A Novel Representation of Vaccine Efficacy Trial Datasets for Use in Computer Simulation of Vaccination Policy

Mohammadamin Tajgardoon[1], Michael M. Wagner[1,2], Shyam Visweswaran[1,2], Richard K. Zimmerman[3]

[1]**Intelligent Systems Program, University of Pittsburgh, Pittsburgh, PA, USA;**
[2]**Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, USA;**
[3]**Department of Family Medicine, University of Pittsburgh, Pittsburgh, PA, USA**

Abstract

Computer simulation is the only method available for evaluating vaccination policy for rare diseases or emergency use of new vaccines. The most realistic simulation of vaccination policy is agent-based simulation (ABS) in which agents have similar socio-demographic characteristics to a population of interest. Currently, analysts use published information about vaccine efficacy (VE) as the probability that a vaccinated agent develops immunity; however, VE trials typically report only a single overall VE, or VE conditioned on one covariate (e.g., age). Thus, ABS's potential to realistically simulate the effects of co-existing diseases, gender, and other characteristics of a population is under-used. We developed a Bayesian network (BN) model as a compact representation of a VE trial dataset for use in ABS of vaccination policy. We compared BN-based VEs to the VEs estimated directly from the dataset. Our evaluation results suggest that VE trials should release statistical models of their datasets for use in ABS of vaccination policy.

1 Introduction

Computer simulation is an important method that allows policy makers to examine the effectiveness of vaccination strategies in a virtual environment before applying them to a real population[1]. It is really the only method available for rare diseases (e.g., study of smallpox vaccination[2]), or emergency use of new vaccines (e.g., study of vaccination policy for Ebola and Zika epidemics[3–5]).

Agent-based simulation (ABS) is the most realistic method for computer simulation of vaccination policy. Briefly, when using an ABS to analyze a vaccination policy *P*, analysts first create a population of agents whose socio-demographic and behavioral characteristics are similar to a population of interest. They then program the simulator to emulate policy *P* and emergence of an epidemic. In particular, analysts use published data about vaccine efficacy (VE) to determine the probability of disease transmission to a vaccinated agent.

There is usually limited information about VE available to ABSs of vaccination policy since VE trials typically report only a single overall VE, or VE conditioned on only one covariate (e.g., age)[6–8]. Additionally, while it is possible to recover other conditional VEs from VE trial datasets, these datasets are not released to the public domain and they are difficult to obtain even under restrictions. Thus, ABSs of vaccination policy are not able to simulate the effects of co-existing diseases, medications, gender, and other socio-demographic characteristics of a population.

Therefore, our objective is to improve the information available about VE. In particular, we developed and evaluated a Bayesian network (BN) model as a compact representation of a VE trial dataset for use in an ABS of vaccination policy. A BN is a probabilistic model that encodes a joint probability distribution over all dataset variables. VE conditioned on any set of covariates can be estimated from a BN model of a VE trial dataset. We discuss the details of estimating VE from a BN model in Section 3.

The VE trial dataset that we used was collected in a study of a trivalent inactivated influenza vaccine (TIV), manufactured by GlaxoSmithKline (GSK) Biologicals, during the 2006–2007 influenza season in the Czech Republic and Finland[8]. We used two score-based BN learning algorithms in combination with multiple score functions to develop several BN models of the VE trial dataset, in addition to a naïve Bayes model as the baseline model. We evaluated the models in 10-fold cross validations to find the model that fits the dataset the best. We compared the VEs estimated from the best BN model to the VEs that we directly estimated from the dataset. Our results suggest that VE trials consider releasing a BN model of their datasets which contains significantly more information about VE compared to the tables they report in their publications.

2   Related Work

Previous studies have addressed the problem that a single VE may not apply to all individuals[9–15]. The issue is usually referred to as *heterogeneity in VE* and has been investigated since 1915 when Greenwood and Yule[9] noticed the problem. Longini et al.[13] categorized the approaches to modeling the heterogeneity in VE into two groups: (1) by using stratification[10,12], and (2) by allowing the susceptibility in vaccinated persons to follow a probability distribution[11,13,16]. In the stratification approach, vaccinated individuals are stratified into a number of mutually exclusive groups and VE is estimated for each group. The number of strata in this approach is usually determined manually and depends on whether the heterogeneity in VE originates from the host characteristics or is vaccine-related[12]. In the second approach, the vaccine effect on individuals follows a probability distribution. For example, Longini et al.[13] defined a parameter, α, as a fraction of the vaccinated persons who may develop complete protection. The remaining fraction, $1 - \alpha$, may receive only partial protection. Their objective was to estimate α and the parameters for the distribution of individuals with partial protection from trial or observational data. To model VE, they used a frailty mixture model which is a class of survival models that allow some vaccinated individuals to be more susceptible than other vaccinees.

