Proactive advising: a machine learning driven approach to vaccine hesitancy

Andrew Bell
Nova School of Business and Economics
Lisbon, Portugal
andrew.bell@novasbe.pt

Dr. Alexander Rich
New York University
New York, NY, USA
asr443@nyu.edu

Melisande Teng
CentraleSupélec/ENS Paris-Saclay
Gif-sur-Yvette, France
melisande.teng@student.ecp.fr

Tin Orešković
IBM
Armonk, NY, USA
tin@ibm.com

Nuno B, Bras
Universidade Autónoma de Lisboa
Lisbon, Portugal
nbras@autonoma.pt

Lénia Mestrinho
Nova School of Business and Economics
Lisbon, Portugal
lenia.mestrinho@novasbe.pt

Srdan Golubovic
Croatian National Institute of Public Health
Zagreb, Crotia
srdan.golubovic@hzjz.hratia

Dr. Ivan Pristas
Croatian National Institute of Public Health
Zagreb, Croatia
ivan.pristas@hzjz.hr

Dr. Leid Zejnilovic
Nova School of Business and Economics
Lisbon, Portugal
leid.zejnilovic@novasbe.pt

***Abstract*—Despite once being nearly eradicated, Measles cases in Europe have surged to a 20-year high with more than 60,000 cases in 2018, due to a dramatic decrease in vaccination rates. The decrease in Measles, Mumps, and Rubella (MMR) vaccination rates can be attributed to an increase in 'vaccine hesitancy', or the delay in acceptance or refusal of vaccines despite their availability. Vaccine hesitancy is a relatively new global problem for which effective interventions are not yet established. In this paper, a novel machine learning approach to identify children at risk of not being vaccinated against MMR is proposed, with the objective of facilitating proactive action by healthcare workers and policymakers. A use case of the approach is the provision of individualized informative guidance to families that may otherwise become or are already vaccine hesitant. Using a LASSO logistic regression model trained on 44,000 child Electronic Health Records (EHRs), vaccine hesitant families can be identified with a higher precision (0.72) than predicting vaccine uptake based on a child's infant vaccination record alone (0.63). The model uses a low number of attributes of the child and his or her family and community to produce a prediction, making it readily interpretable by healthcare professionals. The implementation of the machine learning model into an open source dashboard for use by healthcare providers and policymakers as an Early Warning and Monitoring System (EWS) against vaccine hesitancy is proposed. The EWS would facilitate a wide variety of proactive, anticipatory and therefore potentially more effective public health interventions, compared to reactive interventions taken after vaccine rejections.**

I. INTRODUCTION

Despite once being nearly eradicated, Measles has spread across the globe due to a dramatic decrease in vaccination rates. In 2017, 25.465 measles cases were reported in Europe, a four-fold increase from a year earlier. The number of measles cases hit a record high in 2018 in the European region with 59.578 cases. In fact, in 20 of the 28 European Union countries, Measles, Mumps, and Rubella (MMR) vaccination rates are below the 95% herd immunity threshold, as recommended by the World Health Organization (WHO) [1]. The dramatic decrease in MMR vaccination can be attributed to an increase in "vaccine hesitancy," defined by the WHO's Strategic Advisory Group of Experts on Immunization Working Group (SAGE WG) as "the delay in acceptance or refusal of vaccines, despite their availability [2]."

The primary catalyst for the global surge in vaccine hesitancy is an erosion of public trust in the effectiveness and safety of vaccines [3]. In 2011, the WHO EURO Vaccine Communications Working Group proposed the "3 Cs" model to better explain the determinants of vaccine refusal. The first "C," confidence, refers to the public trust in the effectiveness and safety of vaccines, and the competency and motivations of healthcare workers and policymakers. Second, complacency occurs when the perceived threat of a vaccine-preventable disease is low. Third, convenience refers to the consideration of geographic accessibility, sufficient supply, and the willingness-to-pay of vaccines [4]. In 2015, Betsch, Bohm, and Chapman added a fourth "C" to the model: the rational calculation of the pros and cons of vaccination [5].

