Balancing Baseline Characteristics with SMOTE and Propensity Score Weighting in a Comparative Study of PASCAL Transcatheter Valve Repair and Medical Therapy for Degenerative Mitral Regurgitation

Lisa S Kemp, PhD[1]; Sarah Mollenkopf, MPH[1]; Rebecca Horn, PhD[1]; Michael Ryan, MS[2]; William Irish, PhD[3]

1 Edwards Lifesciences, Irvine, CA, 2 MPR Consulting, Cincinnati, OH, 3 East Carolina University, Greenville, NC

MSR50

OBJECTIVE

- Managing imbalanced datasets is a critical challenge in non-randomized cohort studies
- We applied the Synthetic Minority Over-sampling Technique (SMOTE) to address significant imbalance between patients with degenerative mitral regurgitation (DMR) derived from two datasets.

METHODS

- The treated group received the PASCAL transcatheter valve repair system (Edwards Lifesciences, Irvine, CA) in the CLASP IID randomized trial (NCT03706833)
- A real-world, medically managed cohort of DMR patients was derived from Optum® Market Clarity [Electronic health records (EHR) and Claims] 2016-2023Q3

INCLUSION CRITERIA

CLASP IID Treated Group

- No changes were made from the trial inclusion criteria. Patients had to have at least moderate-severe DMR and be at prohibitive risk for open mitral valve (MV) surgery

Medically-Managed DMR patients

- Patients were selected using a proxy definition of the CLASP IID inclusion criteria. Frailty score was used to define MV surgical risk.

EXCLUSION CRITERIA

CLASP IID Treated Group

- Patients treated at a hospital site outside the United States were excluded for the current baseline comparison

Medically-Managed DMR patients

- 14 exclusion criteria from the trial were mapped to the real-world database. Patients were also excluded with a record of systolic heart failure or ischemia and if their frailty score was low (low MV surgical risk)

Table 1. Assessment of balance before and after propensity score weighting

Covariate	SMD Before Weighting	SMD After Weighting
Age at index date (years)	0.8623	-0.2454
Sex (Male vs Female)	0.5705	0.0445
Race (Non-White vs White)	0.0006	0.0292
Frailty (High vs Low-Moderate)	-0.8010	-0.1985
History cardiac-related comorbidities (Yes vs No):		
Cardiomyopathy	0.2025	-0.0487
Coronary artery disease	-0.3339	0.2981
Pulmonary hypertension	0.5644	-0.1636
History of cardiac valve disease (Yes vs No):		
Aortic valve disease	-0.2091	0.2003
Tricuspid valve disease	-0.0394	-0.3082
History of non-cardiac related comorbidities:		
Chronic obstructive pulmonary disease	-0.2602	0.3343
History of previous cardiac procedures (Yes vs No)		
Percutaneous coronary intervention	0.7083	0.1764
CABG	0.4600	0.1776
History or Current Smoker	0.1788	0.4641
Number of heart failure hospitalization in the 12 months prior to index date	0.4460	-0.0392

STEPS OF SMOTE

1. Select cohort groups using inclusion and exclusion criteria
2. Oversample the cohort with the smaller sample size by generating synthetic samples
3. Generate propensity scores using a logistic regression
4. Calculate inverse probability of treatment weights to adjust for confounding
5. Compare standardize mean difference change before and after balancing to look for residual confounding based on a pre-established threshold

LIMITATIONS

- The Optum® Market Clarity database represents patients with commercial health insurance coverage and some third-party data which may include Medicare Fee-for-Service (FFS). Balancing may look different in a population with more FFS patients.
- This comparator group includes data that is either billed to patient insurance or available in the patient EHR during the study period.

CONCLUSIONS

- SMOTE, followed by PS-based weighting, effectively balanced the baseline characteristics between the PASCAL and comparator groups
- Residual confounding remains a concern, but this technique introduces another option for cohort balancing in non-randomized studies
- Future research should consider SMOTE when dealing with imbalanced datasets to improve the robustness of treatment comparisons in real-world data