffusion of information that is seen as beneficial, compatible, and straightforward (Rogers, 2003; Greenhalgh et al., 2004). However, the translation of information into reportable outcomes remains subject to access, guidance, and system capacity. This discrepancy helps explain the combination of meaningful elasticities with moderate explanatory power. Google Trends captures movement in early-stage cognitive processes, whereas case incidence reflects additional logistical and institutional constraints.

The heterogeneity of adopters further provides context for the results. Innovators and early adopters, who are often opinion leaders, are likely to search first, pulling the adoption curve forward (Rogers, 2003). As broader segments engage, information diffusion approaches an S-curve (Rogers, 2003). However, conversion to clinical outcomes follows different pathways, such as symptom onset, testing, and reporting. Therefore, the Google Trends-to-cases link is best interpreted as an indicator of population learning at a given stage rather than as a direct, immediate link to outcomes across all adopter groups (Rogers, 2003).

Overall, the Austrian setting highlights the importance of context sensitivity in diffusion. The one-week lead is a characteristic of the national communication and reporting system from 2020 to 2023, not a universal truth (Rogers, 2003). Within this framework, Google Trends should be treated as a complementary, stage-specific signal to inform initial messaging, reduce testing barriers, and improve short-term preparedness, while avoiding causal claims or inferences about individuals that go beyond the study's ecological design and data.

Conflicting Explanations and Contextual Moderators

It is important to note that there are several alternative mechanisms that could generate or amplify the observed one-week Google Trends-to-cases lead. Administrative timing artifacts, including reporting delays, backlogs, and evolutions in definitions, have the capacity to generate distortions in the measurement of weekly case growth. Simultaneously, policy announcements and media coverage can act as common factors that lead to an increase in search interest and subsequent testing or health care service utilization without implying a direct information-seeking-to-incidence pathway (Mavragani & Ochoa, 2019; Porcu et al., 2023; Rogers, 2003). Weekly aggregation reduces day-to-day noise but cannot remove these confounds entirely. Therefore, interpretations are limited to short-term association rather than causation (Mavragani & Ochoa, 2019; Porcu et al., 2023).

The context also functions as a moderator, regulating the signal strength and stability. At the national level, subnational heterogeneity in policies, access, and epidemic timing is averaged out (Mavragani & Ochoa, 2019; Porcu et al., 2023; Talic et al., 2021). Representativeness is further affected by internet use and linguistic preferences (Mavragani & Ochoa, 2019; Porcu et al., 2023). These factors shape the individuals captured by search-based indicators (Mavragani & Ochoa, 2019; Porcu et al., 2023). The diffusion theory also underscores the notion that communication structures and social-system attributes influence the rate at which awareness translates into observable outcomes (Rogers, 2003). This highlights that the Austrian one-week lead reflects properties of this setting during 2020–2023 rather than being a universal constant.

Two design features help resolve the ambiguity surrounding these competing explanations, though they do not entirely eliminate it. Initially, variance-stabilizing Δlog modeling, explicit lag screening, and HAC/Newey–West inference are employed to eliminate artificial lead-lag patterns resulting from persistence and short-memory autocorrelation. This approach aligns with recommended practices for digital-trace surveillance as outlined in the research by Mavragani and Ochoa (2019). Secondly, the influenza negative-control analyses demonstrated materially smaller and less explanatory associations in comparison to those observed for terms specific to COVID-19. This finding suggests that construct specificity may take precedence over generic seasonality or media effects (Porcu et al., 2023). However, residual confounding from shared seasonal or policy drivers remains a plausible concern. As a result, Google Trends is positioned as a complementary, short-term indicator alongside administrative and clinical data, not as a replacement (Mavragani & Ochoa, 2019).

Summary of Interpretation for Research Question 1

The evidence from RQ1 suggests that concept-specific Google Trends indices, particularly those oriented toward symptoms and composite measures, appear to provide a lead time of approximately one week for the subsequent increase in cases. The elasticities are positive and practically meaningful, and the pattern is stable across adjacent lags and Δ-level checks, though the explanatory power remains modest. Generally, the observed timing is consistent with the knowledge and persuasion stages of the innovation-decision process (Rogers, 2003). Search activity is indicative of early awareness and consideration, which occur before clinical testing and reporting (Rogers, 2003). Therefore, these activities do not directly modify epidemic dynamics.

From a methodological perspective, the utilization of Δlog transformations, cross-correlation–guided lag selection, and HAC/Newey–West inference aligns with established guidelines for digital-trace surveillance (Mavragani & Ochoa, 2019). As demonstrated by Porcu et al. (2023), negative-control analyses with influenza terms yielded smaller, less explanatory associations than those from searches related to COVID-19. These findings provide support for the construct specificity of the phenomenon under investigation, suggesting that it is not driven by generic seasonality or media effects (Porcu et al. , 2003). Consequently, Google Trends should be used as a complementary indicator with a short-term predictive ability, rather than as a standalone forecasting system (Mavragani & Ochoa, 2019; Greenhalgh et al., 2004).

Interpretation for Research Question 2: Google Trends and Vaccination Uptake

This subsection provides an interpretation of RQ2 by describing the observed synchronous to short-lead associations between vaccination-relevant Google Trends searches, particularly brand-specific terms, and subsequent vaccination uptake, with a focus on the Austrian context. In the context of the knowledge and persuasion stages of the diffusion-of-innovations model, an increase in search activity is interpreted as an early indicator of awareness and intention formation that can occur before near-term adoption when access and eligibility allow (Rogers, 2003). In line with the established principles of integrating digital traces into surveillance methods, the empirical elasticities demonstrate practical significance, though they are accompanied by low-to-moderate explanatory power (Mavragani & Ochoa, 2019; Porcu et al., 2023; Talic et al., 2021). This finding reinforces the idea that Google Trends should not be regarded as a standalone element in operational planning but rather as a complementary component to existing methodologies (Mavragani & Ochoa, 2019; Porcu et al., 2023). When conceptualized in this manner, Google Trends can serve as a guide to the frequency of outreach and the execution of short-term capacity assessments — while remaining constrained by the practical constraints highlighted in the health innovation literature (Greenhalgh et al., 2004).

Direction, Lag Structure, and Effect Magnitude

The findings indicated that vaccination-relevant Google Trends activity demonstrated synchronous short-lead relationships with weekly vaccination rates. The analysis of cross-correlation functions showed primary peaks at 0 to +1 weeks, suggesting that increases in brand-specific and closely related search terms occurred in temporal alignment with or prior to approximately one-week increases in administered doses. The negative-lag profiles, which correspond to vaccinations leading searches, did not exceed the significance thresholds, thereby supporting the temporal ordering used in the subsequent modeling. These patterns motivated specifications in which Google Trends at week $t-1$ predicts vaccination growth at week t and, where appropriate, the inclusion of a synchronous ($k = 0$) term.

Time-lagged Δlog regressions were utilized to validate these screens, thereby confirming that coefficients at $k = +1$ for brand-specific predictors were positive and statistically significant, with effect sizes that are practically meaningful for short-term planning. The elasticities were found to be smaller but still positive for broader vaccination search terms and composite indices. This finding is consistent with the hypothesis that focused searches are more indicative of proximal intention than generic interest. The model fit remained in the low-to-moderate range, indicating that while Google Trends provides useful information about near-term movement, it explains only a portion of the week-to-week variation in vaccinations.

Stability checks confirmed the lag structure and magnitudes. Estimates at adjacent lags (0 and +2) remained directionally positive but were generally weaker than at +1, consistent with the observed decay in the cross-correlation profile beyond the primary peak. The findings were robust to Δ-level specifications and standard error corrections, and there was no evidence that a small number of influential weeks drove the results. The direction, lag structure, and effect sizes collectively support the interpretation of vaccination-relevant Google Trends series, particularly brand-specific searches, as short-term signals suitable for operational awareness rather than as standalone forecasting tools.

Behavioral Translation Pathway

According to the diffusion of innovations perspective, the search activity that is relevant to vaccination serves as an early indicator of the knowledge and persuasion stages (Rogers, 2003). Individuals initially become aware of vaccine options, which are often specific to a particular brand. They then seek clarifying information and form preliminary attitudes before progressing to the decision and implementation stages (Rogers, 2003). In this particular framing, the synchronous to +1-week associations reflect the time required to convert awareness and intention into concrete actions, such as locating an appointment and receiving the actual vaccination. These actions are dependent on access and eligibility, which are contingent on various factors. As with RQ1, search interest should therefore be regarded as an indicator of population learning rather than a direct driver of uptake (Mavragani & Ochoa, 2019).

The translation of intention into vaccination is dependent on the implementation environment (Greenhalgh et al., 2004; Mavragani & Ochoa, 2019). The phases of eligibility, the availability of appointments, the clarity of guidance, and the capacity of the health system can either accelerate or decelerate the transition from persuasion to decision/implementation, thereby generating the moderate explanatory power observed even when elasticities are practically meaningful (Greenhalgh et al., 2004; Mavragani & Ochoa, 2019). Brand-specific searches are informative because they occur during the intention formation stage, when users are more likely to have a solid decision (Greenhalgh et al., 2004; Rogers, 2003). However, the effectiveness of these searches is still influenced by external factors (Greenhalgh et al., 2004; Rogers, 2003).

Alignment With Vaccine Acceptance and Hesitancy Literature

The associations between brand-specific Google Trends activity and vaccination uptake, measured in weeks, are consistent with the knowledge and persuasion stages in diffusion theory (Rogers, 2003). According to this theory, the seeking of information and the development of attitudes precede the decision and implementation of a health behavior (Rogers, 2003). In this perspective, brand-term searches are considered proximate indicators of intention, occurring closer to the decision point than generic interest. Therefore, moderate positive elasticities over brief periods are consistent with a process in which awareness consolidates into action when eligibility and access permit (Rogers, 2003).

At the same time, the low-to-moderate explanatory power that was observed reflects well-documented implementation realities that have been emphasized in the health-innovation literature (Greenhalgh et al., 2004). These include the fact that translation from intention to uptake is conditional on communication quality, perceived attributes of the innovation, organizational capacity, and system context (Greenhalgh et al., 2004). Consequently, even when Google Trends signals indicate rising intention, utilization remains sensitive to bottlenecks in booking pathways, guidance clarity, and delivery capacity, which are factors that diffusion and implementation frameworks identify as common determinants of variable adoption (Greenhalgh et al., 2004; Rogers, 2003).

In the context of digital epidemiology, previous research has indicated that search-based indicators should serve a complementary role to administrative data when interpreting health behaviors (Mavragani & Ochoa, 2019; Porcu et al., 2023). This perspective aligns with the findings of this study, which demonstrate the existence of practically meaningful but limited associations. When considered as a whole, the results align with the principles of acceptance and hesitancy theory by conceptualizing Google Trends as an early indicator of vaccination intention that can facilitate outreach and reduce hesitancy (Greenhalgh et al., 2004; Rogers, 2003; Mavragani & Ochoa, 2019; Porcu et al., 2023). However, it is crucial to acknowledge that structural and organizational factors ultimately influence the translation of intention into action, shaping whether vaccination rates are achieved (Mavragani & Ochoa, 2019; Porcu et al., 2023).

