Causal machine learning for assessing pneumococcal vaccine effectiveness: innovations in real-world data analysis and confounding pathway adjustment

MSR151

View all of Parexel's posters at ISPOR Europe

Gregg M[1], Wilson A[1,2], Streja E[1], Vanderpuye-Orgle J[1], Roessner M[1], Alderden J[3]

[1] Parexel International, Waltham, MA, USA [2] University of Utah, Salt Lake City, UT, USA [3] Boise State University, Boise, ID, USA

Background

Determining real-world effectiveness from observational data requires careful consideration of the data generation process to account for confounding to compare treatment groups properly. This is especially true when estimating effects in situations with strong health behavior aspects, such as in vaccine effectiveness. We explore causal assumptions and innovations in causal discovery, utilizing publicly-available real-world data (RWD), to estimate protective effects of pneumococcal vaccination, which helps prevent pneumococcal disease caused by streptococcus pneumoniae bacteria. We demonstrate the need for causal machine learning (ML) methods to supplement traditional methods to de-confound vaccine effectiveness estimates using RWD.

Methods

We used real world dataset Medical Information Mart for Intensive Care (MIMIC) IV version 2.2 (2). The MIMIC IV is a deidentified electronic health records dataset from Beth Israel Deaconess Medical Center in Boston USA. The dataset captures admissions from 2008-2019 for 299,712 patients.

To estimate the causal effect of vaccination on pneumococcal disease, we employed directed acyclic graphs (DAG) to identify potential biasing pathways and then applied targeted maximum likelihood estimation (TMLE)(1) to calculate the estimates. Analysis was performed using R software (v 4.3.0).

Results

- 158,421 patients were included in this analysis, 42,625 with pneumococcal vaccine and 115,796 with no pneumococcal vaccine. (Figure 1)
- We employed DAG to identify potential biasing pathways. (Figure 2)
- Initial analysis indicates a paradoxically elevated risk of pneumonia among vaccinated individuals, with a crude odds ratio (OR) of 1.35 (95% CI: 1.27-1.43). However, we detected a strong imbalance of covariates (in particular, age) between treatment groups. (Figure 3)
- Propensity score matching (with enforced caliper matching) was necessary to achieve cohort balance. When accounting for this imbalance and leveraging TMLE, our data revealed a significant protective effect of the vaccine against pneumonia (Figure 4); TMLE-adjusted OR = 0.78 (95% CI: 0.72-0.84).

