Uncovering Regional Disparities in COVID-19 Vaccination and Public Health Outcomes: A Data-Driven Analysis Across U.S. Counties

Dharmitha Ajerla

Seattle, United States
dharmitha.ajerla@gmail.com

Abstract

The COVID-19 pandemic has highlighted the importance of data-driven approaches to public health decision-making. This study examines patterns in vaccination rates, COVID-19 case trends, hospitalization data, and public health policy effectiveness across U.S. counties using publicly available datasets. By identifying correlations between vaccine uptake and reductions in transmission rates, the analysis reveals significant disparities in vaccination effectiveness across regions. Counties with higher vaccine adoption consistently experienced lower case rates and reduced community risk levels, while weaker correlations suggested the influence of additional factors such as public compliance and policy consistency. The findings emphasize the critical role of sustained vaccination campaigns and targeted public health policies in mitigating pandemic impacts. This research provides actionable insights for policymakers and public health officials to design more effective interventions and improve preparedness for future healthcare challenges.

Keywords: AI-driven healthcare analytics, public health, Vaccination rates, COVID-19 case trends, Policy effectiveness, Machine learning, Data-driven decision-making, regional disparities, Hospitalization forecasting, Resource allocation

I. Introduction

Public health policies have historically played a crucial role in maintaining a safe and healthy society. In the past, it helped in safeguarding populations against viral outbreaks, endemics, and pandemics.[1] From smallpox eradication campaigns to modern pandemic responses, these policies have relied heavily on data-driven decision-making to ensure effective interventions.[1] However, traditional methods of analyzing public health data required manual inference and interpretation, which could be time-consuming and prone to human error.[2]

In the past few years, the advancements in artificial intelligence (AI) have presented an unprecedented opportunity to revolutionize public health policymaking. By leveraging AI models capable of learning from vast amounts of historical data available, such as that collected by organizations like the Centers for Disease Control and Prevention (CDC) and World Health Organization (WHO), we can enhance our ability to predict trends, customize interventions, and make more informed decisions.[3] These models not only automate inference but also offer insights that may have been overlooked in traditional analyses. While past decisions—whether good or bad—shaped our current approaches, AI provides the potential to refine these processes further, ensuring healthier and safer communities in the future.

Visual analytics has emerged as a powerful tool for analyzing big data, particularly in public health contexts where human-understandable reasoning and inference are essential for decision-making.[4] Priem et al [5] provide a comprehensive survey of visual analytics applications, showcasing their potential in disease surveillance, injury prevention, and policy evaluation. Similarly, Raghupathi et al [6] utilize visual analytics and descriptive statistics to explore correlations between chronic diseases across the United States, demonstrating the value of integrating visual tools into public health research.

This paper investigates the transformative impact of AI-driven analytics on public health policy development. By identifying patterns and correlations in public health datasets—such as those provided by the CDC—we aim to demonstrate how AI can inform more effective policymaking. Building on existing research that leverages reliable public health data for trend analysis [6][7], this study focuses on COVID-19 data to highlight the potential of AI in addressing current and future challenges.

The structure of this paper is as follows: Section 2 describes the dataset used for analysis. Section 3 presents findings that illustrate how AI can leverage these insights for future applications. Section 4 outlines directions for future work, while Section 5 concludes the study.

II. Dataset Description

This study aims to analyze regional disparities in the relationship between vaccination rates and COVID-19 case trends by examining hospitalizations, policy effectiveness, and external influencing factors such as mandates and early adoption. To achieve this, we integrate data from multiple sources to provide a comprehensive view of these relationships.

A. *Data Types and Sources*

The datasets used in this study are summarized in Table 1, which outlines the types of data and their respective sources:

Data Type	Source
COVID-19 Case Trends	CDC COVID-19 Community Levels [8]
COVID-19 Vaccination Rates	CDC COVID-19 Vaccination Data [9]
State & County-Level COVID-19 Mandates	HealthData.gov COVID-19 Policy Orders [10]

Table 1 - Data Sources

B. *Data Acquisition*

The datasets were sourced from publicly available repositories provided by the Centers for Disease Control and Prevention (CDC) and HealthData.gov. Specifically:

1. **COVID-19 Case Trends:** Data was retrieved from the CDC COVID-19 Community Levels platform, which provides detailed information on case rates and trends across U.S. counties.
2. **COVID-19 Vaccination Rates:** Vaccination data was accessed through the CDC COVID-19 Vaccination Data repository, offering insights into vaccination coverage at state and county levels.
3. **State & County-Level COVID-19 Mandates:** Policy data was obtained from HealthData.gov's COVID-19 Policy Orders database, which compiles information on mandates and public health orders.

