RWD195: Estimating the Average Treatment Effect Among Patients With Heart Failure Using Targeted Maximum Likelihood Estimation: Implications From a Large-Scale Health Administrative Database

Fei Zhao[1], Yao Xu[1], Xiaojue Liu[1], Jiayue Tian[1], Koki Idehara[1], Seok-Won Kim[1]
1. IQVIA Solutions Japan G.K., Tokyo, Japan

BACKGROUND AND RATIONAL

Observational studies using real-world data are susceptible to confounding bias if covariate imbalance is not properly addressed. Traditional methods such as inverse probability of treatment weighting (IPTW) and Cox proportional hazards models are widely used for confounder adjustment due to their simplicity, but they rely on strong assumptions (e.g., correct model specification, independent censoring, positivity) and may yield unstable estimates. Recently, more robust approaches like G-computation and targeted maximum likelihood estimation (TMLE) have gained attention for their flexibility in model specification, handling missing data, and addressing non-collapsibility [1]. Heart failure (HF) is a clinical syndrome caused by structural or functional cardiac abnormalities, leading to impaired ventricular filling or ejection and resulting in symptoms such as dyspnea and fatigue. In Japan, currently approved therapies for heart failure include RAAS inhibitors, beta-blockers, MRAs, SGLT2 inhibitors and sGC stimulator, but these treatments may still be associated with substantial residual risk of clinical events [2]. Selecting accurate confounder-adjustment methods are crucial for estimating treatment effects in HF patients, leveraging real-world data to inform clinical decisions.

METHODS

This retrospective observational cohort study used the large-scale and all-age covered IQVIA Claims Plus2 Database from October 2020 to September 2024. Patients diagnosed with heart failure (ICD-10: I50, I11.0) during the study period who initiated Drug Class A or B at age ≥18 between April 2021 and September 2024 were included. The index date was defined as treatment initiation, with follow-up up to 1-year or until death, loss to follow-up, or study end. The primary outcome was 1-year all-cause inpatient mortality. We estimated the hazard ratio and average treatment effects (ATE) between Drug Class A and B using multiple effect measures, including risk difference, survival probabilities at 12 months and restricted mean survival time (RMST). Crude estimates were compared with confounder-adjusted results using IPTW, G-computation (employing Cox PH model), and targeted maximum likelihood estimation with Super Learner (ensemble of general linear model and XGboost). Sensitivity analyses were explored in the subgroup of patients aged ≥65. All analyses were conducted in R (v4.3.3).

OBJECTIVES

To assess the consistency of inpatient mortality risk estimates across different confounder-adjusted methods among HF patients treated with two drug classes.

CONCLUSION

This study showed that all confounder-adjusted methods consistently outperformed unadjusted estimates in estimating inpatient mortality. Methods incorporating the outcome mechanism (such as G-computation and TMLE) were more efficient and robust than those relying solely on the treatment mechanism, like IPTW. Although the three approaches yielded similar conclusions in this study, TMLE may give substantially different results from estimation methods that are not doubly robust or do not exploit data-adaptive model specification [3]. Therefore, validating causal effects using multiple confounder-adjusted methods in observational studies is essential to ensure reliability and to support clinical and policy decision-making.

RESULTS

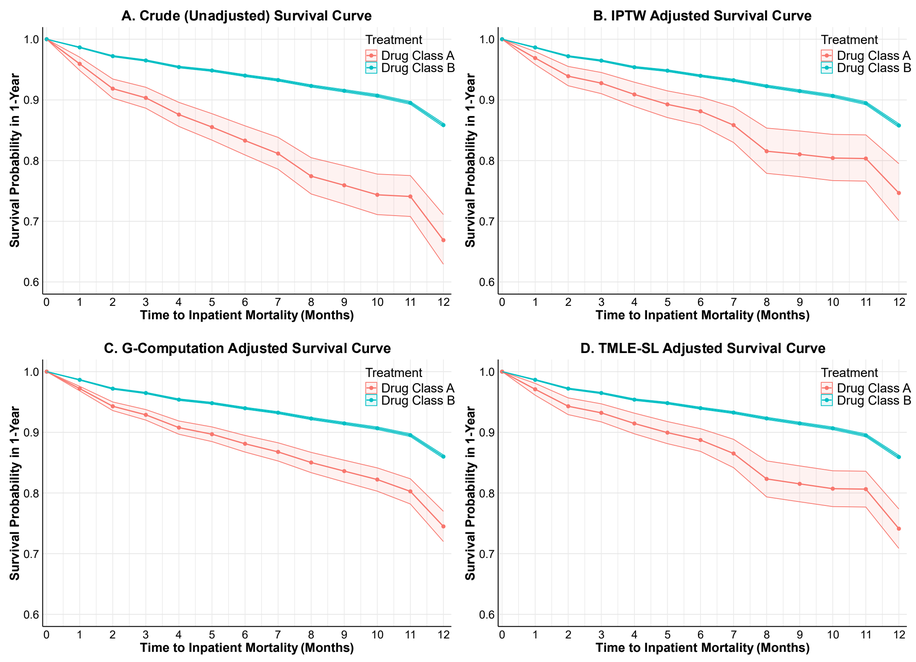
Table 1. Baseline demographic and clinical characteristics

