Best of intent, worst of both worlds: why sequentially combining epidemiological designs does not improve signal detection in vaccine surveillance

Faaizah Arshad, Lana YH Lai, George Hripcsak, Daniel Prieto-Alhambra, Martijn J. Schuemie, Marc A. Suchard

Background

Accurate vaccine safety surveillance requires the use of analytic methods that correctly detect all patients with a particular outcome of interest. Ideally, surveillance methods should be highly sensitive and specific to minimize false positive and negative signals [1]. Sensitivity is the proportion of the total cases that have the outcome of interest in the target population that the system correctly detects as being positive. Specificity is the proportion of the total cases that do not have the outcome of interest in the target population that the system correctly detects as being negative.

While some studies have used simulated data to examine the sensitivity and specificity of epidemiological designs in vaccine safety surveillance, little is known about their comparative performance when used alone versus combined sequentially with real-world evidence. Therefore, we selected one highly sensitive method (historical comparator) and one highly specific method (self-controlled case series) and examined their performance in detecting signals for adverse events following vaccination when they were used alone versus combined sequentially [2,3]. The performance comparison employed known negative outcome controls and imputed positive controls as part of the larger Evaluating Use of Methods for Adverse Event Under Surveillance (EUMAEUS) initiative.

Methods

We defined six vaccination exposures of interest at specific historic start and end dates: H1N1 pdm, seasonal flu (Fluvirin), season flu (Fluzone), seasonal flu (all), Zoster (Shingrix) (first or second shot), and HPV (Gardasil 9) (first or second shot). We retrieved clinical records from four sources (Optum EHR, IBM MarketScan Commercial Claims and Encounters, IBM MarketScan Medicare Supplemental Database, and IBM MarketScan Multi-State Medicaid Database) that captured data using the Observational Medical Outcomes Partnership (OMOP) common data model (CDM). We generated a single set of negative control outcomes that were not expected to be caused by the vaccine exposures of interest and that should not generate a signal. The presence of a signal would indicate false positives (type I error). We also relied on imputed positive controls that were known to be caused by the vaccine to empirically calibrate our estimates, such that only 5% of negative controls return a p-value less than 0.05, therefore reducing random and systematic error [4]. We then assessed outcomes that occurred during the time-at-risk window of 1-28 days following vaccination.

We computed effect size estimates with 95% confidence interval with and without empirical calibration, one-sided p-values, type I error, and type II error across all databases and all exposures for the historical comparator method alone, self-controlled case series method alone, and the historical comparator method followed by the self-controlled case series method.

Results

We observed that type I and type II error varied across all methods for a single specified database, exposure, and time-at-risk. In the figures below, bars to the left of 0 indicated type I error, while bars to the right of 0 indicated type II error. The dotted line to the left of 0 was the threshold for the allowable

amount of type I error, with a nominal cutoff of 0.05. If a bar exceeded the dotted red line, the type I error rate was greater than its nominal level. For a time-at-risk of 1-28 days following H1N1 vaccination, the historical comparator method had high type I error but low type II error. Therefore, it was highly sensitive (produced many false positives) but not highly specific. However, the self-controlled case series method had high type II error and low type I error, indicating high specificity (lots of false negatives) and low sensitivity. While there is a clinical intuition that sequentially combining the methods would increase sensitivity and specificity, we observed that the combined method had high specificity and low sensitivity, suggesting that signal detection was not improved compared to when the methods were used alone.
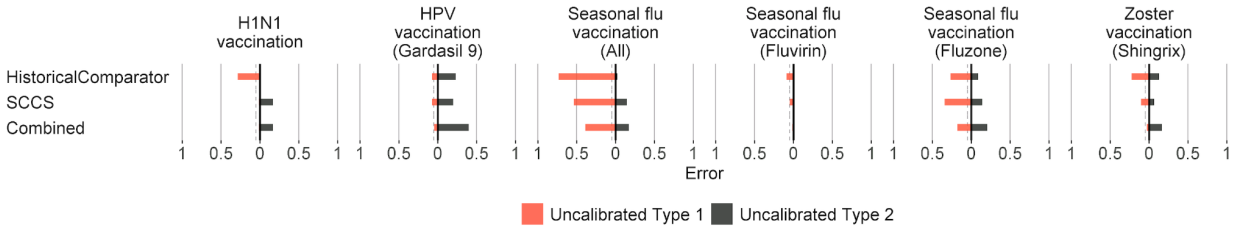