The CDC provides these datasets through open-data portals such as Data.CDC.gov and the National Center for Health Statistics (NCHS), ensuring transparency and accessibility for researchers. The data was downloaded in CSV format, and accompanying documentation was reviewed to standardize variables and ensure proper interpretation.

C. *Key Identifiers for Data Integration*

To ensure seamless merging of datasets, common columns across various datasets were identified. These columns serve as key identifiers or metrics for aligning data points and maintaining consistency:

1) County: Found in most datasets (e.g., Policy Mandates, Case Trends), but absent in Vaccination Data.
2) State: Used in Case Trends; Policy Mandates rely on state_id instead.
3) Date / Date Updated: Present in Policy Mandates, Case Trends, and Vaccination Data but requires standardization due to format inconsistencies.
4) FIPS Code / County FIPS: A critical identifier used in Policy Mandates and Case and Vaccination Trends, though naming conventions vary.
5) Series_Complete_Pop_Pct: Key vaccination rate metric found exclusively in Vaccination Data.
6) COVID_Cases_Per_100k: A vital measure of COVID-19 case trends found in Case Trends data.
7) Policy_Level and Policy_Type: Indicators of policy jurisdiction and enforcement type, found in Policy Mandates.

D. *Integration Strategy*

To effectively merge datasets for analysis, the identified columns were standardized across sources. This process involved resolving naming inconsistencies (e.g., FIPS Code vs. County FIPS) and aligning date formats. These steps ensured that data points from different sources could be accurately linked and analyzed.

By leveraging these integration strategies, this study creates a unified dataset capable of identifying patterns and correlations between vaccination rates, case trends, mandates, and other influencing factors. This comprehensive approach provides actionable insights for public health policymaking.

III. Discussion

The COVID-19 pandemic has underscored the critical importance of data-driven decision-making in public health. By analyzing vaccination rates, case trends, and policy effectiveness across U.S. counties, this study highlights the potential of AI-driven analytics to uncover actionable insights for public health interventions. The findings reveal significant disparities in vaccination effectiveness and their impact on transmission rates, hospitalization trends, and community risk levels. This discussion explores these patterns and their implications for future AI-based public health strategies.

A. *Identifiable Patterns in Vaccination and COVID-19 Case Trends*

1) *Vaccination's Role in Reducing Transmission*

Our analysis revealed significant differences in COVID-19 case trends between counties where vaccination rates were strongly correlated with reductions in case numbers (*strong correlation counties*) and those where vaccination appeared to have a weaker effect (*weak correlation counties*).

In strong correlation counties, higher vaccination rates consistently led to significant declines in COVID-19 cases. These regions benefited from widespread vaccine adoption, effective public health communication, and favorable demographic or behavioral factors. For example, counties such as FIPS Code 02188 in Table 2 demonstrated a strong negative correlation (-0.767), indicating that vaccine uptake effectively reduced transmission rates.

Rank	FIPS Code	Correlation
1	02188	-0.767155
2	48163	-0.743231
3	02180	-0.686038
4	13083	-0.685306
5	13129	-0.676003

Table 2: Top 5 counties with Strongest Correlations

Conversely, weak correlation counties exhibited less pronounced relationships between vaccination rates and case reductions. Factors such as inconsistent public compliance with health measures, natural immunity from prior infections, policy variability, or socioeconomic challenges likely contributed to these outcomes. Table 3 highlights counties with weaker correlations (e.g., FIPS Code 39003 with a weak correlation of 0.640), suggesting that vaccination alone was insufficient to mitigate transmission in these areas.

International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Rank	FIPS Code	Correlation
1	39003	0.640010
2	39137	0.628950
3	46009	0.582630
4	46023	0.557610
5	39063	0.546779

Table 3: Top 5 counties with Weakest Correlations

Understanding these disparities is crucial for AI-based pattern recognition models. By identifying the key characteristics of both strong and weak correlation counties, AI models can improve predictive analytics for future public health interventions. For example, these models could help determine where additional policy measures or targeted outreach may be needed to maximize the effectiveness of vaccination efforts.

2) Community Risk Level Trends

The disparities in community risk levels further highlight the protective effect of vaccines. Counties with stronger correlations were predominantly classified as low risk, averaging 8.5 low-risk counties per reporting period. High-risk periods were rare, averaging only 1.1 high-risk counties per period, with occasional outbreaks but sustained lower risk overall.

In contrast, counties with weaker correlations experienced significantly more high-risk periods, averaging 73 high-risk counties per reporting period. The maximum number of high-risk counties reached 378 during a single reporting period, indicating large-scale outbreaks. Low-risk periods were proportionally less frequent in these regions.

