medRxiv preprint doi: https://doi.org/10.1101/2023.01.31.23285300; this version posted February 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Global generalisability of AI-driven COVID-19 vaccination policies: a cross-sectional observational study

Raghav Awasthi[1*] MSc, Aditya Nagori[1*] PhD, Bouchra Nasri[2+] PhD

1. PathCheck Foundation, Cambridge, MA , 2. Université de Montréal, Canada

***Equal contribution, +Corresponding author**

Abstract:

Objective: To discover global determinants of vaccine uptake behavior and to develop a generalizable machine learning model to predict the vulnerability of vaccine uptake behavior at the individual level.

Methodology: 23135 Respondents across the 23 countries were interviewed for the survey questionnaire, after preprocessing and cleaning data, we performed Bayesian networks and generalized linear models to identify the key determinants of vaccine uptake. Markov Blankets obtained from the Bayesian networks were used to estimate the important predictors of the vaccine uptake. These variables were then used to build the models. To build generalizable models, we used country-wise data splitting. Model evaluation is assessed for the prediction performance on the new countries. We also developed income specific models cross validated within the income group.

Results: We found 16 important predictors of vaccine uptake using the Bayesian network and Markov Blanket approach. We found that the trust of the central government (Log-Odds 0.55[0.25, 0.84] (p= 0.0002)), Vaccination restriction for national and international travel

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2023.01.31.23285300; this version posted February 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

(Log-Odds 0.4[0.14, 0.65] (p= 0.0034)) as the key determinants of Vaccine uptake. Our Generalized mixed effects model approach achieved an AUC of 89%, Precision 90% and Recall of 82% on the prediction task on new countries, thus, generalizing to new countries.

Introduction

Vaccination is a crucial public health measure that helps to prevent the spread of infectious diseases and protect individuals and communities. However, the success of vaccination programs depends on the willingness and ability of individuals to receive vaccines [1]. Formally, Vaccine hesitancy according to the *Strategic Advisory Group of Experts on Immunization (SAGE)* Working Group [1,2] is defined as a delay in accepting or refusing vaccination [3,4]**.** Upon the release of COVID-19 vaccines, it was essential that a substantial percentage of the population need to get vaccinated [3] [7] [5] to achieve herd immunity. For the decision-makers it becomes crucial to comprehend and recognise a person's vaccination-related behavior. The vaccination hesitation behavior of an individual is influenced by variables such as vaccine accessibility, geographical location, cost, socio-economic factors, quality of healthcare, and most importantly, trust in vaccines and health experts. Artificial intelligence (AI) techniques, which involve processing large amounts of data, identifying patterns and trends, could be helpful in making predictions that might guide the understanding and the implementation of efficient vaccination strategies. Several researches have shown that AI-based techniques may be used to identify important vaccination uptake factors and predict vaccine uptake behavior at the individual level [6–10]. However, all these articles were instigating vaccine hesitancy on a local scale (for a specific region or country) and none of them studied whether or not vaccination policies can be implemented on a global scale. The success of vaccination programmes may vary between geographical location and socio-demographic factors of a specific population. The capacity of AI

medRxiv preprint doi: https://doi.org/10.1101/2023.01.31.23285300; this version posted February 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

algorithms to effectively predict the effects of vaccination programmes in diverse countries and people may be impacted by factors such as cultural, economic, and political disparities [9,11–13]. Additionally, there may be variations in data availability and quality, as well as in the knowledge and resources available to create and deploy AI algorithms, among locations and populations[13,14].

In this study, we aim to analyze the global generalizability of AI-driven vaccination policies, with the focus of identifying the key determinants of vaccine uptake behavior and developing a generalizable machine learning model to predict the vulnerability of vaccine uptake behavior at the individual level. To achieve these goals, we used open data from a cross-sectional observational study involving 23,135 respondents across 23 countries about COVID-19 vaccine hesitancy. We used survey questionnaire data to identify the key determinants of vaccine uptake and used Bayesian networks and generalized linear models to build predictive models. To ensure the generalizability of our models, we used country-wise data splitting to train, validate, and test our models. We also tested the generalizability of our models across different income groups for vaccine uptake. The findings of this study will provide valuable insights into the global determinants of COVID-19 vaccine hesitancy and the generalizability of prediction models for COVID-19 vaccine hesitancy. These insights will be useful for policymakers, public health officials, and other stakeholders in their efforts to address COVID-19 vaccine hesitancy and promote vaccination uptake. Finally, as part of this study, a AI-driven tool is developed to help researchers and policymakers in analyzing the vaccine uptake behavior based on survey data for a specific disease or one/multiple geographical locations.

medRxiv preprint doi: https://doi.org/10.1101/2023.01.31.23285300; this version posted February 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Data and Methods

Data

We have utilized publicly available survey data for our analysis [15]. Data consists of 23,135 participants from 23 countries who participated in the survey. About half (50.2%) were female and resided in LMICs (52.2%), three-fifths (59.9%) were between the ages of 30 and 59, and one-fifth (22.4%) were college graduates. The data represents 23 countries: Brazil, Canada, China, Ecuador, France, Germany, Ghana, India, Italy, Kenya, Mexico, Nigeria, Peru, Poland, Russia, Singapore, South Africa, South Korea, Spain, Sweden, Turkey, the UK, and the US. The survey questionnaire was made by a panel of experts based on COVID-19 vaccine acceptance studies as well as studies on pandemic control measures and vaccination intentions [16,17]. The questionnaire is composed of 31 questions (***See S*upplementary information of** [15]). These questions underline mainly the three following concepts: 1) Questions representing perceptions of risk (q1) (q3), efficacy (q2), safety (q4), and trust (q5 and q6) as important individual determinants of COVID-19 vaccine hesitancy and of routine immunization; 2) questions regarding the status of vaccine regarding the first dose (q7) and the vaccine hesitancy which based on taking or no vaccine when available(q8) 3) questions related to socio-economic factors like education (q29), income (q30), and geographical location (q31).