Figure 1 Patient Sample

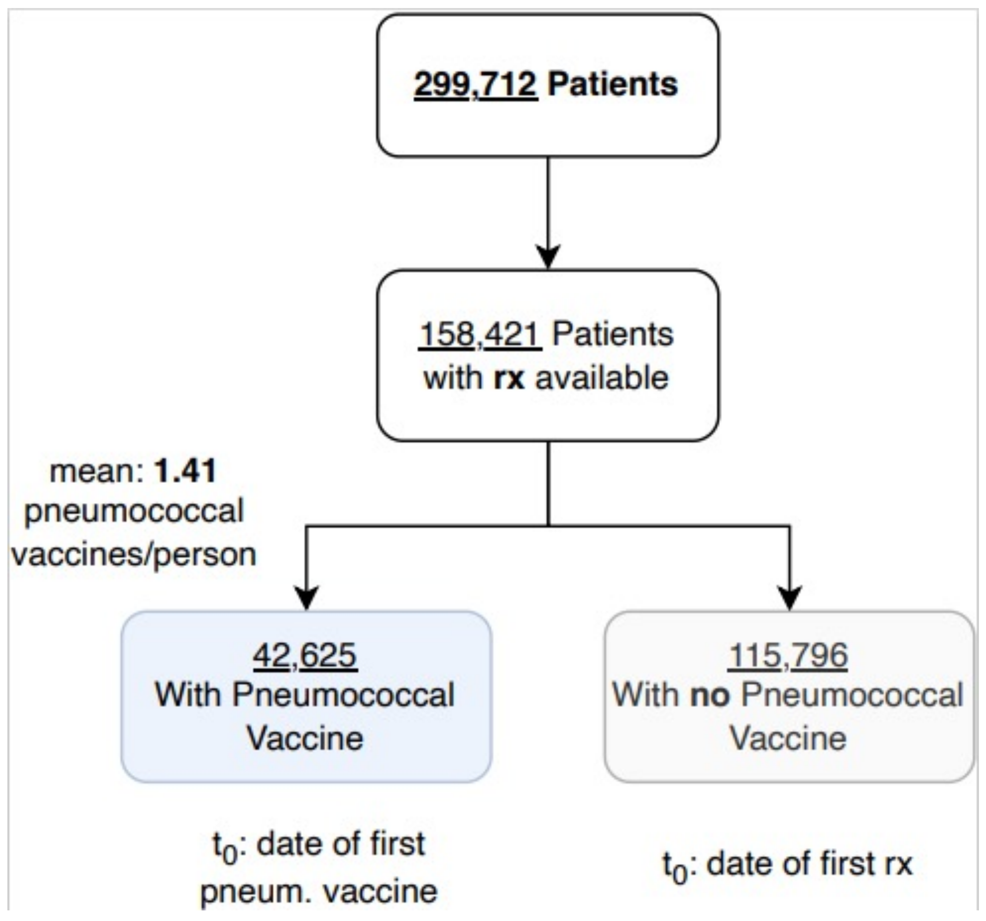


Figure 2 DAG

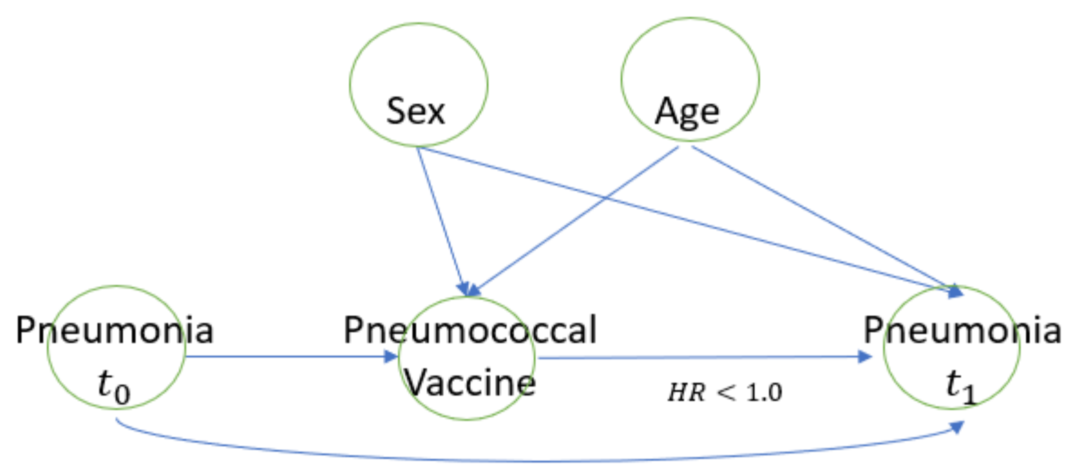


Figure 3 Crude Odds Ratio (before balancing)

