A machine learning approach to predict the risk of fall for elderly patients using physiological attributes from the Market Clarity Database

Verma V, Dawar V, Brooks L, Ashra P, **Gaur A**, Kukreja I, Rastogi M, Sanyal S, Gupta A, Kumar K, Chawla S, Nayyar A

RWD151

Optum Life Sciences

Introduction

- The elderly population is most susceptible to fall risk and related injuries, which can lead to medical complications and financial burden.

  1 out of 4 older adults fall each year

  | Unintentional injuries due to fall | **2ⁿᵈ** | Leading cause of death worldwide in 2021 |

  39,443 (2019) → Fall-Related Deaths → 44,686 (2021)

- Predicting the likelihood of falls can prevent incidents and accurately identify the correlated risk factors.

Objective

- To design and develop a prediction model for assessing the risk of falls among the elderly population in the US, using extensively Optum® claims and EHR data.

- To identify and utilize the vast array of all clinical symptoms from EHR data that can potentially lead to falls, along with claims data for building a more robust model.

- To calibrate the variation in results with only claims data and integrated claims with EHR data.
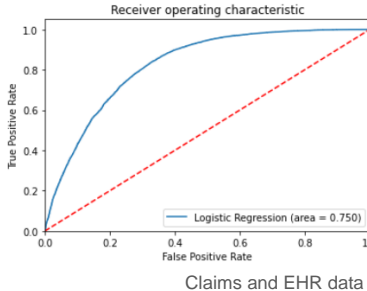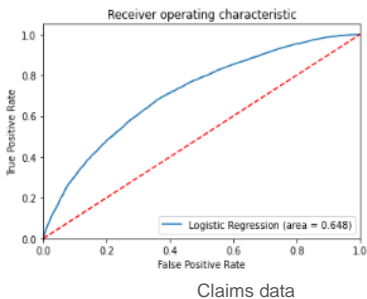
Method

- A retrospective study was conducted using Optum® de-identified Market Clarity database, where the intrinsic factors for predicting fall risk were accrued from both claims and EHR data.

- To reduce possible bias, a propensity score matching (PSM) technique was applied with age and gender as co-variates.

- For event patients we considered the condition of at least two Signs Disease and Symptoms (SDS) terms, while non-event patients were considered randomly from the non-event cohort.

- A total of 29,506 members aged 60 years and older were considered for the index period 2019-2021, with 12,912 incident subjects identified in the event cohort.

- An initial model was created using comorbidities and demographic variables (using ICD10 codes) from the claims database. This model was further developed by matching clinical symptoms (SDS terms) from the EHR database.

- Supervised ML techniques were used to develop the algorithm for fall prediction, which included logistic regression, random forest, XGBoost, Gradient Boost and Decision tree models.

- Using the claims and EHR database, 39 predictors were used for testing and training purposes, considering 80:20 ratio for training the model.

- Null or duplicate values were eliminated from our analysis.

Results

- Using only claims data, the logistic model classifier depicted an accuracy level of 65.48%, while an accuracy level of 75.07% was achieved using claims and EHR data. These results were substantiated by the rest of the models used in our analysis.

Claims data — Receiver operating characteristic, Logistic Regression (area = 0.648)

Claims and EHR data — Receiver operating characteristic, Logistic Regression (area = 0.750)

- In the logistic regression, it was observed that musculo-skeletal disorders, vertigo, brain related disorders, edema and mental disorders have a significantly greater impact on the probability to fall, compared to other predictors. The overall analysis using claims and EHR provided better prediction in comparison to claims only data.

5 most significant co-morbidities	Claims only		Claims and EHR	
	Odds Ratio	P-value	Odds Ratio	P-value
Musculo-skeletal disorders	1.39	0.000	2.47	0.000
Vertigo	1.13	0.073	2.21	0.000
Brain related disorders	2.31	0.000	2.08	0.000
Edema	1.51	0.000	1.85	0.000
Mental disorders	1.21	0.002	1.64	0.000

Conclusions

- Various physiological factors can significantly increase the likelihood of fall in the elderly population.

- Using Linked EHR and claims data has provided a broader spectrum of information for a more in-depth analyses on the prediction of future falls.

Optum