Moreover, in a recent study, Mehtälä et al.[15] investigated how to estimate the efficacy of a vaccine with heterogeneous protection in a SIS (Susceptible-Infected-Susceptible) model. They modeled VE using three parameters: the proportion of vaccinated completely protected from infection, the relative attack rates in acquisition phase, and the relative attack rates in clearance phase, for a vaccinated person not completely protected as compared to an unvaccinated individual.

In this paper we propose a different approach for modeling the vaccine effect in the presence of heterogeneity, which to our knowledge has not been tried previously. We use machine-learning algorithms to fit a joint probability distribution over all variables in a vaccination trial dataset including socio-demographic and disease characteristics, and vaccination-related covariates. We expect that the machine-learning algorithms will automatically capture the host or vaccine-related heterogeneity in the vaccine effect, as long as the relevant covariates are measured in the vaccination trial.

3   Method

We propose a machine-learning based method to fit a joint probability distribution over variables in a VE trial dataset and then use the distribution to estimate VE conditioned on any set of covariates. In the following sections, we first describe BNs and the learning algorithms we use. Then, we explain how to estimate VE from a BN model of a VE trial dataset.

3.1   Bayesian Networks (BN)

A BN over a set of random variables V is a pair $BN = (G, P)$, where G is a directed acyclic graph (DAG) and P is a joint probability distribution of V. Each node in the DAG is assigned a conditional probability distribution as a function of the node's parents. Nodes with no parents have a marginal probability distribution. Arcs represent dependency relationships between variable pairs[17]. Learning a BN model of a dataset includes two steps: (1) DAG construction, in which a BN structure-learning algorithm is used to find a DAG from data, and (2) parameter estimation, which is estimating the probability distribution of nodes in the DAG.

We used two score-based structure-learning methods: hill-climbing (HC) with random restart and tabu search. These algorithms start from an initial network structure (e.g., an empty graph) and add, delete, and reverse arcs until a score function is no longer improved. Tabu search algorithm maintains a list of recent operations (i.e., tabu list) to avoid repeating the most recent actions and becoming stuck in a suboptimal solution[18]. We tested several score functions that are common in the literature: Bayesian information criterion (BIC)[19], Bayesian Dirichlet equivalence uniform (BDeu)[17], and K2[20].

Once a DAG is learned by a BN structure-learning algorithm, a parameter estimation method is used to learn the conditional probability distributions of the nodes given the DAG and data. We use the maximum a posteriori (MAP) parameter estimation method in our analysis.

After learning a BN model from data, i.e., learning the structure and estimating the parameters, we use inference algorithms to obtain the probability of an event of interest (e.g., disease contraction) conditioned on a set of evidence (e.g., vaccinated and female). This process is called inference or probabilistic reasoning. BN inference algorithms are categorized into exact and approximate inference algorithms[21]. Exact inference algorithms use Bayes theorem and Markov condition to derive exact values of conditional probabilities and are limited to small BNs. Approximate inference algorithms use Monte Carlo sampling to estimate conditional probabilities from a large set of dataset instances generated from the BN. These algorithms perform better on large BNs. In our experiments we use logic sampling[21] which is an approximate inference algorithm.

3.2 Estimate VE from a BN Model

In this section, we explain how to estimate VE from a joint probability distribution that is obtained by a BN learning algorithm. In particular, we formulate VE as a function of two conditional probabilities that are inferred from a joint probability distribution.

VE is measured as the proportional reduction in disease incidence in vaccinated group compared to unvaccinated group[22]. When we do not condition on any other covariates, e.g., age, gender, etc., VE is referred to as *Overall VE* and is calculated using the following equation:

$$VE = \frac{ARU - ARV}{ARU} = 1 - \frac{ARV}{ARU} \tag{1}$$

where ARV and ARU are attack rates in the vaccinated and unvaccinated groups, respectively.