Methods

We first based our analysis on a probabilistic graphical model approach, which aims to determine the global factor of vaccine hesitancy. Then using a supervised machine learning approach, we have examined the generalisability of vaccination uptake prediction models across

medRxiv preprint doi: https://doi.org/10.1101/2023.01.31.23285300; this version posted February 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

countries. **Figure 1** summaries the steps used in the analysis. First we cleaned and preprocessed the data-in order to handle missing information.. Our outcome variable regarding vaccine hesitancy corresponds to answers related to q8 of the survey which is *'I will take the COVID-19 vaccine when it is available to me*'. We trained a bootstrapped Bayesian network in order to quantify the interaction between data variables and learn the determinants of global vaccination hesitancy. Using the learnt Bayesian network, we computed Markov blanket variables for outcome variables. Further Markov blankets were used to predict the outcome variable. Finally, we calculated the generalizability of the prediction model across countries.

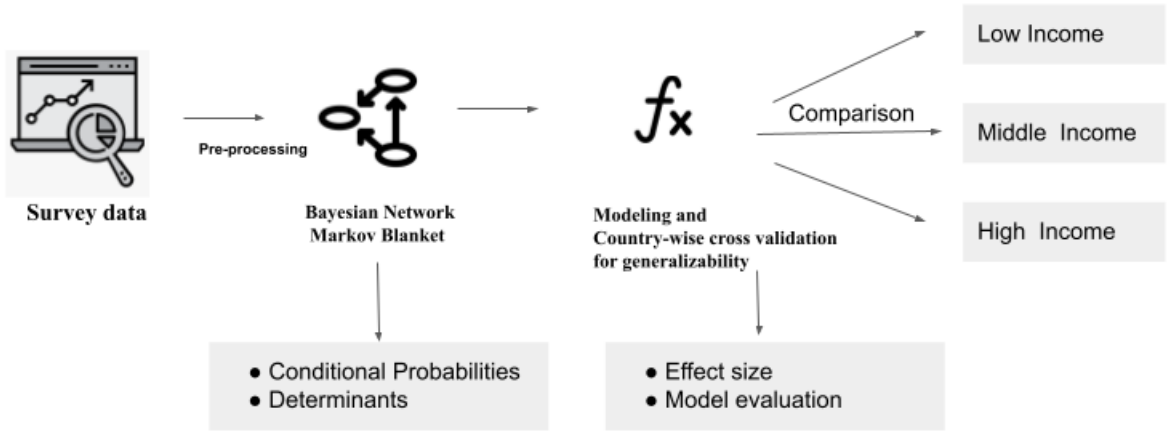


Figure 1: *Summary figure of the overall analysis, survey data was pre-processed to find the determinants of vaccine uptake using bayesian network markov blanket approach, conditional probabilities were analyzed for inference. The important determinants were used to build generalizable models and cross-validated on new countries, finally the effect sizes of determinants were computed and analyzed with income group stratification.*

medRxiv preprint doi: https://doi.org/10.1101/2023.01.31.23285300; this version posted February 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Bayesian Network Analysis for Learning Global Determinants of Vaccine Hesitancy:

A Bayesian network [18] is a probabilistic graphical model that depicts the joint probability distribution of a group of random variables with the potential for mutual causation. The Bayesian network is a directed acyclic graph with nodes representing random variables, edges between pairs of nodes expressing the causal link between these nodes, and a conditional probability distribution inside each node. The primary objective of Bayesian network analysis is to model complex systems and understand the relationships between different variables [19,20]**.** In our study, a Bayesian network model will be used to identify the key determinants of vaccine hesitancy and understand how these determinants interact with each other. We used a robust structure of the learnt Bayesian network (BN) by including a bootstrapping and ensemble averaging of edge directions in order to overcome the complex dependencies with potential confounding effects, mediation, and intercausal dependency regarding our outcome [21–24]. The best probabilistic graphical model that adequately explained the data was chosen using the hill climbing optimizer and a score based on the Akaike information criteria [25]. Over 11 BNs, we used bootstrapped learning with majority voting to learn Bayesian networks [20] . The package *wiseR* in R was used to do the analysis [26]. Furthermore,for community (clusters among nodes in BN) detection in the learnt network, we have used *linkcomm* R package[27]**.**

Predictive modeling

medRxiv preprint doi: https://doi.org/10.1101/2023.01.31.23285300; this version posted February 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Predictive vaccine uptake behavior might be a helpful tool for public health professionals, as it can help identify populations at risk of vaccine hesitancy and allow targeted interventions to be implemented in order to increase vaccine uptake. We used the GLMM to predict vaccine uptake among the survey respondents [28,29]. Variable Country was used as the random effect . We used the variables obtained in the markov blanket bayesian network analysis as the important predictors. The model equation is given by

$$y = x\boldsymbol{\beta} + Z\boldsymbol{\mu} + \boldsymbol{\epsilon}$$

Where, y is nx1 matrix of binary responses, i.e. vaccine uptake agree or disagree, x is nxp matrix of predictors, important survey questionnaire as obtained from the markov blanket, $\boldsymbol{\beta}$ is the p x1 vector of regression coefficients, Z is the Nxq for q random effects, i.e. Country. $\boldsymbol{\mu}$ is qx1 vector of random effects and $\boldsymbol{\epsilon}$ is a Nx1 vector of residuals that of y is not explained by model.