Researchers from a range of different fields have tested interventions aimed at decreasing vaccine hesitancy. The impact of these interventions has varied greatly [6]. Behavioral economists observed an increase in vaccine uptake using notification systems that send out reminders to families who have missed recommended vaccination target dates; however, there is no evidence to suggest the uptake was among vaccine hesitant groups [7,8,9,10,11]. Researchers have also found that broad-based information interventions implemented via public health communication channels, such as distributing educational

978-1-5386-9138-0/19/$31.00 ©2019 IEEE

pamphlets at health clinics, had no significant effect on vaccine hesitancy. Paradoxically, in some cases, researchers observed broad-based messaging reinforced vaccine hesitancy in already vaccine hesitant individuals [12]. These findings emphasize the complexity of the problem, and the need for carefully selecting who to message, the message content, and the message delivery method and timing. The first step in vaccine hesitancy research is to identify vaccine hesitant groups and individual children at risk.

There is limited success so far in that direction. Social science researchers modeled the probability that an individual child will be vaccinated using logistic regression models, and found that the educational attainment of the child's parents, whether or not the child was the first born, race, age of mother, and geographic region were all significant at the 5% level [13]. The findings however, are not yet associated with any specific intervention.

Machine learning approaches have been used to analyze discussions among vaccine hesitant groups on social media platforms and online blogs. The results of such research suggest that these groups form an anti-vaccine narrative framework, such as the alleged connection to autism, and proliferate the narrative through their networks [14,15]. Researchers in Pakistan have proved the feasibility of using machine learning models, such as Support Vector Machines and Random Forests, to make individual level predictions for whether or not a child will default on any immunization in the recommend Pakistani vaccination schedule. The research was carried out using individual-level immunization records, which contained features such as the child's gender, place of residency, vaccination history, and date of birth. The models had an AUCs ranging from 0.782 to 0.791 [16]. A major limitation of all the machine learning analyses conducted to date is that while they have shown some success in monitoring and detecting vaccine hesitancy, again, they are not associated with any specific intervention aimed at increasing vaccine uptake.

In this paper, an action-driven analytics approach is proposed, where the analysis and the outcomes are tightly linked with an action from key stakeholders that may introduce a required change. The objectives of the research have two distinguished strengths, as compared to other research in the field: first, vaccine hesitancy models were built using Electronic Health Record data, and second, the an intervention aimed at increasing vaccination rates is proposed.

II. Research Context

While the threat of vaccine hesitancy is a concern across Europe and the globe, falling vaccination rates have been among the most alarming in a specific country in Southern Europe[1]. For instance, in County 1 of Country, vaccination rates for the infant dose of MMR dropped from 95% in 2010 to below 60% in 2017. As a result, this research project was formed from a partnership between the country's Institute of Public Health (IPH).

The country's healthcare officials follow the WHO's recommended vaccine schedule for MMR: one dose administered by pediatricians at infancy, typically around the age of 1 (referred to as "primovaccination"), and a second dose administered by school and adolescent medicine doctors to children around age 6 (referred to as "revaccination"). In some cases, if a child has missed MMR primovaccination, they receive a double dose of the MMR vaccine at the intended time of the revaccination dose. This makes revaccination a critical time point in ensuring immunization against MMR.

The current response to vaccine rejection in the region is a notification system, where doctors and healthcare administrators contact the families of children who have missed the mandatory vaccine schedule. A major weakness of this system is that it is entirely reactionary. Studies indicate that once a child is delayed from their normal vaccine schedule, they are much less likely to be vaccinated at all [17]. Moreover, despite legal repercussions for refusing mandatory vaccines, many parents are able to circumvent the law through various loopholes. Doctors in the region have also expressed concerns that enforcing legal procedures following a rejection of vaccination will harm the doctor-patient relationship, and cause parents to avoid going to the doctor altogether. Currently, healthcare workers in the country confront vaccine hesitancy by promoting scientific research on vaccination.

Here we partner with the IPH to test two hypotheses. First, to assess whether or not effective machine learning models can be built to predict if an individual child is at risk of not receiving the MMR vaccination using EHRs – in other words, their risk of being vaccine hesitant. Second, that the models can be deployed as an Early Warning and Monitoring System (EWS) that will allow healthcare workers and policymakers to more accurately, and preemptively, intervene on vaccine-hesitant families.