According to the relative frequency definition of probability[23], ARV can be defined as the probability of disease incidence in vaccinees, i.e., $Pr(disease|vaccinated)$, and similarly, we can formulate ARU as the probability of contacting the disease in non-vaccinees, i.e., $Pr(disease|unvaccinated)$. Therefore, by replacing attack rates with probabilities in Equation 1, we can re-write VE as a function of two conditional probabilities:

$$VE = 1 - \frac{Pr(disease|vaccinated)}{Pr(disease|unvaccinated)} \tag{2}$$

We can extend Equation 2 to obtain VE conditioned on a set of covariates (e.g., age, gender). Let $X = \{X_1, X_2, ..., X_n\}$ be a set of covariates in a VE trial dataset, which are initialized with values $x = \{x_1, x_2, ..., x_n\}$. VE conditioned on $X = x$ is formulated as follows:

$$[VE|X = x] = 1 - \frac{Pr(disease|vaccinated, X = x)}{Pr(disease|unvaccinated, X = x)} \tag{3}$$

4 Dataset Description

The dataset that we used to develop and evaluate BN models was acquired in a randomized, double-blind, placebo-controlled study of a trivalent inactivated split-virus influenza vaccine (TIV) manufactured by GlaxoSmithKline (GSK) Biologicals[8]. The study was conducted during the $2006-2007$ influenza season in the Czech Republic and Finland. The primary objective of the study was to assess the efficacy of the TIV in the prevention of culture-confirmed influenza due to strains antigenically matched to the vaccine.

The study subjects were 7652 healthy adults between 18 and 64 years old. 5103 of the subjects were vaccinated with 1 dose of TIV and the remaining 2549 received 1 dose of placebo in their first visit. Table 1 demonstrates the demographics of vaccine and placebo groups. Due to the randomization in group assignments, the distribution of demographic characteristics is similar in both groups.

The dataset contains 38 tables which fall into two categories of *administrative* (30 tables) and *analysis* (8 tables). Administrative tables are populated with the study meta-data, including the list of codes for activities and events, information on dataset fields, and administrative details (e.g., consent forms, eligibility and elimination codes, protocol

violations). The analysis tables contain the information related to vaccination, lab results, demographic characteristics, and medical conditions (Table 2).

We used only variables from six analysis tables. We excluded the two remaining analysis tables, influenza-like-illness (ILI) episodes and serology information, for two reasons: (1) ILI variables, e.g., symptoms and medications during ILI episodes, were post-incidence events that could not influence the VE, and (2) serology tests were performed for only a subset of participants in the study. The outcome variable that we used for model evaluations and VE estimations is *influenza due to vaccine antigenically matched strains*, which is the same outcome variable that the VE trial used for their primary analysis[8].

For our analysis we created a single table with 133 variables including demographics, medical conditions, concomitant vaccinations, non-ILI-episode medications, vaccine-related variables, and the outcome variable (influenza due to vaccine antigenically matched strains). All variables were either binary or categorical with a few categories except for Age which had integer values between 18 and 64. We discretized the Age variable to reduce its number of possible values. Specifically, at each iteration of a 10-fold cross-validation we applied a supervised discretization method to categorize the Age variable into a few bins.

Table 1: Demographic characteristics of the VE trial subjects. Second and third columns list the number and percentage of each demographic characteristic within vaccine and placebo groups, respectively. For Age, the average age and the range for each group is listed.

Characteristic	Vaccine (n=5103)	Placebo (n=2549)
Age, mean years (range)	39.94 (18-64)	39.74 (18-64)
Sex, no. (%)		
Female	3069 (60.14)	1542 (60.49)
Race, no. (%)		
African Heritage or African American	1 (0.02)	0 (0)
Asian - East Asian	1 (0.02)	0 (0)
White - Arabic or North African	3 (0.06)	1 (0.02)
White - Caucasian or European	5097 (99.88)	2547 (99.92)
Other	1 (0.02)	1 (0.04)
Ethnicity, no. (%)		
American Hispanic or Latino	5 (0.10)	1 (0.04)
Not American Hispanic or Latino	5098 (99.90)	2548 (99.96)
Country, no. (%)		
Czech Republic	2664 (52.20)	1332 (52.26)
Finland	2439 (47.80)	1217 (47.74)

Table 2: List of 8 analysis tables and their variables. The VE trial dataset contained 38 tables including 30 administrative and 8 analysis tables. We used only the analysis tables to create a single table with 133 variables.