Baseline characteristics	Cohorts		P-value	SMD
	Drug Class A N=1,258	Drug Class B N=108,126		
Gender (*n, %*)				
Female	456 (36.2)	47,539 (44.0)	<0.001	0.158
Male	802 (63.8)	60,587 (56.0)		
Age [*Mean (SD)*]	77.90 (11.91)	76.99 (12.46)	0.01	0.075
Age category (*n, %*)				
18 ≤ Age < 65	157 (12.5)	16,709 (15.5)	0.012	0.088
Age in 65-75	212 (16.9)	18,384 (17.0)		
Age ≥75	888 (70.6)	73,002 (67.5)		
Prior medical history (*n, %*)				
Dialysis	46 (3.7)	2,786 (2.6)	0.021	0.062
Cardiovascular procedure	11 (0.9)	342 (0.3)	0.001	0.073
Heart failure event	1,064 (84.6)	62,331 (57.7)	<0.001	0.624
Comorbidities (*n, %*)				
Myocardial infarction	490 (39.0)	25,683 (23.8)	<0.001	0.332
Stroke	262 (20.8)	21,796 (20.2)	0.581	0.017
Pulmonary embolism	46 (3.7)	2,855 (2.6)	0.032	0.058
Peripheral artery disease	425 (33.8)	34,232 (31.7)	0.114	0.045
Aortic dissection	1,018 (80.9)	73,548 (68.0)	<0.001	0.299
Hypertension	1,107 (88.0)	103,892 (96.1)	<0.001	0.302
Coronary artery disease	932 (74.1)	61,318 (56.7)	<0.001	0.372
Atrial fibrillation	714 (56.8)	38,008 (35.2)	<0.001	0.444
Diabetes mellitus	1,012 (80.4)	80,394 (74.4)	<0.001	0.146
Anemia	536 (42.6)	38,096 (35.2)	<0.001	0.152
Hyperkalemia	147 (11.7)	8,377 (7.7)	<0.001	0.133
Chronic Obstructive Pulmonary Disease (COPD)	109 (8.7)	5,266 (4.9)	<0.001	0.151
Chronic kidney disease (CKD)	406 (32.3)	23,801 (22.0)	<0.001	0.232
Severe CKD	130 (10.3)	6,895 (6.4)	<0.001	0.143
Prior therapies* history (*n, %*)				
None	314 (25.0)	44,277 (40.9)	<0.001	0.589
Monotherapy	308 (24.5)	36,601 (33.9)		
Dual therapy	316 (25.1)	18,605 (17.2)		
Triple therapy	247 (19.6)	6,889 (6.4)		
Quadruple therapy	73 (5.8)	1,754 (1.6)		

*: Foundation therapies include RAAS, Beta-Blockers, MRA, and SGLT2i;
Abbreviation: SD=standard deviation; SMD=standard mean difference;

Almost all baseline covariates showed an imbalance between the two drug classes. Unadjusted estimates indicate significantly higher probability and hazard of inpatient mortality for drug class A compared to drug class B. Adjusting for baseline confounders using three approaches shifts all estimated treatment effects downward, reducing both the magnitude of the difference and the ratio between the two classes.

Figure 1. Estimated 1-year Inpatient Survival Curve



The 1-year risk difference for inpatient mortality was estimated as 0.045 (0.025, 0.066) by TMLE-SL and 0.047 (0.032, 0.063) by G-computation, both lower than the IPTW estimate of 0.054 (0.031, 0.077). TMLE-SL and G-computation produced similar point estimates, while G-computation yielded narrower 95% confidence intervals except for RMST difference. Similar patterns were observed in sensitivity analyses.