III. Data and Methods

A. Data

Two data sources were used. The first was Electronic Health Records (EHRs) collected by public health clinics, namely the school-medicine centers in Country, typically as a child is about to enter the first grade of primary school. Access to this data is a distinguishing feature of this research paper. To our knowledge, no other machine learning models have been used to model vaccine hesitancy using individual level EHRs. The IPH provided anonymized data for approximately 37,000 children from County 1 and 7,000 children from a second county (County 2). This encompasses all children who received a health check up from a public health clinic prior to First Grade enrollment, after 2011. Our modeling approaches used features from the EHRs in the following areas: (1) demographic and personal information: the parents education level, age, work status, marital status, smoking status, child's educational attainment, living situation, number of siblings, whether the child attended preschool or kindergarten, and speaks a foreign language; (2) geographic information: the child's settlement of residence , and assigned health center; (3) visits to a doctor - the number of recorded visits a child has made to the doctor; (4) infant vaccination record: vaccination record, for Polio (four doses), DTP (five doses), PPD (tuberculosis), BCG (tuberculosis), HBV (hepatitis B, three doses), Hib (four doses), MMR primovaccination dose and age at which it was received; (5) sibling vaccination history—specifically the number and proportion of siblings who received the doses of the MMR vaccine; (6) Personal medical history—specifically an

[1] At the time of writing this paper, the country name must remain anonymous under the project's GDPR-compliant data sharing agreement. Henceforth, the country will simply be referred to as Country.

indication of complications at birth, the age at which the child first sat, walked, teethed, and spoke (words and sentences).

The second data source used was publicly available census data which included the municipality population and the proportion of the population under 20 years of age.

B. Data preprocessing

To convert the raw EHR and census data into a form usable for modeling, several preprocessing steps were required. Continuous variables, such as the age at which a child began to walk, were rescaled to have a mean of zero and standard deviation of one. Binary variables, such as whether the child received their infant MMR vaccination, were coded as 1 for true and 0 for false. Categorical variables with more than two options, such as the parents' marital status, were encoded using a one-hot encoding scheme.

EHR data can be incomplete and error-prone. In particular, we found that many patients had significant missing data. For some children, personal medical history or family demographics was missing entirely. We created separate binary variables that indicated when this kind of systematic missingness occurred. Individual variables were also missing in some cases. For the purposes of modeling, simple imputation methods were used: missing variables were filled using the variable's mean value for continuous variables, using a value of 0.5 for binary variables, and a value of 0 for one-hot encoded variables. In the future, logic-based imputation methods could be explored, including performing a sensitivity analysis testing the effects of different methods on evaluation metrics. Along with missing data, some variables were also entered incorrectly. We attempted to correct for entry errors on an ad hoc basis by removing outliers, for example by replacing parent ages less than 12 or greater than 80 with the variable's mean value and setting a maximum age for early developmental milestones (e.g., sitting up, crawling) at 48 months.

C. Modeling

Consistent with best practices in applied machine learning, a gamut of models was tested. Our approach was model agnostic, in the sense that we valued high interpretability and high performance on the prediction metric more than a priori preference towards a specific model. To this end, a grid search method was implemented to select the optimal model and hyperparameters from among several model choices, including two linear models (ridge and LASSO logistic regression) and a nonlinear model (gradient-boosted decision trees). Separate models were used for each county for two reasons: first, each county has an independent data storage system, and second, to account for hidden regional features that may influence the probability of vaccine hesitancy. This practice has been used in other applied machine learning settings [18]. Note that in practice, this means that feature importance and hyperparameters may vary by region.

The goal was to select a model that was highly interpretable while simultaneously being effective at identifying children at high risk of vaccine hesitancy. For each model, the following evaluation metrics were recorded: AUC, accuracy, average log likelihood, precision and recall. Performance was compared across models, and to a baseline model which assumed a family's vaccination choice for revaccination will match their choice for primovaccination. In general, models were compared using AUC, and the precision at k, where k is a percentage of the population that healthcare workers have sufficient resources to carry out interventions for. After evaluating all models, the LASSO regression was selected as the optimal model due its high performance while using few features, making it more easily interpretable.