Theoretical Implications for Awareness and Decision Stages

The synchronous associations between vaccination-relevant Google Trends activity and output map onto the awareness–persuasion–decision sequence in the innovation decision process (Rogers, 2003). Brand-specific searches indicate that potential adopters have progressed beyond general awareness toward evaluation and intention (Rogers, 2003). In this theoretical framework, Google Trends functions as a stage-specific indicator of movement through early phases rather than a causal driver of behavior (Mavragani & Ochoa, 2019). This provides a rationale for the observation that associations are short-term and practically meaningful but not indicative of weekly uptake (Mavragani & Ochoa, 2019).

The moderate degree of explanatory power that was observed, including in cases where elasticities are positive, corresponds with perspectives on implementation that underscore the role of contextual factors and constraints in translating intention into action (Greenhalgh et al., 2004). The transition from persuasion to decision/implementation is influenced by factors such as eligibility rules, booking processes, message clarity, and delivery capacity (Greenhalgh et al., 2004; Rogers, 2003). Consequently, Google Trends can serve as an indicator of readiness in the population, while actual vaccinations are influenced by organizational and systemic factors (Greenhalgh et al., 2004; Rogers, 2003).

Conceptually, the findings contribute to the operationalization of early diffusion stages with a national-scale digital trace and to the bounding of the expected timing of intention-to-action conversion in this setting. In practice, Google Trends is employed as a stage-aware cue to augment targeted messaging, streamline access pathways, and align short-term capacity with emerging demand (Rogers, 2003; Mavragani & Ochoa, 2019; Greenhalgh et al., 2004). These actions are consistent with the theory that emphasizes communication channels and perceived attributes of the innovation, while treating Google Trends as a complementary input rather than a substitute for administrative indicators (Rogers, 2003; Mavragani & Ochoa, 2019; Greenhalgh et al., 2004).

Attenuation Points, Contextual Constraints, and Practical Significance

Short-term associations between vaccination-relevant searches and outcomes are mitigated as intentions meeting implementation challenges. The transition from persuasion to decision/implementation is influenced by a variety of factors, including eligibility rules, booking and access barriers, message clarity, and delivery capacity (Greenhalgh et al., 2004; Rogers, 2003). Therefore, even when brand-term interest rises, the achievement of actual vaccinations is contingent on organizational and system conditions rather than information alone (Greenhalgh et al., 2004; Rogers, 2003). This phenomenon can be explained by the empirical combination of positive elasticities with low-to-moderate explanatory power (Greenhalgh et al., 2004). Google Trends has been employed to trace movement in early-stage cognition, while uptake is additionally shaped by logistics, guidance, and capacity constraints (Greenhalgh et al., 2004).

The strength of the signal is further limited by context. At the national weekly resolution, subnational heterogeneity in access, policy timing, and communication channels is averaged out (Mavragani & Ochoa, 2019; Porcu et al., 2023). Representativeness may also be affected by differential internet use and language choice, which shape who appears in search-based indicators (Mavragani & Ochoa, 2019). The diffusion theory highlights that such social system characteristics influence the rate at which awareness translates into action (Rogers, 2003). Consequently, the Austrian lead of approximately one week should be interpreted as a characteristic of this environment during the 2020–2023 period rather than as a universal constant (Rogers, 2003).

In practice, these constraints support the use of Google Trends as a stage-aware, short-term cue to facilitate the timing of outreach initiatives, streamline the booking process, and verify near-term capacity (Greenhalgh et al., 2004; Mavragani & Ochoa, 2019; Porcu et al., 2023; Talic et al, 2021). This approach involves triangulating with administrative and clinical data, while avoiding an overreliance on a single signal (Greenhalgh et al., 2004; Mavragani & Ochoa, 2019; Porcu et al., 2023; Talic et al, 2021). The study's findings, including the observation of smaller effects for influenza-related terms, support the interpretation that vaccination-relevant queries are best understood as construct-specific rather than merely seasonal or media-driven (Porcu et al., 2023). This interpretation strengthens the value of these searches for bounded operational use (Porcu et al., 2023).

Summary of Interpretation for Research Question 2

Overall, vaccination-relevant Google Trends activity, particularly brand-specific searches, exhibited synchronous to approximately one-week lead associations with weekly vaccination uptake. When interpreted through the diffusion of innovations lens, these signals function as stage indicators of movement from awareness to persuasion and toward decision, rather than as causal drivers of uptake (Rogers, 2003). The positive elasticities are practically meaningful for short-term planning (Mavragani & Ochoa, 2019; Porcu et al, 2023; Talic et al, 2021). However, the low-to-moderate explanatory power underscores that Google Trends should inform, but not determine, operational choices (Mavragani & Ochoa, 2019; Porcu et al, 2023; Talic et al, 2021).

Methodologically and substantively, the pattern aligns with a cautious integration of digital data into public health workflows. Specifically, brand-term searches are associated with the development of intentions, but the conversion of these intentions into actual vaccinations remains contingent on factors such as eligibility, the ease of booking, the clarity of messaging, and delivery capacity (Greenhalgh et al., 2004; Mavragani & Ochoa, 2019; Porcu et al., 2023). The application of specificity checks, incorporating smaller and less explanatory associations for influenza-related terms, supports the interpretation of the observed Google Trends-to-vaccination relationships as construct-specific rather than artifacts of generic seasonality or media cycles (Porcu et al., 2023). In summary, Google Trends is best positioned as a short-term, stage-aware complement to administrative indicators for synchronizing outreach cadence and near-term capacity checks (Rogers, 2003; Mavragani & Ochoa, 2019; Greenhalgh et al., 2004). However, Google Trends is subject to limitations related to context and the study's ecological design (Rogers, 2003; Mavragani & Ochoa, 2019; Greenhalgh et al., 2004).

Interpretation for Research Question 3: Short-Run Case–Vaccination Dynamics

This section provides an interpretation of the RQ3 results, which demonstrate a minimal, and unstable, short-term relationship between the growth of COVID-19 cases and the rate at which vaccines are administered. The emphasis is on interpreting the implications of the near-zero contemporaneous association, the flat lead–lag profiles within ±8 weeks, and the absence of robust directional effects for understanding program dynamics in Austria during 2020–2023. The absence of tight, short-horizon coupling is theoretically coherent when considered from the perspectives of diffusion and implementation (Rogers, 2003; Greenhalgh et al., 2004). Awareness and persuasion do not directly translate into program output, and vaccination throughput is influenced by eligibility phases, booking friction, guidance clarity, and delivery capacity (Rogers, 2003; Greenhalgh et al., 2004). Therefore, the evidence aligns more closely with distinct process milestones, like wave-driven incidence versus episodic campaign logistics, than with direct, short-term causal relationships.

Empirical Pattern and Substantive Meaning

The findings from RQ3 suggest minimal contemporaneous association and flat lead-lag profiles between the growth of cases of COVID-19 and the output of vaccinations across a period of ±8 weeks, with no significant directional effects. This pattern suggests the presence of distinct process cycles. Epidemic intensity is predominantly wave-driven, while program output is influenced by campaign cadence, eligibility windows, booking dynamics, and delivery capacity. Therefore, the absence of short-term relationships is theoretically consistent with diffusion and implementation perspectives, in which movement from awareness to action is guided by organizational and system factors rather than being closely aligned with epidemiological curves (Greenhalgh et al., 2004; Rogers, 2003).

These results also caution against attributing observed short-term fluctuations in output to simultaneous case dynamics. The observed variation in vaccinations from 2020 to 2023 is likely indicative of a multifaceted set of factors, including policy staging, such as the implementation of phased eligibility criteria, resource allocation dynamics, and communication cycles within the health care system (Dearing & Cox, 2018; Greenhalgh et al., 2004; Rogers, 2003). This interpretation aligns with the principles of implementation theory, which underscore the significance of contextual factors and organizational readiness in shaping health care practices (Dearing & Cox, 2018; Greenhalgh et al., 2004; Rogers, 2003). Additionally, it draws parallels with the findings of research in the field of diffusion studies, which emphasize the dynamics of adoption across social systems (Dearing & Cox, 2018; Greenhalgh et al., 2004; Rogers, 2003). Therefore, the unstable, isolated estimate at longer lags is more accurately interpreted as the result of shared external drivers rather than as evidence of direct causal pathways (Dearing & Cox, 2018; Greenhalgh et al., 2004; Rogers, 2003).

From a surveillance and planning perspective, the substantive meaning is straightforward. It is not reasonable to expect same-week or short-term correlation between cases and vaccinations, even during periods of heightened activity such as waves or vaccination campaigns. Instead, case indicators and vaccination signals should be regarded as complementary yet independent operational streams (Mavragani & Ochoa, 2019; Talic et al., 2021). Messaging and capacity checks should be aligned with intention cues, such as brand-term search interest, while acknowledging that throughput will be constrained by program design and access conditions (Mavragani & Ochoa, 2019; Talic et al., 2021). This positioning is consistent with the negative-control and specificity checks, which support a cautious, context-aware interpretation of digital and administrative signals rather than overly linking them in the short term (Porcu et al., 2023).

Plausible Mechanisms (Eligibility, Mandates, Supply, Reporting Lags)

A number of programmatic and policy mechanisms have been identified as having the potential to decouple short-term vaccination output from concurrent case dynamics. The establishment of eligibility criteria and mandates generates distinct decision-making periods. The introduction of new groups into the eligible population or the announcement of mandates can prompt a rapid escalation in intention. However, the actual uptake is subject to the scheduling and accessibility of resources in the following days and weeks. These impacts are primarily communicated and influenced through social networks, aligning with the diffusion theory's explanation of how policy directives influence adopters' transition from awareness and persuasion to decision-making (Rogers, 2003; Dearing & Cox, 2018; Talic et al., 2021). However, the translation of these impacts into observable outcomes is influenced by implementation constraints rather than epidemiological curves (Rogers, 2003; Dearing & Cox, 2018; Talic et al., 2021).

Supply and delivery capacity further influence the conversion of intention into action. Organizational readiness, process simplicity, and local resource availability have a significant impact on the conversion of heightened demand, signaled by brand-term interest, into completed vaccinations (Greenhalgh et al., 2004). However, bottlenecks in booking pathways, staffing, or site availability can attenuate or delay effects, even when intention is high (Greenhalgh et al., 2004). In this context, the modest explanatory power in combination with positive elasticities is a theoretically consistent proposition. Google

Trends captures early cognitive movement, while output remains constrained by system factors identified in implementation frameworks (Greenhalgh et al., 2004; Rogers, 2003).

