The performance of overlap weighting vs inverse probability of treatment weights to minimize confounding and systematic error in COVID-19 vaccine effectiveness: an empirical evaluation study

Martí Català[1], Edward Burn[1], Trishna Rathod-Mistry[1], Antonella Delmestri[1], Daniel Prieto-Alhambra[1,2], Annika M. Jödicke[1]

[1]Center for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, UK; [2]Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, Netherlands

Background

Following the start of the COVID-19 vaccination programs, routinely collected data are being widely used to evaluate the effectiveness and safety of COVID-19 vaccines. Careful consideration of how to account for confounding is required when comparing vaccinated and unvaccinated people. While several methods were used in previous studies, a rigorous assessment of their ability to resolve confounding was not completed[1].

Our study provides an empirical evaluation of the comparative performance of two propensity score methods to minimise confounding in the study of COVID-19 vaccine/s effectiveness: overlap weighting (OW)[2], and inverse probability of treatment weights (IPTW)[3].

Methods

Data: We used primary care data from the UK Clinical Practice Research Datalink (CPRD) AURUM, mapped to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM)[4].

Study population: All people aged ≥ 75 years, who were not previously infected with or vaccinated against SARS-CoV-2 and were registered in CPRD AURUM England practices for >= 180 days before study start date (January 4th, 2021) were eligible. Subsequently, individuals were assigned to the "vaccinated" and "unvaccinated" cohorts, based on whether they were vaccinated against COVID-19 between 4th and 28th of January 2021. Index dates for people in the vaccinated cohort was defined as their vaccination date. Meanwhile, index dates for unvaccinated people were randomly assigned following the distribution of index dates in the vaccinated cohort, as can be seen in Figure 1. After assignment of the respective index dates, individuals with a recording of COVID-19 infection before or at index date were excluded.

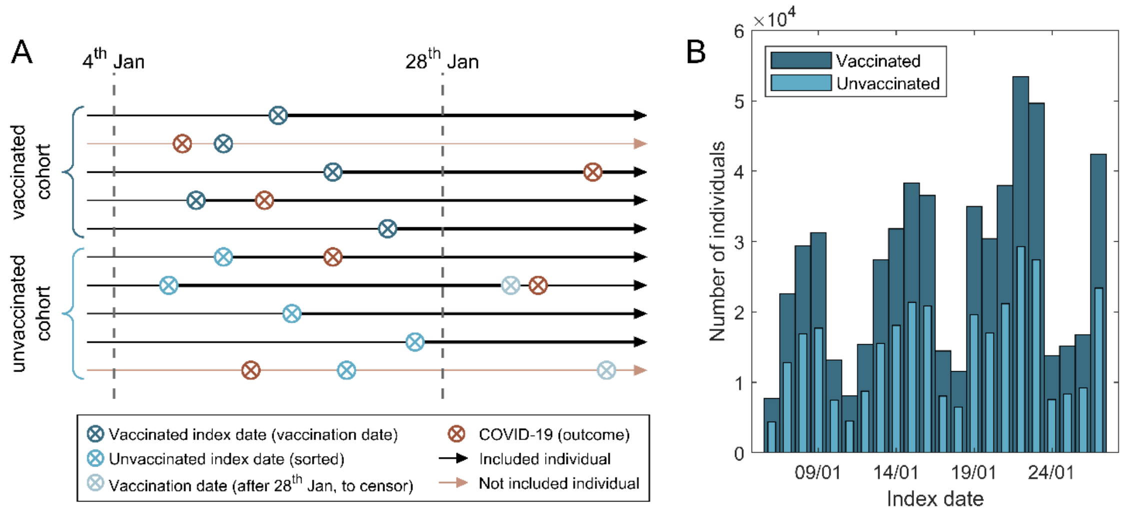

Figure 1. Unvaccinated sorting index date. (A) Diagram to show difference between index date and vaccination date for both cohorts; and prior COVID-19 exclusion criteria. Follow-up period is highlighted with a thick line. (B) Distribution of index dates for vaccinated and unvaccinated cohorts.

Methods to account for confounding: Large-scale Propensity Scores (PS) were used to minimise observed confounding based on baseline covariates.

PS estimation: Covariates to be included in the PS equation were extracted, including condition occurrences for three different time windows (1 to 30 days before index date, 31 to 180 days before index date and 181 days to any time before index date), and drug exposures for 2 time periods (1 to 30 days, 31 to 180 days before index date). Subsequently, all covariates with a frequency >0.5% were included in a lasso regression, which was used to identify relevant covariates to be included in the large-scale PS. In addition, the following variables were forced into the PS equation: location (region identifier; 9 different ones; or General Practice (GP) surgery identifier; 1357 different GP); age as categorical (5-year bands) and as a continuous variable (using a 2-degree polynomial for non-linearity); prior observation years; regional vaccination, testing and incidence rates on index date. PS were computed using a logistic regression model, with 3 different representations of location: without location (PS_{base}), location defined as region (PS_{reg}) or de-identified GP surgery (PS_{GP}).

PS Weighting: We used and compared two different weighting methods: Inverse Probability of Treatment Weighting (IPTW) with trimming at [0.05-0.95], and Overlap Weighting (OW).

Metrics: The following metrics were used to assess the performance of both weighting methods to minimise bias:

(1) Covariate imbalance as a proxy of measured confounding was assessed by calculating standardized mean differences (SMD) between vaccinated vs unvaccinated cohorts after PS weighting
(2) The association between vaccination status and Negative Control Outcomes (NCO) was estimated using Cox proportional hazard regression to detect unmeasured confounding.