Table name	Variable names (possible values)
Demographics	age, gender, race, ethnicity, country, vaccination center
Medical condition	diagnosis, status (past, current, both)
Concomitant vaccination	vaccine name, code, vaccination date
Medication	name, code, start date, end date, daily dose
Vaccine	vaccine (Fluarix, placebo), vaccine vial-type (original, replacement), vaccination route, vaccine dose
Lab results	influenza type (A, B), virus strain (H1N1, H3N2, B), influenza due to vaccine antigenically matched strains (positive, negative), test type (HAI, QPCR)
Influenza-Like Illness (ILI) episodes	ILI number, start date, end date, symptoms, medications
Serology results (for a subset of 504 subjects)	seropositivity status

4.1 Bivariate Analysis of Covariates

We explored the statistical dependencies to the outcome variable using bivariate logistic regressions. In particular, for each variable X_i in the dataset, we fit a logistic regression model with the outcome as the dependent variable and X_i as the independent variable. Only a few variables demonstrated significant correlation to the outcome variable (i.e., p-value < 0.05) (Table 3). These variables are likely to have an arc to the outcome node in the BN structure of the dataset. Most of the correlated variables are vaccine-related (i.e., vaccine, vaccine lot, vaccine vial type, and ATP vaccination).

Table 3: List of variables with significant association with the outcome, influenza due to antigenically matched strains. For each variable in the dataset, we used a bivariate logistic regression with the outcome variable as the dependent variable. Here we list variables with coefficient p-value < 0.05 in their corresponding logistic regression model.

Variable	no. of influenza cases within each group (out of 124 total cases)	Estimated coefficient (CI. 95%)	Coefficient p-value
age [a]		-0.02 (-0.03 -0.01)	<0.001
country [b]	81	0.55 (0.18 0.93)	0.003
vaccine [c]	49	-1.14 (-1.5 -0.78)	<0.001
vaccine lot [d]			
lot1	25	-1.12 (-1.59 -0.67)	<0.001
lot2	24	-1.16 (-1.64 -0.71)	<0.001
vaccine vial type [e]	122	-2.32 (-3.64 -0.46)	0.002
ATP vaccination [f]	2	1.63 (-0.2 2.86)	0.027

[a] *age* was analyzed as a continuous variable.
[b] variable values are: 0=Finland (FI), 1=Czech Republic (CZ)
[c] variable values are: 0=placebo, 1=Flaurix vaccine
[d] variable values are: "placebo", "lot1", "lot2"
[e] variable values are: 0= replacement vaccine (due to unusability of study vaccine), 1=study vaccine
[f] variable values are: 0= vaccine not administered According To Protocol (ATP), i.e., vaccine received in the dominant arm or a replacement vial was administered, 1=vaccine administered ATP

5 Evaluation Method

5.1 Model Evaluation

To find the best BN model of the dataset, we performed model selection among several BN models. In particular, we evaluated the goodness of fit and the probability errors of the models in 10-fold cross validations. The following sections introduce the evaluation metrics.

5.1.1 Goodness of Fit Test

We evaluated the goodness of fit of BN models by two statistical tests which we call *global* test and *local* test[24]. The global test examines the null hypothesis that the observed data instances are occurring with the probabilities stated by the model. We constructed a test statistic using a logarithmic score as follows. Let Y be a binary outcome variable, y_i be the observed value of Y_i for instance i, and $p_i(Y_i = y_i)$ be the probability of $Y_i = y_i$ according to the model. Then, a logarithmic score of instance i is:

$$S_i = -\log p_i(Y_i = y_i).$$

For N data instances the cumulative logarithmic score S would be the sum of the logarithmic scores over N instances, i.e.,

$$S = \sum_{i=1}^{N} S_i = -log \prod_{i=1}^{N} p_i(Y_i = y_i). \tag{4}$$

We can use S to construct a standardized test statistic Z:

$$Z = \frac{S - \sum_{i=1}^{N} E_i}{\sqrt{\sum_{i=1}^{N} V_i}}, \tag{5}$$

where

$$E_i = - \sum_{k \in \{0,1\}} p_i(Y_i = k) \log p_i(Y_i = k), \tag{6}$$

$$V_i = \sum_{k \in \{0,1\}} p_i(Y_i = k) \log^2 p_i(Y_i = k) - E_i^2. \tag{7}$$

For large N, Z is approximately distributed as a standard normal under the null hypothesis that the model fits to the data[25]. Therefore, we can calculate the probability of seeing values as extreme or more extreme than Z as a measure of the fitness of the model.