Additionally, the presence of reporting lags and administrative artifacts has the potential to introduce ambiguity to short-term relationships. Batch uploads, weekend effects, and changes in definitions can lead to fluctuations in recorded doses across weeks (Mavragani & Ochoa, 2019; Porcu et al., 2023). Concurrent media cycles can generate increased search interest without immediate logistical action (Mavragani & Ochoa, 2019; Porcu et al., 2023). The specificity checks, which represent smaller, less explanatory associations for influenza-related terms, have been shown to reduce the likelihood that results merely reflect generic seasonality or media attention (Porcu et al., 2023). This suggests that interpretation of vaccination-relevant Google Trends signals as construct-specific within these operational constraints is valid (Porcu et al., 2023).

Implications for Modeling Behavior During Epidemics

The almost-absent contemporaneous association and the flat lead–lag profiles between cases and vaccinations suggest that short-term models should not assume a strong, direct relationship. When viewed on a weekly national scale, the incidence and program output appear to be influenced by distinct process dynamics, like wave dynamics and episodic campaign logistics. Therefore, specifications that demand contemporaneous feedback, such as cases leading to vaccinations within the same week, may lead to inaccurate specifications. A more robust approach involves the systematic structuring of distinct but interconnected subsystems through the incorporation of external policy and implementation inputs, such as eligibility phases, mandates, and changes in access (Rogers, 2003; Greenhalgh et al., 2004; Dearing & Cox, 2018). This approach involves the stratification of subsystems based on program phases, with allowance for structural breaks that are consistent with diffusion and implementation perspectives (Rogers, 2003; Greenhalgh et al., 2004; Dearing & Cox, 2018). These perspectives emphasize the role of communication channels, organizational capacity, and the social-system context (Rogers, 2003; Greenhalgh et al., 2004; Dearing & Cox, 2018).

Digital evidence should be used as a short-term, stage-aware indicator rather than as the sole driver of behavior. In practice, this involves the constrained use of Google Trends predictors within narrow lead windows (approximately 0–1 weeks), the application of variance-stabilizing transformations and robust inference methods, and the triangulation of results with administrative data (Mavragani & Ochoa, 2019; Porcu et al., 2023). Additionally, the use of negative controls, such as influenza-related terms, is employed to probe construct specificity (Mavragani & Ochoa, 2019; Porcu et al., 2023). Given the implementation challenges and the heterogeneity of policies, the evaluation process should prioritize decision-relevant, rolling, and near-term performance metrics (Greenhalgh et al., 2004; Rogers, 2003). This approach aligns with the notion that intention signals facilitate timing, but they do not guarantee uptake without the presence of conducive system conditions (Greenhalgh et al., 2004; Rogers, 2003).

Summary of Interpretation for Research Question 3

In summary, RQ3 suggests that there were minimal, unstable short-term relationships between case growth and vaccination output at the national level on a weekly basis during the period from 2020 to 2023. In general, this pattern is most consistent with distinct process dynamics, like wave-driven incidence versus episodic program logistics, rather than with a direct, short-term causal relationship. The diffusion and implementation perspectives offer a theoretical framework to understand this observation. The transition from awareness and persuasion to decision-making and implementation is influenced by factors such as eligibility staging, booking friction, guidance clarity, and delivery capacity (Rogers, 2003; Greenhalgh et al., 2004). Consequently, program outputs may not necessarily align with epidemiologic curves on a weekly basis (Rogers, 2003; Greenhalgh et al., 2004).

Methodologically, the findings suggest that models which assume contemporaneous feedback between cases and vaccinations over short horizons may be inaccurate. Instead, incidence and output should be conceptualized as distinct but interconnected elements, incorporating external policy and capacity inputs, as well as allowance for phase transitions (Mavragani & Ochoa, 2019; Porcu et al., 2023). Digital signal indicators should be used as near-term, stage-aware prompts that can be integrated with administrative data to enhance understanding (Mavragani & Ochoa, 2019; Porcu et al., 2023). Specificity checks defined as smaller, less explanatory associations for influenza-related terms support the interpretation that the observed pattern is construct-specific rather than an artifact of generic seasonality or media cycles (Porcu et al., 2023). Overall, these findings highlight the complementary nature of search-based signals and administrative indicators, aligning with theoretical frameworks and being constrained by contextual factors (Rogers, 2003; Greenhalgh et al., 2004).

Specificity and Robustness of Findings

This section's objective is to evaluate the construct-specificity and robustness of the detected Google Trends associations to reasonable analytic variations. The specificity of the terms is assessed using an influenza negative control. The smaller and less explanatory associations observed for these terms in comparison to those specific to COVID-19 support a targeted signal rather than generic respiratory seasonality (Porcu et al., 2023). Robustness is examined through a combination of variance-stabilizing Δlog transformations, cross-correlation–guided lag selection with adjacent-lag checks, and time-lagged regressions estimated with HAC/Newey–West inference (Mavragani & Ochoa, 2019; Talic et al., 2021). This approach is designed to address short-memory autocorrelation and heteroscedasticity, aligning with recommended practices in digital epidemiology as outlined by Mavragani and Ochoa (2019) and Porcu et al. (2023). As no explicit media-attention covariate was incorporated into the model, residual confounding from contemporaneous media events cannot be excluded (Mavragani & Ochoa, 2019; Porcu et al., 2023). Therefore, Google Trends is interpreted as a complementary, short-term indicator rather than a standalone signal (Mavragani & Ochoa, 2019; Talic et al., 2021; Porcu et al., 2023).

Evidence From Influenza Sensitivity Analyses

Using the identical preprocessing and modeling workflow as for the term sets related to COVID-19, the influenza negative-control analyses demonstrated only a minimal, epidemiologically plausible co-movement with the growth of cases of COVID-19 at short leads (approximately +1 to +2 weeks) and no short-term association with vaccinations. The effect sizes and explanatory power were found to be substantially smaller in comparison to those seen for searches related to COVID-19, suggesting that the most significant signals occur when searches directly refer to the specific construct under investigation (e.g., symptoms associated with the virus, brands of vaccines) rather than generic respiratory themes (Porcu et al., 2023; Mavragani & Ochoa, 2019).

These results align with construct specificity principles, suggesting that if broad seasonality or general health attention were the primary drivers, influenza terms would have generated associations similar to those observed for the terms related to COVID-19. However, given the lack of direct proxies for media volume in this study, residual confounding by contemporaneous media events cannot be excluded. Accordingly, the use of Google Trends indicators specific to COVID-19 is regarded as a complementary, though limited, contribution to surveillance efforts (Mavragani & Ochoa, 2019). This integration is most effectively achieved through the triangulation of administrative and clinical data, in addition to the application of negative controls (Porcu et al., 2023).

Model Diagnostics and Stability Across Specifications

The results of the diagnostics indicated that the model fit was appropriate and that stable inferences could be made under reasonable analytic variations. The implementation of Δlog transformations led to a reduction in persistence, while HAC/Newey–West inference addressed mild short-lag autocorrelation and heteroscedasticity. Additionally, influence checks were performed, which revealed that no single weeks were identified as drivers of the results. The primary effects were found to be stable under adjacent lags (0, +2), Δ-level specifications, and alternative variance estimators. In addition, the direction and significance patterns showed consistency across RQ1 and RQ2, and the null findings remained consistent in RQ3. Overall, these results align with established guidelines for digital-trace analyses and support the interpretation of the estimates as both robust and constrained (Mavragani & Ochoa, 2019). This provides informative insights over brief time frames, exhibiting a low-to-moderate explanatory capability (Mavragani & Ochoa, 2019).

Residual Uncertainty and Boundaries of Inference

Given the observational and ecological nature of the study, the detected associations cannot be interpreted causally. Unmeasured, time-varying factors, such as media attention, policy changes, or shifts in testing access, may influence both search behavior and outcomes. While negative-control analyses with influenza-related terms support the concept of construct specificity, they do not entirely eliminate the possibility of confounding variables, and administrative artifacts have the potential to distort week-to-week relationships. The modeling choices (Δlog transformation, short leads, HAC/Newey–West inference) were selected to reduce bias and improve inference (Mavragani & Ochoa, 2019; Porcu et al., 2023). However, the limited explanatory power suggests that Google Trends should remain a complementary signal rather than a standalone predictor (Mavragani & Ochoa, 2019; Porcu et al., 2023).

The external validity of the model is similarly limited. The findings refer to national, weekly aggregates in Austria during 2020 to 2023 and may not be applicable to subnational settings, different phases, or populations with distinct digital access and language use. In line with diffusion and implementation perspectives, Google Trends is most appropriately used as a stage-specific indicator of awareness and intention rather than as a proxy for individual behavior or a tool that can determine outcomes with certainty (Rogers, 2003; Greenhalgh et al., 2004; Mavragani & Ochoa, 2019). Therefore, its practical application should involve triangulation with administrative and clinical data, as well as careful consideration of context (Rogers, 2003; Greenhalgh et al., 2004; Mavragani & Ochoa, 2019).

Limitations of the Study

This section explores the limitations of the study's findings, considering the characteristics of search-based indicators, an observational ecological design, and a context-dependent implementation of innovations. The limitations of measurement can be attributed to the nature of Google Trends as a proxy for attention, which is influenced by platform normalization, term selection, language, and user coverage (Mavragani & Ochoa, 2019; Porcu et al., 2023). These factors may result in discrepancies between the underlying constructs and the proxies, despite careful consideration in design (Mavragani & Ochoa, 2019; Porcu et al., 2023). Design-related constraints emerge from working with national, weekly aggregates due to the influence of time-varying confounders and organizational factors on both attention and outcomes (Greenhalgh et al., 2004; Dearing & Cox, 2018). This limits causal interpretation and individual-level inference (Greenhalgh et al., 2004; Dearing & Cox, 2018). Despite the efficacy of time-series procedures in mitigating autocorrelation and associated risks, the potential impact of administrative artifacts and phase changes on short-term associations remains a concern (Mavragani & Ochoa, 2019). Finally, external validity is constrained by the Austrian setting, time period, and social-system features emphasized in the diffusion theory (Rogers, 2003). These characteristics suggest that timing and magnitudes are context-sensitive (Rogers, 2003). All these considerations suggest that Google Trends should be used as a complementary, short-term signal whose efficacy is contingent upon triangulation with administrative and clinical data as opposed to its use as a standalone predictor (Mavragani & Ochoa, 2019; Greenhalgh et al., 2004).

Measurement and Construct Validity

Google Trends provides a normalized relative search volume (RSV) rather than absolute counts. The RSV is reflective of the behavior of users with internet access who select specific search terms. As a result, the concept of "public attention" is evaluated using coverage and composition limits, with particular attention to sensitivity to term selection, linguistic nuances, and ambiguity in queries (Mavragani & Ochoa, 2019). While the pre-specifying of symptom and brand terms, in addition to the use of composite indices, enhances validity, fluctuations in search platform practices or user behavior can still influence RSV, independent from epidemiologic or programmatic shifts (Mavragani & Ochoa, 2019).

The influenza sensitivity checks supported construct specificity by demonstrating smaller and less explanatory associations compared to those observed for terms specific to COVID-19. However, negative controls are unable to fully eliminate all potential off-target influences (Porcu et al., 2023). For instance, the occurrence of concurrent respiratory seasons, the temporal alignment of news cycles, or the presence of crossover curiosity may persist in impacting the measured construct, even when average effects are reduced (Porcu et al., 2023).

