International Journal forComputational Biology (IJCB)
Vol.9, Dec 2020, pp. 01~06
ISSN: 2278-8115 ❒1

Machine Learning Analysis of National Vaccination Schedules and Rates of Compliance Reveals Correlation with COVID19 Mortality

Malik Yousef[1,2], Louise Showe[3], Izhar Ben Shlomo[1,4*]
[1]Zefat Academic College, Safed, Israel,
[2]Galilee Digital Health Research Center,Safed, Israel
[3]The Wistar Institute, Philadelphia, PA, USA,
[4]Program of Emergency Medicine, Safed, Israel

Article Info

Article history:

Received Aug 8th, 2020
Revised Sept15th, 2020
Accepted Sept 23th, 2010

Keyword:

National vaccination schedule
COVID_19
Mortalit rate

ABSTRACT

We questioned whether previous vaccination histories, as reflected by the national vaccination schedules and compliance rates, might correlate with mortality rates from COVID-19.Using a machine learning algorithm, we aligned the WHO data on national vaccinations schedules and compliance rates with national mortality rates, as published daily by official health authorities and posted on the WHO website. Vaccination schedule and the national rates of compliance were significantly associated with national mortality rates. The five most influential vaccines were Rubella, inactivated Polio, DTP (Diphtheria, Tetanus, Pertussis) and the Pneumococcal vaccine. These vaccinations are more likely to be associated with younger generations whereas mortality is reported mainly in older people. Our findings suggest that further studies exploring the possible mechanism(s) underlying this association with vaccination may provide important information in dealing with an ongoing pandemic. Supporting our conclusions, a recent computational study predicted an epitope similarity between the viruses causing COVID_19, measles, and rubella, the latter two covered by the triple vaccine MMR identified in our studies. This suggests that childhood vaccinations contribute to the differential clinical presentation of COVID_19 in young and older populations. Some recent publication also hypothesized that there would be a non-specific protective effect for the MMR vaccine.

Corresponding Author:

Izhar Ben Shlomo,
Program of Emergency Medicine,
Zefat Academic College,
Safed, Israel.
Email: izharb@zefat.ac.il

How to Cite:

Malik Y*et. al.* Machine Learning Analysis of National Vaccination Schedules and Rates of Compliance Reveals Correlation with COVID_19 Mortality. IJCB.2020; Volume 9, Page 01-06.

1. INTRODUCTION

The COVID-19 pandemic caught the world unprepared to deal with a virus where there was no history of infection by any closely related virus in the general population. Since there was none or limited cross reactivity with previous corona virus infections it appeared there was no population immunity to provide even partial protection from infection. .

The human adaptive immune system is activated and modulated in a significant way by interactions with the commensal microbiome, particularly in the alimentary tract [1][2][3]. While training of the immune system is well established with regard to bacteria, fungi, protozoa or helminths, it is less clear whether a similar protection exists to infection by a myriad of unrelated viruses, although there is some evidence for acquired protection for viral infections for related species (Li et al. 2016). It is also well known that people who grew up in different environments, such as rural vs. urban, or different countries display differing sets of "soft spots" for

the development of allergies, and the composition of their gut microbiota (De Filippo et al. 2010) [4] [5]. It has long been recognized that exposure, such as to Bacillus Calmette-Guerin (BCG) vaccine, primarily used against tuberculosis because of its relation to Mycobacterium tuberculosis, can also serve to enhance the immune response to some malignant tumors [6]. Jensen et al. (2016) [7] referred to the controversial concept that some vaccines can have an impact beyond that for the specific pathogen for which the vaccine was designed as non-specific effects (NSE) of past immunologic exposure. In addition, Sánchez-Ramón et al. (2018) [8] described trained immunity-based vaccines (TIbV), which can provide protection from new invaders to which the organism has not been previously exposed.

Following the hypotheses behind NSE and TIbV, we proposed that the previous vaccination histories, as reflected by the national vaccination schedules and immunization rates, could affect the severity of the clinical presentation of COVID-19 as reflected by mortality rates. We repeated the test on data collected at atwo week interval on a global dataset, and carried out a re-test on the same basis a month later, to verify the robustness of the model over time and the validity of its findings.

