Inverse probability of treatment weighting versus propensity score matching in the Sentinel System: a test case

Marie C Bradley, PhD, MPharm, MScPH[1]; Talia Menzin MS[2]; Joy Kolonoski, MPH[2]; Mayura Shinde, DrPH[2]; Rongmei Zhang, PhD[3]; Efe Eworuke PhD, MSc[1]; David J Graham, MD, MPH[1]; Laura Hou, MS[2]; Adebola Ajao, PhD, MPH[1]; Po-Yin Chang, PhD[1]; John G Connolly, ScD[2]

[1]Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, Food and Drug Administration. Silver Spring, MD

[2]Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA

[3]Office of Biostatistics, Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, MD

Acknowledgment: The authors thank the Sentinel Data Partners who provided data used in the analysis. This project was supported by Task Order 75F40119F19002 under Master Agreement 75F40119D10037 from the U.S. Food and Drug Administration (FDA). The views expressed are those of the authors and are not intended to convey official US Food and Drug Administration policy or guidance.

Background and Objective

- **Propensity score matching (PSM)** is widely used for confounding adjustment in real world evidence (RWE) studies evaluating outcomes associated with drug use in routine care
- However, it has an important limitation of removing unmatched observations from the analysis, potentially leading to increased covariate imbalance, reduced sample size and limited generalizability
- In contrast, **weighting on the propensity score** has several advantages, including increased precision through retaining most observations and flexibility for targeting specific populations for inference
- The FDA Sentinel System recently added inverse probability of treatment weighting (IPTW) for confounding adjustment in comparative safety and effectiveness studies

Objective: To evaluate IPTW in the Sentinel System by comparing adjusted effect estimates obtained using a PSM approach versus an IPTW approach for a study of stroke and bleeding risk in patients aged 65 years or older initiating nonvitamin K oral anticoagulants (NOAC) for non-valvular atrial fibrillation (NVAF)

Methods

- Retrospective new user cohort study among standard dose NOAC users with NVAF, aged ≥ 65 years between October 19, 2010, to September 30, 2015, in the Sentinel Medicare data partner only
- New initiators of standard dose apixaban, dabigatran, rivaroxaban, with a diagnosis of NVAF in the previous 183 days were identified
- Three pairwise NOAC-NOAC comparisons: Rivaroxaban vs. Dabigatran, Rivaroxaban vs. Apixaban, Dabigatran vs. Apixaban
- PSM (1:1) and IPTW with stabilized average treatment effect weights were applied separately for each pairwise comparison. Cox proportional-hazards regression was used to estimate adjusted hazard ratios (HR) and 95% confidence intervals (CI) for the outcomes of thromboembolic stroke (stroke), intracranial hemorrhage (ICH), major extracranial bleeding (MEB), and GI bleeding (GIB) comparing each NOAC with each other
- Outcomes were defined using previously validated algorithms based on *ICD-9-CM* diagnosis codes

