Visit level machine learning imputation of uncoded self-harm in major mental illness and characterization of incidence of self-harm

Praveen Kumar[1,2]; Anastasiya Nestsiarovich, M.D., Ph.D.[1]; Stuart J. Nelson, M.D.[3]; Berit Kerner, M.D.[4]; Douglas J. Perkins, PhD[1]; Christophe G. Lambert, Ph.D.[1,2,5*]

[1]Center for Global Health, Department of Internal Medicine, University of New Mexico Health Sciences Center, Albuquerque, New Mexico, USA; [2]Department of Computer Science, University of New Mexico, Albuquerque, New Mexico, USA; [3]Biomedical Informatics Center. George Washington University, Washington, DC, USA; [4]Semel Institute for Neuroscience and Human Behavior, David Geffen School of Medicine, University of California, Los Angeles, California, USA; [5]Translational Informatics Division, Department of Internal Medicine, University of New Mexico Health Sciences Center, Albuquerque, New Mexico, USA

*Corresponding author's email: cglambert[at]unm.edu.

Abstract

Incomplete coding of self-harm in US administrative claims data poses a major challenge to perform time-to-event comparative effectiveness pharmacotherapy studies with self-harm as the outcome, as well as to estimate the prevalence of these events. We aimed to use machine learning (ML) based model to impute uncoded self-harm in administrative claims data of individuals with major mental illness (MMI), characterize self-harm incidence, and identify factors associated with coding bias. Our ML imputation results show that only a small fraction (1/19) of self-harm events were coded in claims data for individuals with MMI and ML models can effectively identify the uncoded events.

Introduction

Suicide is one of the ten leading causes of death in the United States[1,2] and suicide attempts/self-harm are common manifestations of MMI (bipolar disorder, schizophrenia, schizoaffective disorder, major depressive disorder)[3]. Prior studies have reported incomplete coding of suicidality/self-harm in US administrative claims data[4,5], which has posed a major obstacle in having sufficient power to estimate event prevalence and to perform time-to-event comparative effectiveness pharmacotherapy studies. It was shown that suicidal ideation was only noted in 25% of patient charts[6], with only 3% of suicidal ideation and 19% of suicide attempts coded[7]. We present our machine learning approach to imputing self-harm at the visit-level. We also report coded versus imputed incidence of self-harm and factors associated with self-harm coding discrepancies. To our knowledge, this is the first study describing coded versus imputed incidence of self-harm.

Methods

We used the *IBM Health Analytics MarketScan®* commercial claims and encounters database transformed to the OMOP common data model to analyze data of 10,120,030 commercially insured US individuals (age <= 65 years) with ≥2 diagnostic codes for MMI during the observation period 2003-2016. We combined consecutive inpatient, emergency room (ER), and outpatient visits with no gap of >1 day into "meta-visits" to get the complete information related to one clinical event. All meta-visits consisting of only outpatient visit(s) were excluded because outpatient visits rarely had a self-harm billing code and a total of 20,783,244 meta-visits were selected for analysis, corresponding to 6,037,479 unique patients (31.9% males and 68.1% females).

To label a meta-visit as self-harm (class "1"), these codes were used: ICD-10CM codes X7[1-9]*, X8[0-3]*, ICD-9CM codes E95[0-9]*, SNOMED codes 4244894, 439235, 4303690, and their descendants. Using all 20M meta-visits, a total of 185,234 unique covariates were identified, including patient age, gender, meta-visit start year, and nine feature classes: *Manually Curated, Procedure, Condition, Drug, Billing Code Position, Device, Observation, Measurement, and Ancestor terms*. On average, each meta-visit had 115 features.

The XGboost classification models were developed using different subsets of 20M meta-visits (full dataset). Four additional ML algorithms were also tested on the balanced dataset comprising all 83,113 class "1" meta-visits and randomly selected 83,113 class "0" meta-visits. Accuracy (ACC), Matthews correlation coefficient (MCC), and Area under curve - receiver operating characteristic (AUC-ROC) were reported using 5-fold cross-validation. Using the balanced dataset, for all possible combinations of 9 classes of covariates, we performed 5-fold cross-validation using the XGboost model ("Full-factorial-models") and computed accuracy, MCC, and AUC-ROC for each combination.

To assess the effectiveness of the ML model, the ML assigned labels of randomly selected 200 meta-visits (50 with coded and imputed self-harm, 50 with coded but not imputed self-harm, 50 with imputed but not coded self-harm, and 50 with neither coded nor imputed self-harm) were compared with a experts-driven "Gold standard". Half of the class "1" meta-visits were randomly mislabeled in full and balanced dataset as class "0" and the XGboost models ("Mislabeled-full-data-model" and "Mislabeled-data-model" respectively) were built using the mislabeled data, reporting classification performance using the original labels. The XGboost model (“Mislabeled-data-model”) was also tested by randomly mislabeling half of the class “0”as class “1” in the balanced dataset.

To verify that our classification models did not overfit due to within-individual information leakage, we ran the XGboost model (“Per-person-model”) with 5-fold cross-validation on the dataset comprising one randomly selected meta-visit per person (6,037,479 meta-visits).

To understand the self-harm coding and uncoding pattern, the incidence of coded and imputed self-harm was computed as a function of patient age, sex, meta-visit start year, state of residence, and MMI type.