2. RESEARCHMETHOD

2.1. a. Data

Information about national vaccination schedules and compliance rates in 194 countries were downloaded from the WHO organization website[6] (data as of 10-December-2019). The vaccinations data contains information for 46 different vaccination types and schedules with data collected from 1966 to 2018. The distribution of the number of vaccination types and schedules has increased annually, from just 3 vaccination types in 1966 to 36 types and schedules in 2018.

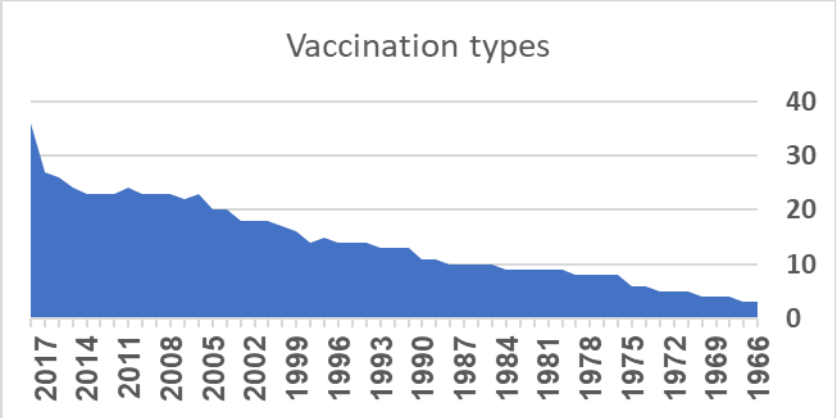

Figure 1. The annual number of available vaccination types and schedules from 1966 until 2017

The rate of compliance for each vaccination type is reported. We calculated the average compliance rate over the reported years with missing values being set to -1.

The second, third and fourth COVID-19 datasets were also downloaded from the WHO website[7], updated: April 11th, 2020, 11:00am CEST, on April 24th, 2020, 11:00am and on June 3rd 11:00am. As of April 11th, 1,569,504 confirmed cases of COVID-19 were globally reported to WHO, including 95,269 deaths. On April 24th, there were 2,631,839 confirmed cases and 182,100 deaths. On June 3rd there were 6,194,533 confirmed cases and 376,320 deaths. The COVID-19 data report covers 212 countries, including cumulative confirmed cases and cumulative deaths for each country. While these are likely under-estimations of cases and deaths in several countries, they are the available confirmed numbers. The COVID-19 data was used to categorize the vaccination data. We calculated the mortality rate (MR) over the total confirmed cases for each country, in countries with more than 2,000 cases of infection recorded (n=50 for Apr. 11th, n=62 for Apr. 24th and n=90 for June 3rd). We additionally tested different thresholds of MR to analyze the data. For example, we set an MR ratio of 1%, 2%, and 3% so that countries with MR higher or equal to the specific percentage fall into the category of “neg”, whereas those with lower MR fall into the category of “pos” .

Table 1. Part of the data created from the vaccinations' data. The rows contain the countries' names while the columns contain the vaccinations' type or name. The "class" column is the label of the country by the mortality rate.