The interpretation of search interest as movement through diffusion stages (knowledge/persuasion) is contingent upon the establishment of a link between population-level data and conceptual stages that are inherently individual and socially embedded. The diffusion theory cautions that communication channels, perceived attributes, and social-system context influence the allocation of attention toward evaluation and action (Rogers, 2003). Therefore, measurement error is expected, even in circumstances where directionality is plausible (Dearing & Cox, 2018).

Design and Statistical Conclusion Validity

The study analyzes national, weekly aggregate data and can therefore not support causal claims or individual-level inferences. Time-varying confounders, like policy announcements, access changes, or media attention, may co-influence search behavior and outcomes, producing associations that are substantively informative but not causal (Dearing & Cox, 2018; Greenhalgh et al., 2004). The literature on diffusion and implementation highlights the significance of contextual shocks in understanding adoption dynamics (Greenhalgh et al., 2004). These shocks are challenging to isolate in observational settings (Greenhalgh et al., 2004).

Additionally, it is important to acknowledge the potential for inflation of observed relationships due to autocorrelation, non-stationarity, and contemporaneous shocks, if these issues remain unaddressed. The study employed variance-stabilizing Δlog transformations, cross-correlation–guided lag selection, and HAC/Newey–West inference to mitigate these issues. However, residual dependence and model misspecification remain possible, and multiple comparisons across lags can raise Type I error risk if not interpreted conservatively (Mavragani & Ochoa, 2019). The models' limited explanatory power across different specifications suggests that they only capture a portion of the underlying process (Mavragani & Ochoa, 2019). This finding supports the notion of making cautious claims about their predictive ability (Mavragani & Ochoa, 2019).

Lastly, despite the implementation of negative controls, alternative pathways, including alterations in testing and reporting practices or eligibility phases, have the capacity to mimic or distort short-term relationships between attention and outcomes. The implementation frameworks that have been developed to anticipate such interactions between organizational capacity, guidance clarity, and user pathways complicate attribution in aggregate time series (Greenhalgh et al., 2004; Dearing & Cox, 2018).

External Validity and Transferability

The findings are applicable to Austria, at the national weekly level, during the period 2020 to 2023. The transfer of research findings to subnational settings or other countries requires caution because adoption processes, digital access, language use, and reporting mechanisms differ across social systems (Rogers, 2003; Dearing & Cox, 2018). The temporal dynamics of diffusion, and consequently the point at which attention transitions into behavior, can be influenced by communication infrastructures and institutional frameworks (Rogers, 2003). This variability limits the ability to make extrapolations (Rogers, 2003).

Additionally, the Google Trends data set offers insights into the behavior of internet users who are using particular languages or search terms. Variability in access and literacy can influence whose attention is observed, which can affect representativeness and potentially miss capturing marginalized populations (Mavragani & Ochoa, 2019). This uncertainty can lead to challenges when generalizing results to the entire population (Mavragani & Ochoa, 2019). Implementation research has also demonstrated that organizational and socio-technical contexts influence transferability, indicating that even consistent short-term patterns should undergo validation at the local level (Greenhalgh et al., 2004).

Finally, the study examines the impact of the pandemic across multiple phases, including waves, campaigns, and boosters. The diffusion theory emphasizes that stage, salience, and system readiness evolve, which can alter both effect sizes and lead times (Rogers, 2003). The approximately 1-week signals documented here should not be assumed to be constant across future variants or program configurations (Dearing & Cox, 2018).

Execution Constraints and Overall Impact on Inference

The availability of data and unmodeled covariates are additional factors that must be considered. The analysis used archival national series with fixed windows; however, direct proxies for media volume, mobility, or subnational access were not included (Mavragani & Ochoa, 2019; Porcu et al., 2023). This leaves potential residual confounding that the negative control cannot fully exclude (Mavragani & Ochoa, 2019; Porcu et al., 2023). Additionally, modifications in case definitions, reporting practices, or vaccine delivery logistics may have resulted in the introduction of artifacts that are only partially addressed by weekly aggregation (Mavragani & Ochoa, 2019).

The term lists were pre-specified, and the transformations and lags were selected using transparent criteria. However, reasonable alternatives, such as more comprehensive distributed-lag or structural models, could potentially reveal different nuances. The existing literature suggests the use of iterative validation when transitioning from exploratory signals to operational use (Greenhalgh et al., 2004). This recommendation emphasizes that the results of such validation processes should serve as a source of information, rather than as the sole determinant of decisions (Greenhalgh et al., 2004).

Taken together, these constraints suggest the importance of limited claims. The study provides evidence that concept-specific search interest offers short-term, construct-specific signals of case growth and vaccination intention. However, these signals are complementary to administrative and clinical indicators, rather than being substitutes for them (Mavragani & Ochoa, 2019; Porcu et al., 2023). The theoretical framework and prior diffusion/implementation work support the interpretation of Google Trends as a stage-aware indicator whose value is contingent on context, access, and system capacity (Rogers, 2003; Dearing & Cox, 2018; Greenhalgh et al., 2004).

Recommendations

The following recommendations are grounded in three principles that stand out from the study's findings. First, stage-aware integration of digital evidence, recognizing that search activity is an index of awareness and persuasion rather than completed behavior. Second, context-sensitive implementation, acknowledging that adoption is limited by eligibility, access, and organizational capacity. Third, methodological integrity and transparency, so that any operational use is validated prospectively and defined by evidence. In practical terms, this means that Google Trends should be used as a complementary, short-term indicator that can help plan outreach and capacity checks while avoiding causal assumptions and excessive reliance on a single signal (Mavragani & Ochoa, 2019; Porcu et al., 2023). It also implies aligning actions with diffusion and implementation guidance, using communication channels to consolidate intention, reducing delay in pathways to action, and tailoring strategies to the social-system context in which adoption happens (Rogers, 2003; Greenhalgh et al., 2004; Dearing & Cox, 2018). Finally, the recommendations emphasize three key principles: reproducibility, equity, and governance. These principles include pre-specifying queries, documenting decision rules, and monitoring for unintended consequences (Greenhalgh et al., 2004; Mavragani & Ochoa, 2019). By following these recommendations, the potential benefits of early awareness can be realized responsibly (Greenhalgh et al., 2004; Mavragani & Ochoa, 2019). Furthermore, these benefits can be generalized or adapted with confidence (Greenhalgh et al., 2004; Mavragani & Ochoa, 2019).

Recommendations for Future Research

To develop this body of knowledge further, it is essential to improve the methods of measurement. Extending research beyond national, language-specific aggregates to subnational levels with regionally tailored, multilingual query sets has the potential to reduce construct error tied to linguistic and access patterns (Mavragani & Ochoa, 2019; Rogers, 2003; Dearing & Cox, 2018). This approach could help to determine whether the observed timing generalizes across social systems with different communication infrastructures (Mavragani & Ochoa, 2019; Rogers, 2003; Dearing & Cox, 2018).

Methodological development should prioritize the use of prospective evaluation and more advanced time-series designs. Real-time assessments over short periods, in conjunction with distributed-lag models, can provide a quantitative estimation of operational lead time without the need for overfitting retrospective patterns. The implementation of negative controls, including both topic-adjacent and topic-distant controls, will contribute to the exploration of construct specificity under alternative specifications and seasonal conditions (Mavragani & Ochoa, 2019; Porcu et al., 2023).

In order to better understand the mechanisms and heterogeneity involved, future studies should consider stratifying their analyses by program phase, like eligibility expansions or booster campaigns, and by equity-relevant subgroups. Additionally, these studies should pair search signals with pathway indicators such as vaccination booking ease or guidance clarity. Such designs align with diffusion and implementation perspectives that position adoption as the result of the interaction of communication, perceived attributes, and organizational capacity (Rogers, 2003; Greenhalgh et al., 2004; Dearing & Cox, 2018).

In observational settings, addressing confounding more explicitly remains important. The incorporation of policy and eligibility series, access measures, and media-attention proxies has been found to facilitate the differentiation of commonly occurring shocks from signals of genuine intention (Dearing & Cox, 2018; Greenhalgh et al., 2004). Structured comparisons across policy changes, such as staggered rollouts, have been shown to enhance the temporal ordering without exaggerating the causal relationship (Greenhalgh et al., 2004).

Finally, the principles of transparency and reproducibility should be considered as standard practices. Prior to implementation, it is crucial to establish query lists and analytic plans, version code and data, and document decision rules for thresholds and escalation (Mavragani & Ochoa, 2019; Greenhalgh et al., 2004; Talic et al., 2021). Additionally, it is essential to establish guardrails for equity and communication risk (Smith et al., 2020). Such practices support responsible generalization and adaptation across contexts (Mavragani & Ochoa, 2019; Greenhalgh et al., 2004; Talic et al., 2021).

Recommendations for Practice and Policy

Google Trends can be used as a short-term, stage-aware indicator, but it should not be considered a standalone predictor. To operationalize this, it is necessary to define pre-specified thresholds for validated symptoms and vaccine-brand terms. These thresholds should be adjusted to account for the specific context of the situation. Additionally, the implementation of these thresholds should be preceded by an assessment of administrative indicators. Following this assessment, if the indicators suggest that action is required, the implementation of the thresholds should be executed. It is also recommended that each threshold be mapped to a limited set of responses that do not result in significant adverse consequences. This approach leverages Google Trends' alignment with awareness/persuasion stages while avoiding causal overreach (Rogers, 2003; Mavragani & Ochoa, 2019; Greenhalgh et al., 2004).

The transformation of intention signals into specific actionable strategies that result in the mitigation of friction is a critical component of this process. It is important that the booking pathways for the brand-terms are simplified and made clear, and that the eligibility criteria are made more accessible to the public. Additionally, the outreach schedule should be adjusted to align with the observed approximately 0–1-week window. In situations where policy stages or mandates are anticipated, it is essential to allocate time, communication, and capacity to these decision points. This ensures that intentions are seamlessly converted into uptake, aligning with diffusion and implementation perspectives that highlight the role of communication channels, perceived attributes, and organizational capacity (Dearing & Cox, 2018; Greenhalgh et al., 2004; Rogers, 2003).

The application of rigor, transparency, and governance in routine use is critical. To ensure the accuracy and precision of the data analysis, it is essential to pre-specify query lists, document trigger rules, disclose uncertainty ranges in dashboards, and maintain negative-control monitors to detect any deviations from the construct specificity (Mavragani & Ochoa, 2019; Porcu et al., 2023; Greenhalgh et al., 2004). It is also crucial to periodically re-validate thresholds, particularly in the aftermath of significant policy alterations (Mavragani & Ochoa, 2019; Porcu et al., 2023; Greenhalgh et al., 2004). Moreover, it is important to thoroughly document the decisions associated with signals to facilitate learning processes (Mavragani & Ochoa, 2019; Porcu et al., 2023; Greenhalgh et al., 2004). These measures are consistent with the established guidelines for digital-trace indicators, contributing to the maintenance of Google Trends as a complementary component to clinical and administrative data streams (Mavragani & Ochoa, 2019; Porcu et al., 2023; Greenhalgh et al., 2004).