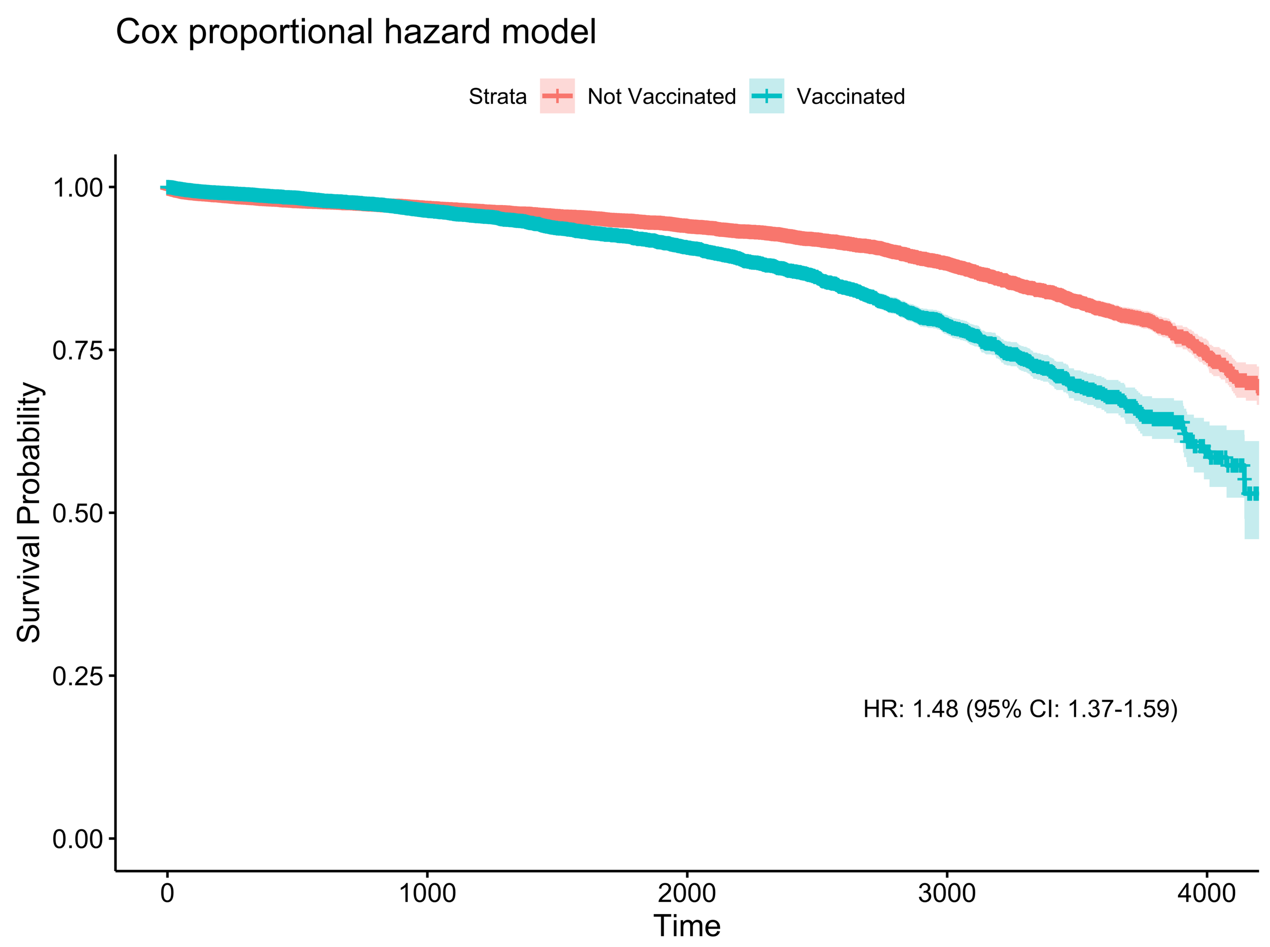
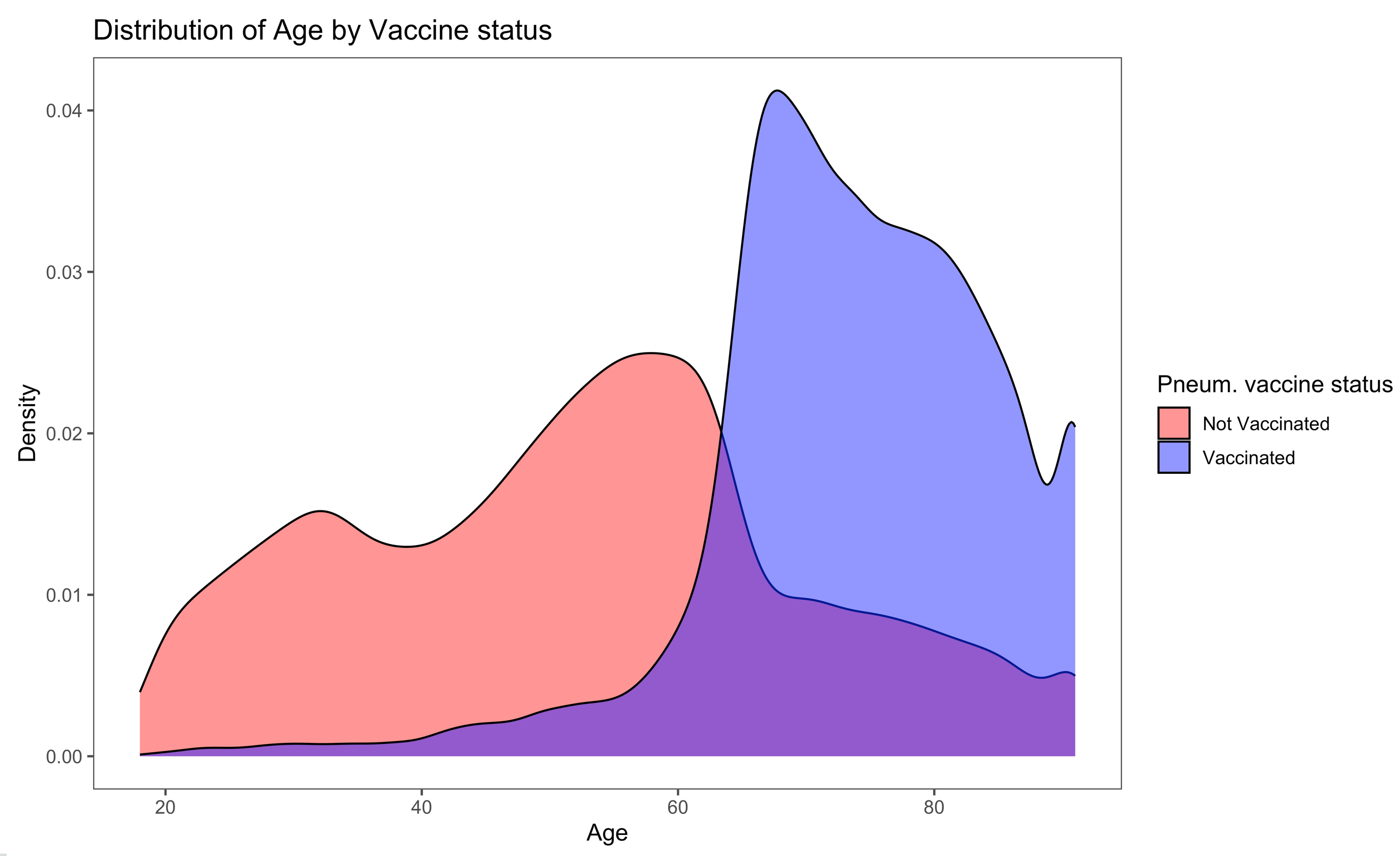


Figure 4 Balancing Cohort by Age



Conclusions

- Real-world datasets and ML models can provide robust vaccine effectiveness estimates and give insights into the causal relationship between vaccination and disease when properly accounting for confounding pathways.
- ML allows for the inclusion of multiple data sets, data types, and can capture interactions for a large number of variables, therefore generating more robust evidence on our understanding vaccine effectiveness.
- Future Research:
  - Analysis of hospital vs. community-acquired and accounting for the type of pneumonia (bacterial vs. viral) will help to better inform personal and policy decisions on vaccination.
- ML methods support the inclusion of social determinants of health in model; an important, and often excluded, group of variables influencing vaccine effectiveness.

REFERENCES

[1] Gruber, S., & Laan, M. van der. (2012). tmle: An R Package for Targeted Maximum Likelihood Estimation. Journal of Statistical Software, 51, 1–35. https://doi.org/10.18637/jss.v051.i13

[2] Johnson, A., Bulgarelli, L., Pollard, T., Horng, S., Celi, L. A., & Mark, R. (2023). MIMIC-IV (version 2.2). PhysioNet. https://doi.org/10.13026/6mm1-ek67.

BOISE STATE UNIVERSITY

THE UNIVERSITY OF UTAH®

parexel®

www.parexel.com