For testing the performance of models on new countries, we divided data into country-wise 5 fold splits, with 80% countries in the model development and 20% countries in the testing set in each split. The model evaluation resulted in generalized performance in the new countries. We also developed the income specific models and computed the effects sizes and model performances.

Results:

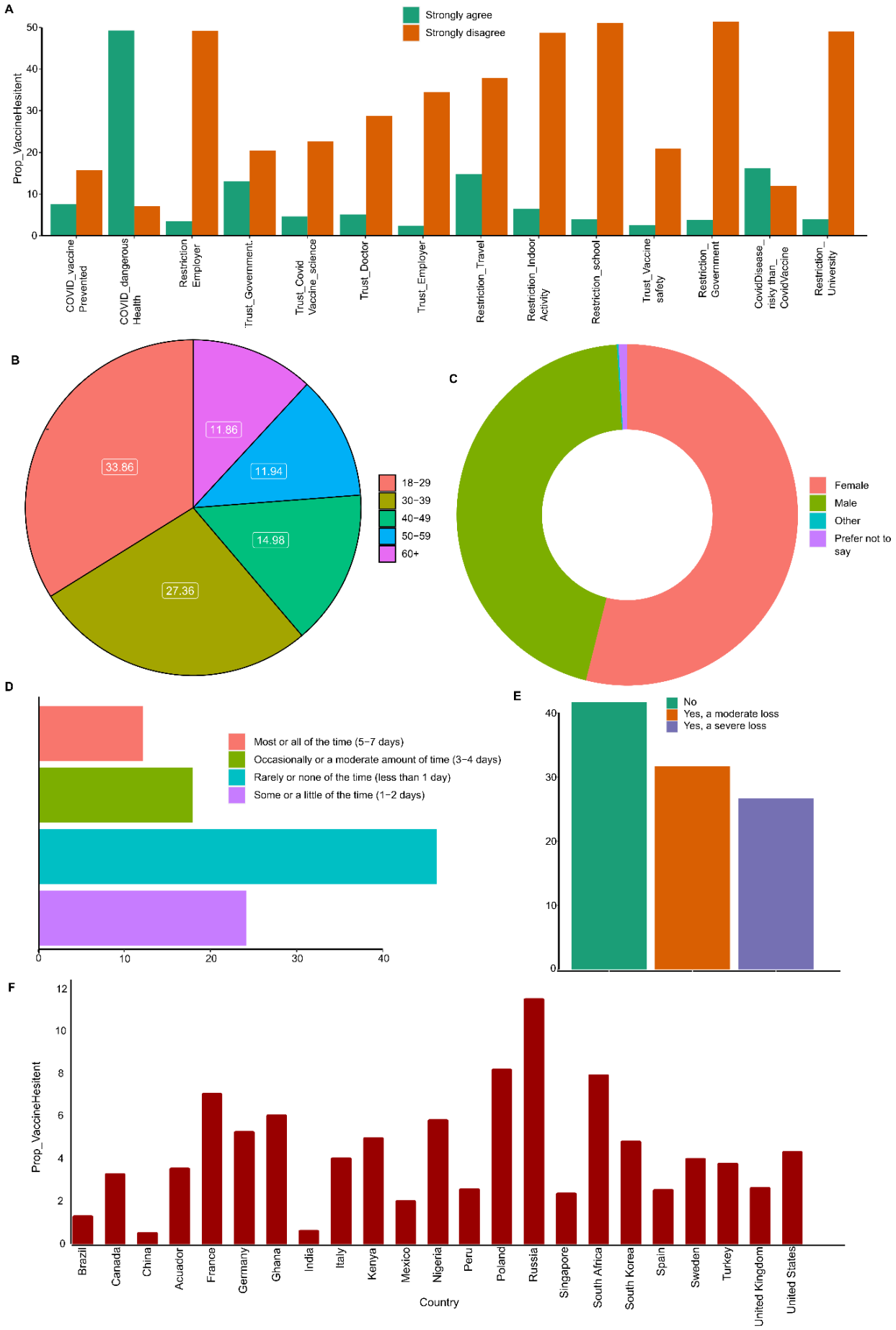
Data characteristics:

The descriptive statistical analysis showed some pattern of vaccine hesitancy on socioeconomic factors, opinions on COVID-19 and COVID-19 vaccine, trust on government, doctors and employer, and necessity of COVID-19 vaccine on college, school, travel and employment. A

medRxiv preprint doi: https://doi.org/10.1101/2023.01.31.23285300; this version posted February 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

descriptive association between behavioral factors, socioeconomic factors and vaccine hesitancy and key determinants of policy is given in **Figure 2**. We found that the proportion of respondents who are hesitant to receive vaccines decreases with age. Similarly, respondents with higher anxiety disorders tend to have a lower probability of vaccine hesitancy. Change in household income is a strong indicator of behavior change. In fact, we found that people whose household income is severely affected during COVID-19 are less likely to have vaccine hesitancy. Overall data depicts the disparities in vaccine hesitancy among people of various ages, mental health conditions, trust-related behavior indicators, and impact of COVID-19 related indicators . Based on the prevalence of vaccination hesitancy across countries, we found that Russia has the highest rate (11.51%), followed by Poland (8.23%) and South Africa (7.95%). Moreover, we found that India has the lowest rate of vaccination hesitancy, at 0.57%.

