Developing and Validating Predictive Models of Vaccine Hesitancy Among Parents in the United States

Yi Zheng[1], PhD, MPH; Paula M. Frew[1], PhD, MA, MPH; Dong Wang[1], PhD; Yan Song[2], PhD; Oscar Patterson-Lomba[2]; Arshya Feizi[2], PhD; Tayler Li[2], MS; Amanda L. Eiden*, PhD, MPH, MBA

[1]Merck & Co., Inc., Rahway, NJ, USA; [2]Analysis Group, Inc, Boston, MA, USA
*Corresponding author

Background

- Vaccines are among the most cost-effective clinical preventative services and are a core component of any preventative services package, with a high economic and humanistic return on investment
- However, in many parts of the world, people challenge the existing evidence of the value, efficacy, and safety of vaccines and refuse vaccination for themselves or their children
- Vaccine hesitancy is multifaceted and a context-specific challenge that varies by time, location, and vaccine type, affecting vaccination uptake
- There are a lack of studies comprehensively examining the simultaneous role of multiple predictors in vaccine hesitancy among parents of children less than 18 years of age

Objective

To better understand the factors that drive parental decision-making with respect to vaccinating their children using conceptually different machine learning algorithms to analyze survey responses

Methods

Data source

- A cross-sectional online survey of parents (N=692) of children under 18 years old in the US in 2022, a sample from the National Health and Wellness Survey (NHWS)

Outcome measurement

- Vaccine hesitancy:

  "Overall, how hesitant about childhood shots (vaccines) would you consider yourself to be?"
  Yes: "Very hesitant," "Somewhat hesitant," and "Not sure" (49.7%)
  No: "Not too hesitant" and "Not at all hesitant" (50.3%)
- Vaccine literacy:

  "How familiar are you with the vaccines your child should receive?"
  Low: "Not at all familiar," "Slightly familiar," or "Somewhat familiar" (25.6%)
  High: "Moderately familiar" and "Extremely familiar" (74.4%)

Predictors

- More than 500 predictors were evaluated including factors characterizing information seeking behavior, attitudes, and beliefs towards children's vaccines, and access to care
- Maximal information coefficient (MIC), which captures linear and non-linear relationships between the outcomes and all the features, was used to rank and select features. Models were trained using the top 10, 25, 40, 50, 80, and 100 features based on MIC scores

Model development and validation

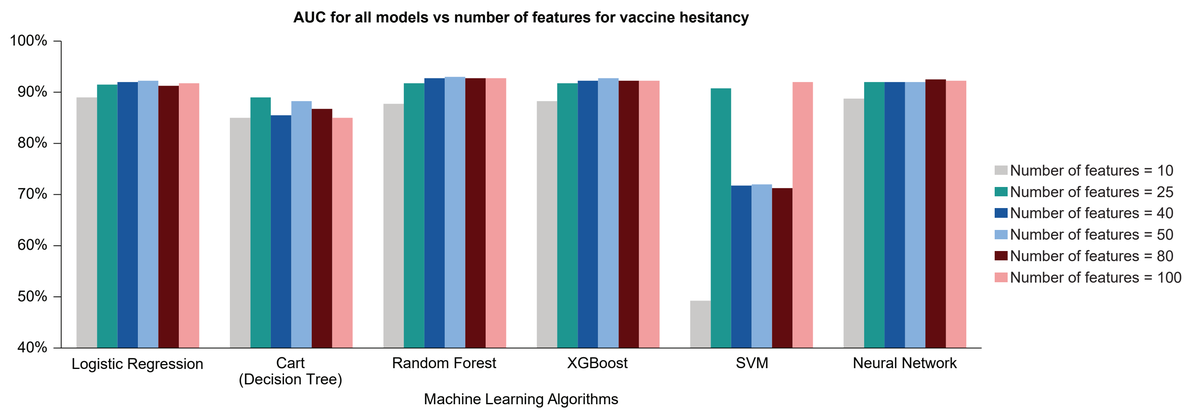
- Data were split into training (80%) and testing (20%) sets
- We trained the model using the following algorithms:
  - Logistic regression (benchmark model)
  - Basic decision tree (Classification And Regression Tree, CART)
  - Random forest
  - Extreme Gradient Boosting (XGBoost)
  - Support Vector Machine (SVM)
  - Neural network (Multilayer perceptron, MLP)
- Bayesian optimization approach was used for hyperparameter tuning, maximizing precision-recall AUC
- Each training was completed using a 10-fold cross-validation approach
- Prediction performance was evaluated in the testing set using accuracy, precision, recall, F1 score, area under the receiver operating characteristic curve (ROC-AUC), and area under the precision-recall curve (PR-AUC) for imbalanced data (ie, vaccine literacy)
- Important features were extracted from the model with highest performance based on SHAP (SHapley Additive exPlanations) values

