The use of data-driven vs. clinical based propensity score in COVID-19 vaccine safety research:

Association between thrombosis with thrombocytopenia syndrome (TTS) or thromboembolic events (TE), and COVID-19 vaccines

PRESENTER: **Xintong** Li

INTRO

- Propensity score (PS) have been widely used in observational studies to reduce confounding by indication
- Clinical knowledge based vs. data-driven PS

METHODS

Data source: OMOPed data from 5 European counties: France, Germany, Netherlands, Spain, and the United Kingdom) and two from the United States.

Cohort study:

- Target: adenovirus-based

- Comparator: mRNA

Analysis:

- miniPS: clinically-driven
- Large-scale PS: data-driven, L1 regularized logistic regression
- 1-to-4 matching

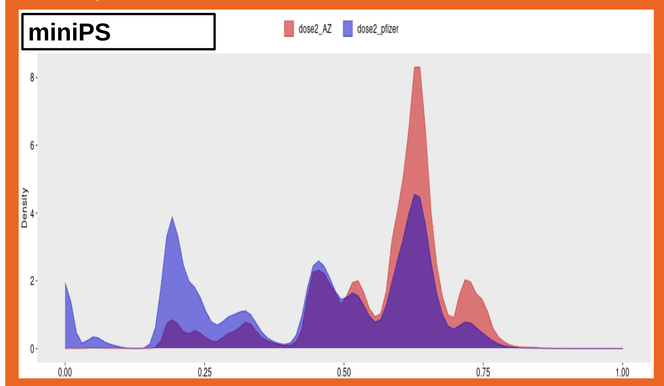
Diagnostics:

1. Measured confounding: Covariate balance after propensity score matching (SMD < 0.1)

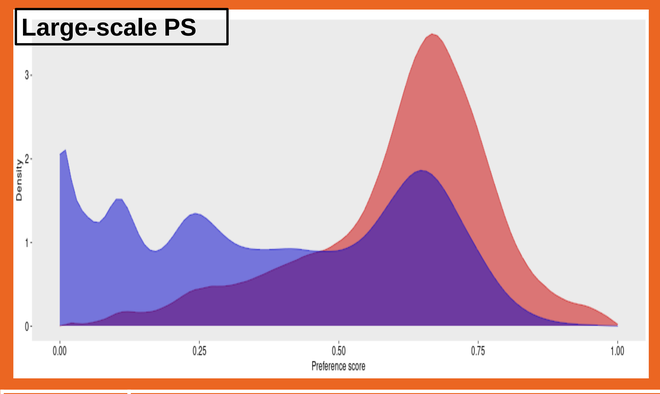
2. Power: minimal detectable relative risk in the matched cohorts

3. Systematic error: using negative control outcomes

Figure 1. Propensity score distribution covariates with top 6 absolute values of Beta, 2nd dose Vaxzevria and Comirnaty cohorts, UK CPRD data.



Beta	Covariate
-7.12	index month: 1
-4.10	index month: 12
-2.84	index month: 2
-1.26	index month: 3
-0.81	age group: 10 - 19
-0.76	age group: 20 - 29



Beta	Covariate
-7.05	observation distinct concept count during day -180 through -4 concept_count relative to index
-6.61	index month: 1
-2.21	index month: 2
1.95	observation during day -180 through -4 days relative to index: Housebound
-1.79	procedure_occurrence during day -180 through -4 days relative to index: Administration of vaccine
-1.77	index month: 12

While selected confounders were balanced in **clinical-based PS** after matching, other potentially relevant covariates remained unbalanced, suggesting **residual confounding**

Figure 2. Before and after matching SMD, 2nd dose Vaxzevria and Comirnaty cohorts, UK CPRD data