Feature selection was based on both previous literature and the available data. Past research has shown that whether or not the child was the first born, the age of the mother, geographic region, gender, age, and date of birth are all strong indicators for predicting vaccine compliance. Additionally, features were selected from the 6 data categories mentioned earlier in this paper. Over 70 features were selected for modeling.

Many of the features used by the model, such as pediatric vaccination history and demographic information, are generally stable over time. However, other features, such as the number of times a child has visited a health clinic change over time. Thus, it was important to train the model to make predictions at multiple time points, taking into consideration changes in the data. The model was trained on historical data from, and created predictions for, four time points for each child: January 1 and April 1 in the year before a child enters 1st grade, and September 1 and February 1 during the school year. For all prediction timepoints, the quantity being predicted was whether the child would be vaccinated by July 1 following the end of the 1st grade school year.

Each model was trained on data from students who began first grade between 2011 and 2015, using the cohort of students who began school in 2016 as a validation set to determine the best model hyperparameters through grid search, based on AUC value. To avoid comparing students across different time points when calculating AUC, the AUC score was calculated within data from each time point, and a weighted average of these scores was used for grid search evaluation. Each model's best hyperparameter settings were then used to produce predictions on the 2017 student cohort for final evaluation and comparison.

D. EWS Implemenation

While researchers have had success using machine learning to model vaccine hesitancy, there is a research gap with respect to connecting findings to a specific intervention. To address this gap, here we prototyped an EWS that can be used by healthcare workers and policymakers to support existing interventions, by allowing for preemptive rather than reactive interventions, as well as to allow novel, well-targeted messaging. The EWS was built using the Flask web framework for Python, and has the following features: (1) displays the average calculated risk-score for the students assigned to each health center in the country, (2) displays the individual risk score of each student in a given health clinic, and (3) displays each child's value for the features used in the final predictive model.

Children are assigned into one of four levels of risk not receiving MMR revaccination (low, medium, high, and very high) based on the percentile calculated risk score from each counties' LASSO regression model, as early as January 1st before the child enters first grade (8 months before the expected revaccination date). For the purposes of prototyping the EWS, cutoffs for the four risk levels were arbitrarily decided as such: the bottom quartile of students is assigned to the "low" risk category, and the highest quartile assigned to the "very high" risk category (the middle two quartiles assigned to the "medium" and "high" categories, respectively). These categories are not meant to not indicate risk of non-vaccination in an absolute sense (e.g., indicating a particular probability of not being vaccinated) but rather indicate the risk relative to other children, which is the most relevant information in the context of limited resources and a pending policy decision of whom to dedicate time to or intervene on.

In practice, these risk levels would be defined by health clinic administrators, doctors, and nurses, who will use the EWS to view the percentage of children in each risk category at their health clinic, as well view the feature values of each child in the clinics. The EWS can help doctors prepare additional vaccine information prior to medical visits with the families of high-risk children. Administrators and policymakers can use the EWS to better target public messaging and policy interventions. The EWS allows administrators to observe which regions have dense concentrations of children at high risk for not vaccinating.

IV. Results

A. Modeling

All models tested outperformed the baseline on AUC. For County 1, the baseline AUC was 0.72, and the AUC for the ridge regression, LASSO regression, and gradient-boosted trees models were found to be 0.81, 0.81, and 0.83, respectively. For County 2, the baseline AUC was 0.63, and the AUC for the ridge regression, LASSO regression, and gradient-boosted trees models were found to be 0.73, 0.73, and 0.76, respectively. Findings for County 1 and 2 are found in Figure 1 and 2.

Figure 1. AUC for County 1, selected model in blue

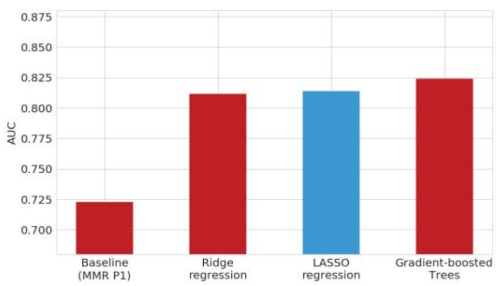
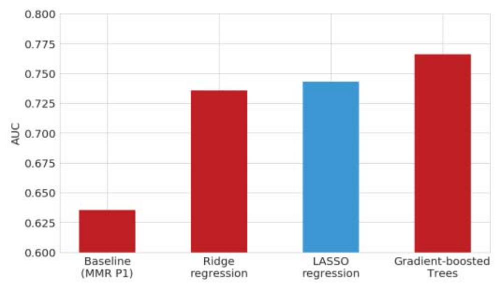