medRxiv preprint doi: https://doi.org/10.1101/2023.01.31.23285300; this version posted February 2, 2023. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .



medRxiv preprint doi: https://doi.org/10.1101/2023.01.31.23285300; this version posted February 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Figure 2: A) *Proportion of vaccine hesitancy characterized based on opinions* ***B)*** *Distribution of vaccine hesitancy proportion on different age-groups.* ***C)*** *Distribution of vaccine hesitancy proportion on different genders:* ***D)*** *Vaccine hesitancy and recent trends of the number of days that a respondent felt depressed.* ***E)*** *Vaccine hesitancy and household income loss during COVID-19* ***F)*** *Vaccine Hesitancy distribution by Country.*

Bayesian Network structure learning and network community/cluster detection to visualize global determinants:

In the learned bootstrapped network, we have only considered the nodes' connections that have edge strengths greater than 0.51(majority of voting in bootstrapped network). Then, in the bootstrapped Bayesian network, we first performed node clustering to visualize the set of key indicators that have a high impact on the outcome variable.. Our algorithm detected 11 clusters of nodes, and the largest cluster contains 15 nodes (**Figure 3A**). Our outcome variable was found in the 5 clusters (Clusters # 2,3,4,7,8) among all 11 identified clusters. Every cluster includes some specific features. For example, cluster #1 compromise about socio-demographic information such as age, gender, and monthly income. Similarly, cluster#8 has 15 nodes, making it the largest cluster. This cluster includes variables associated to COVID-19 and COVID-19 vaccinations, and household income changes during COVID-19. With trust in stakeholders (Doctor, Employer, Government), COVID-19 vaccination necessity for international travel, and indoor activities. Since communities reveal the clusters of variables, it is also necessary, from a policy perspective, to determine which variables are mainly affecting vaccine hesitancy. Next,

medRxiv preprint doi: https://doi.org/10.1101/2023.01.31.23285300; this version posted February 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

we calculated the markov blanket of outcome variables in order to derive conclusions about the main factors influencing our outcome variable.

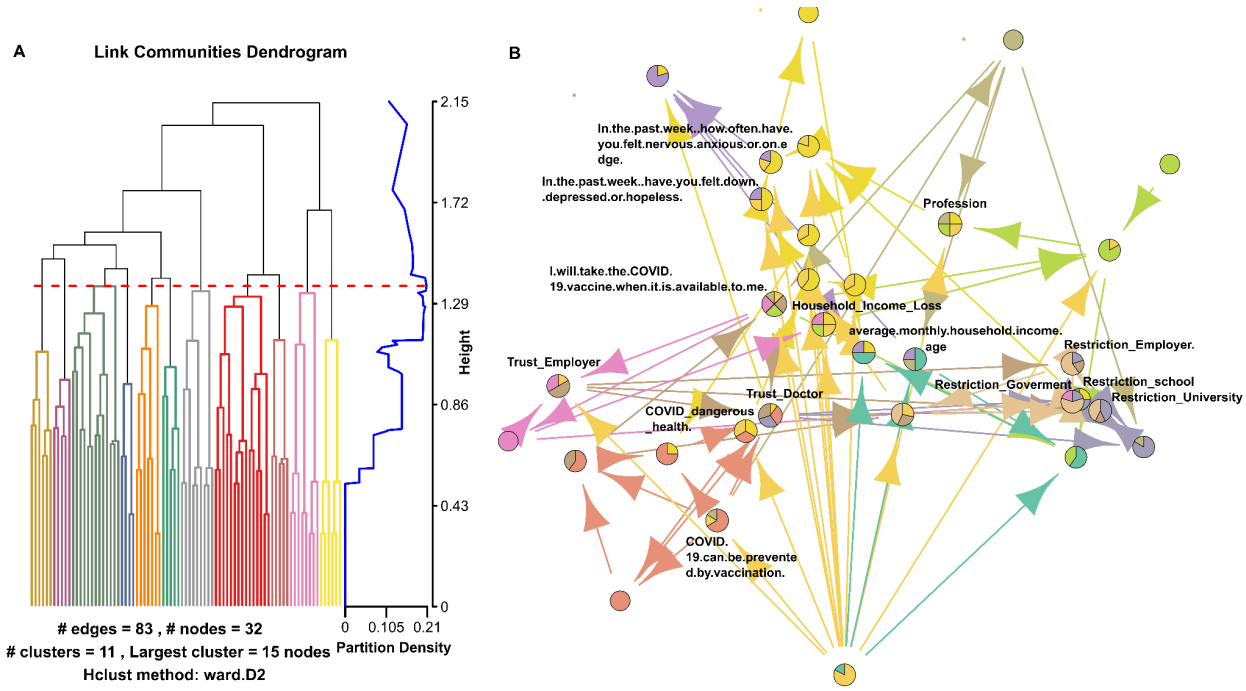


Figure 3: A) extracting link communities from the entire Bayesian network: *11 clusters/communities were identified in the Bayesian network with the largest cluster of 15 nodes.* ***B) Network Structure with learned community:*** *To aid the visualization of key nodes we can limit the display of nodes that belong to 3 or more communities*

From the markov blanket analysis, we found that trust and restriction are the major categories of intervention that impact vaccine hesitancy. Furthermore, our analysis will be focused on these two categories..

medRxiv preprint doi: https://doi.org/10.1101/2023.01.31.23285300; this version posted February 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Association between trust towards stakeholders and vaccine uptake: The results showed that the public's trust in COVID-19 vaccines and vaccination is a key for vaccines' success **(Figure 4B)**. Our Bayesian network analysis found that people's faith in their doctors, government officials, and the science behind COVID-19 had a substantial effect on their vaccination behavior. Our analysis revealed a 23% increase in the likelihood that a person would get a vaccination if they had confidence in physicians. Similarly, we discovered a 23% increase in the likelihood of vaccination acceptance when individuals had confidence in government authorities. The Bayesian network also identified confidence in the science behind COVID-19 as a vital element for vaccine uptake behavior, and a 19% increase in the likelihood of vaccine uptake behavior if individuals had faith in the research underlying COVID-19. Also, the results have shown that a person's likelihood of getting a vaccination goes up by 18% if they trust their employer.