The local goodness of fit test evaluates the conditional probability distribution of the outcome node in a BN model. Let Y be a binary outcome variable, Y_i be the outcome variable for instance i in the data, y_i be the observed value of Y_i, and $X_{pa_i} = \rho$ be the instantiation of parents of node Y in the data instance i. The local score of the outcome node in a BN model is computed via Equation 8,

$$LS_i = - \log p_i(Y_i = y_i | X_{pa_i} = \rho) \tag{8}$$

where $p_i(Y_i = y_i | X_{pa_i} = \rho)$ is the probability of $Y_i = y_i$ given the configuration of parents of Y ($X_{pa_i} = \rho$). We can compute p_i directly from the conditional probability distribution of Y learned in the parameter learning step without doing inference. The standardized local score is obtained similar to the global test by use of Equations 4, 6, 7, and 5, respectively.

5.1.2 Outcome Probability Error

We compare the BN models in 10-fold cross validations with respect to their accuracy in estimating the probability of the outcome. We use mean squared error (MSE) and two probability calibration error metrics[26]. MSE is the average of squared differences between the estimated probabilities and the corresponding value of the outcome variable (i.e., contracting influenza due to vaccine antigenically matched strains $\in \{0, 1\}$), and is calculated using the following formula:

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (e_i - y_i)^2 \tag{9}$$

where e_i is the estimated probability of outcome for data instance i, y_i is the actual value of outcome for data instance i, and N is the total data instances.

Probability calibration error is the discrepancy between probabilities inferred from a model and probabilities obtained from the dataset. To compute this error, we sort the estimated probabilities in an increasing order and divide them into k number of bins. The model is perfectly calibrated if the mean of estimated probabilities are equal to the actual fraction of positive outcomes in each bin. The difference between the mean of probabilities and the fraction of positives in each bin is the base for probability calibration error metrics. We used maximum calibration error (MCE) and expected calibration error (ECE) metrics. MCE (Equation 11) measures the maximum calibration error over all bins, and ECE (Equation 10) calculates the average calibration error over all bins:

$$ECE = \sum_{i=1}^{k} P(i) \cdot |o_i - e_i| \tag{10}$$

$$MCE = \max_{i=1}^{k}(|o_i - e_i|) \tag{11}$$

where k is the number of bins, o_i is the fraction of observed positive outcomes in bin i, and e_i is the mean of the estimated probabilities in bin i.

5.2 Evaluating VEs estimated from a BN Model

We evaluate the VEs estimated from the BN model by comparing them to the same VEs acquired directly from the dataset. To estimate a VE from the BN model, we obtain two conditional probabilities and use Equation 3 to calculate the VE. For example, to estimate VE for females, two conditional probabilities are inferred from the BN model: the probability of disease contraction given vaccinated and being female, $Pr(disease|vaccinated, female)$, and the probability of disease contraction given unvaccinated and being female, $Pr(disease|unvaccinated, female)$. Each VE estimation is repeated 100 times. The average and standard deviation (SD) of 100 VEs is used as the estimate and the uncertainty interval, respectively. For estimating a VE directly from the dataset, we use Risk ratio test and report a 95% confidence interval. P-values are calculated using two-sided Fisher's exact test.

6 Evaluation Results

6.1 Model Evaluation Result

We performed two goodness of fit tests followed by probability error measurements. All evaluations were implemented in a 10-fold cross validation. Table 4 shows the result of goodness of fit tests for different BN models and naïve Bayes. According to the global test there is significant evidence that the naïve Bayes model does not fit to the dataset. However, there is no evidence that BN models misfit the dataset. The local test demonstrates that there is no significant disagreement between the conditional probability distribution of the outcome node and the observed distribution in the dataset. We did not repeat the local test for naïve Bayes model as its overall fitness was already rejected by the global test.

Table 5 demonstrates the outcome probability error measures computed in a 10-fold cross-validation. mean square error (MSE), maximum calibration error (MCE), and expected calibration error (ECE) was calculated based on the probability of the outcome in each test instance. All BN models have low calibration errors compared to the naïve Bayes model. MSE is similar for all models.

Table 4: Goodness of fit test. We performed two tests to evaluate the models' goodness of fit in 10-fold cross validations. In the global test, we tested the null hypothesis that the model fits to the data. The local test evaluates the null hypothesis that the probability distribution learned for the outcome variable fits to the data. None of the BN models was rejected according to the p-values, except for the baseline model (naïve Bayes). We did not perform the local test for the naïve Bayes model.