The primary objective is to ensure that the implementation of these measures is equitable and promotes positive social change. It is critical to tailor queries and messaging to the language used in specific communities, prioritize outreach in regions with limited digital access, and closely monitor differential effects across various regions and populations (Rogers, 2003; Dearing & Cox, 2018; Greenhalgh et al., 2004). This focus on the social system's context enhances representativeness and fosters fair adoption, aligning with the findings of research on heterogeneous adoption groups and implementation guidance on organizational and contextual fit (Rogers, 2003; Dearing & Cox, 2018; Greenhalgh et al., 2004).

Implications

The implications of this study extend to various theoretical, methodological, and practical aspects. When interpreting search activity as a stage-specific indicator of population awareness and intention, rather than a determinant of epidemiologic outcomes, a new paradigm for how digital traces can inform public health action can be developed. The findings suggest that modest and transparent integration of Google Trends into short-term decision-making is appropriate, while acknowledging that adoption and uptake are contingent on communication channels, perceived attributes of innovations, and organizational capacity (Rogers, 2003; Greenhalgh et al., 2004; Mavragani & Ochoa, 2019).

Theoretical Implications

The observed short lead from concept-specific searches to cases and vaccinations supports a stage-based view of diffusion, positioning Google Trends as an indicator of movement through the knowledge and persuasion phases rather than as a driver in the implementation phase (Rogers, 2003). This provides insight into the operational usefulness of associations, which are limited in nature. Information seeking occurs prior to behavior, yet conversion to testing or vaccination is contingent on context and capacity.

Seen more broadly, the study contributes to the advancement of diffusion perspectives by operationalizing early-stage dynamics with a national-scale digital footprint and by emphasizing social-system contingency. The observed timing and magnitude of effects appear to reflect the Austrian communication and reporting structures during 2020 to 2023, suggesting that system features shape the pace at which awareness becomes action, as indicated by the diffusion framework (Dearing & Cox, 2018; Rogers, 2003).

Methodological Implications

From a methodological perspective, the findings support the use of digital traces with short-term, specification-aware applications. This approach involves the implementation of limited lead windows of approximately 0–1 weeks, the application of variance-stabilizing transformations, robust inference for short-memory autocorrelation, and the incorporation of negative controls to assess construct specificity (Mavragani & Ochoa, 2019; Porcu et al., 2023). Additionally, the findings also emphasize the significance of reporting practices, such as the use of pre-specified term lists, transparent lag selection, and sensitivity diagnostics, in facilitating reliable inferences with Google Trends data (Mavragani & Ochoa, 2019).

Practice and Policy Implications

According to Mavragani and Ochoa (2019), Google Trends should be used as a supplementary tool to synchronize outreach and capacity checks with emerging intention, not as a replacement for clinical or administrative surveillance. Tiered triggers linked to validated query sets can facilitate steps that support low-regret readiness, such as pre-cleared messaging, simplified vaccinations booking, or provisional staffing, within the documented short term.

For policy design, a diffusion- and implementation-aware approach is recommended. This approach involves the communication of attributes that are important for adoption, such as clarity, compatibility, and relative advantage and it also involves the reduction of obstacles on the path from intention to action (Rogers, 2003; Greenhalgh et al., 2004). In the context of planned eligibility phases or mandates, the alignment of communication and access changes with these decision windows was seen to improve the conversion of intention signaled by search interest into actual uptake (Greenhalgh et al., 2004).

Finally, the re-evaluation of thresholds and term lists in the aftermath of significant policy shifts or phase transitions is essential. Additionally, the incorporation of Google Trends indicators alongside negative controls like influenza terms is crucial to ensure the maintenance of construct specificity and the identification of deviations (Porcu et al., 2023).

Implications for Positive Social Change

Early awareness can be used to reduce disparities if it is applied to time outreach and access improvements for populations with higher barriers to care. Tailoring queries and messages to groups with linguistic specificities and aligning intention signals with targeted capacity is consistent with implementation guidance on organizational fit and equity in adoption (Greenhalgh et al., 2004; Dearing & Cox, 2018).

It is imperative to position Google Trends as a supplementary and not determinative input to avoid the overreliance on digital behaviors that may underrepresent certain demographic groups. Transparent communication regarding uncertainty and representativeness is an integral component of responsible implementation for social benefit (Mavragani & Ochoa, 2019).

Ethical Considerations and Safeguards

The ethical use of Google Trends involves transparency, proportionality, and oversight. This entails disclosing that Google Trends is a supplementary indicator with a short-term value, documenting trigger rules and uncertainty, and incorporating topic-relevant negative controls to protect against artificial seasonality or media effects (Mavragani & Ochoa, 2019; Porcu et al., 2023). These safeguards support the legitimate and accountable use of digital evidence in public health, while respecting the limits of the evidence available.

Conclusion

This study's objective was to examine whether concept-specific Google Trends searches in Austria can serve as an early indicator of subsequent increases in cases of COVID-19 and the uptake of vaccinations, as well as whether there are short-term dynamics between cases and vaccinations. The evidence indicates that symptom-oriented and vaccination-relevant search activity provides a brief but operationally significant lead, typically ranging from one to three epidemiological weeks. In contrast, cases and vaccinations show minimal, unstable correlation over short periods. The findings are interpreted as construct-specific and statistically reliable within the study's design, though the explanatory power remains limited. This interpretation is supported by negative control analyses and multiple robustness checks.

When interpreted through the study's theoretical framework, Google Trends functions as an indicator of movement through the initial diffusion stages of awareness and persuasion. In this context, Google Trends does not serve as a determinant of epidemic dynamics or program outcomes. Consequently, Google Trends is best positioned as a complementary, short-term indicator to facilitate communication and low-regret readiness steps, rather than as a standalone forecasting system. The study's findings are constrained by the limitations associated with an observational, ecological design, measurement via search behavior, and national-level aggregation. Therefore, the claims are limited to short-term timing within Austria's 2020 to 2023 context.

This research builds on these findings and offers targeted recommendations for future research, including improved measurement, prospective evaluation, explicit handling of confounding, and transparency. It also offers recommendations for practice and policy, such as stage-aware use of Google Trends, triangulation with administrative data, and attention to equity, governance, and re-validation. When considered as a whole, the research provides an empirically grounded, theory-consistent explanation of how digital search signals can support public health surveillance when used in a responsible manner. The fundamental conclusion of this study is that search interest can provide early awareness, which can improve preparedness and communication. However, the value of this awareness is only realized when it is interpreted within the context of the relevant data and used to guide proportionate, transparent action.

References

Abate, B. B., Tilahun, B. D., & Yayeh, B. M. (2024). Global COVID-19 vaccine acceptance level and its determinants: An umbrella review. *BMC Public Health*, *24*(1), 5. https://doi.org/10.1186/s12889-023-17497-4

Adu, P., Popoola, T., Medvedev, O. N., Collings, S., Mbinta, J., Aspin, C., & Simpson, C. R. (2023). Implications for COVID-19 vaccine uptake: A systematic review. *Journal of Infection and Public Health*, *16*(3), 441–466. https://doi.org/10.1016/j.jiph.2023.01.020

Aguiar-Oliveira, M. de L., Campos, A., R. Matos, A., Rigotto, C., Sotero-Martins, A., Teixeira, P. F. P., & Siqueira, M. M. (2020). Wastewater-based rpidemiology (WBE) and viral detection in polluted surface water: A valuable tool for COVID-19 surveillance—A brief review. *International Journal of Environmental Research and Public Health*, *17*(24), Article 24. https://doi.org/10.3390/ijerph17249251

Ahmed, W., Bath, P. A., Sbaffi, L., & Demartini, G. (2019). Novel insights into views towards H1N1 during the 2009 Pandemic: A thematic analysis of Twitter data. *Health Information & Libraries Journal*, *36*(1), 60–72. https://doi.org/10.1111/hir.12247

Alafif, T., Tehame, A. M., Bajaba, S., Barnawi, A., & Zia, S. (2021). Machine and deep learning towards COVID-19 diagnosis and treatment: survey, challenges, and future directions. *International Journal of Environmental Research and Public Health*, *18*(3), Article 3. https://doi.org/10.3390/ijerph18031117

Al-Dmour, H., Masa'deh, R., Salman, A., Abuhashesh, M., & Al-Dmour, R. (2020). Influence of social media platforms on public health protection against the COVID-19 pandemic via the mediating effects of public health awareness and behavioral changes: Integrated model. *Journal of Medical Internet Research*, *22*(8), e19996. https://doi.org/10.2196/19996

Ali, G. G. M. N., Rahman, M. M., Hossain, M. A., Rahman, M. S., Paul, K. C., Thill, J.-C., & Samuel, J. (2021). Public perceptions of COVID-19 vaccines: Policy implications from US spatiotemporal sentiment analytics. *Healthcare*, *9*(9), Article 9. https://doi.org/10.3390/healthcare9091110

Alo, U. R., Nkwo, F. O., Nweke, H. F., Achi, I. I., & Okemiri, H. A. (2022). Non-pharmaceutical interventions against COVID-19 pandemic: Review of contact tracing and social distancing technologies, protocols, apps, security and open research directions. *Sensors*, *22*(1), Article 1. https://doi.org/10.3390/s22010280

AlShurman, B. A., Khan, A. F., Mac, C., Majeed, M., & Butt, Z. A. (2021). What demographic, social, and contextual factors influence the intention to use COVID-19 vaccines: A scoping review. *International Journal of Environmental Research and Public Health*, *18*(17), Article 17. https://doi.org/10.3390/ijerph18179342

Alvarez, E., Bielska, I. A., Hopkins, S., Belal, A. A., Goldstein, D. M., Slick, J., Pavalagantharajah, S., Wynfield, A., Dakey, S., Gedeon, M.-C., Alam, E., & Bouzanis, K. (2023). Limitations of COVID-19 testing and case data for evidence-informed health policy and practice. *Health Research Policy and Systems*, *21*(1), 11. https://doi.org/10.1186/s12961-023-00963-1

Arora, V. S., McKee, M., & Stuckler, D. (2019). Google Trends: Opportunities and limitations in health and health policy research. *Health Policy*, *123*(3), 338–341. https://doi.org/10.1016/j.healthpol.2019.01.001

Athey, S., Grabarz, K., Luca, M., & Wernerfelt, N. (2023). Digital public health interventions at scale: The impact of social media advertising on beliefs and outcomes related to COVID vaccines. *Proceedings of the National Academy of Sciences*, *120*(5), e2208110120. https://doi.org/10.1073/pnas.2208110120

Awijen, H., Ben Zaied, Y., & Nguyen, D. K. (2022). COVID-19 vaccination, fear and anxiety: Evidence from Google search trends. *Social Science & Medicine*, *297*, 114820. https://doi.org/10.1016/j.socscimed.2022.114820

Ayinde, K., Lukman, A. F., & Arowolo, O. (2015). Robust regression diagnostics of influential observations in linear regression model. *Open Journal of Statistics*, *05*(04), 273–283. https://doi.org/10.4236/ojs.2015.54029

Ayouni, I., Maatoug, J., Dhouib, W., Zammit, N., Fredj, S. B., Ghammam, R., & Ghannem, H. (2021). Effective public health measures to mitigate the spread of COVID-19: A systematic review. *BMC Public Health*, *21*(1), 1015. https://doi.org/10.1186/s12889-021-11111-1

Babbie, E. (2017). *Basics of social research* (7th ed.). Boston, MA: Cengage Learning.