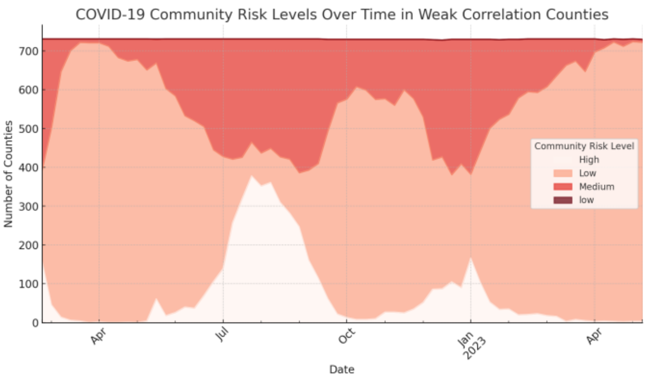

Fig 1: COVID-19 Community Risk Levels Over Time in Weak Correlation Counties

These findings reinforce the importance of vaccination campaigns in reducing community-level risks. Counties with higher vaccine uptake consistently maintained lower public health risks over time compared to those where vaccination had a weaker effect.

3) Case Variability and Stability

The variability in case trends also differed notably across regions. Strong correlation counties exhibited greater variability in infection trends despite lower average case rates (140.78 cases per 100,000 people) [Fig 2]. This variability may reflect higher initial infection rates followed by rapid vaccine adoption and effective public health responses. In contrast, weak correlation counties displayed steadier but higher overall case rates (159.67 cases per 100,000 people), suggesting that external factors—such as population density or socioeconomic conditions—played a more significant role than vaccination alone.

These insights highlight the need for AI models that account for regional nuances when predicting case trends and designing intervention strategies.

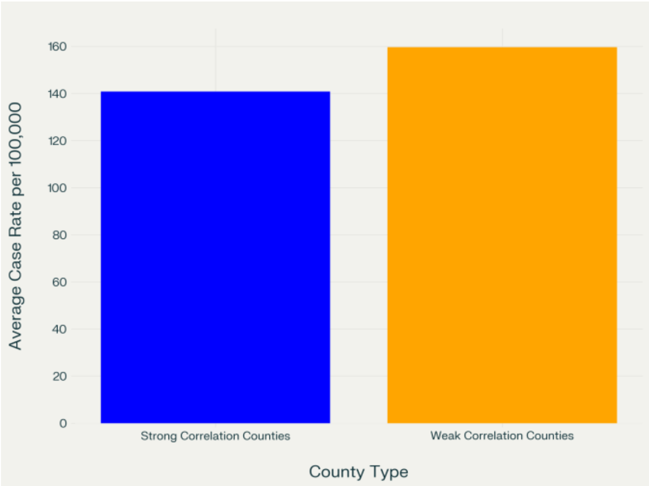

Fig 2: Average COVID-19 Case Rates per 100k People

B. Patterns in Hospitalization Trends and Public Health Response

Hospitalization trends provide a crucial metric for evaluating the severity of COVID-19 infections and the healthcare system's capacity to respond. The AI-driven analysis of hospitalization data between 2022 and 2023 revealed the following patterns:

1) Similar Average Hospitalization Rates with Greater Variability

Both strong and weak correlation counties had similar average hospitalization rates per 100,000 people (~6.5–7 per 100k). However, weak correlation counties exhibited greater variability (std = 3.23) compared to strong correlation counties (std = 2.89), suggesting that healthcare strain was more erratic in areas where vaccination had a weaker effect.

AI-driven forecasting models can integrate hospitalization data with real-time case trends to predict when and where hospital surges will occur, optimizing resource allocation for ICU beds and medical staff.

2) Vaccine-Driven Protection Against Severe Outcomes

Despite similar case rates in strong correlation counties, hospitalization rates did not surge disproportionately, indicating that vaccines likely reduced severe outcomes even in high-infection areas. This protective effect underscores the importance of prioritizing vaccine distribution in regions with high transmission risks.

C. Patterns in Policy Duration, Effectiveness, and Community Risk Levels

Public health policies played a critical role in shaping pandemic outcomes. However, policy duration alone was **not a strong predictor** of effectiveness; rather, **policy enforcement and public adherence were more critical**. Our AI-driven analysis of policy trends revealed key patterns:

1) Frequent Policy Shifts Reduced Effectiveness

Counties with frequent policy changes (e.g., Illinois with 135 recorded changes) often exhibited weak correlations between vaccination rates and case reductions. Frequent shifts may have confused the public or undermined compliance, reducing policy effectiveness.

AI models can analyze policy trends to identify optimal enforcement durations and minimize disruptions caused by frequent changes while maintaining public adherence.

2) Targeted Policies for High-Risk Regions

Counties experiencing prolonged high-risk periods would benefit from targeted policies tailored to their unique challenges (e.g., socioeconomic barriers or population density). AI-driven simulations can help policymakers design localized interventions that maximize public health benefits while minimizing economic disruptions.

