A Comparative Analysis of Survival Models to Predict COVID-19 Booster Vaccine Uptake in Nueces County: The Influence of Geography and Demography

Hossein Naderi, **Ziba Abbasian**, **Yuxia Huang**

ABSTRACT: This study focuses on Nueces County, analyzing factors influencing the timing of COVID-19 booster vaccine uptake, particularly assessing the role of geographical and demographic characteristics. Utilizing a complete dataset from the county's public health department, this research investigates the association between these factors and the interval for receiving booster doses. Various survival models, including Kaplan Meier, Random Survival Forest (RSF), DeepSurv, and Cox proportional hazards model, were employed to evaluate the efficacy of these models in predicting vaccine uptake. The study highlighted significant associations between urban residency and distance to vaccination sites with the timing of booster vaccinations. The Cox proportional hazards model and the DeepSurv model performed better in analyzing these associations, as indicated by the concordance index (c-index). The findings highlight the importance of considering geographical and demographic factors in public health strategies for vaccine distribution and underscore the potential of advanced survival models in enhancing risk assessment and personalized treatment plans in various medical fields. This study provides critical insights for public health policymakers in the strategic planning and execution of vaccination campaigns, emphasizing the need for equitable distribution and accessibility of vaccines.

KEYWORDS: COVID-19 vaccination, Geographical factors, Survival Analysis models, DeepSurv model, Healthcare equity.

Introduction

COVID-19 has emerged as a global health crisis with profound social, economic, and health implications, claiming almost 7 million lives worldwide as of December 22, 2023(WHO 2023). The response to this pandemic, particularly the implementation of vaccination programs, has been a cornerstone in managing the spread of the virus (Prevention. 2022). However, the decline in vaccine efficacy over time, particularly against new variants, has raised concerns (Sang, Zhang et al. 2022) so the next vaccination doses are recommended by the World Health Organization (WHO) and the Drug and Food Organization (El Adam, Zou et al. 2022). Determining the optimal interval for administering COVID-19 vaccine doses is pivotal in enhancing its efficacy and ensuring sustained immunity. As research evolves, understanding the ideal timing for booster doses becomes crucial, especially considering the diminishing efficacy over time and against emerging variants like Delta and Omicron (Vasireddy, Vanaparthy et al. 2021).

A popular topic in public health research is studying the probability of an event of interest over time. (Bajema, Rowneki et al. 2023) in their cohort study that focused on what was the uptake of and factors associated with COVID-19 primary and booster on-time vaccination, suggested that targeted outreach to younger, rural veterans may improve COVID-19 vaccination rates. In similar research (Ioannou, Green et al. 2021) also stated that geographical and demographic factors significantly associated with second dose timely vaccination.

A study employed survival models to predict depression relapse risk. The results showed that DeepSurv outperformed in comparison among different survival analyses conducted on health outcomes (Garcia, Hirao et al. 2021).
(Seidi et al., 2023) proposed an innovative approach to improve the performance of current survival models by focusing on datasets containing geographic location-based public health information. They introduced a novel method for generating datasets oriented towards public health, utilizing geographic location-based features in survival analysis. This approach emphasizes the potential of leveraging geographic data to enhance survival model predictions (Seidi, Tripathy et al. 2024).

In this study, we introduce an approach focused on identifying geospatial and demographic factors that are pivotal for timely vaccine uptake. Additionally, we aim to assess and compare the efficacy of various survival analysis models in their ability to examine these factors and their impact on comprehensive vaccination plans. Our testing of these models serves a dual purpose: firstly, to optimize statistical analysis methods, and secondly, to evaluate equitable geographical access to vaccines. This dual focus is significant within the public health context. By understanding these associations and methodological differences, health policymakers and researchers can be better equipped to enhance vaccination rates, fortify community immunization, and advance the development of more effective scientific analysis methods.

Method

Data description

We obtained a complete dataset of administrated vaccinations from the county's public health department. The dataset includes clients' characteristics, vaccination dates, and vaccination sites. The baseline included demographic and spatial characteristics: sex (Female and Male), age, race (American Indian or Alaska Native, Asian, Native Hawaiian or Pacific Islander, Black, White, and two or more races), ethnicity (Hispanic or Latino and not Hispanic or Latino), and home address that obtained from self-reported by individuals. The residency region (urban vs rural), and the Social Vulnerability Index (SVI) (CDC 2020) were also considered for individuals who received the vaccine. Administrated vaccine doses included Pfizer, Moderna, and AstraZeneca, except for Janssen vaccines, from December 14, 2020, to February 6, 2022.