Association between social restrictions and vaccine hesitancy: COVID-19 has halted the socio-economic vehicles of the entire world, which was required to lessen the acceleration of viral spread. Multiple sectors, including travel, employment, colleges, social gatherings, and events, have mandated comprehensive COVID-19 immunization since its introduction. This mandate has raised the vaccination rate among individuals. We found that foreign travel limitations have boosted the likelihood of vaccination uptake by 15%. Mandatory vaccines in workspace and in schools/universities (colleges) limits in job colleges and schools have raised the likelihood of vaccination uptake by 15% and 14%, respectively. With the advent of cinema and concerts, it is believed that these activities would become the focal point of social gatherings and in certain countries, vaccination were also made obligatory. We found that vaccination requirement criteria had led to a 14% rise in the decision to get vaccinated **(Figure 4C)**.

medRxiv preprint doi: https://doi.org/10.1101/2023.01.31.23285300; this version posted February 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

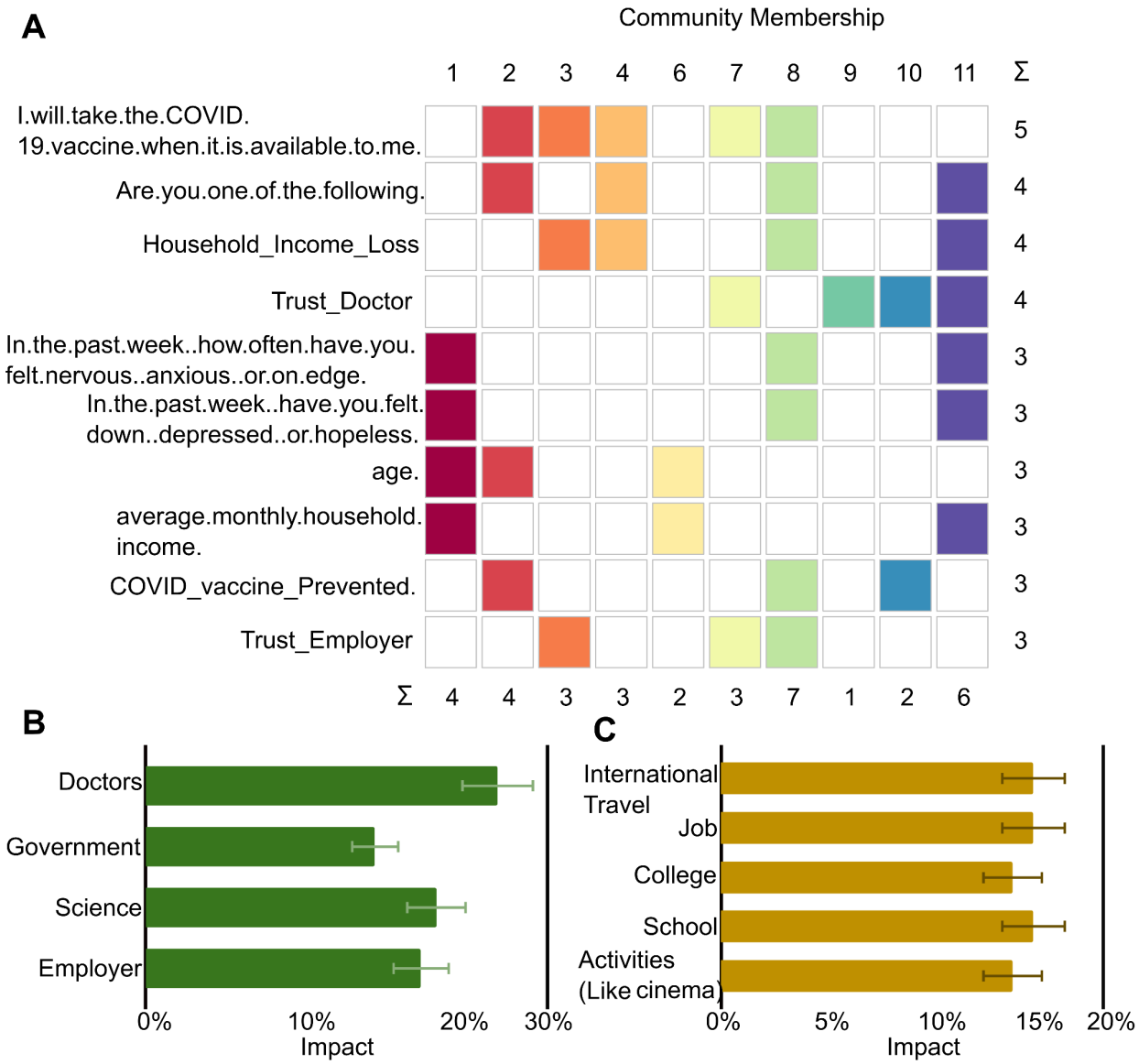


Figure 4: A) *Node community membership for the top-connected nodes: B) Association of vaccine uptake with trust on stakeholders C) Association of vaccine uptake with public activity.*

medRxiv preprint doi: https://doi.org/10.1101/2023.01.31.23285300; this version posted February 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Prediction of vulnerable individuals to maximize vaccine uptake