Results

Table 1. Distribution of baseline characteristics among parents by vaccine hesitancy and literacy (n=692)

	Vaccine hesitancy			Vaccine literacy			Total
	Yes	No	*P*-value	Low	High	*P*-value	
	N (%)						
Overall	344 (49.7)	348 (50.3)		177 (25.6)	515 (74.4)		
Age			**<0.001**			**0.786**	
18-26	19 (5.5)	10 (2.9)		9 (5.1)	20 (3.9)		29 (4.2)
27-35	117 (34.0)	71 (20.4)		45 (25.4)	143 (27.8)		188 (27.2)
36-45	172 (50.0)	182 (52.3)		94 (53.1)	260 (50.5)		354 (51.2)
46+	36 (10.5)	85 (24.4)		29 (16.4)	92 (17.9)		121 (17.5)
Gender			**0.040**			**0.039**	
Female	182 (52.9)	212 (60.9)		113 (63.8)	281 (54.6)		394 (56.9)
Male	162 (47.1)	136 (39.1)		64 (36.2)	234 (45.4)		298 (43.1)
Race/ethnicity			**0.004**			**0.021**	
Non-Hispanic White	233 (67.7)	226 (64.9)		104 (58.8)	355 (68.9)		459 (66.3)
Non-Hispanic Black	36 (10.5)	33 (9.5)		22 (12.4)	47 (9.1)		69 (10.0)
Asian	15 (4.4)	35 (10.1)		21 (11.9)	29 (5.6)		50 (7.2)
Hispanic	54 (15.7)	38 (10.9)		26 (14.7)	66 (12.8)		92 (13.3)
Others	6 (1.7)	16 (4.6)		4 (2.3)	18 (3.5)		22 (3.2)
Education			**0.035**			**0.011**	
High school or less	55 (16.0)	45 (12.9)		36 (20.3)	64 (12.4)		100 (14.5)
Some college	70 (20.3)	81 (23.3)		45 (25.4)	106 (20.6)		151 (21.8)
College	101 (29.4)	130 (37.4)		54 (30.5)	177 (34.4)		231 (33.4)
Graduate school	118 (34.3)	92 (26.4)		42 (23.7)	168 (32.6)		210 (30.3)
Income (USD)			**0.950**			**0.006**	
<$50,000	73 (21.4)	74 (21.7)		53 (29.9)	94 (18.6)		147 (21.6)
$50,000 - $100,000	124 (36.4)	120 (35.2)		54 (30.5)	190 (37.6)		244 (35.8)
≥$100,000	144 (42.2)	147 (43.1)		70 (39.5)	221 (43.8)		291 (42.7)
Region			**0.497**			**0.51**	
Northeast	59 (17.5)	72 (21.0)		32 (18.4)	99 (19.5)		131 (19.2)
Midwest	77 (22.8)	83 (24.2)		48 (27.6)	112 (22.1)		160 (23.5)
South	133 (39.3)	118 (34.4)		59 (33.9)	192 (37.9)		251 (36.9)
West	69 (20.4)	70 (20.4)		35 (20.1)	104 (20.5)		139 (20.4)
Urbanity			**<0.001**			**0.002**	
Urban	211 (61.3)	128 (36.8)		67 (37.9)	272 (52.8)		339 (49.0)
Suburban	98 (28.5)	166 (47.7)		84 (47.5)	180 (35.0)		264 (38.2)
Rural	35 (10.2)	54 (15.5)		26 (14.7)	63 (12.2)		89 (12.9)