	class	BCG	DTP1	DTP3	DTP4	HepB3	Hib3	IPV1	MCV1	MCV2	PCV1	PCV2	PCV3	Pol3	Rota1
Afghanistan	neg	49.9	83.3	44.3	93.0	79.8	83.0	82.0	43.8	41.5	94.4	87.2	78.8	44.1	75.0
Algeria	neg	94.4	97.0	84.5	95.0	91.7	94.8	98.8	80.0	93.2	96.0	95.7	95.0	84.7	-1.0
Andorra	neg	-1.0	99.0	96.3	96.3	89.1	93.2	99.0	96.2	87.8	99.0	97.0	93.8	96.7	-1.0
Argentina	neg	92.8	93.5	81.4	77.8	87.8	89.4	79.0	89.9	85.9	89.4	81.9	78.0	81.0	85.8
Armenia	neg	92.6	97.1	90.5	94.5	88.5	94.3	73.0	93.3	93.2	83.8	98.0	81.0	94.1	94.5
Australia	pos	-1.0	92.0	87.5	93.7	93.5	91.8	-1.0	88.4	89.1	91.0	91.0	92.3	87.3	84.0
Austria	neg	90.0	77.8	87.5	60.0	78.9	83.3	94.8	67.1	62.8	-1.0	-1.0	-1.0	87.8	67.0
Azerbaijan	pos	96.3	95.3	94.4	97.3	95.8	87.4	93.0	95.5	97.3	80.8	93.0	89.2	96.4	-1.0
Bahrain	pos	82.6	98.2	93.4	98.8	95.7	95.1	98.6	87.6	98.4	94.5	99.1	90.6	93.4	87.0
Belarus	neg	96.6	98.4	95.9	97.5	95.4	84.7	98.2	90.1	98.0	98.4	95.4	96.0	93.2	-1.0
Belgium	neg	-1.0	98.4	94.4	92.2	79.9	86.9	98.4	84.0	83.2	97.8	97.1	91.4	96.4	88.0
Brazil	neg	89.3	94.8	81.5	78.0	78.0	95.6	94.9	86.9	74.4	96.4	93.7	79.9	77.9	93.6
Bulgaria	neg	98.0	95.1	95.7	81.5	92.6	91.2	93.6	95.0	90.6	93.6	93.7	89.2	95.5	36.0
Cameroon	neg	65.0	84.6	57.9	-1.0	83.4	84.1	66.0	55.5	-1.0	91.0	85.3	82.6	58.5	80.8
Canada	neg	-1.0	91.6	90.3	77.0	55.9	90.3	87.0	91.6	85.9	82.4	94.3	56.5	89.9	-1.0
Chile	neg	95.9	95.7	94.0	88.3	93.5	93.5	98.3	94.0	88.2	94.4	94.7	84.6	94.0	-1.0
China	pos	89.9	96.6	90.8	99.0	92.6	-1.0	99.0	92.5	95.6	-1.0	-1.0	-1.0	93.0	-1.0
Colombia	neg	86.5	87.7	75.9	88.8	87.3	83.7	93.0	77.7	81.6	92.5	91.3	82.4	77.2	85.6
Costa Rica	neg	86.3	91.1	89.4	90.7	89.0	86.3	94.8	86.9	88.9	95.6	95.3	83.9	87.1	-1.0
Croatia	pos	98.0	96.8	86.8	90.7	95.3	94.2	-1.0	89.2	93.6	-1.0	-1.0	-1.0	87.5	96.0
Cuba	neg	98.5	96.9	94.1	98.0	97.7	96.0	99.0	92.8	94.9	-1.0	-1.0	-1.0	96.7	-1.0

We used the random forest (RF) [11] classifier implemented by KNIME [12] .RF also provides a score that indicates the significance of each feature. Features in this analysis are each type of the vaccine. Significant features are those most important in building the final model. In addition, the level of compliance with vaccination were used by the algorithm to define thresholds, which in turn served to differentiate between the categories. Features with a very low score are discarded from the final model through a process in machine learning called feature selection. One aim of feature selection is the simplification of models to make them easier for interpretation and to eliminate noisy features that do not add information.

2.2. Training and Evaluation of the Model

The RF classifier is an ensemble of decision trees. For visual simplicity, we present only the decision tree model applied on the whole data (see Figure 2). Additionally, there will be differences in the weight of each vaccine type (feature) based on 100 iterations, as explained ahead.

For each experiment we have generated the decision tree (DT) based on the vaccination data. The DT model can be expressed as a set of if-then-else decision rules. The DT is a tree with decision nodes and leaf nodes. A decision node has two or more branches, a leaf node represents a classification or decision ("pos" or "neg" label). The topmost decision node in a tree which corresponds to the best predictor is called the root node. For interpretation of DT one needs to start from the root node, moving to the next node based on the edges' expression. This process is repeated until reaching the leaf node, the leaf node tells the prediction outcome ("pos" or "neg"). One can consider the path from the root to the leaf as rules connected by an 'and' relationship.