We found based on the GLMM that the Trust on the central governments as the one the key predictors with log-odds ratio greater than 0.55[0.25, 0.84] (p-value< 0.0002), is equally significant across different income groups as shown in Figure 5b. We also found that the restrictions on international travel as important trigger for vaccine uptake with log-odds of 0.4[0.15, 0.65] (p-value= 0.0034) for higher income countries, 0.64[0.35, 0.93] (p-value=0.00016) for lower middle income countries, and 0.79[0.54, 1.04] (p-value= $3.69*10^{-10}$) for upper middle income countries. Age group-wise, individuals with age-group 60+ have higher positive odds of vaccine uptake in lower middle income countries contrary to higher and upper middle income countries **(Figure 5A, 5B)**. Income groups having higher than the country median income have significant odds of vaccine uptake.

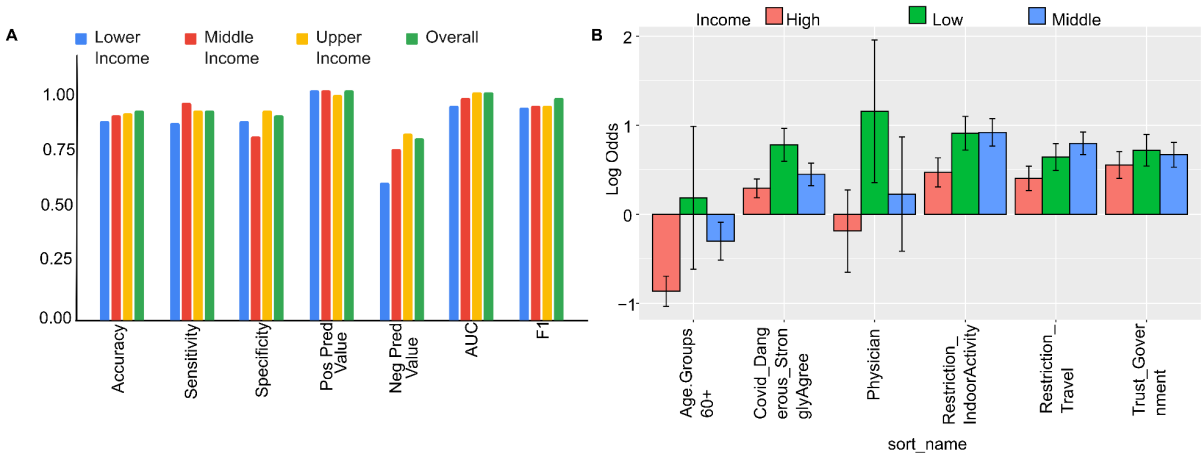


Figure 5: A) Prediction model results : *Modeling results using Generalized mixed effects models, model achieved an AUC of 89% for the generalized prediction across different countries with Positive predictive value of ~90% with Sensitivity of 82%.* **B) Top 5 features in prediction** : *The top 5 features were selected from the Mixed effects model based on their coefficient*

medRxiv preprint doi: https://doi.org/10.1101/2023.01.31.23285300; this version posted February 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

magnitude in the model for low, middle, and high income countries. We plotted the coefficient estimate(Log odds) of the top features' union (Low Income, Middle Income, and High Income), and the error bar represents the standard error. In the prediction model, we again find the trust and social restriction indicators to be the most important

Discussion:

Multiple stakeholders, including government decision-makers, community, health personnel, media sources, and internet platforms, impact vaccination rates [30]. These factors may encourage or discourage vaccination by fostering more or less permissive environments. Consequently, it is essential to evaluate how the health system factors may impact public behavior. Several data-driven models and theoretical studies have been done to uncover vaccine uptake indicators. Other machine-learning models were also built to predict the vaccine adherence paradigm [31–34].

Here, we analyzed data from various countries to determine the variables influencing vaccination rates. We also evaluated the performance of machine learning prediction models in high-income, lower-middle-income, and high-middle-income nations. We found that our models outperformed similar efforts in the literature with a high margin of accuracy compared to studies in the literature[35-37]; however, there were few differences concerning the choice of data and methods. Previous studies like Lincoln et al surveyed individuals from five different high income countries to understand factors driving vaccine hesitancy during the COVID-19 pandemic. Results showed that only 57.4% of participants indicated they would definitely or probably get vaccinated. A machine learning model was used to identify predictors of vaccine hesitancy, such as vaccination conspiracy beliefs, mistrust, age, income, and population density, with accuracy

medRxiv preprint doi: https://doi.org/10.1101/2023.01.31.23285300; this version posted February 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

(82% sensitivity and 79-82% specificity) [35]. Carrieri et al. used the supervised random-forest machine-learning algorithm on area-level indicators of institutional and socioeconomic backgrounds to predict the vaccine hesitancy rate for Italian local authorities, which helped public-health practitioners run targeted awareness campaigns. The study found that non-clinical features like employment rate, proportion of waste recycling had the highest predictive powers in the random-forest algorithm, with an AUC of 0.836 [36] . Santillana et al. used Google searches, Twitter microblogs, nearly real-time hospital visit records, and data from a participatory surveillance system data with ensemble methods of stacking regression that combine separate outcomes from each model of different statistical classifiers to predict vaccinations and achieved a testing accuracy of 85.71% in the 10-fold cross-validation [37]. This study examines the state-level features and policies that are most important in achieving a threshold level vaccination rate using a decision tree algorithm on a dataset of all states in the US. The study finds that workplace travel is the most important predictor of vaccination rates, The study also uses alternative algorithms to confirm these findings, with accuracy ranging from 80-88% and sensitivity from 92.5-100%. The findings provide actionable policy insights to increase vaccination rates and combat the COVID-19 pandemic [38].