Figure 2. AUC for County 2, selected model in blue



When evaluating the models based on precision, k was chosen to be 20%, meaning that the results show the precision for the top 20% most at-risk individuals. This reflects the assumption that health workers would only have resources to implement interventions for 20% of children, and that thus the goal would be to maximize the proportion of families within that 20% who truly would not otherwise vaccinate (the assumed k will need to be updated once true resource constraints are known). The reported precision score is the weighted average over all time points, where the weight corresponds to the number of unvaccinated individuals in the sample in each time point.

For County 1, the baseline precision was 0.63, and the precision for the ridge regression, LASSO regression, and gradient-boosted trees models were found to be 0.72, 0.72, and 0.73. For County 2, the baseline precision was 0.39, and the precision for the ridge regression, LASSO regression, and gradient-boosted trees models were 0.54, 0.56, and 0.62, respectively. Precision findings for County 1 and 2 are found in Figure 3 and 4, respectively.

Figure 3. Precision at 20% for County 1, selected model in blue

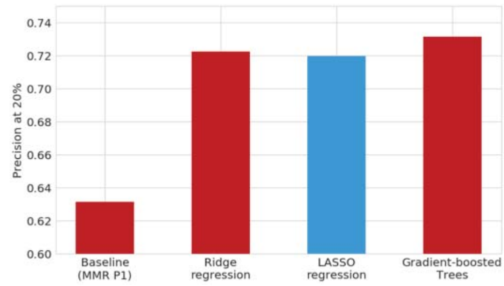
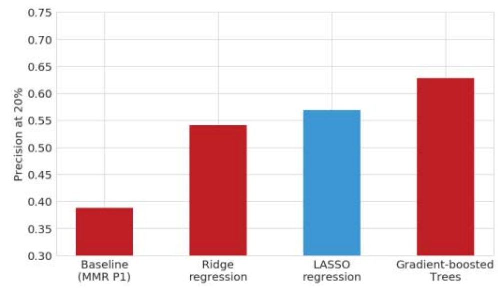


Figure 4. Precision at 20% for County 2, selected model in blue



With respect to AUC, the overall best-performing model was the gradient-boosted trees model; however, performance was only marginally better than that of the regression models (less than 3%). Furthermore, similar performances were in observed in the ridge and LASSO regression models, specifically, an 11% increase in performance as compared to the baseline model.

For precision, again, the best-performing model as compared to the baseline was the gradient-boosted trees model. While only a marginal improvement was observed in County 1, the gradient-boosted trees performed an average 7% better than the regression models in County 2, and 23% over the baseline.

While the gradient-boosted trees model performed slightly better in terms of AUC and precision, the LASSO model was far more easily interpretable. It used only 25 features, combined in a linear manner, while the gradient boosted trees model used all features in an opaque, nonlinear manner. Whether the child had a personal anamnesis on file was the strongest selected feature. This suggests that pupils who have received medical care during their childhood are more likely to be vaccinated than the others. Whether they had received the five-year-old Diphteria, Tetanus and Polio (DTP) vaccine dose was also listed among the strongest features. Note that the aforementioned DTP vaccine dose is the last infant vaccine a child is supposed to receive before the MMR dose. EWS Implementation

An EWS was built in the form of a web dashboard. The dashboard contains views at the county, health center, and child levels, and shows risk scores based on the LASSO regression model. The architecture of the EWS is such that it is built directly on top of the IPH servers, meaning that it is robust in responding to real-time data changes. Risk scores can be recalculated at a user-specified interval, as often as several times in one day. The EWS has been built as a proof-of-concept and is currently awaiting deployment in several health centers across the country.

The simple solution produced by the LASSO model makes it ideal for use in an EWS, because the exact attributes that caused a certain prediction to be made can be reviewed and understood by healthcare providers. For example, besides predicting that children who received their infant vaccinations will receive the vaccine, the model also predicts that children who attended kindergarten are more likely to be vaccinated, while those that learned to walk later than peers are less likely to get vaccinated. While the current model can draw no causal links between these features and vaccine hesitancy, future research may lead to individualized interventions based on the attributes of a given child.