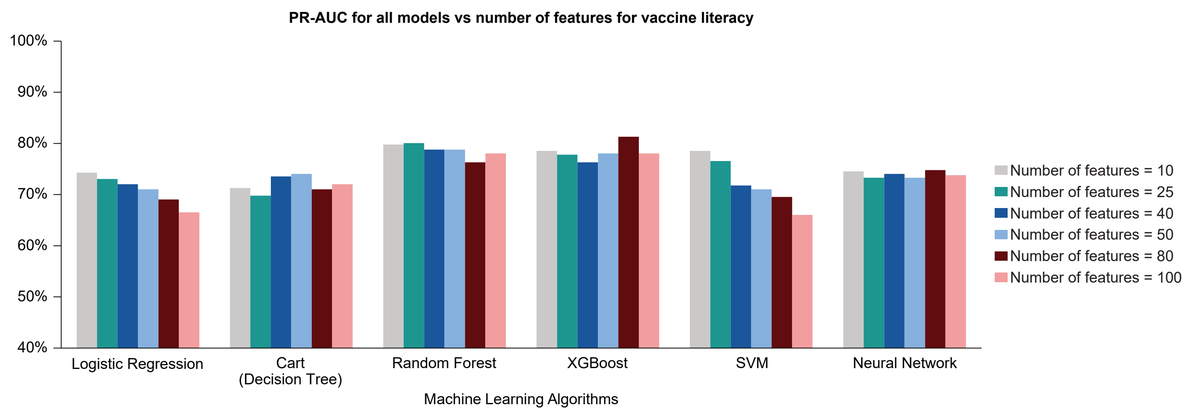
Figure 1. Performance of predictive models for vaccine hesitancy



For vaccine hesitancy, the highest AUC score (93.0%) was achieved using the random forest model using 50 features (Figure 1). Below are the top 5 predictors:

1. Disagreement with the sentiment that there is no need for my child to get vaccinated because everybody else does predicts *lower* hesitancy
2. Disagreement with the sentiment that I do not like the idea of vaccines for my child predicts *lower* hesitancy
3. Disagreement with the sentiment that children get more vaccinations that are good for them predicts *lower* hesitancy
4. Disagreement with the sentiment that healthy children do not need vaccinations predicts *lower* hesitancy
5. Saying "yes" to wanting their new infant to get all the recommended shots predicts *lower* hesitancy

Figure 2. Performance of models for vaccine literacy



For vaccine literacy, the highest PR-AUC score was 81.25% for the XGBoost model using 80 features (Figure 2). Below are the top 5 predictors:

1. A low familiarity with the vaccine schedules is highly predictive of limited vaccine literacy
2. A medium/low reported influence in the vaccine schedules is predictive of limited vaccine literacy
3. Parents not knowing whether their children receive their vaccines during specified times (or not allowing the child to be vaccinated) is predictive of limited vaccine literacy
4. Understanding information on vaccines is predictive of higher vaccine literacy
5. Knowledge of chickenpox/varicella is predictive of higher vaccine literacy

Summary of findings

- Overall, more hesitant parents were more likely to be younger, male, Hispanic/Latino, and reside in urban areas
- Random forest model performed the best in predicting vaccine hesitancy (F1 score = 0.86, ROC-AUC = 93.00%), followed by XGBoost (F1 score = 0.84, ROC-AUC = 92.75%)
- The belief in "no need for their children to get vaccinated because everybody else does" contributed significantly to higher hesitancy, followed by beliefs related to a "lack of trust in vaccines," children "getting too many vaccines," and "healthy children don't need vaccines"
- The model also reflected that information-seeking challenges and concerns over safety, efficacy, and side effects were strong predictors of attitudinal shaping
- XGBoost model performed the best in predicting limited vaccine literacy (PR-AUC = 81.25%)
- Low familiarity with vaccine schedules, medium/low parental influence on vaccine schedules, not allowing children to be vaccinated, and not understanding vaccination information received were predictive of limited vaccine literacy
- This study introduced an effective machine learning approach to help providers and policy makers understand and monitor factors that shape attitudes and influence behaviors towards vaccination and disentangle how parents interpret information discussed in shared clinical decision-making

References

1. Vanderslott S, et al. Our World in Data. Vaccination. https://ourworldindata.org/vaccine-preventable-diseases.
2. Larson HJ, et al. *Vaccine*. 2014;32(19):2150-2159.
3. Carrico J, et al. *Pediatrics*. 2022;150(3):e2021056007.
4. Frew PM, et al. *Vaccine*. 2016;34(46):5689-5696.

Contact information
Amanda Eiden, PhD, MBA, MPH (She/Her)
Email: amanda.eiden@merck.com

https://bit.ly/4fA29wI

Presented at ISPOR Europe; Barcelona, Spain; 17-20 November 2024.

W6005444_2024_MRL_IPOSR_Europe_Poster_1_Vaccine_Hesitancy-eidena_V1.0211/07/2024•OutputSize:33.1"x46.8"Scale:100%