This research derived a distance index to assess how far driving can influence the time to receive a booster dose. This index should also consider the accessibility of different areas of the county to have a fair comparison of the distance traveled to vaccination centers. We used the two-step floating catchment area (2SFCA) method (Tao, Cheng et al. 2020) to access vaccination centers from the center of the population of each block group. We calculated the difference between the normalized actual travel time and the normalized accessibility value (Naderi, Abbasian et al. 2023), as the distance index.

Statistical Analysis

The Kapalan Meier, Random Survival Forest (RSF), and DeepSurv were used to compare the performance of the prediction with Coxph, the models using the Concordance Index (C-index) metric that measures a degree of compromise between the predicted and observed survival times (Seidi, Tripathy et al. 2024).

Cox proportional hazard model (Cox 1972) was used to analyze the association between sex, age, race, ethnicity, residence area, Social Vulnerability Index (SVI), and distance indices to receive the COVID-19 vaccines, as the demographic and spatial factors, and time to receive a booster dose. The proportional hazard model assumption was evaluated graphically using the Schoenfeld residuals.

Results

The primitive result in Table 1 reveals the analysis concordance index in different methods. The c-index measures the predictive accuracy of the model. Higher values indicate a model's better predictive ability for the time until an event of interest. Table1 would show which models performed better according to the c-index metric. It is obvious that the DeepSurv model represents the best score, while the Kaplan Meier demonstrates the least value.

Table 1: C-index score comparison

Model	*Concordance Index*
DeepSurv	0.61
Coxph	0.6
RSF	0.58
Kaplan Meier	0.5

Considering the initial results, we decided to opt for the Coxph model to investigate the association between geographical and demographic factors and the time to receive the vaccine. Figure 1 represents the results of a Cox proportional hazards model analyzing various factors for their impact on a survival-related outcome, the time to receive a COVID-19 booster shot. Geographical factors like residence region, SVI, and distance indices, generally have significant associations with the time to receive a COVID-19 booster dose. Living in urban regions, areas with high SVI, and low distance indices represent increasing HRs for a stronger association with the receiving vaccine.

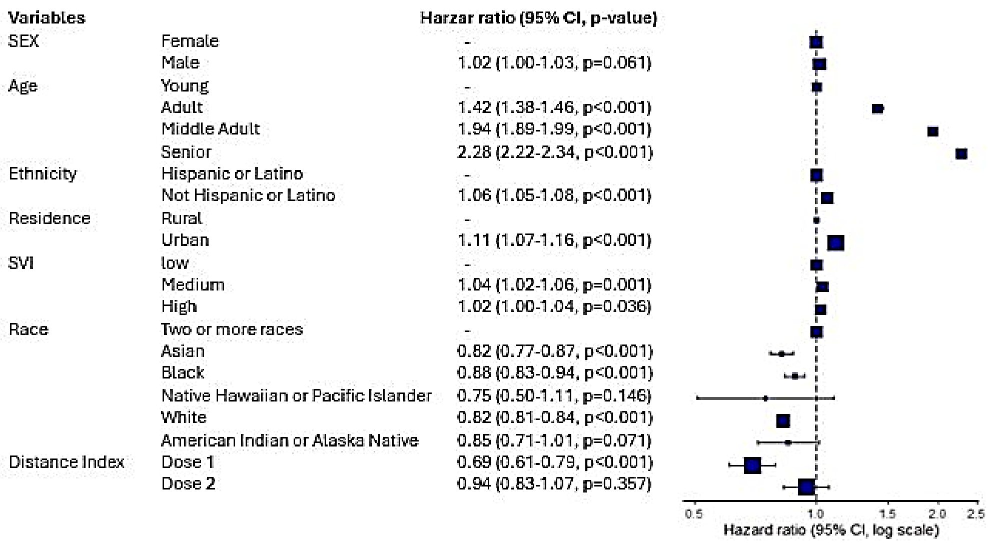

Figure 1: Forest diagram of hazard ratios (HRs) from the Cox proportional hazards model.

Conclusions

In conclusion, our study offers a comprehensive analysis of factors influencing booster vaccine uptake in Nueces County. We found that geographical and demographic characteristics significantly affect the timing of booster vaccinations. Notably, older age groups, urban residents, and non-Hispanic or Latino individuals are more likely to receive a booster dose. Racial disparities also emerged, with Asians and Whites showing lower booster uptake compared to individuals of multiple races. The impact of social vulnerability and travel time to vaccination sites were identified as key factors as well.

Our comparison of the concordance index (c-index) across various survival models underscores the effectiveness of the DeepSurv and Coxph models in decoding vaccination patterns. The findings of a study showed that DeepSurv and Coxph performed better than conventional survival hazards models and had encouraging potential to enhance risk assessment and personalized treatment recommendations across a range of medical specialties.