Algorithm	Score function	Global score	P-value	Local score	P-value
Tabu	BIC	-0.31	0.38	-0.09	0.22
Tabu	K2	0.16	0.44	0.06	0.22
Tabu	BDeu	0.35	0.36	0.12	0.2
HC	BIC	-0.31	0.38	-0.09	0.22
HC	K2	0.16	0.44	0.06	0.22
HC	BDeu	0.35	0.36	0.12	0.2
Naïve Bayes	-	13.1	<.001	-	-

Table 5: Probability error measures via 10-fold cross validation. For each model, mean square error (MSE), maximum calibration error (MCE), and expected calibration error (ECE) computed for the outcome variable.

Algorithm	Score function	MSE	MCE	ECE
Tabu	BIC	0.126	0.009	0.003
Tabu	K2	0.127	0.008	0.003
Tabu	BDeu	0.127	0.010	0.005
HC	BIC	0.126	0.009	0.003
HC	K2	0.127	0.008	0.003
HC	BDeu	0.127	0.010	0.005
Naïve Bayes	-	0.141	0.072	0.015

6.1.1 Final BN Model

All BN learning algorithms performed equally well (except for naïve Bayes) according to the model evaluations. We decided to continue the analysis with the tabu search algorithm and BDeu score function. We used a model averaging technique to build the final BN model[27,28]. In particular, we trained 100 BN models over 100 bootstrap samples of the dataset. Then, we built a single BN model from the 100 BNs by selecting the arcs with proportional frequency of greater that or equal to 0.5. Since the resulted BN's structure was too large to fit in here, Figure 1 illustrates a sub-graph of the structure which includes the outcome node (influenza due to antigenically matched strains) and its first and second level neighbors.
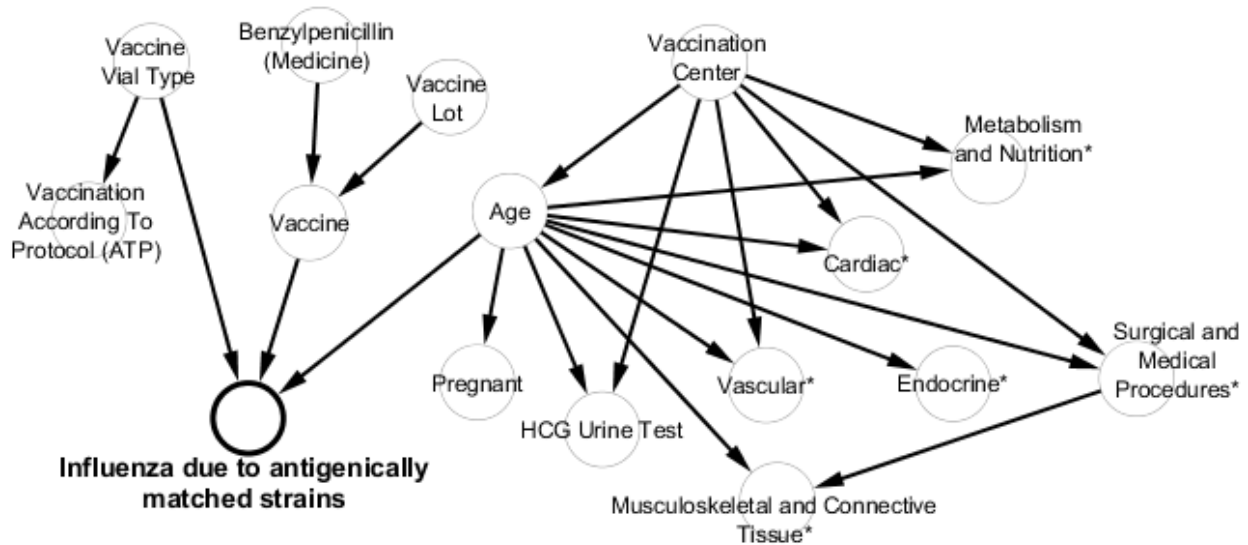

Figure 1: A sub-graph of the average BN structure learned by the tabu search algorithm. 100 instances of BN models learned from 100 bootstrap samples of the data. An average BN model was built from arcs with proportional frequency ≥ 0.5. The illustrated sub-graph includes the outcome variable (influenza due to antigenically matched strains) and its first and second level adjacent variables. Three variables (age, vaccine, vaccine vial type) form the Markov blanket of the outcome variable. The outcome variable is independent of other variables given its Markov blanket[21].
* symbol in the figure indicates a diagnosis variable.