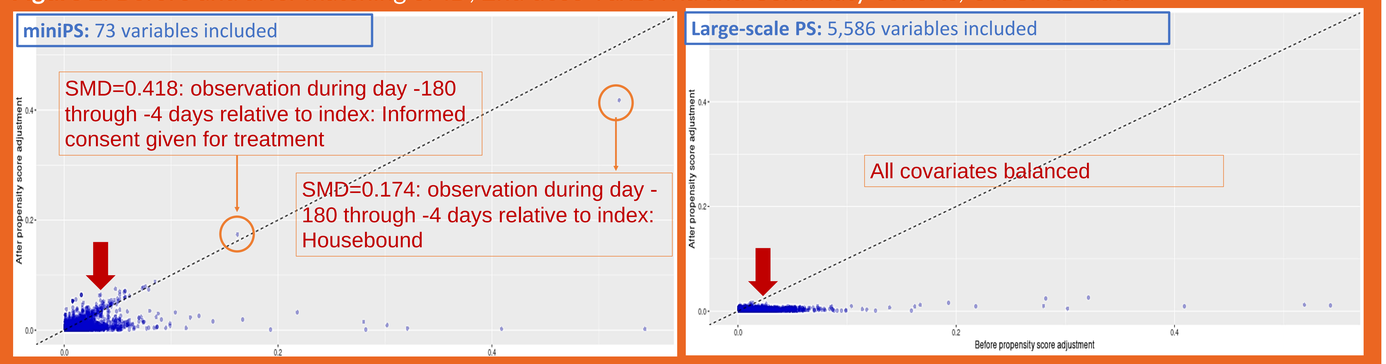
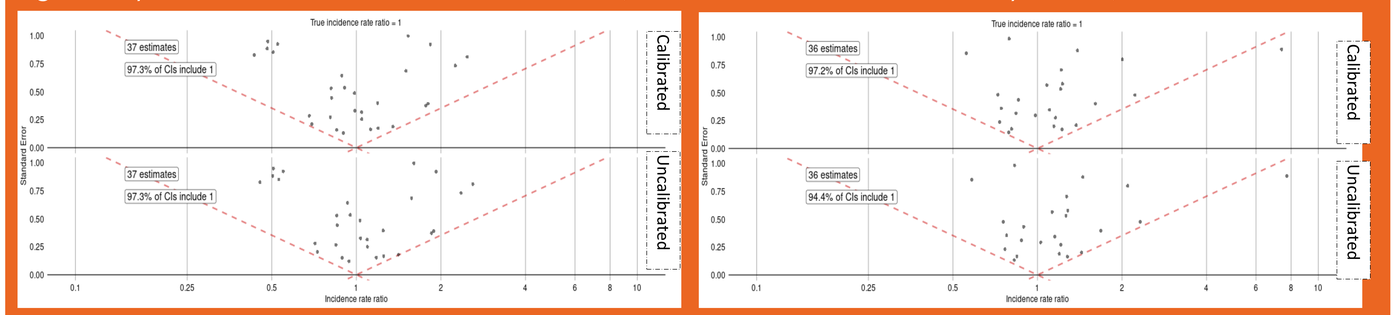


Figure 3. Systematic errors before and after calibration, 2nd dose Vaxzevria and Comirnaty cohorts, UK CPRD data .



RESULTS

- PS distribution
- Before and after matching SMD
- Systematic error using negative control outcomes

CONCLUSIONS

- Index month and age have high impact for both clinical based and data-driven propensity scores.
- Clinical-based PS: balanced on selected variables, but not other covariates
- Large-scale PS: all covariates were well-balanced after matching
- Performance on controlling systematic errors were similar
- Cons of large-scale: computing time (30mins vs. 6 hours on a 250,000 down sampling cohort)

Table 1. Summary of the covariate balance for both propensity scores.

			No SMD > 0.1 after matching	
Database	**Target**	**Comparator**	**Mini PS**	**Large-scale PS**
UK CPRD Aurum	Vaxzevria 1st	Comirnaty 1st	×	√
UK CPRD Aurum	Vaxzevria 2nd	Comirnaty 2nd	×	√
Germany DA	Janssen	Comirnaty 1st	×	√
NL IPCI	Vaxzevria 1st	Comirnaty 1st	×	√
US OpenClaims	Janssen	Comirnaty 1st	√	√
US OpenClaims	Janssen	Spikevax 1st	√	√

*CPRD AURUM: Clinical Practice Research Datalink (CPRD) Aurum, United Kindom; IPCI: Integrated Primary Care Information (IPCI), The Netherlands; DA Germany: IQVIA Disease Analyser (DA) Germany; US OpenClaims: Medical and Institutional Claims (Dx and Hx); SMD: standardized mean difference.

Xintong Li[1], Edward Burn[1,2], Prieto-Alhambra[1,3]

1.Centre for Statistics in Medicine (CSM), Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (NDROMS), University of Oxford, UK
2.Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain
3. Research Programme on Biomedical Informatics, Hospital del Mar Medical Research Institute, Faculty of Health and Life Sciences, Universitat Pompeu Fabra, Barcelona, Spain

NDORMS NUFFIELD DEPARTMENT OF ORTHOPAEDICS, RHEUMATOLOGY AND MUSCULOSKELETAL SCIENCES · UNIVERSITY OF OXFORD · OHDSI