The classifier was trained and tested with a split into 90% of the data used for training and 10% of the data used for testing. We used a 100-fold Monte Carlo cross-validation (MCCV)[13] for model establishment. Additional test was applied with split of 80%-20%.

The features (vaccine type) for each RF model were recoded over all 100 iterations. The average of those scores were calculated to assign a final score to each feature. The higher the score of a feature means it is the more significant to the model.

In order to allow evaluation of the performance of the RF classifier, a set of the following measures were considered: (1) Sensitivity (SEN) which represents the true positive rate, (2) Specificity (SPE) which represents the true negative rate (complement of sensitivity), (3) Precision (PREC) which represents the ability to correctly predict positive target condition to the total, (4) Accuracy (ACC) represents the classifier ability to predict the target condition correctly, (5) F-measure represents the classifier ability to predict the target condition correctly (compared to AUC, it is more informative in the case of an imbalanced data set, since it considers both PR and SE).

All reported performance measures refer to the average performance of a 100-fold Monte Carlo Cross Validation (MCCV)[13].

3. RESULTS AND ANALYSIS

Table 2 summarizes the results of our analysis when testing our model with different MR thresholds. While the MR threshold with a value 1% yields the best performance indices, the number of positive countries (low MR) at this value is small. Repeating the process with data collected two weeks and a month and a half later provided basically the same results, i.e. the performance of the model is best for the 1% threshold. However, with the later run it is evident that at a threshold of 3% the model is no longer informative for the individual countries. The line in the table designated Random presents the results of the algorithm when countries were labeled randomly.

Table 2.Summary results for different national mortality rate thresholds. Acc - Accuracy, Sen - sensitivity, Spe – specificity. The results associated with the line 80%-20% is the results with splitting the data into 80% training and 20% testing..

	MR threshold	**#pos**	**#neg**	**Acc**	**Sen**	**Spe**
Apr. 11, 2020	1%	7 80%-20%	43	0.89 0.85	1.00 1.00	0.78 0.82
	2%	16	34	0.81	0.91	0.80
	3%	26	24	0.78	0.96	0.78
	Random	25	25	0.20		0.43
Apr. 24, 2020	1%	9 80%-20%	53	0.90 0.88	1.00 0.86	0.73 0.88
	2%	16	46	0.83	0.91	0.74
	3%	28	34	0.66	0.88	0.66
June 3, 2020	2%	27	63	0.71	0.89	0.63

Figure 2 present the decision trees applied on the data of MR thresholds 1% (a) and 2% (b).

It is clear from the figure and Table 2, that the five top features are sufficient to classify the results, suggesting that these 5 vaccines are the most important with regards to the national MR.

Among the 62 countries that were included in the study on Apr. 24th, 9 had MR<1% and 16 had MR<2% (labeled "pos") whereas those with higher MR were 53 and 46, respectively. Although the first node for the 1% threshold is DiphCV6 and for the 2% threshold it is HepB3, Table 2 shows that the vaccine that is most important in the correlation to the MR is HepB3 in both cases, with 0.8 in the 1% threshold and 1.0 in the 2% threshold. This difference is probably attributable to the high importance of all four of top features in the more stringent 1% threshold analysis. The poor performance of the algorithm at the level of 3% can be explained by the fact that 3% is closer to the global rate of about 5%-7%. For June 3rd we only performed the 2% threshold analysis and found the model somewhat weaker, but still indicating there is a basic correlation between vaccination schedules and compliance to national mortality rates. A reanalysis, dividing the total data into 80%/:20% showed essentially the same results.