Badri, S., Sardá, V., Moncada, J. S., Merçon, M., Rezai, K., Weinstein, R. A., & Trick, W. E. (2021). Disparities and temporal trends in COVID-19 exposures and mitigating behaviors among Black and Hispanic adults in an urban setting. *JAMA Network Open*, *4*(9), e2125187.

https://doi.org/10.1001/jamanetworkopen.2021.25187

Bashir, W., Abdelrazek, E., Mohamed Khair, A., Higazy, O., Mohammed, A., Salih, S., Fadlalmola, H. A., Hammad, A., & Fadlalmola. (2021). Public awareness and compliance with preventive measures for the Novel Coronavirus (COVID-19) Pandemic in Jazan Area, KSA. *International Journal of Nursing Education*, *13*, 87. https://doi.org/10.37506/ijone.v13i4.16593

Becerra-García, J. A., Sánchez-Gutiérrez, T., Barbeito, S., & Calvo, A. (2023). COVID-19 pandemic and mental health in Spain: An analysis of their relationship using Google Trends. *Spanish Journal of Psychiatry and Mental Health*, *16*(4), 215–220. https://doi.org/10.1016/j.rpsm.2021.05.001

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bento, A. I., Nguyen, T., Wing, C., Lozano-Rojas, F., Ahn, Y.-Y., & Simon, K. (2020). Evidence from internet search data shows information-seeking responses to news of local COVID-19 cases. *Proceedings of the National Academy of Sciences*, *117*(21), 11220–11222. https://doi.org/10.1073/pnas.2005335117

Bentotahewa, V., Hewage, C., & Williams, J. (2021). Solutions to big data privacy and security challenges associated with COVID-19 surveillance systems. *Frontiers in Big Data*, *4*. https://doi.org/10.3389/fdata.2021.645204

Bird, J. M., Karageorghis, C. I., & Hamer, M. (2021). Relationships among behavioural regulations, physical activity, and mental health pre- and during COVID–19 UK lockdown. *Psychology of Sport and Exercise*, *55*, 101945. https://doi.org/10.1016/j.psychsport.2021.101945

Bivins, A., North, D., Ahmad, A., Ahmed, W., Alm, E., Been, F., Bhattacharya, P., Bijlsma, L., Boehm, A. B., Brown, J., Buttiglieri, G., Calabro, V., Carducci, A., Castiglioni, S., Cetecioglu Gurol, Z., Chakraborty, S., Costa, F., Curcio, S., de los Reyes, F. L. I., … Bibby, K. (2020). Wastewater-based epidemiology: Global collaborative to maximize contributions in the fight against COVID-19. *Environmental Science & Technology*, *54*(13), 7754–7757. https://doi.org/10.1021/acs.est.0c02388

Blasche, G., Ekmekcioglu, C., & Schernhammer, E. (2022). Vaccination rates are associated with COVID-19 incidence in Austria. *Deutsches Ärzteblatt*, *119*. https://www.aerzteblatt.de/int/archive/article?id=224146

BMSGPK. (n.d.-a). *Bundesministerium für Arbeit, Soziales, Gesundheit, Pflege und Konsumentenschutz*. Bundesministerium für Arbeit, Soziales, Gesundheit, Pflege und Konsumentenschutz. Retrieved April 20, 2025, from https://www.sozialministerium.gv.at/Themen/Gesundheit.html

BMSGPK. (n.d.-b). *COVID-19 (SARS-CoV-2)*. Bundesministerium für Arbeit, Soziales, Gesundheit, Pflege und Konsumentenschutz. Retrieved April 20, 2025, from https://www.sozialministerium.gv.at/Themen/Gesundheit/Uebertragbare-Krankheiten/Infektionskrankheiten-A-Z/Neuartiges-Coronavirus.html

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series*

Analysis: Forecasting and Control (5th ed.).

Brunori, P., & Resce, G. (2020). *Searching for the Peak Google Trends and the COVID-19 Outbreak in Italy* (SSRN Scholarly Paper 3569909). https://doi.org/10.2139/ssrn.3569909

Burkholder, G. J., Cox, K. A., Crawford, L. M., & Hitchcock, J. H. (Eds.). (2020). *Research, design and methods: An applied guide for the scholar-practitioner*. SAGE Publications.

Cao, A., Ueta, M., Uchibori, M., Murakami, M., Kunishima, H., Santosh Kumar, R., Prommas, P., Tomoi, H., Gilmour, S., Sakamoto, H., Hashizume, M., Naito, W., Yasutaka, T., Maruyama-Sakurai, K., Miyata, H., & Nomura, S. (2024). Trust in governments, public health institutions, and other information sources as determinants of COVID-19 vaccine uptake behavior in Japan. *Vaccine*, *42*(17), 3684–3692. https://doi.org/10.1016/j.vaccine.2024.04.081

Charles-Smith, L. E., Reynolds, T. L., Cameron, M. A., Conway, M., Lau, E. H. Y., Olsen, J. M., Pavlin, J. A., Shigematsu, M., Streichert, L. C., Suda, K. J., & Corley, C. D. (2015). Using social media for actionable disease surveillance and outbreak management: A systematic literature review. *PLOS ONE*, *10*(10), e0139701. https://doi.org/10.1371/journal.pone.0139701

Cheong, S. H. R., Ng, Y. J. X., Lau, Y., & Lau, S. T. (2022). Wearable technology for early detection of COVID-19: A systematic scoping review. *Preventive Medicine*, *162*, 107170. https://doi.org/10.1016/j.ypmed.2022.107170

Coetzee, B. J., & Kagee, A. (2020). Structural barriers to adhering to health behaviours in the context of the COVID-19 crisis: Considerations for low- and middle-income countries. *Global Public Health*, *15*(8), 1093–1102. https://doi.org/10.1080/17441692.2020.1779331

Dai, Y., & Wang, J. (2020). Identifying the outbreak signal of COVID-19 before the response of the traditional disease monitoring system. *PLOS Neglected Tropical Diseases*, *14*(10), e0008758. https://doi.org/10.1371/journal.pntd.0008758

Daniel, W. W., & Cross, C. L. (2019). *Biostatistics: A foundation for analysis in the health sciences* (11th ed.). Wiley.

data Austria. (n.d.). *COVID-19: Zeitreihe der verabreichten Impfungen der Corona-Schutzimpfung - Open Government Data Austria*. Retrieved April 20, 2025, from https://www.data.gv.at/katalog/dataset/6475a5ce-d9a4-4a14-ac3e-58eadc25bb25

Dearing, J. W., & Cox, J. G. (2018). Diffusion Of Innovations Theory, principles, and practice. *Health Affairs*, *37*(2), 183–190. https://doi.org/10.1377/hlthaff.2017.1104

Desjardins, M. R. (2020). Syndromic surveillance of COVID-19 using crowdsourced data. *The Lancet Regional Health – Western Pacific*, *4*. https://doi.org/10.1016/j.lanwpc.2020.100024

Donia, A., Hassan, S., Zhang, X., Al-Madboly, L., & Bokhari, H. (2021). COVID-19 Crisis creates opportunity towards global monitoring & surveillance. *Pathogens*, *10*(3), Article 3. https://doi.org/10.3390/pathogens10030256

Durbin, J., & Watson, G. S. (1951). Testing for serial correlation in least squares regression. II. *Biometrika*, *38*(1–2), 159–178.

Effenberger, M., Kronbichler, A., Shin, J. I., Mayer, G., Tilg, H., & Perco, P. (2020). Association of the COVID-19 pandemic with internet search volumes: A Google TrendsTM analysis. *International Journal of Infectious Diseases*, *95*, 192–197. https://doi.org/10.1016/j.ijid.2020.04.033

Fagherazzi, G., Goetzinger, C., Rashid, M. A., Aguayo, G. A., & Huiart, L. (2020). Digital health strategies to fight COVID-19 worldwide: Challenges, recommendations, and a call for papers. *Journal of Medical Internet Research*, *22*(6), e19284. https://doi.org/10.2196/19284

Farooq, F., & Rathore, F. A. (2021). COVID-19 Vaccination and the challenge of infodemic and disinformation. *Journal of Korean Medical Science*, *36*(10). https://doi.org/10.3346/jkms.2021.36.e78

Folotiya, M., & Ngoma, C. (2024). Social media influence on COVID-19 vaccine perceptions among University students: A Malawi case study. *BMC Public Health*, *24*(1), 1312. https://doi.org/10.1186/s12889-024-18764-8

Frankfort-Nachmias, C., Leon-Guerrero, A., & Davis, G. (2021). *Social statistics for a diverse society* (9th ed.). SAGE Publications Inc.

Garett, R., & Young, S. D. (2022). Ethical views on sharing digital data for public health surveillance: Analysis of survey data among patients. *Frontiers in Big Data*, *5*, 871236. https://doi.org/10.3389/fdata.2022.871236

Gasser, U., Ienca, M., Scheibner, J., Sleigh, J., & Vayena, E. (2020). Digital tools against COVID-19: Taxonomy, ethical challenges, and navigation aid. *The Lancet Digital Health*, *2*(8), e425–e434. https://doi.org/10.1016/S2589-7500(20)30137-0

Gendy, M. E. G., & Yuce, M. R. (2023). Emerging technologies used in health management and efficiency improvement during different contact tracing phases against COVID-19 pandemic. *IEEE Reviews in Biomedical Engineering*, *16*, 38–52. IEEE Reviews in Biomedical Engineering. https://doi.org/10.1109/RBME.2022.3219433

Germani, F., Pattison, A. B., & Reinfelde, M. (2022). WHO and digital agencies: How to effectively tackle COVID-19 misinformation online. *BMJ Global Health*, *7*(8). https://doi.org/10.1136/bmjgh-2022-009483

Google News Initiative. (n.d.). *Basics of Google Trends—Google News Initiative*. Retrieved April 20, 2025, from https://newsinitiative.withgoogle.com/en-gb/resources/trainings/basics-of-google-trends/

Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., & Kyriakidou, O. (2004). Diffusion of Innovations in service organizations: Systematic review and recommendations. *The Milbank Quarterly*, *82*(4), 581–629. https://doi.org/10.1111/j.0887-378X.2004.00325.x