Figure 2. ATE and Hazard Ratio in Sensitivity Analysis

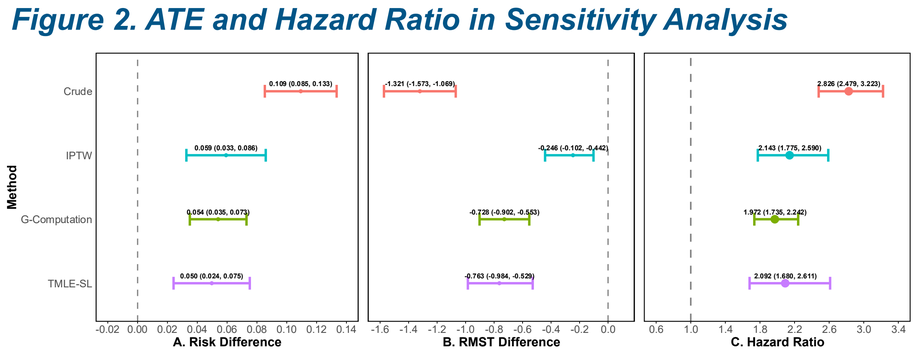


Table 2. Treatment Effect Estimates on 1-Year All-Cause Inpatient Mortality

	Crude (Unadjusted) Estimation		IPTW Adjusted Estimation		G-Computation Adjusted Estimation		TMLE-SL Adjusted Estimation	
	Drug Class A N=1,257	Drug Class B N=108,095	Drug Class A ESS=101,439	Drug Class B ESS=109,355	Drug Class A N=1,257	Drug Class B N=108,095	Drug Class A ESS=NA	Drug Class B ESS=NA
Patient number of inpatient mortality (*n, %*)	234 (18.6%)	9,146 (8.5%)	14,070 (13.9%)	9,303 (8.5%)	234 (18.6%)	9,146 (8.5%)	NA	NA
All-cause inpatient mortality during 1 year after treatment initialization (binary outcome)								
Risk ratio (*95% CI*)	2.20 (1.96, 2.47)		1.63 (1.38, 1.93)		1.56 (1.38, 1.74)		1.56 (1.33, 1.80)	
Risk different (*95% CI*)	0.102 (0.080, 0.123)		0.054 (0.031, 0.077)		0.047 (0.032, 0.063)		0.045 (0.025, 0.066)	
Time from treatment initialization to All-cause inpatient mortality (time-to-event outcome)								
Survival probability at 12 month	0.669 (0.629, 0.711)	0.859 (0.856, 0.861)	0.747 (0.701, 0.797)	0.858 (0.855, 0.861)	0.750 (0.720, 0.770)	0.860 (0.857, 0.863)	0.741 (0.709, 0.774)	0.859 (0.856, 0.862)
RMST [*Mean (95% CI)*]	10.18 (9.95,10.40)	11.34 (11.32, 11.35)	11.64 (11.47, 11.80)	11.85 (11.84, 11.86)	10.71 (10.55, 10.88)	11.34 (11.32, 11.35)	10.66 (10.50, 10.78)	11.34 (11.33, 11.35)
RMST difference (*95% CI*)	-1.16 (-1.39, -0.94)		-0.21 (-0.064, -0.30)		-0.63 (-0.79, -0.47)		-0.70 (-0.84, -0.56)	
Hazard ratio (*95% CI*)	2.83 (2.48, 3.22)		2.30 (1.93, 2.74)		1.95 (1.66, 2.29)		2.06 (1.69, 2.52)	

Abbreviation: ATE=average treatment effect; IPTW=inverse probability of treatment weighting; ESS=effective sample size; TMLE-SL=targeted maximum likelihood estimation with supper learner;
CI=confidential interval; RMST=restricted mean survival time; RR=risk ration; HR=hazard ratio; Estimates adjusted by all baseline covariates listed in Table 1 in each corresponding models;

LIMITATION

Due to computational burden of TMLE-SL model, only common and simple models with 5-fold cross-validation were employed, which may not obtain the most optimal estimate.

REFERENCE

[1] Talbot, Denis, et al. "Guidelines and Best Practices for the Use of Targeted Maximum Likelihood and Machine Learning When Estimating Causal Effects of Exposures on Time-To-Event Outcomes." Statistics in Medicine 44.6 (2025): e70034.
[2] Marcondes-Braga, Fabiana G., et al. "Emerging Topics in Heart Failure: New Era of Pharmacological Treatment." Arquivos Brasileiros de Cardiologia 115 (2020): 956-960.
[3] Decker, Anna L., et al. "Semiparametric estimation of the impacts of longitudinal interventions on adolescent obesity using targeted maximum-likelihood: accessible estimation with the ltmle package." Journal of Causal Inference 2.1 (2014): 95-108.

DISCLOSURES: All authors declare that they have no competing interests. All authors are employees in IQVIA.