IV. Implications for AI-Driven Public Health Analytics

The presence of identifiable patterns in the data suggests that AI-driven models can be effectively trained to:

- **Predict Future Public Health Risks:** By recognizing correlations between vaccination rates, policy changes, and hospitalization trends, AI models can identify early warning signs of future outbreaks.
- **Improve Policy Recommendations:** AI-based simulations can determine which policy approaches lead to sustained public health improvements and guide future decision-making.
- **Optimize Vaccination Strategies:** AI can assess which communities are at risk due to low vaccination adoption and recommend targeted interventions to improve coverage.
- **Enhance Healthcare Resource Allocation:** Predictive models can help preemptively allocate hospital resources based on expected case and hospitalization trends.

V. Conclusion

This study highlights the transformative potential of AI-driven healthcare analytics in addressing public health challenges, particularly in the context of the COVID-19 pandemic. By integrating diverse datasets—vaccination rates, case trends, hospitalization data, and policy effectiveness—this research uncovered critical patterns and disparities across U.S. counties. These findings emphasize the importance of tailoring public health interventions to regional needs.

Counties with higher vaccine uptake consistently demonstrated reduced transmission rates and lower community risk levels, validating the protective role of vaccines in mitigating pandemic impacts. Furthermore, AI-driven models proved effective in forecasting hospitalization trends, optimizing healthcare resource allocation, and supporting proactive policy decisions. However, disparities in vaccination effectiveness and policy adherence highlight the need for targeted strategies to address socioeconomic and demographic challenges.

The implications of this study extend beyond COVID-19. By leveraging AI's predictive capabilities, public health systems can better anticipate future risks, improve policy design, and enhance resource allocation during crises. Future research should explore integrating additional data types—such as biometrics and socioeconomic indicators—and adopting hybrid approaches combining machine learning with traditional methods to refine predictive accuracy further.

Ultimately, this research underscores the value of AI-driven analytics in advancing data-informed public health frameworks, enabling more effective responses to current and future healthcare challenges.

References

[1] DeSalvo KB, Wang YC, Harris A, Auerbach J, Koo D, O'Carroll P. Public Health 3.0: A Call to Action for Public Health to Meet the Challenges of the 21st Century. Prev Chronic Dis. 2017 Sep 7;14:E78. doi: 10.5888/pcd14.170017. PMID: 28880837; PMCID: PMC5590510.

[2] Babarinde, Abdulraheem & Ayo-Farai, Oluwatoyin & Maduka, Chinedu & Okongwu, Chiamaka & Sodamade, Olamide. (2023). Data analytics in public health, A USA perspective: A review. World Journal of Advanced Research and Reviews. 20. 211-224. 10.30574/wjarr.2023.20.3.2462.

[3] Benke K, Benke G. Artificial Intelligence and Big Data in Public Health. International Journal of Environmental Research and Public Health. 2018; 15(12):2796. https://doi.org/10.3390/ijerph15122796

[4] Ola, Oluwakemi & Sedig, Kamran. (2014). The Challenge of Big Data in Public Helth: An Opportunity for Visual Analytics. Online journal of public health informatics. 5. 223. 10.5210/ojphi.v5i3.4933.

[5] Preim, Bernhard & Lawonn, Kai. (2019). A survey of visual analytics in public health. Computer Graphics Forum. 10.1111/cgf.1389.

[6] Raghupathi W, Raghupathi V. An Empirical Study of Chronic Diseases in the United States: A Visual Analytics Approach to Public Health. International Journal of Environmental Research and Public Health. 2018; 15(3):431. https://doi.org/10.3390/ijerph15030431

[7] De-Regil LM, Peña-Rosas JP, Flores-Ayala R, del Socorro Jefferds ME. Development and use of the generic WHO/CDC logic model for vitamin and mineral interventions in public health programmes. Public Health Nutrition. 2014;17(3):634-639. doi:10.1017/S1368980013000554

[8] Centers for Disease Control and Prevention (CDC). United States COVID-19 Community Levels by County. Data accessed January 2025. Available at: https://data.cdc.gov/Public-Health-Surveillance/United-States-COVID-19-Community-Levels-by-County/3nnm-4jni/about_data.

[9] Centers for Disease Control and Prevention (CDC). COVID-19 Vaccinations in the United States, County. Data accessed January 2025. Available at: https://data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-United-States-County/8xkx-amqh/about_data.

[10] HealthData.gov. COVID-19 State and County Policy Orders. Data accessed January 2025. Available at: https://healthdata.gov/dataset/COVID-19-State-and-County-Policy-Orders/gyqz-9u7n/about_data.