Güemes, A., Ray, S., Aboumerhi, K., Desjardins, M. R., Kvit, A., Corrigan, A. E., Fries, B., Shields, T., Stevens, R. D., Curriero, F. C., & Etienne-Cummings, R. (2021). A syndromic surveillance tool to detect anomalous clusters of COVID-19 symptoms in the United States. *Scientific Reports*, *11*(1), 4660. https://doi.org/10.1038/s41598-021-84145-5

Harris, J. E. (2022). Timely epidemic monitoring in the presence of reporting delays: Anticipating the COVID-19 surge in New York City, September 2020. *BMC*

Public Health, *22*(1), 871. https://doi.org/10.1186/s12889-022-13286-7

HealthManagement.org. (n.d.). *Radiology management, ICU management, healthcare IT, cardiology management, executive management*. Retrieved December 21, 2024, from https://healthmanagement.org/c/hospital/issuearticle/the-austrian-healthcare-system?utm_source=chatgpt.com

Higgins, T. S., Wu, A. W., Sharma, D., Illing, E. A., Rubel, K., Ting, J. Y., & Alliance, S. F. (2020). Correlations of online search engine trends with Coronavirus Disease (COVID-19) incidence: Infodemiology study. *JMIR Public Health and Surveillance*, *6*(2), e19702. https://doi.org/10.2196/19702

Huang, C., Yang, L., Pan, J., Xu, X., & Peng, R. (2022). Correlation between vaccine coverage and the COVID-19 pandemic throughout the world: Based on real-world data. *Journal of Medical Virology*, *94*(5), 2181–2187. https://doi.org/10.1002/jmv.27609

Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, *35*(1), 73–101. https://doi.org/10.1214/aoms/1177703732

Husnayain, A., Fuad, A., & Su, E. C.-Y. (2020). Applications of Google Search Trends for risk communication in infectious disease management: A case study of the COVID-19 outbreak in Taiwan. *International Journal of Infectious Diseases*, *95*, 221–223. https://doi.org/10.1016/j.ijid.2020.03.021

Ibrahim, N. K. (2020). Epidemiologic surveillance for controlling COVID-19 pandemic: Types, challenges and implications. *Journal of Infection and Public Health*, *13*(11), 1630–1638. https://doi.org/10.1016/j.jiph.2020.07.019

Jiang, J.-Y., Zhou, Y., Chen, X., Jhou, Y.-R., Zhao, L., Liu, S., Yang, P.-C., Ahmar, J., & Wang, W. (2021). COVID-19 surveiller: Toward a robust and effective pandemic surveillance system based on social media mining. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *380*(2214), 20210125. https://doi.org/10.1098/rsta.2021.0125

Jun, S.-P., Yoo, H. S., & Choi, S. (2018). Ten years of research change using Google Trends: From the perspective of big data utilizations and applications. *Technological Forecasting and Social Change*, *130*, 69–87. https://doi.org/10.1016/j.techfore.2017.11.009

Katz, E., Levin, M. L., & Hamilton, H. (1963). Traditions of research on the Diffusion of Innovation. *American Sociological Review*, *28*(2), 237–252. https://doi.org/10.2307/2090611

Khubchandani, J., Sharma, S., Price, J. H., Wiblishauser, M. J., Sharma, M., & Webb, F. J. (2021). COVID-19 vaccination hesitancy in the United States: A rapid national assessment. *Journal of Community Health*, *46*(2), 270–277. https://doi.org/10.1007/s10900-020-00958-x

Kostkova, P., Saigí-Rubió, F., Eguia, H., Borbolla, D., Verschuuren, M., Hamilton, C., Azzopardi-Muscat, N., & Novillo-Ortiz, D. (2021). Data and digital solutions to support surveillance strategies in the context of the COVID-19 pandemic. *Frontiers in Digital Health*, *3*. https://doi.org/10.3389/fdgth.2021.707902

Krishnamurthi, R., Gopinathan, D., & Kumar, A. (2021). Wearable devices and COVID-19: State of the art, framework, and challenges. In F. Al-Turjman, A. Devi, & A.

Nayyar (Eds.), *Emerging Technologies for Battling COVID-19: Applications and Innovations* (pp. 157–180). Springer International Publishing. https://doi.org/10.1007/978-3-030-60039-6_8

Kuehn, M., LaMori, J., DeMartino, J. K., Mesa-Frias, M., Doran, J., Korrapati, L., Bhojwani, R., Lefebvre, P., & Kirson, N. (2022). Assessing barriers to access and equity for COVID-19 vaccination in the US. *BMC Public Health*, *22*(1), 2263. https://doi.org/10.1186/s12889-022-14636-1

Lazarus, J. V., Wyka, K., White, T. M., Picchio, C. A., Rabin, K., Ratzan, S. C., Parsons Leigh, J., Hu, J., & El-Mohandes, A. (2022). Revisiting COVID-19 vaccine hesitancy around the world using data from 23 countries in 2021. *Nature Communications*, *13*(1), 3801. https://doi.org/10.1038/s41467-022-31441-x

Li, H., Zhang, N., Ma, X., Wang, Y., Yang, F., Wang, W., Huang, Y., Xie, Y., & Du, Y. (2025). Google Trends applications for COVID-19 pandemic: A bibliometric analysis. *DIGITAL HEALTH*, *11*, 20552076241310055. https://doi.org/10.1177/20552076241310055

Lin, Y.-H., Liu, C.-H., & Chiu, Y.-C. (2020). Google searches for the keywords of "wash hands" predict the speed of national spread of COVID-19 outbreak among 21 countries. *Brain, Behavior, and Immunity*, *87*. https://doi.org/10.1016/j.bbi.2020.04.020

Liu, G., & Rusling, J. F. (2021). COVID-19 antibody tests and their limitations. *ACS Sensors*, *6*(3), 593–612. https://doi.org/10.1021/acssensors.0c02621

Luo, S., Xin, M., Wang, S., Zhao, J., Zhang, G., Li, L., Li, L., & Lau, J. T. (2021).

Behavioural intention of receiving COVID-19 vaccination, social media exposures and peer discussions in China. *Epidemiology & Infection*, *149*, e158. https://doi.org/10.1017/S0950268821000947

Ma, S., & Yang, S. (2022). COVID-19 forecasts using Internet search information in the United States. *Scientific Reports*, *12*(1), 11539. https://doi.org/10.1038/s41598-022-15478-y

Machado, de A. V. M., Roberts, B., Wong, B. L. H., van Kessel, R., & Mossialos, E. (2021). The relationship between the COVID-19 pandemic and vaccine hesitancy: A scoping review of literature until August 2021. *Frontiers in Public Health*, *9*. https://doi.org/10.3389/fpubh.2021.747787

Madhavan, S., Bastarache, L., Brown, J. S., Butte, A. J., Dorr, D. A., Embi, P. J., Friedman, C. P., Johnson, K. B., Moore, J. H., Kohane, I. S., Payne, P. R. O., Tenenbaum, J. D., Weiner, M. G., Wilcox, A. B., & Ohno-Machado, L. (2021). Use of electronic health records to support a public health response to the COVID-19 pandemic in the United States: A perspective from 15 academic medical centers. *Journal of the American Medical Informatics Association*, *28*(2), 393–401. https://doi.org/10.1093/jamia/ocaa287

Malik, Y. S., Sircar, S., Bhat, S., Ansari, M. I., Pande, T., Kumar, P., Mathapati, B., Balasubramanian, G., Kaushik, R., Natesan, S., Ezzikouri, S., El Zowalaty, M. E., & Dhama, K. (2021). How artificial intelligence may help the COVID-19 pandemic: Pitfalls and lessons for the future. *Reviews in Medical Virology*, *31*(5), e2205. https://doi.org/10.1002/rmv.2205

Mardian, Y., Kosasih, H., Karyana, M., Neal, A., & Lau, C.-Y. (2021). Review of Current COVID-19 diagnostics and opportunities for further development. *Frontiers in Medicine*, *8*. https://doi.org/10.3389/fmed.2021.615099

Mark, K., Bradley, J., Chute, C., Sumpter, C., Adil, M., & Crooks, G. (2021). Using innovation to develop digital tools for public health during the COVID-19 pandemic. *European Medical Journal*, 50–60. https://doi.org/10.33590/emj/20-00227

Maugeri, A., Barchitta, M., & Agodi, A. (2022). Using Google Trends to Predict COVID-19 vaccinations and monitor search behaviours about vaccines: A retrospective analysis of Italian data. *Vaccines*, *10*(1), 119. https://doi.org/10.3390/vaccines10010119

Mavragani, A., & Gkillas, K. (2020). COVID-19 predictability in the United States using Google Trends time series. *Scientific Reports*, *10*(1), 20693. https://doi.org/10.1038/s41598-020-77275-9

Mavragani, A., & Ochoa, G. (2019a). Google Trends in infodemiology and infoveillance: methodology framework. *JMIR Public Health and Surveillance*, *5*(2), e13439. https://doi.org/10.2196/13439

Mavragani, A., & Ochoa, G. (2019b). Google Trends in Infodemiology and Infoveillance: Methodology framework. *JMIR Public Health and Surveillance*, *5*(2), e13439. https://doi.org/10.2196/13439

Mavragani, A., Ochoa, G., & Tsagarakis, K. P. (2018). Assessing the methods, tools, and statistical approaches in Google Trends research: Systematic review. *Journal of*

Medical Internet Research, *20*(11), e9366. https://doi.org/10.2196/jmir.9366

Mello, M. M., & Wang, C. J. (2020). Ethics and governance for digital disease surveillance. *Science*, *368*(6494), 951–954. https://doi.org/10.1126/science.abb9045

Memedovich, A., Orr, T., Hollis, A., Salmon, C., Hu, J., Zinszer, K., Williamson, T., & Beall, R. F. (2024). Social network risk factors and COVID-19 vaccination: A cross-sectional survey study. *Vaccine*, *42*(4), 891–911. https://doi.org/10.1016/j.vaccine.2024.01.012

Ming, L. C., Untong, N., Aliudin, N. A., Osili, N., Kifli, N., Tan, C. S., Goh, K. W., Ng, P. W., Al-Worafi, Y. M., Lee, K. S., & Goh, H. P. (2020). Mobile health apps on COVID-19 launched in the early days of the pandemic: Content analysis and review. *JMIR mHealth and uHealth*, *8*(9), e19796. https://doi.org/10.2196/19796

Mondal, P., Sinharoy, A., & Su, L. (2021). Sociodemographic predictors of COVID-19 vaccine acceptance: A nationwide US-based survey study. *Public Health*, *198*, 252–259. https://doi.org/10.1016/j.puhe.2021.07.028

Müller, M., Bartsch, S., & Zinn, J. O. (2021). Communicating the unknown: An interdisciplinary annotation study of uncertainty in the coronavirus pandemic. *International Journal of Corpus Linguistics*, *26*(4), 498–531. https://doi.org/10.1075/ijcl.21096.mul

Naseem, M., Akhund, R., Arshad, H., & Ibrahim, M. T. (2020). Exploring the potential of artificial intelligence and machine learning to combat COVID-19 and existing opportunities for LMIC: A scoping review. *Journal of Primary Care &*