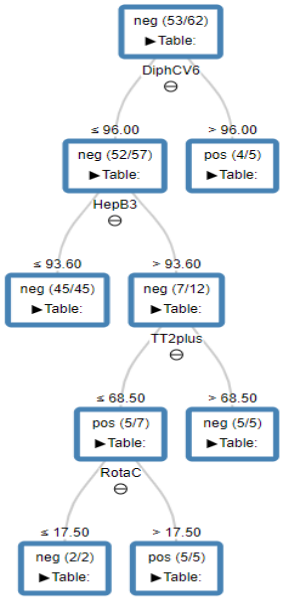

Figure 2. Decision tree models representation of the relation of national vaccination schedules and rates of compliance, to national mortality rates from COVID-19 in 62 countries with more than 2,000 verified infected persons. (a) The tree obtained with mortality ratio of 1%. (b) The tree obtained with mortality ratio of 2%. Abbreviations: "pos"- countries with MR higher or equal to the percentage; "neg" – countries with MR lower than the percentage, respectively. DiphCV6 – Diphteria containing vaccine 6th dose; HepB3 – Hepatitis B 3rd dose; PCV1 - Polio Conjugate Vaccine 1st dose; TT2plus - Tetanus toxoid vaccine beyond 2nd dose; RotaC – Vaccine against Rota virus; HPVFaem – Human papilloma virus vaccination for girls; Hib3 - Haemophilus influenza vaccine 3rd dose.

The observation that a previous exposure to immune-provoking microorganism can be associated with enhanced response to new microorganisms/antigens has primarily been discussed in relation to bacterial infections. Recent publications have hypothesized that a similar process may be associated with viral infections/vaccinations. Our analysis suggests that the global data on national immunization schedules and rates of compliance to those schedules are related to national mortality rates from COVID_19. We have used machine learning approach to test this hypothesis. Machine learning or artificial intelligence (AI) processes are the best unsupervised, hence unbiased, analysis tools for this type of study and are not influenced by a-priori assumptions of the researcher. In this regard it is important to note that when we labeled the countries randomly, the algorithm did not reveal any significant correlation.

As we were analyzing our data a recent publication on more than 4500 consecutive patients positive for COVID-19 found that hospitalization risk was reduced with prior influenza vaccination [11]. Wider scope, worldwide database analysis revealed that countries where BCG vaccination is given at birth have shown a lower contagion rate and fewer COVID-19-related deaths, suggesting that BCG vaccine may induce trained immunity that could confer some protection from COVID-19 [12, 13]. On a bioinformatics, computational grounds Sidiq et al.[14] found a 30 amino acid sequence homology between the SARS-CoV-2 Spike (S) glycoprotein (PDB: 6VSB) of both the measles virus fusion (F1) glycoprotein (PDB: 5YXW_B) and the rubella virus envelope (E1) glycoprotein (PDB: 4ADG_A), and hypothesized that a cross immunity on this basis may attenuate the clinical severity of COVID-19 among children. Along the same line of thought, some authors suggested that live attenuated vaccines, such as MMR (measles, mumps, rubella) may confer some protection against the more severe aspects of COVID-19 [14,15]. A novel trial even tested the effect of MMR vaccination on people who already contracted the virus and found an attenuation of the disease severity [16].

Our findings provide insights for clinical and laboratory studies which may explain the phenomenon. Clinical studies that can plot personal severity of the disease against the respective personal vaccinations received in the past, can be carried out in those countries where a registration exists. More basic studies that assess correlations of levels of serum antibodies and immune cells responsiveness to the Corona virus and personal vaccination histories may lead to a further understanding of this protection which appears to be related to having had access to vaccinations against unrelated organisms. The most promising candidates for such

studies would be the history of vaccinations with any of the 5 that were identified in our analysis as being the most essential modifiers of COVID-19 severity.

Our study has some obvious shortcomings. First, reports on national mortality rates are subject to limitations, beginning from difficulty in attributing death to the pandemic and through purposeful intervention with numbers due to political considerations. In addition, different countries use different criteria to identify infected individuals. Yet, having no better source of information, we used the one available.

Second, we used mean vaccination rates, which in some countries represent the last 5-10 years, while in others it represents 30 or more years. On the other hand, using only recent data would hardly capture the immunity of older portions of the populations. Indeed, the current pandemic affects most severely the elderly. An additional issue that is not addressed by our results, is "how vaccinations, mostly administered during the last 50 years or less, affect severity of the clinical picture in older people".