However, previous studies have focused on specific geographical areas and have not investigated the global factors that influence vaccination rates or the generalizability of machine learning models for predicting vaccine hesitancy. In this study, using a survey questionnaire with 23135 respondents across 23 countries. We used Bayesian networks and generalized linear models to identify the critical determinants of vaccine uptake. We then used Markov Blankets obtained from the Bayesian networks to estimate the vital predictors of vaccine uptake, which were used to build our models. We tested the generalizability of our models across countries and income groups. Our study found that trust in doctors had a greater impact on vaccination behavior than

medRxiv preprint doi: https://doi.org/10.1101/2023.01.31.23285300; this version posted February 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

trust in other stakeholders. Factors such as faith in vaccine efficacy and safety, and trust in institutions and government decisions also play a role in vaccination uptake. The use of social constraints, such as restrictions on certain activities for those who are not vaccinated, may also influence behavior [39-40]. Further, our generalized mixed effects model approach achieved an AUC of 89%, Precision of 90%, and Recall of 82% on the prediction task on new countries, indicating its potential effectiveness in predicting vaccine uptake in contemporary populations.

Our study has several strengths, including large sample size and a rigorous methodology for identifying and predicting vaccine uptake behavior. Our approach also has the potential to be applied on a global scale, making it a valuable tool for policymakers looking to improve vaccination programs. However, our study has some limitations, like the fact that it is challenging to derive causal influences from observational data, and our study does not claim to do that. However, the usefulness of Bayesian networks as probabilistic reasoning systems is well known but little applied in public health settings, including we believe that additional data is required to explore in more detail and determine the feasibility of AI-driven techniques for enhancing global vaccination programmes. In addition, since the study deals with sensitive public health issues, the responses to the survey questions may have some degree of erroneousness. Hence, it is difficult to corroborate the associations that were discovered in a data-driven way.

Finally, as a conclusion, despite the development of numerous machine learning models in public health, there are only a handful of models that are explainable and generalizable, In this study, we provided a framework that not only identified explainable policy solutions for COVID-19 vaccination but could also be adopted for any data-driven interventional modeling studies.

.

medRxiv preprint doi: https://doi.org/10.1101/2023.01.31.23285300; this version posted February 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Funding: This work was supported by Pathcheck foundation's data informatics center for Epidemiology(DICE), computational health fellowship program.

Acknowledgement: We acknowledge Graham Dodge , president at PathCheck foundation and Bethany LoMonaco, Director, strategy and development for their constant support. We acknowledge the funding support by Pathcheck foundation awarded to Dr. Aditya Nagori and Mr. Raghav Awasthi.

References:

1. Domachowske J, Suryadevara M. Vaccines: A Clinical Overview and Practical Guide. Springer Nature; 2020.

2. MacDonald NE, SAGE Working Group on Vaccine Hesitancy. Vaccine hesitancy: Definition, scope and determinants. Vaccine. 2015;33: 4161–4164.

3. National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Population Health and Public Health Practice, Board on Health Sciences Policy, Committee on Equitable Allocation of Vaccine for the Novel Coronavirus. Framework for Equitable Allocation of COVID-19 Vaccine. National Academies Press; 2020.

4. Vashishtha VM, Kalra A. IAP Textbook of Vaccines. Jaypee Brothers Medical Publishers; 2020.

5. Sallam M, Al-Sanafi M, Sallam M. A Global Map of COVID-19 Vaccine Acceptance Rates per Country: An Updated Concise Narrative Review. J Multidiscip Healthc. 2022;15: 21.

medRxiv preprint doi: https://doi.org/10.1101/2023.01.31.23285300; this version posted February 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

6. Roshchina Y, Roshchin S, Rozhkova K. Determinants of COVID-19 vaccine hesitancy and resistance in Russia. Vaccine. 2022. pp. 5739–5747. doi:10.1016/j.vaccine.2022.08.042

7. Jung Y, Lee S. Trump vs. the GOP: Political Determinants of COVID-19 Vaccine Hesitancy. SSRN Electronic Journal. doi:10.2139/ssrn.3966799

8. Piltch-Loeb R, Silver D, Kim Y, Norris H, McNeill E, Abramson D. Determinants of the COVID-19 Vaccine Hesitancy Spectrum. doi:10.1101/2021.08.05.21261675

9. Fernandes N, Costa D, Costa D, Keating J, Arantes J. Predicting COVID-19 Vaccination Intention: The Determinants of Vaccine Hesitancy. Vaccines. 2021. p. 1161. doi:10.3390/vaccines9101161

10. Van Khuc Q. Determinants of COVID-19 vaccine hesitancy and implications for improving vaccination strategy and public health policy. doi:10.31219/osf.io/685a9

11. Rosental H, Shmueli L. Integrating Health Behavior Theories to Predict COVID-19 Vaccine Acceptance: Differences between Medical Students and Nursing Students. doi:10.1101/2021.05.18.21257416

12. Choi KH, Denice PA, Ramaj S. Vaccine and COVID-19 Trajectories: Equal vaccine rates do not reduce inequality in COVID-19 rates. doi:10.31235/osf.io/27dvz

13. Attwell K, Hannah A, Leask J. COVID-19: talk of “vaccine hesitancy” lets governments off the hook. In: Nature Publishing Group UK [Internet]. 22 Feb 2022 [cited 6 Jan 2023]. doi:10.1038/d41586-022-00495-8

14. Rana NP, Slade EL, Sahu GP, Kizgin H, Singh N, Dey B, et al. Digital and Social Media

medRxiv preprint doi: https://doi.org/10.1101/2023.01.31.23285300; this version posted February 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Marketing: Emerging Applications and Theoretical Development. Springer Nature; 2019.