B. *EWS Implementation*

An EWS was built in the form of a web dashboard. The dashboard contains views at the county, health center, and child levels, and shows risk scores based on the LASSO regression model. The architecture of the EWS is such that it is built directly on top of the IPH servers, meaning that it is robust in responding to real-time data changes. Risk scores can be recalculated at a user-specified interval, as often as several times in one day. The EWS has been built as a proof-of-concept and is currently awaiting deployment in several health centers across the country. The EWS for this project was built as standalone software, in-part because the EHR system does not allow for adding external features developed by third parties. However, alternative implementation methods exist, such as integrating the EWS directly into an EHR data system. Several open source EHR systems, like OpenEMS, allow for the direct integration of custom applications into their platform.

V. Discussion

The results show that effective models for predicting vaccine hesitancy at the individual level can be built using EHRs. Despite achieving slightly lower precision than that of other models, the LASSO regression model was implemented into the dashboard, because it uses fewer than 25 features to generate risk scores for both counties, as compared to over 70 features used by the gradient-boosted decision tree and ridge regression models. The payoff for having fewer features is that it minimizes the "black box" effect often observed in utilizing machine learning models. The importance of having a transparent, interpretable model that can easily be explained to healthcare workers and policymakers will likely increase uptake of the dashboard. As a result, the EWS can be fully leveraged to facilitate the efficacy of interventions meant to address vaccine hesitancy, such as informational phone calls and one-on-one sessions with members of the at-risk population. Delivered with the appropriate consideration and early enough, the downsides of acting upon information about who the at-risk population consists of seem to be low, as persuasion of those who have already formed a strong opposing opinion is unlikely but is also unlikely to be found provocative by those at risk but still bereft of strong explicit stances.

The model is well equipped to answer a variety of policy and healthcare questions by offering predictions at four time points. Policy makers, for instance, may be concerned about the risk score based on April 1 before the start of 1st grade, which is the deadline by law for receiving the second, revaccination dose of the MMR vaccine; while doctors might be interested in the risk of not being vaccinated by the start of first grade, when a child starts spending a significant amount with their peers, some of whom could be carriers.

Evaluating models on both AUC and precision at *k* also has important policy implications. For instance, if health clinics only have the resources to intervene on a limited number of families per week, the model can be tuned to maximize the precision of the most at-risk children. Alternatively, if the IPH plans to implement targeted policy interventions against vaccine hesitancy, it is more useful to maximize AUC.

There are few remaining technical hurdles to implementing the EWS in a real-world healthcare setting. By building the prototype directly into the IPH servers, the machine learning model is always trained on the most up-to-date dataset. Furthermore, few challenges are faced with respect to scalability. The dashboard could easily be implemented to many regions throughout the Country because the models were based on nationally-standardized EHRs, despite models varying by region. When a nationally-standardized EHR system is available, this robustness with respect to scalability is a unique strength of the proposed EWS. However, in the absence of a standardized EHR system, the robust scalability may be dramatically undermined. For example, if certain variables that proved to be strong features in one region are non-existent in a different region, the EWS cannot be generalized. Another limitation of the EWS is that it must remain connected to the PHI servers to provide accurate risk score predictions, which could provide significant security concerns. Future work for this research includes a Randomized Control Trial (RCT) that would evaluate the effect of using the EWS on vaccine uptake among vaccine hesitant groups, as compared to business-as-usual. Such an experiment might also involve comparisons with the efficacy of yet to be developed elements of interventions meant to address vaccine hesitancy, in addition to that of making the existing routine interventions proactive by implementing them preceding explicit expressions of vaccine hesitancy.

