Fairness of Machine Learning Readmission Predictions Following Open Ventral Hernia Repair

Tyler Zander[1,2], Melissa A. Kendall[1,2], Emily A. Grimsley[1,2], Rachel L. Wolansky[1,2], Rajavi Parikh[1,2], Joseph Sujka[1], Paul C. Kuo[1,3]

1. Department of Surgery, University of South Florida, Tampa, FL; 2. Department of Surgery, Moffitt Cancer Center, Tampa, FL; 3. Bay Pines Veterans Affairs Health Care System, Bay Pines, FL

INTRODUCTION

- Few models have predicted readmission following open ventral hernia repair (open VHR)
- None have evaluated fairness.
- Fairness evaluation compares model performance between demographic groups
- Interpretable machine learning can identify if influential predictors are similar between demographic groups

METHODS

- 2018-2021 NSQIP
  - Adults 18-89 years old
  - Open VHR (initial, recurrent, reducible, incarcerated)
  - Excluded length of stay > 10 days
- Univariable Analysis
  - Readmitted vs. non-readmitted
- XGBoost
  - Predict readmissions within 30 days of surgery
  - Fairness assessed by gender, ethnicity, and race
  - Interpretation with SHapley Additive exPlanations (SHAP)
  - SHAP quantifies the impact that predictors have on individual predicted log odds of readmission

RESULTS

Fairness Metrics

	Statistical Parity Difference	Equal Opportunity Difference	Average Odds Difference	Disparate Impact
Sex	-0.031	-0.093	-0.033	1.212
Ethnicity	0.034	-0.028	-0.032	0.812
Race	0.005	-0.004	-0.005	0.971

Statistical parity, equal opportunity, and average odds differences near 0 indicate similar performance. Positive values favor non-protected groups (male, non-Hispanic, White), and negative values favor protected groups (women, Hispanic, non-White). Disparate impacts near 1 indicate similar performance. Disparate impacts below 1 favor non-protected groups, and above 1 favors protected groups.

Top Mean Absolute SHAP Values

	Overall	Male	Female	Hispanic	non-Hispanic	White	non-White
OR to DC, days	0.40	0.41	0.39	0.41	0.39	0.41	0.39
Morbidity probability	0.28	0.27	0.29	0.31	0.28	0.30	0.28
Operative time	0.26	0.27	0.25	0.25	0.25	0.24	0.25
Mortality probability	0.15	0.14	0.16	0.19	0.14	0.18	0.14
Age, years	0.08	0.07	0.08	0.07	0.08	0.08	0.08
Preop BUN	0.07	0.07	0.07	0.06	0.07	0.07	0.07
Admission status	0.06	0.07	0.06	0.07	0.06	0.06	0.07
Work RVU	0.06	0.06	0.07	0.07	0.06	0.07	0.06
Preop HCT	0.06	0.06	0.06	0.06	0.06	0.06	0.06
Weight, pounds	0.06	0.06	0.06	0.05	0.06	0.06	0.06
Preop WBC count	0.06	0.06	0.06	0.06	0.06	0.06	0.06
Preop sodium	0.06	0.06	0.06	0.06	0.06	0.06	0.06

OR – operating room. DC – discharge. BUN – blood urea nitrogen. RVU – relative value unit. HCT – hematocrit. WBC – white blood cell

RESULTS

- 59,482 patients - 5.5% readmission rate.
- Readmits had more comorbidities ($p<0.01$)
- Readmits had more complications ($p<0.01$)
- Model performance: accuracy = 0.82, area under the curve = 0.72, specificity = 0.85, recall = 0.38, precision = 0.13, F1 score = 0.19, Brier score = 0.16
- Fairness metrics reveal marginal performance differences across demographic groups, with sex having the largest performance disparity

CONCLUSIONS

- Interpretable machine learning predicts fair readmission risk after open VHR
- SHAP identified similar influential features across demographics
- Early identification of high-risk patients could inform discharge and follow-up decisions
- Surgical machine learning models should assess fairness through fairness metrics and interpretation of predictions

Research reported in this presentation was supported by the Ruth L. Kirschstein Institutional National Research Service Award of the National Institutes of Health under award number **T32GM144274**.