6.2 Estimate VE from the BN Model

We estimated a number of VEs from both the final BN model and the dataset. Table 6 compares VEs estimated from the BN model to VEs directly estimated from the dataset. VEs obtained from the two approaches are the same or very close to each other. The uncertainty intervals for the BN-based estimates are also relatively similar to the confidence intervals of the direct estimates, i.e., when the confidence intervals become wider, the uncertainty intervals follow the same pattern (see the last row in Table 6).

7 Discussion and Conclusion

We developed and evaluated BN models as more complete representations of VE trial datasets compared to tables in the publications reported by VE trials. We performed model evaluations using the goodness of fit tests and probability error metrics to select the best BN model. We compared the VEs estimated from the BN model to the VEs estimated directly from the dataset. The results were promising as the BN-based estimates were very similar to VEs that we obtained directly from the dataset (Table 6).

Table 6: Comparing VE estimations from the selected BN model to the VEs estimated directly from the dataset. For **VE estimated from the BN model**, we estimated each VE 100 times using logic sampling inference algorithm and Equation 3. The average and standard deviation (SD) of 100 VEs used as the VE estimate and its uncertainty interval. For **VE estimated from data**, risk ratio test was used to obtain the VE estimates and 95% confidence intervals. P-values calculated by Fisher's exact test.

Condition	VE estimated from the BN model		VE estimated from data		
	VE	$[\text{VE}-2*SD, \text{VE}+2*SD]$	VE	95% conf. interval	p-value
Overall	0.67	[0.59, 0.75]	0.67	[0.53, 0.77]	$< .001$
Female	0.67	[0.57, 0.77]	0.67	[0.48, 0.79]	$< .001$
Male	0.66	[0.52, 0.80]	0.68	[0.41, 0.83]	$< .001$
AGE ≤ 45	0.72	[0.65, 0.80]	0.69	[0.45, 0.83]	$< .001$
AGE > 45	0.46	[0.17, 0.74]	0.47	[0.13, 0.76]	0.15

Although we did not validate our results by reproducing them on other VE trial datasets, we expect that our proposed approach will perform similarly well; this is mainly because the machine learning algorithm could find the underlying statistical dependencies despite the little evidence in the data (only 1.6% influenza attack rate). We expect the proposed method to perform well on the datasets with higher attack rates. In addition to having a low attack rate, the dataset did not include children and very elderly subjects. This might be a reason for low Influenza incidents in the study as children and old population are the most vulnerable to influenza. The BN model of such datasets may not be usable in ABSs of vaccination policy since these simulations usually include agents from all age ranges.

The significance of this research is that groups that conduct VE trials now have a method to release a statistical model of their data from which vaccination policy makers can obtain VEs conditioned on any set of individuals' characteristics without the need to access the study dataset. Therefore, an ABS of vaccination policy would be able to simulate the effects of a vaccine in presence of all characteristics of a population. We expect that an improvement in information about VE needed for vaccination policy analysis will lead to more effective use of vaccines, and ultimately improvements in health.

In the future, we will use an ABS of vaccination policy to compare the new representation of VE trial datasets with the traditional method (using tables from publications). In particular, we will examine how the the use of a statistical model of the VE trial dataset changes the results of an ABS of vaccination policy that uses VEs from tables in the study publication.

Acknowledgements

References

1. Van de Velde N, Brisson M, Boily MC. Understanding differences in predictions of HPV vaccine effectiveness: A comparative model-based analysis. Vaccine. 2010;28(33):5473–5484.

2. Burke DS, Epstein JM, Cummings DA, Parker JI, Cline KC, Singa RM, et al. Individual-based computational modeling of smallpox epidemic control strategies. Academic Emergency Medicine. 2006;13(11):1142–1149.

3. Kurahashi S. A health policy simulation model of ebola haemorrhagic fever and zika fever. In: Agent and Multi-Agent Systems: Technology and Applications. Springer; 2016. p. 319–329.

4. Chowell G, Kiskowski M. Modeling ring-vaccination strategies to control ebola virus disease epidemics. In: Mathematical and Statistical Modeling for Emerging and Re-emerging Infectious Diseases. Springer; 2016. p. 71–87.