VI. CONCLUSION

This study explores the use of a novel machine-learning based approach on EHR data to predict which families will be hesitant to vaccinate their child against MMR, for the purpose of enabling proactive informational guidance. Evaluated on both AUC and precision at *k*, machine learning insights predicted that a child will not receive revaccination of MMR at higher rates than the status quo methods, and the model findings were successfully implemented into the prototype of an EWS. This is an important step towards helping healthcare workers reach the right families at the right time to encourage potentially life-saving vaccinations. Nevertheless, it is our belief that this is not a silver bullet against vaccine hesitancy. Instead, the research findings shown in this paper provide the most benefit when used in conjunction with the professional opinion of healthcare workers, public health officials, and policymakers, as well as careful considerations of which interventions are suitable given the predictive insights provided by the EWS.

REFERENCES

[1] A. E. Francis, et al, "Vaccination greatly reduces disease, disability, death and inequity worldwide," Bulletin of the World Health Organization 86. 2008, pp.140-146.

[2] The World Health Organization, "Report of the SAGE Working Group on Vaccine Hesitancy." 2014.

[3] A, Ammon and X. P. Monné, "Vaccines, trust and European public health. Eurosurveillance," Eurosurveillance. 2018, pp.18-21.

[4] N. E. MacDonald, "Vaccine hesitancy: Definition, scope and determinants," Vaccine. 2015, pp.4161-4641.

[5] C. Betsch, R. Böhm, and G. B. Chapman, "Using Behavioral Insights to Increase Vaccination Policy Effectiveness," Policy Insights from the Behavioral and Brain Sciences, vol 2.1. 2015, pp.61-73.

[6] E. Dubéa, D. Gagnona, and N. E. MacDonald, "Strategies intended to address vaccine hesitancy: Review of published literatures," Vaccine, vol 33.34. 2015, pp.3191-4023.

[7] "Community Preventive Service Task Force. Guide to community preventive services. Increasing appropriate vaccination." 2014. Available from http://www. thecommunityguide.org/vaccines/index.html [accessed 22.02.19].

[8] K. Ward, MY Kwan Chow, C. King, and J. Leak, "Strategies to improve vaccination uptake in Australia, a systematic review of types and effectiveness," Austr N Z J Public Health, vol. 36.4. 2012, pp.369–77.

[9] N. Williams, H. Woodward, A. Majeed, and S Saxena, "Primary care strategies to improve childhood immunisation uptake in developed countries: systematic review," JRSM Short Rep, vol. 2.10. 2011, pp.81.

[10] A. G. Fiks, R. W Grundmeier, S. Mayne, L. Song, K. Feemster, and D. Karavite, "Effectiveness of decision support for families, clinicians, or both on HPV vaccine receipt," Pediatrics, vol. 131.6. 2013 pp.1114–1124.

[11] B. Muehleisen, G. Baer, U. B. Schaad, and U. Heininger, "Assessment of immunization status in hospitalized children followed by counseling of parents and primary care physicians improves vaccination coverage: an interventional study," J Pediatr, vol. 151.6. 2007, pp.704–706.

[12] B. Nyhan, J. Reifler, S. Richey, and G. L. Freed, "Effective messages in vaccine promotion: a randomized trial," Pediatrics, vol. 133.4. 2014, pp. 1–8.

[13] E. Crouch and L. A. Dickes, "A prediction model of childhood immunization rates," Applied health economics and health policy, vol. 13.2. 2015, pp.243-251.

[14] T. R. Tangherlini, et al, "'Mommy Blogs' and the vaccination exemption narrative: results from a machine-learning approach for story aggregation on parenting social media sites," JMIR Public Health and Surveillance, vol. 2.2. 2016.

[15] J. Radzikowski, et al, "The measles vaccination narrative in Twitter: a quantitative analysis," JMIR Public Health and Surveillance, vol. 2.1. 2016.

[16] S. Chandir, et al. "Using Predictive Analytics to Identify Children at High Risk of Defaulting From a Routine Immunization Program: Feasibility Study," JMIR Public Health and Surveillance, vol. 4.3. 2018.

[17] J. Hambidge, et al, "Timely versus delayed early childhood vaccination and seizures," Pediatrics, vol. 133.6. 2014, pp.e1492-e1499.

[18] Í. Martínez de Rituerto de Troya, R. Chen, L. O. Moraes, P. Bajaj, J. Kupersmith, R. Ghani, N. B. Brás, and L. Zejnilovic, "Predicting, explaining, and understanding risk of long-term unemployment," 32nd Conference on Neural Information Processing Systems. 2018.