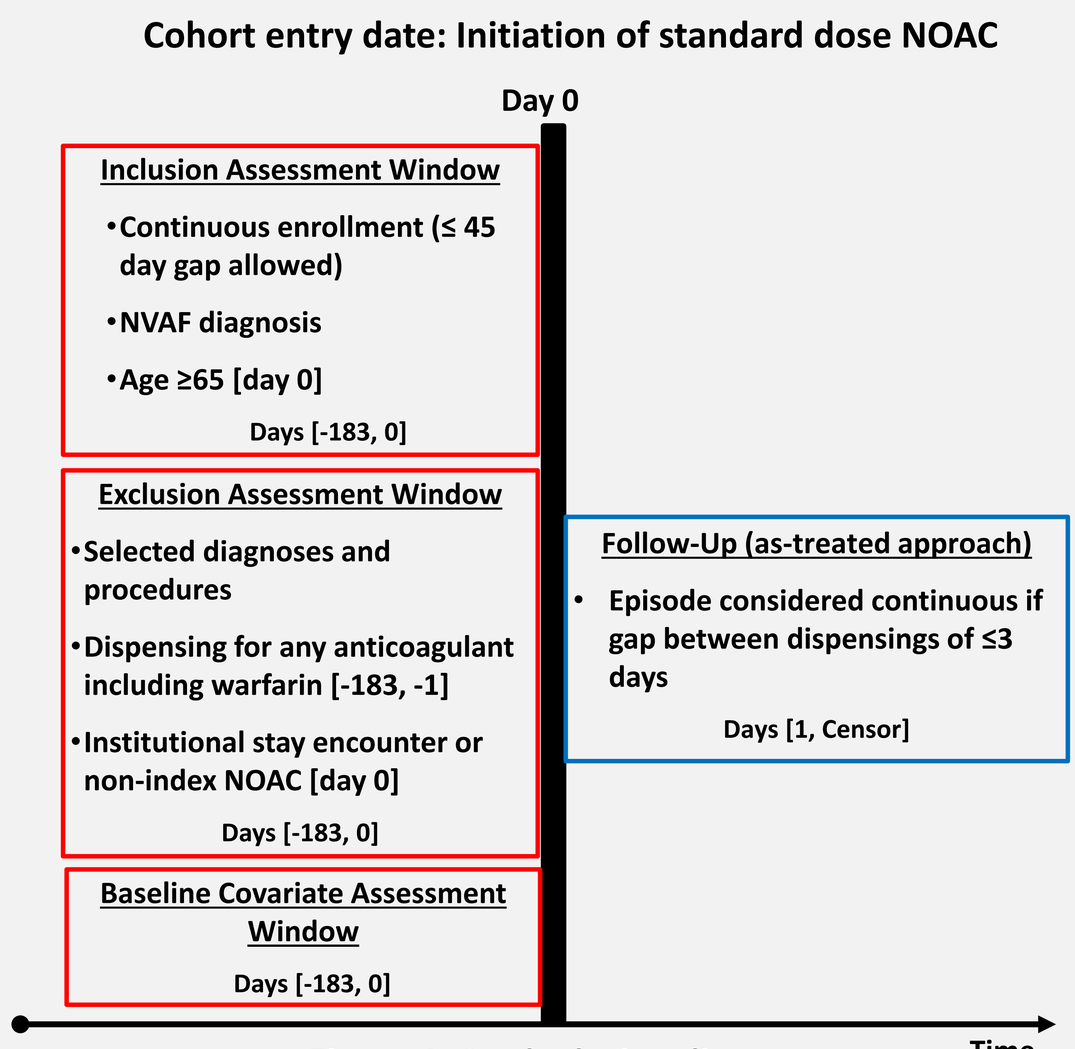

Figure 1: Study design diagram

Inclusion criteria

- Continuous enrollment for ≥183 days
- NVAF diagnosis
- Age ≥65 years

Exclusion criteria

- Dialysis, kidney replacement, deep vein thrombosis, pulmonary embolism, joint replacement, mitral stenosis, valve replacement or repair
- Other anticoagulant dispensing
- Institutional stay encounter

Baseline Covariates

- Demographics
- Medical conditions and medication use
- Stroke and bleeding risk scores
- Health care utilization

Censoring Criteria

- Death, query end date, disenrollment, any outcome event, end of exposure episode, comparator drug dispensing, low-dose of current exposure, warfarin dispensing, other NOAC dispensing, kidney transplant or dialysis, institutional stay encounter

Results

- Overall, the point estimates and 95% CIs were similar between the analyses using PSM and IPTW
- For example, there was no difference in the risk of stroke comparing rivaroxaban versus dabigatran after adjustment using PSM (HR [95% CI]: 0.89 [0.74, 1.07]) or IPTW (0.90 [0.76, 1.06])
- Similarly, rivaroxaban use was associated with a numerical but non-significant increased risk of ICH compared to apixaban in both analyses [PSM HR (95% CI) 1.28 (0.99, 1.67) and IPTW 1.23 (0.96, 1.58)]
- An increased risk of MEB and GIB was observed comparing rivaroxaban and apixaban users after both PSM and IPTW adjustment

Table 1: Adjusted hazard ratios (95% confidence intervals) for each NOAC pairwise comparison and thromboembolic stroke, intracranial hemorrhage, major extracranial (including major gastrointestinal) bleeding, and major GI bleeding using PSM and IPTW approaches

	HR (95%CI)			
	Thromboembolic stroke	**Intracranial hemorrhage**	**Major extracranial bleed**	**Major GI bleed**
PSM				
Rivaroxaban vs. Dabigatran (N= 82,326)	0.89 (0.74, 1.07)	1.67 (1.29, 2.17)	1.21 (1.12, 1.32)	1.17 (1.08, 1.28)
Rivaroxaban vs. Apixaban (N= 75,889)	1.00 (0.82, 1.22)	1.28 (0.99, 1.67)	2.29 (2.06, 2.55)	2.32 (2.07, 2.59)
Dabigatran vs. Apixaban (N= 69,054)	1.15 (0.93, 1.40)	0.75 (0.55, 1.03)	1.96 (1.75, 2.20)	2.04 (1.81, 2.31)
IPTW				
Rivaroxaban vs. Dabigatran (N= 110,111 vs. 84,481)	0.90 (0.76, 1.06)	1.58 (1.23, 2.03)	1.20 (1.11, 1.30)	1.16 (1.07, 1.25)
Rivaroxaban vs. Apixaban (N= 111,814 vs. 77,234)	0.99 (0.82, 1.19)	1.23 (0.96, 1.58)	2.33 (2.11, 2.58)	2.35 (2.11, 2.61)
Dabigatran vs. Apixaban (N= 84,600 vs. 76,863)	1.13 (0.93, 1.37)	0.74 (0.55, 1.00)	1.93 (1.73, 2.15)	2.01 (1.79, 2.26)

Discussion and Conclusions

- This study demonstrated the feasibility of a newly implemented distributed approach to IPTW in the Sentinel System
- Effect estimates after applying PSM and IPTW were similar for all comparisons despite differences in sample size
- Similarities may be explained by relatively small reduction in the number of outcomes after applying PSM compared to IPTW
- Further studies to compare IPTW and alternative propensity score adjustment approaches in the Sentinel System are needed, especially, where PSM limitations are evident, such as reduction in sample size and limited generalizability