Figure 1. Here, we depict uncalibrated type I error and uncalibrated type II errors for the OptumEhr database using one exposure ID (H1N1 pdm), one outcome ID (74816), a maximum period of observation of 9 months, and a time-at-risk of 1-28 days following vaccination. For the H1N1 pdm vaccination, the historical comparator method had very low type II error (false negative probability), indicating high sensitivity (large number of true positives). However, the self controlled case series method had low type I error (false positive probability), suggesting high specificity. While clinical intuition would therefore predict the combined method to be highly sensitive and highly specific, results demonstrate that the combined method had high specificity and low sensitivity. Therefore, the combined method did not increase accuracy of signal detection.
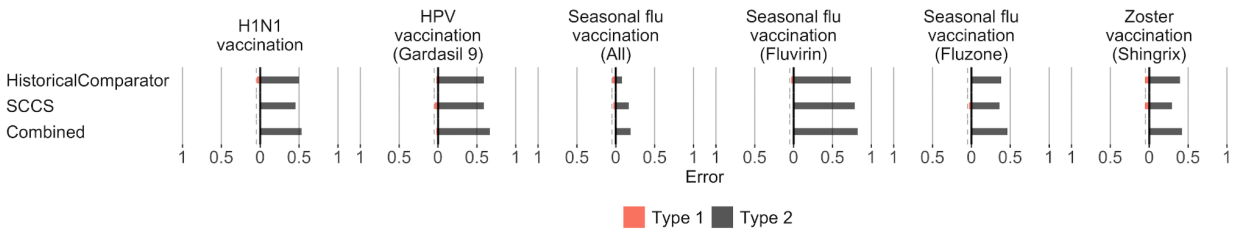

Figure 2. Using the OptumEhr database across all exposures, outcomes, a maximum period of observation, and a time-at-risk of 1-28 days, we calibrated for type I error and type II error, which restored nominal operating characteristics. While results were not consistent across all exposures, the Zoster vaccination showed that the historical comparator method was more sensitive than the self-controlled case series method. However, the combined method did not increase sensitivity and specificity.

Conclusion

Using real-world evidence, we found that it was not desirable to first use a highly sensitive method and then a highly specific method for signal detection of adverse events following vaccination. Particularly, the historical comparator method had low specificity and thus many false positives, while the self-controlled case series method had low sensitivity and many false negatives. Combining methods did not compensate for the individual flaws of each method. Moving forward, clinical tests and vaccine safety monitoring systems should reconsider the accuracy of a sequence of methods to achieve high sensitivity and high specificity when diagnosing an outcome in a population.

References/Citations

1. Shih, M. C., Lai, T. L., Heyse, J. F., & Chen, J. (2010). Sequential generalized likelihood ratio tests for vaccine safety evaluation. *Statistics in medicine*, *29*(26), 2698–2708.
2. Li, X., Lai, L.YH, Ostropolets, A., Arshad, F., Tan, E. H., Casajust, P., et al. (2021). Bias, precision and timeliness of historical (background) rate comparison methods for vaccine safety monitoring: an empirical multi-database analysis. *British Medical Journal*. Not published yet.
3. Simpson, S. E., Madigan, D., Zorych, I., Schuemie, M. J., Ryan, P. B., & Suchard, M. A. (2013). Multiple self-controlled case series for large-scale longitudinal observational databases. *Biometrics*, *69*(4), 893–902.
4. Schuemie, M. J., Hripcsak, G., Ryan, P. B., Madigan, D., & Suchard, M. A. (2018). Empirical confidence interval calibration for population-level effect estimation studies in observational healthcare data. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(11), 2571–2577.