15. Lazarus JV, Wyka K, White TM, Picchio CA, Rabin K, Ratzan SC, et al. Revisiting COVID-19 vaccine hesitancy around the world using data from 23 countries in 2021. Nat Commun. 2022;13: 1–14.

16. Lazarus JV, Binagwaho A, El-Mohandes AAE, Fielding JE, Larson HJ, Plasència A, et al. Keeping governments accountable: the COVID-19 Assessment Scorecard (COVID-SCORE). Nat Med. 2020;26: 1005–1008.

17. White TM, Cash-Gibson L, Martin-Moreno JM, Matesanz R, Crespo J, Alfonso-Sanchez JL, et al. COVID-SCORE Spain: public perceptions of key government COVID-19 control measures. Eur J Public Health. 2021;31: 1095–1102.

18. Pearl J. BAYESIAN INFERENCE. Probabilistic Reasoning in Intelligent Systems. 1988. pp. 29–75. doi:10.1016/b978-0-08-051489-5.50008-4

19. Awasthi R, Rakholia V, Agrawal S, Dhingra LS, Nagori A, Kaur H, et al. Estimating the impact of health systems factors on antimicrobial resistance in priority pathogens. J Glob Antimicrob Resist. 2022;30: 133–142.

20. Jha IP, Awasthi R, Kumar A, Kumar V, Sethi T. Learning the Mental Health Impact of COVID-19 in the United States With Explainable Artificial Intelligence: Observational Study. JMIR Ment Health. 2021;8: e25097.

21. Darwiche A. Modeling and Reasoning with Bayesian Networks. Cambridge University Press; 2009.

medRxiv preprint doi: https://doi.org/10.1101/2023.01.31.23285300; this version posted February 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

22. Pearl J. Causality. Cambridge University Press; 2009.

23. Pearl J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Elsevier; 2014.

24. Pearl J, Mackenzie D. The Book of Why: The New Science of Cause and Effect. Penguin UK; 2018.

25. Yehezkel R. Bayesian Network Structure Learning Using Recursive Autonomy Identification. 2004.

26. Maheshwari S, Pahwa K, Sethi T. WiseR: An end-to-end structure learning and deployment framework for causal graphical models. 2021 [cited 17 Nov 2022]. doi:10.48550/arXiv.2108.07046

27. Kalinka AT, Tomancak P. linkcomm: an R package for the generation, visualization, and analysis of link communities in networks of arbitrary size and type. Bioinformatics. 2011;27: 2011.

28. Faraway JJ. Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models. CRC Press; 2016.

29. Wu L. Mixed Effects Models for Complex Data. CRC Press; 2009.

30. Erchick DJ, Gupta M, Blunt M, Bansal A, Sauer M, Gerste A, et al. Understanding determinants of vaccine hesitancy and acceptance in India: A qualitative study of government officials and civil society stakeholders. PLoS One. 2022;17. doi:10.1371/journal.pone.0269606

medRxiv preprint doi: https://doi.org/10.1101/2023.01.31.23285300; this version posted February 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

31. Chen C. Dissemination of Vaccine Misinformation on Twitter and Its Countermeasures. 2021.

32. Chang V, Kaur H, Fong SJ. Artificial Intelligence and Machine Learning Methods in COVID-19 and Related Health Diseases. Springer Nature; 2022.

33. Oliva D, Hassan SA, Mohamed A. Artificial Intelligence for COVID-19. Springer Nature; 2021.

34. Romate J, Rajkumar E, Greeshma R. Using the integrative model of behavioural prediction to understand COVID-19 vaccine hesitancy behaviour. Sci Rep. 2022;12: 1–13.

35. Lincoln TM, Schlier B, Strakeljahn F, Gaudiano BA, So SH, Kingston J, et al. Taking a machine learning approach to optimize prediction of vaccine hesitancy in high income countries. Sci Rep. 2022;12: 1–12.

36. Carrieri V, Lagravinese R, Resce G. Predicting vaccine hesitancy from area-level indicators: A machine learning approach. Health Econ. 2021;30. doi:10.1002/hec.4430

37. Santillana M, Nguyen AT, Dredze M, Paul MJ, Nsoesie EO, Brownstein JS. Combining Search, Social Media, and Traditional Data Sources to Improve Influenza Surveillance. PLoS Comput Biol. 2015;11: e1004513.

38. Predictors of COVID-19 vaccination rate in the USA: A machine learning approach. Machine Learning with Applications. 2022;10: 100408.

39. Chen X, Lee W, Lin F. Infodemic, Institutional Trust, and COVID-19 Vaccine Hesitancy: A Cross-National Survey. Int J Environ Res Public Health. 2022;19.

medRxiv preprint doi: https://doi.org/10.1101/2023.01.31.23285300; this version posted February 2, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

doi:10.3390/ijerph19138033

40. Wells CR, Galvani AP. The interplay between COVID-19 restrictions and vaccination. Lancet Infect Dis. 2021;21: 1053–1054.