Results

Out of 20,783,244 meta-visits recorded over 29,799,203 years of patient observation, the probabilities of class “1” summed to 1,592,703 (7.66%), corresponding to an overall imputed annual incidence of 5.34%. The annual coded incidence was 0.28%. Out of 83,113 meta-visits coded for self-harm, 79,882 (96.11%) had class “1” probability >0.5. When the full dataset was used to build the XGboost model with 5-fold cross-validation and 10 repetitions (“full-data-model”), the accuracy was 0.960±$4x10^{-3}$, MCC 0.297±$2x10^{-4}$, AUC-ROC 0.990±$4x10^{-4}$. When the balanced dataset was used to build the XGboost model with 5-fold cross-validation and 100 repetitions (“balanced-data-model”), the accuracy was 0.964±$2x10^{-3}$, MCC 0.928±$4x10^{-4}$, AUC-ROC 0.991±$4x10^{-4}$. For “Full-factorial-models”, the AUCs were 0.988 for conditions, hand-curated: 0.977, code position: 0.875, observations: 0.813, procedures: 0.801, measurements: 0.594, drugs: 0.586, devices: 0.514. Adding ancestor terms had a negligible effect with the exception of procedures (using the ICD-10-PCS ancestors) which improved the AUC to 0.828. The classification results for the “Mislabeled-full-data-model” with 5-fold cross-validation were ACC: 0.974, MCC:0.347, AUC-ROC:0.991. When half of the class "1" meta-visits were mislabeled in the balanced dataset, the classification results for the “Mislabeled-data-model” with 5-fold cross-validation were ACC: 0.962, MCC:0.924, AUC-ROC:0.989. When half of the class "0" meta-visits were mislabeled in the balanced dataset, the classification results for the “Mislabeled-data-model” with 5-fold cross-validation were ACC: 0.963, MCC:0.926, AUC-ROC:0.991. The classification results for the “Per-person-model” were ACC:0.966, MCC:0.334, AUC:0.991.

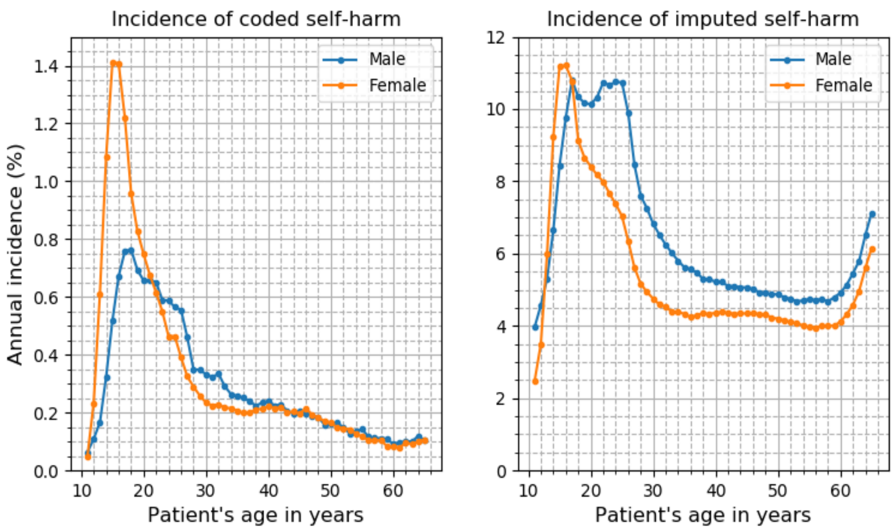

Figure 1: Incidence of self-harm for patients with MMI of different age and sex. The left graph shows the annual percentage incidence of coded self-harm in males (blue line) and females (orange line). The right graph shows the annual percentage incidence of machine learning-imputed self-harm in males (blue line) and females (orange line).

The agreement between ML and the "Gold standard" for different subsets of 200 meta-visits varied from 54% to 100% whereas inter-expert agreement varied from 50% to 98%. The overall agreement between ML and “Gold standard” was high (84%).

The fraction of coded self-harm was higher in young individuals versus older ones, and in females versus males. For all age groups, patients with multiple comorbid MMI had two-fold higher self-harm incidence. Both coded and imputed self-harm increased continuously from 2006 onwards. All but two US states (Montana and Utah) coded less than 10% of the imputed self-harm. Figure 1 shows the large difference in incidence between coded and imputed self-harm by age and sex.

Conclusion

ML methods could effectively infer uncoded self-harm events which were vastly underreported (~18/19) in US claims data of individuals with MMI. The underreporting of self-harm varied for different genders and ages, which suggests potential coding bias related to patient sex and age. Males were more likely to be uncoded for self-harm than females, with coding rates continuously decreasing with age. Both coded and imputed incidence of self-harm had considerably different patterns by age, gender, and among MMI categories.

References

1. WISQARS Leading Causes of Death Reports. https://webappa.cdc.gov/sasweb/ncipc/leadcaus10_us.html (accessed 26 Mar 2019).
2. WISQARS Fatal Injury Reports. https://webappa.cdc.gov/sasweb/ncipc/mortrate10_us.html (accessed 26 Mar 2019).
3. Canner JK, Giuliano K, Selvarajah S, Hammond ER, Schneider EB. Emergency department visits for attempted suicide and self harm in the USA: 2006–2013. Epidemiology and psychiatric sciences. 2018 Feb;27(1):94-102.
4. Bethell J, Rhodes AE. Identifying deliberate self-harm in emergency department data. Health reports. 2009 Jun 1;20(2):35.
5. LeMier M, Cummings P, West TA. Accuracy of external cause of injury codes reported in Washington State hospital discharge records. Injury Prevention. 2001 Dec 1;7(4):334-8.
6. Kemball RS, Gasgarth R, Johnson B, Patil M, Houry D. Unrecognized suicidal ideation in ED patients: are we missing an opportunity?. The American journal of emergency medicine. 2008 Jul 1;26(6):701-5.
7. Anderson HD, Pace WD, Brandt E, Nielsen RD, Allen RR, Libby AM, West DR, Valuck RJ. Monitoring suicidal patients in primary care using electronic health records. J Am Board Fam Med. 2015 Jan 1;28(1):65-71.