Results

The two cohorts were identified with 582,223 individuals in the vaccinated cohort and 322,114 in the unvaccinated cohort. A total of 29 covariates with an SMD > 0.1 were identified before PS weighting: GP practice (0.74) and region (0.14); age (0.31) and age group (0.53); variables related to the number and occurrence of GP visits, and COVID-19 tests; and other covariates such as pulse rate measurement. Location identifiers were clearly the most important ones.

Figure 2 depicts SMD values for the weighted cohorts. We can see that despite PS weighting GP practice covariate was only balanced using OW with PS_{GP}. IPTW did not yield sufficient balance for GP practice in any of the estimated PS. Similar findings were seen for region, which was not balanced in any of the weighting methods unless location was included in the model. OW performed better than IPTW in terms of covariate balance, with lower minimum SMD for all covariates.

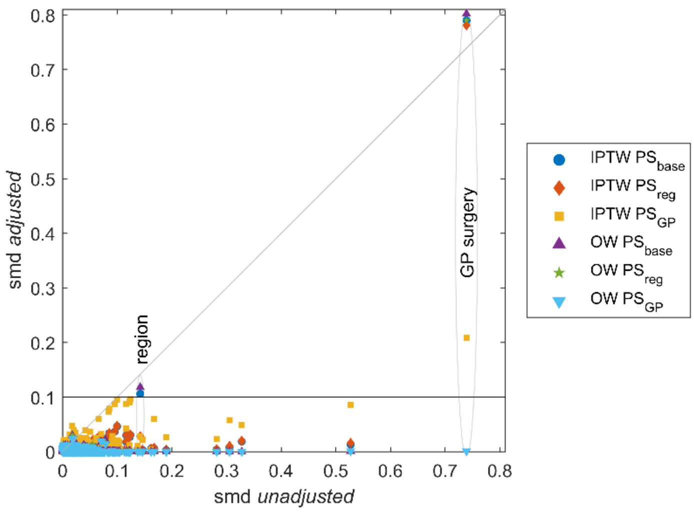

Figure 2. SMD values for the weighted cohorts compared to the unadjusted ones.

In figure 3 we can observe the results of negative control outcome analyses. In general, OW showed lower systematic error than IPTW in most scenarios. Unweighted analyses show, as expected, clear evidence of one-sided systematic error, with many negative control outcomes positively associated with vaccine status (HR>1).

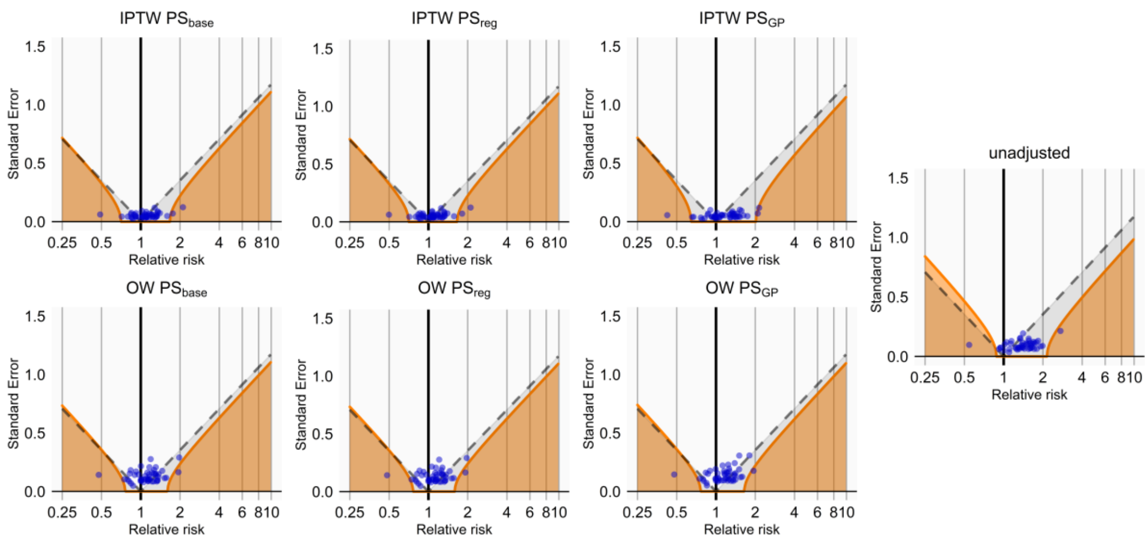

Figure 3. Negative control outcomes (NCO) hazard ratios and standard deviation. Each blue dot represents a different NCO. Gray dashed (without calibration) and orange lines mark significance thresholds. Calibration results are observed in orange. To elaborate this plot EmpiricalCalibration OHDSI package was used[5].

Conclusion

We demonstrate for the first time that OW are preferable to IPTW to minimize observed and unobserved confounding in COVID-19 vaccine effectiveness research. Additionally, our findings illustrate the need to incorporate patient location (e.g. GP practice identifier or region of residence) and related variables (e.g. testing and transmission rates) to minimize community- rather than patient-level confounding in COVID-19 vaccine effectiveness.

References

1. World Health Organization. Evaluation of COVID-19 vaccine effectiveness: interim guidance, 17 March 2021. World Health Organization; 2021.
2. Thomas LE, Li F, Pencina MJ. Overlap Weighting: A Propensity Score Method That Mimics Attributes of a Randomized Clinical Trial. JAMA. 2020;323(23):2417–2418.
3. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Statistics in medicine. 2015 Dec 10;34(28):3661-79.
4. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. J Am Med Inform Assoc. 2012;19(1):54-60.
5. Schuemie MJ, Ryan PB, DuMouchel W, Suchard MA, Madigan D. Interpreting observational studies: why empirical calibration is needed to correct p-values. Statistics in medicine. 2014 Jan 30;33(2):209-18.