This research has significant implications for public health policy and the strategic planning of vaccination campaigns, particularly in the context of geographical and temporal factors. Its relevance is heightened when evaluating traditional survival regression models.

In summary, by analyzing the intervals between COVID-19 vaccine doses, especially booster shots, our study contributes crucial insights into vaccine efficacy, long-term immunity, and equitable distribution across different regions. We specifically focus on identifying and understanding the geographical and temporal factors that influence booster dose administration in Nueces County, thus offering valuable guidance for future public health strategies.

References

Bajema, K. L., M. Rowneki, K. Berry, A. Bohnert, C. B. Bowling, E. J. Boyko, T. J. Iwashyna, M. L. Maciejewski, A. M. O'Hare, T. F. Osborne, E. M. Viglianti, D. M. Hynes and G. N. Ioannou (2023). "Rates of and Factors Associated With Primary and Booster COVID-19 Vaccine Receipt by US Veterans, December 2020 to June 2022." JAMA Network Open **6**(2): e2254387-e2254387.

CDC. (2020, October 28, 2022). "CDC SVI Documentation 2020." Retrieved 09/26/2023, from https://www.atsdr.cdc.gov/placeandhealth/svi/documentation/SVI_documentation_2020.html.
Cox, D. R. (1972). "Regression Models and Life-Tables." Journal of the Royal Statistical Society. Series B (Methodological) **34**(2): 187-220.

El Adam, S., M. Zou, S. Kim, B. Henry, M. Krajden and D. M. Skowronski (2022). "SARS-CoV-2 mRNA Vaccine Effectiveness in Health Care Workers by Dosing Interval and Time Since Vaccination: Test-Negative Design, British Columbia, Canada." Open Forum Infectious Diseases **9**(5): ofac178.

Garcia, F. C. C., A. Hirao, A. Tajika, T. A. Furukawa, K. Ikeda and J. Yoshimoto (2021). Leveraging Longitudinal Lifelog Data Using Survival Models for Predicting Risk of Relapse among Patients with Depression in Remission. 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC).

Ioannou, G. N., P. Green, E. R. Locke and K. Berry (2021). "Factors associated with early receipt of COVID-19 vaccination and adherence to second dose in the Veterans Affairs healthcare system." PLOS ONE **16**(12): e0259696.

Naderi, H., Z. Abbasian and Y. Huang (2023). Measuring spatial accessibility of covid-19 vaccination sites International Conference On Recent Advances In Medical, Medicine And Health Sciences. Houston, Texas, Institute for Technology and Research (ITRESEARCH)**:** 25-28.
Prevention., C. f. D. C. a. (2022). "CDC Strategy for Global Response to COVID-19 (2020-2023)." from https://www.cdc.gov/coronavirus/2019-ncov/global-covid-19/global-response-strategy.html

Sang, Y., Z. Zhang, E. Li, H. Lu, J. Long, Y. Cao, C. Yu, T. Wang, J. Yang and S. Wang (2022). "An mRNA vaccine with broad-spectrum neutralizing protection against Omicron variant sublineages BA.4/5 -included SARS-CoV-2." Signal Transduction and Targeted Therapy **7**(1): 362.

Seidi, N., A. Tripathy and S. K. Das (2024). Using Geographic Location-Based Public Health Features in Survival Analysis. Proceedings of the 8th ACM/IEEE International Conference on Connected Health: Applications, Systems and Engineering Technologies. <conf-loc>, <city>Orlando</city>, <state>FL</state>, <country>USA</country>, </conf-loc>, Association for Computing Machinery**:** 80–91.

Tao, Z., Y. Cheng and J. Liu (2020). "Hierarchical two-step floating catchment area (2SFCA) method: measuring the spatial accessibility to hierarchical healthcare facilities in Shenzhen, China." International Journal for Equity in Health **19**(1): 164.

Vasireddy, D., R. Vanaparthy, G. Mohan, S. V. Malayala and P. Atluri (2021). "Review of COVID-19 Variants and COVID-19 Vaccine Efficacy: What the Clinician Should Know?" Journal of Clinical Medicine Research; Vol. 13, No. 6, Jun 2021.

WHO (2023). COVID-19 epidemiological update – 22 December 2023. 162. WHO Team**:** 26.

Hossein Naderi, College of Engineering and Computer Science, Texas A&M University-Corpus Christi, 6300 Ocean Dr, Corpus Christi, TX

Ziba Abbasian, College of Engineering and Computer Science, Texas A&M University-Corpus Christi, 6300 Ocean Dr, Corpus Christi, TX

Yuxia Huang, College of Engineering and Computer Science, Texas A&M University-Corpus Christi, 6300 Ocean Dr, Corpus Christi, TX