Evaluation of Machine Learning, Statistical and the NSQIP Surgical Risk Calculators for Predicting Length of Stay in Patients Undergoing Major Oral Cavity Cancer Surgery

Hedyeh Ziai[1], Alex Gabinet[1], Konrado Deutsch[1], Frank Rudzicz[2], Wei Xu[1], Shuja Khalid[2], David Goldstein[1], Ralph Gilbert[1], Jonathan Irish[1], Danny Enepekides[1], Kevin Higgins[1], Antoine Eskander[1], John de Almeida[1]

[1]University of Toronto [2]Vector Institute

Background

- Head and Neck patients often have complex perioperative pathways with prolonged hospital stays
  - **Length of stay** (LOS) is related to costs and patient quality of life
- Surgical risk calculators can predict outcomes after major surgery
  - Existing tools (i.e. **ACS-NSQIP**) have poor performance in H&N patients
  - Machine Learning (ML) models have excellent predictive ability for outcomes in non-H&N patients

Objectives

- To develop, validate, and compare predictive models for **LOS** using both **statistical** and **machine learning (ML)** models for patients undergoing major oral cavity cancer surgery
- To evaluate model performance in comparison to the ACS-NSQIP surgical risk calculator

Materials and Methods

- Retrospective, multicenter study
  - University Health Network and Sunnybrook Health Sciences Centre
  - January 1, 2007 to April 30, 2018
- Inclusion criteria:
  - Age ≥18
  - Patients surgically treated for oral cavity cancer (involving neck dissection and/or regional/free flap reconstruction)
- 239 Predictor variables collected:
  - Preoperative, Intraoperative and Postoperative.
- Primary outcome measure: **length of stay**
- Data was pooled and split into training and validation datasets
  - with 10-fold cross validation
- Prediction models were developed

Methods (Models and analysis)

- **Statistical models:** MVA and LASSO models
  - p-value calculated for each variable
- **Machine Learning model:** random forest regression
- Predicted LOS from ACS-NSQIP calculator
- R^2 values were calculated for the developed models (ML/stats) and for the ACS-NSQIP calculator
- Performance was evaluated by comparing **predicted** vs. **actual** LOS and percent accuracy within 4 days of actual LOS

Results

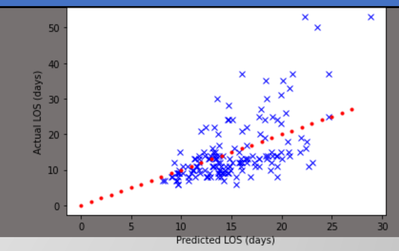
- **1097 patients**
  - Mean age: 62.5 years, 64% male
  - Oral tongue (44%), FOM (23%), Buccal/RMT (15%)
  - SCC (95%)- T3/4 (52%), N1-3 (45%)
  - Current/previous smokers (64%), comorbidities (67%)
  - Flap (81%)- soft tissue (71%), bone (29%)
- **Mean LOS** was **13.2** [SD 8.6] days
  - median was 11 days (range 1-57 days)

Model Performance

Model	Training		Validation						
					Accuracy Within				
	N	R^2	N	R^2	0.5 days	1 day	2 days	3 days	4 days
MVA (stats)	725	0.60	357	0.40	8%	16%	34%	46%	57%
LASSO (stats)	725	0.55	357	0.46	9%	18%	34%	50%	62%
ML	725	0.55	357	**0.47**	13%	31%	46%	57%	**69%**
NSQIP				-0.31	6%	10%	24%	36%	46%

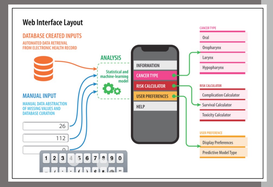
Results

Statistical (LASSO) model
78 variables selected by the model
ML model
all predictive variables were used to develop the model (Figure)



Discussion

- Performance for developed models and ML were fair
  - All developed models were superior to ACS-NSQIP's performance
  - Strongest model was machine learning model?
  - *Model accuracies* in H&N literature: 60-75%
- Composition of the models varied
- Aid in counselling patients, quality improvement initiatives and identifying high risk patients

Limitations:
- Small cohort for ML models
- Retrospective review
- External validation pending

Next steps:
- External validation
- Other ML models
- **Web-based decision support tool**
- Models for other **perioperative outcomes**