It is true that socio-economic status correlates with compliance to vaccination[17] and the quality of medical systems in the various countries. We also have not been able to control for the fact that environmental exposers may differ significantly between nations[18]. However, the correlation with vaccination compliance and severity of infection exists across countries with significant economic differences, with countries with strong economies like Belgium, Austria and Canada in the negative class (Table 1) and less developed countries such as Azerbaijan and Bahrain being in the positive class. Compliance with HepB3 is very high in the later 2 countries but less so in the 3 strong economies with compliance in Canada only being 55.9%.

The stability of our findings on data collected over one-and-a-half-month time lapse is a further measure of the robustness, and validity of our findings.

4. CONCLUSION

Despite the data limitations we have identified, our results showing a strong correlation between compliance for 5 vaccinations and MRs provides some insight into one parameter that correlates with the pathogenicity patterns of COVID-19.

REFERENCES

1. Genuneit J. Exposure to farming environments in childhood and asthma and wheeze in rural populations: A systematic review with meta-analysis. Pediatr. Allergy Immunol. 2012.
2. Schröder PC, Li J, Wong GWK, Schaub B. The rural-urban enigma of allergy: What can we learn from studies around the world? Pediatr. Allergy Immunol. 2015.
3. Annels NE, Simpson GR, Pandha H. Modifying the Non-muscle Invasive Bladder Cancer Immune Microenvironment for Optimal Therapeutic Response. Front. Oncol. 2020.
4. Jensen KJ, Benn CS, van Crevel R. Unravelling the nature of non-specific effects of vaccines—A challenge for innate immunologists. Semin. Immunol. 2016.
5. Sánchez-Ramón S, Conejero L, Netea MG, Sancho D, Palomares Ó, Subiza JL. Trained Immunity-Based Vaccines: A New Paradigm for the Development of Broad-Spectrum Anti-infectious Formulations. Front. Immunol. 2018.
6. WHO vaccine-preventable diseases: monitoring system. 2019 global summary [Internet]. Available from: https://apps.who.int/immunization_monitoring/globalsummary
7. World HelathOrganization,Coronavirus (COVID-19) [Internet]. Available from: https://covid19.who.int/
8. Breiman L. Random Forests. Mach Learn. 2001;45:5–32.
9. Berthold MR, Cebron N, Dill F, Gabriel TR, Kötter T, Meinl T, et al. KNIME - The Konstanz Information Miner. SIGKDD Explor [Internet]. 2009;11:26–31. Available from: http://centaur.reading.ac.uk/6139/
10. Picard RR, Cook RD. Cross-validation of regression models. J Am Stat Assoc. 1984;
11. Jehi L, Ji X, Milinovich A, Erzurum S, Merlino A, Gordon S, Young JB, Kattan MW. Development and validation of a model for individualized prediction of hospitalization risk in 4,536 patients with COVID-19. PLoS One. 2020;11;15.
12. Covián C, Retamal-Díaz A, Bueno SM, Kalergis AM. Could BCG Vaccination Induce Protective Trained Immunity for SARS-CoV-2? Front Immuno. 2020 May 8;11:970.
13. Klinger D, Blass I, Rappoport N, Linial M. Significantly Improved COVID-19 Outcomes in Countries with Higher BCG Vaccination Coverage: A Multivariable Analysis. Vaccines (Basel). 2020;11;8:E378.
14. Sidiq KR, Sabir DK, Ali SM, Kodzius R. Does Early Childhood Vaccination Protect Against COVID-19? Front Mol Biosci . 2020;5;7:120.
15. Fidel PL Jr, Noverr MC. Could an Unrelated Live Attenuated Vaccine Serve as a Preventive Measure To Dampen Septic Inflammation Associated with COVID-19 Infection? mBio. 2020;11:e00907-20.
16. Larenas-Linnemann DE, Rodríguez-Monroy F. Thirty-six COVID-19 cases preventively vaccinated with mumps-measles-rubella vaccine: all mild course. Allergy. 2020 Sep 7.
17. Falagas ME, Zarkadoulia E. Factors associated with suboptimal compliance to vaccinations in children in developed countries: A systematic review. Curr. Med. Res. Opin. 2008.
18. Gupta VK, Paul S, Dutta C. Geography, ethnicity or subsistence-specific variations in human microbiome composition and diversity. Front. Microbiol. 2017.