Community Health, *11*, 2150132720963634. https://doi.org/10.1177/2150132720963634

Nguyen, K. H., Srivastav, A., Razzaghi, H., Williams, W., Lindley, M. C., Jorgensen, C., Abad, N., & Singleton, J. A. (2021). COVID-19 vaccination intent, perceptions, and reasons for not vaccinating among groups prioritized for early vaccination—United States, September and December 2020. *American Journal of Transplantation*, *21*(4), 1650–1656. https://doi.org/10.1111/ajt.16560

Njororai, F. J., Amulla, W., Nyaranga, C. K., Cholo, W., & Adekunle, T. (2024). A Qualitative exploration of factors associated with COVID-19 vaccine uptake and hesitancy in selected rural communities in Kenya. *COVID*, *4*(6), Article 6. https://doi.org/10.3390/covid4060048

Nuti, S. V., Wayda, B., Ranasinghe, I., Wang, S., Dreyer, R. P., Chen, S. I., & Murugiah, K. (2014). The use of Google Trends in health care research: A systematic review. *PLOS ONE*, *9*(10), e109583. https://doi.org/10.1371/journal.pone.0109583

Ostermann, H., Habimana, K., Bobek, J., & Bachner, F. (2017). Health system governance via health targets: Insights from the Austrian health reform: Herwig Ostermann. *European Journal of Public Health*, *27*(suppl_3), ckx187.189. https://doi.org/10.1093/eurpub/ckx187.189

Piltch-Loeb, R., Savoia, E., Goldberg, B., Hughes, B., Verhey, T., Kayyem, J., Miller-Idriss, C., & Testa, M. (2021). Examining the effect of information channel on COVID-19 vaccine acceptance. *PLOS ONE*, *16*(5), e0251095. https://doi.org/10.1371/journal.pone.0251095

Porcu, G., Chen, Y. X., Bonaugurio, A. S., Villa, S., Riva, L., Messina, V., Bagarella, G., Maistrello, M., Leoni, O., Cereda, D., Matone, F., Gori, A., & Corrao, G. (2023). Web-based surveillance of respiratory infection outbreaks: Retrospective analysis of Italian COVID-19 epidemic waves using Google Trends. *Frontiers in Public Health*, *11*. https://doi.org/10.3389/fpubh.2023.1141688

Prado, T., Fumian, T. M., Mannarino, C. F., Resende, P. C., Motta, F. C., Eppinghaus, A. L. F., Chagas do Vale, V. H., Braz, R. M. S., de Andrade, J. da S. R., Maranhão, A. G., & Miagostovich, M. P. (2021). Wastewater-based epidemiology as a useful tool to track SARS-CoV-2 and support public health policies at municipal level in Brazil. *Water Research*, *191*, 116810. https://doi.org/10.1016/j.watres.2021.116810

Ramsey, J. B. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *31*(2), 350–371. https://doi.org/10.1111/j.2517-6161.1969.tb00796.x

Richter, L., Schreml, S., & Heidinger, T. (2022). Ready for vaccination? COVID-19 vaccination willingness of older people in Austria. *Frontiers in Public Health*, *10*. https://doi.org/10.3389/fpubh.2022.859024

Robinson, E., Sutin, A. R., Daly, M., & Jones, A. (2022). A systematic review and meta-analysis of longitudinal cohort studies comparing mental health before versus during the COVID-19 pandemic in 2020. *Journal of Affective Disorders*, *296*, 567–576. https://doi.org/10.1016/j.jad.2021.09.098

Robinson, L., Schulz, J., Khilnani, A., Ono, H., Cotten, S. R., McClain, N., Levine, L., Chen, W., Huang, G., Casilli, A. A., Tubaro, P., Dodel, M., Quan-Haase, A., Ruiu, M., Ragnedda, M., Aikat, D., & Tolentino, N. (2020). Digital inequalities in time of pandemic: COVID-19 exposure risk profiles and new forms of vulnerability. *First Monday*, *25*(10), Article 10. https://doi.org/10.5210/fm.v25i7.10845

Rogers, E. M. (2003). *Diffusion of Innovations* (5th ed.). Free Press.

Romero, R. A., & Young, S. D. (2022). Ethical perspectives in sharing digital data for public health surveillance before and shortly after the onset of the COVID-19 pandemic. *Ethics & Behavior*, *32*(1), 22–31. https://doi.org/10.1080/10508422.2021.1884079

Rovetta, A. (2021). Reliability of Google Trends: Analysis of the limits and potential of web infoveillance during COVID-19 pandemic and for future research. *Frontiers in Research Metrics and Analytics*, *6*, 670226. https://doi.org/10.3389/frma.2021.670226

Ruiz, J. B., & Bell, R. A. (2021). Predictors of intention to vaccinate against COVID-19: Results of a nationwide survey. *Vaccine*, *39*(7), 1080–1086. https://doi.org/10.1016/j.vaccine.2021.01.010

Saegner, T., & Austys, D. (2022). Forecasting and surveillance of COVID-19 spread using Google Trends: Literature review. *International Journal of Environmental Research and Public Health*, *19*(19), Article 19. https://doi.org/10.3390/ijerph191912394

Sallam, M. (2021). COVID-19 Vaccine hesitancy worldwide: A concise systematic review of vaccine acceptance rates. *Vaccines*, *9*(2), Article 2. https://doi.org/10.3390/vaccines9020160

Sato, K., Mano, T., Iwata, A., & Toda, T. (2021). Need of care in interpreting Google Trends-based COVID-19 infodemiological study results: Potential risk of false-positivity. *BMC Medical Research Methodology*, *21*(1), 147. https://doi.org/10.1186/s12874-021-01338-2

Satterfield, B. A., Dikilitas, O., & Kullo, I. J. (2021). Leveraging the electronic health record to address the COVID-19 pandemic. *Mayo Clinic Proceedings*, *96*(6), 1592–1608. https://doi.org/10.1016/j.mayocp.2021.04.008

Seale, H., Dyer, C. E. F., Abdi, I., Rahman, K. M., Sun, Y., Qureshi, M. O., Dowell-Day, A., Sward, J., & Islam, M. S. (2020). Improving the impact of non-pharmaceutical interventions during COVID-19: Examining the factors that influence engagement and the impact on individuals. *BMC Infectious Diseases*, *20*(1), 607. https://doi.org/10.1186/s12879-020-05340-9

Sen-Crowe, B., McKenney, M., & Elkbuli, A. (2021). Disparities in global COVID-19 vaccination rates & allocation of resources to countries in need. *Annals of Medicine and Surgery*, *68*. https://doi.org/10.1016/j.amsu.2021.102620

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference* (2nd ed.). Cengage Learning.

Singh, H. J. L., Couch, D., & Yap, K. (2020). Mobile health apps that help with COVID-

19 management: Scoping review. *JMIR Nursing*, *3*(1), e20596. https://doi.org/10.2196/20596

Statista. (n.d.). *Coronavirus tägliche Fallzahlen Österreich 2023*. Statista. Retrieved April 20, 2025, from https://de.statista.com/statistik/daten/studie/1150777/umfrage/entwicklung-der-taeglichen-fallzahl-des-coronavirus-in-oesterreich/

Statistik Austria. (n.d.). *ICT usage in households*. STATISTICS AUSTRIA. Retrieved April 20, 2025, from https://www.statistik.at/en/statistics/research-innovation-digitalisation/digital-economy-and-society/ict-usage-in-households

Suh, J., Horvitz, E., White, R. W., & Althoff, T. (2022). Disparate impacts on online information access during the COVID-19 pandemic. *Nature Communications*, *13*(1), 7094. https://doi.org/10.1038/s41467-022-34592-z

Szklo, M., & Nieto, F. J. (2019). *Epidemiology beyond the basics* (4th ed.). Jones & Bartlett Learning, LLC.

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (Seventh edition). Pearson.

Talic, S., Shah, S., Wild, H., Gasevic, D., Maharaj, A., Ademi, Z., Li, X., Xu, W., Mesa-Eguiagaray, I., Rostron, J., Theodoratou, E., Zhang, X., Motee, A., Liew, D., & Ilic, D. (2021). Effectiveness of public health measures in reducing the incidence of covid-19, SARS-CoV-2 transmission, and covid-19 mortality: Systematic review and meta-analysis. *BMJ*, *375*, e068302. https://doi.org/10.1136/bmj-2021-068302

Tan, S. B., Chiu-Shee, C., & Duarte, F. (2022). From SARS to COVID-19: Digital infrastructures of surveillance and segregation in exceptional times. *Cities*, *120*, 103486. https://doi.org/10.1016/j.cities.2021.103486

Tucker, C. E., & Wang, Y. (2021). *The role of delayed data in the COVID-19 pandemic* (SSRN Scholarly Paper 3867021). https://doi.org/10.2139/ssrn.3867021

Valente, T. W. (1996). Social network thresholds in the diffusion of innovations. *Social Networks*, *18*(1), 69–89. https://doi.org/10.1016/0378-8733(95)00256-1

Venkatesh, U., & Gandhi, P. A. (2020). Prediction of COVID-19 outbreaks using Google Trends in India: A retrospective analysis. *Healthcare Informatics Research*, *26*(3), 175–184. https://doi.org/10.4258/hir.2020.26.3.175

Verger, P., & Dubé, E. (2020). Restoring confidence in vaccines in the COVID-19 era. *Expert Review of Vaccines*, *19*(11), 991–993. https://doi.org/10.1080/14760584.2020.1825945

Viswanath, K., Bekalu, M., Dhawan, D., Pinnamaneni, R., Lang, J., & McLoud, R. (2021). Individual and social determinants of COVID-19 vaccine uptake. *BMC Public Health*, *21*(1), 818. https://doi.org/10.1186/s12889-021-10862-1

Wesson, P., Hswen, Y., Valdes, G., Stojanovski, K., & Handley, M. A. (2022). Risks and opportunities to ensure equity in the application of big data research in public health. *Annual Review of Public Health*, *43*(Volume 43, 2022), 59–78. https://doi.org/10.1146/annurev-publhealth-051920-110928

Wilcox, R. R. (2022). *Introduction to robust estimation and hypothesis testing* (5th ed.). Academic Press.

Zang, S., Zhang, X., Xing, Y., Chen, J., Lin, L., & Hou, Z. (2023). Applications of social media and digital technologies in COVID-19 vaccination: Scoping review. *Journal of Medical Internet Research*, *25*(1), e40057. https://doi.org/10.2196/40057

Zayed, B. A., Talaia, A. M., Gaaboobah, M. A., Amer, S. M., & Mansour, F. R. (2023). Google Trends as a predictive tool in the era of COVID-19: A scoping review. *Postgraduate Medical Journal*, *99*(1175), 962–975. https://doi.org/10.1093/postmj/qgad012

Zhang, Y., Zhang, L., Guan, H., Hao, R., & Liu, W. (2024). The impact of internet health information seeking on COVID-19 vaccination behavior in China. *BMC Public Health*, *24*(1), 89. https://doi.org/10.1186/s12889-024-17638-3