5. Bellan SE, Pulliam JR, Pearson CA, Champredon D, Fox SJ, Skrip L, et al. Statistical power and validity of Ebola vaccine trials in Sierra Leone: a simulation study of trial design and analysis. The Lancet Infectious Diseases. 2015;15(6):703–710.

6. Victor JC, Lewis KD, Diallo A, Niang MN, Diarra B, Dia N, et al. Efficacy of a Russian-backbone live attenuated influenza vaccine among children in Senegal: a randomised, double-blind, placebo-controlled trial. The Lancet Global Health. 2016;4(12):e955–e965.

7. Jackson LA, Gaglani MJ, Keyserling HL, Balser J, Bouveret N, Fries L, et al. Safety, efficacy, and immunogenicity of an inactivated influenza vaccine in healthy adults: a randomized, placebo-controlled trial over two influenza seasons. BMC infectious diseases. 2010;10(1):71.

8. Beran J, Vesikari T, Wertzova V, Karvonen A, Honegr K, Lindblad N, et al. Efficacy of inactivated split-virus influenza vaccine against culture-confirmed influenza in healthy adults: a prospective, randomized, placebo-controlled trial. Journal of Infectious Diseases. 2009;200(12):1861–1869.

9. Greenwood M, Yule GU. The statistics of anti-typhoid and anti-cholera inoculations, and the interpretation of such statistics in general. Proceedings of the Royal Society of Medicine. 1915;8:113–194.

10. Halloran ME, Haber M, Longini Jr IM. Interpretation and estimation of vaccine efficacy under heterogeneity. American Journal of Epidemiology. 1992;136(3):328–343.

11. Brunet RC, Struchiner CJ, Halloran ME. On the distribution of vaccine protection under heterogeneous response. Mathematical biosciences. 1993;116(1):111–125.

12. Longini IM, Halloran ME, Haber M. Estimation of vaccine efficacy from epidemics of acute infectious agents under vaccine-related heterogeneity. Mathematical biosciences. 1993;117(1-2):271–281.

13. Longini IM, Halloran ME. A Frailty Mixture Model for Estimating Vaccine Efficacy. Applied Statistics. 1996;45(2):165.

14. Halloran ME, Longini IM, Struchiner CJ, Longini IM. Design and analysis of vaccine studies. Springer; 2010.

15. Mehtälä J, Dagan R, Auranen K. Estimation and interpretation of heterogeneous vaccine efficacy against recurrent infections. Biometrics. 2016;72(3):976–985.

16. Struchiner CJ, Brunet RC, Halloran ME, Massad E, Azevedo-neto RS. On the use of state-space models for the evaluation of health interventions. Journal of Biological Systems. 1995;3(03):851–865.

17. Heckerman D. A tutorial on learning with Bayesian networks. In: Learning in graphical models. Springer; 1998. p. 301–354.

18. Bouckaert RR. Bayesian belief networks: from construction to inference; 2001.

19. Schwarz G, et al. Estimating the dimension of a model. The annals of statistics. 1978;6(2):461–464.

20. Cooper GF, Herskovits E. A Bayesian method for the induction of probabilistic networks from data. Machine learning. 1992;9(4):309–347.

21. Koller D, Friedman N. Probabilistic graphical models: principles and techniques. MIT press; 2009.

22. Orenstein WA, Bernier RH, Dondero TJ, Hinman AR, Marks JS, Bart KJ, et al. Field evaluation of vaccine efficacy. Bulletin of the World Health Organization. 1985;63(6):1055.

23. Papoulis A, Pillai SU. Probability, random variables, and stochastic processes. Tata McGraw-Hill Education; 2002.

24. Cowell RG, Lauritzen SL, David AP, Spiegelhalter DJ. Probabilistic networks and expert systems. 1st ed. Nair V, Lawless J, Jordan M, editors. Secaucus, NJ, USA: Springer-Verlag New York, Inc.; 1999.

25. Seillier-Moiseiwitsch F, Dawid A. On testing the validity of sequential probability forecasts. Journal of the American Statistical Association. 1993;88(421):355–359.

26. Naeini MP, Cooper GF, Hauskrecht M. Obtaining well calibrated probabilities using Bayesian binning. In: AAAI; 2015. p. 2901–2907.

27. Friedman N, Goldszmidt M, Wyner A. Data analysis with Bayesian networks: A bootstrap approach. In: Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc.; 1999. p. 196–205.

28. Nagarajan R, Scutari M, Lèbre S. Bayesian networks in R. Springer. 2013;122:125–127.