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, Ph.D.[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: *Inadequate 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 models 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. Self-harm undercoding was higher in males than females and increased with age. Except for years 2010 and 2013, the incidence of both coded and imputed self-harm continuously increased after 2006.*

Background

Suicide is one of the top 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, and major depressive disorder)[3]. Prior studies have shown underreporting and 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 for 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

- *IBM Health Analytics MarketScan®* commercial claims and encounters database.
- 10,120,030 (32.9% males and 67.1% females) commercially insured US individuals (age <= 65 years) with ≥2 diagnostic codes for MMI during the observation period 2003-2016.
- Data transformed to the OMOP common data model (CDM v5).
- **"meta-visits"**: combination of: (a) consecutive inpatient, (b) emergency room (ER), and (c) outpatient visits with no gap of >1 day.
- All meta-visits consisting of only outpatient visit(s) were excluded. Thus, only 20,783,244 (4.0%) out of 519,590,773 unique meta-visits were selected for analysis.
- Self-harm (Class "1") meta-visits had one of these codes: ICD-10CM codes X7{1-9}*, X8{0-3}*, ICD-9CM codes E95{0-9}*, SNOMED codes 4244894, 439235, 4303690, and their descendants. Class '0' meta-visits had none of these codes.
- A total of 185,234 unique covariates included 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*.

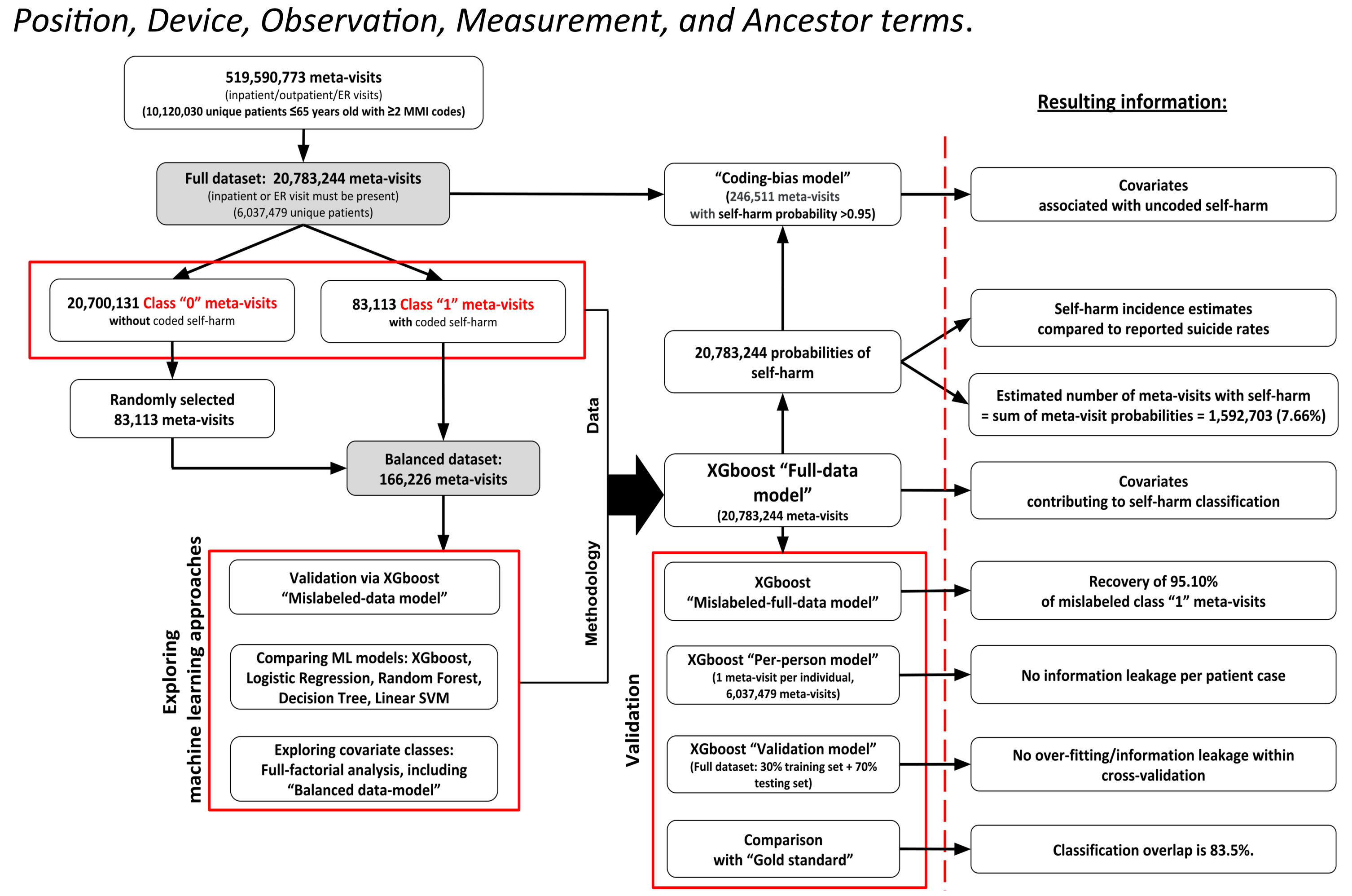

Figure 1. Overall study design. The "full dataset" and the "balanced dataset" were created for ML models using meta-visits. Five different ML algorithms were tested on the balanced dataset, whereas only the XGboost algorithm was chosen to build the "Full-data-model". Because our study was limited to a single dataset, several experimental approaches ("Mislabeled-data-model", "Mislabeled-full-data-model", "per-person-model", and "Validation-model") were used to validate ML models and to verify that ML models were not overfitted. The classes of covariates that contributed most to the model performance were determined by the "Full-factorial-models" and features associated with uncoded self-harm via building the "Coding-bias-model". A clinician-derived "Gold-standard" was also used to validate the predictions of the "Full-data-model".

Methods

- All ICD-9Proc and CPT4 codes were mapped to ICD-10PCS equivalents.
- All ICD-9CM and ICD-10CM diagnosis codes were mapped to SNOMED equivalents.
- Only the XGBoost model was built on the "full dataset" comprising all 20M meta-visits.
- XGBoost, Logistic Regression, Random Forest, Decision Tree, and LinearSVC were trained and tested on a smaller "Balanced dataset" comprising all 83,113 class "1" meta-visits and a randomly selected 83,113 class "0" metavisits.
- To validate ML models and to verify we were not overfitting, multiple approaches were used (Figure1)
  - **"Per-person model"**: to ensure no within-individual information leakage was occurring.
  - **"Validation model"**: to confirm classification performance was not due to overfitting.
  - **"Mislabeled-data-model"** and **"Mislabeled-full-data-model"**: to validate recovery of uncoded self-harm.
  - **"Coding-bias-model"**: to find which variables were associated with highly certain imputed self-harm cases not being coded.
  - **"Full-factorial-models"**: to determine the classes of covariates that contributed most to the classification performance of ML models.
  - **"Gold Standard"**: to validate the ML classifications using the expertise of 3 clinicians.
- Detailed comparisons of coded versus imputed self-harm incidence were made by patient age, sex, MMI category, and US state of residence.

Results

- Selected 20M meta-visits had 6,037,479 unique patients (31.9% males and 68.1% females).
- The XGboost "Full-data-model" probabilities of self-harm (Class "1") summed to 1,592,703 (7.66%).
- Overall imputed annual incidence was 5.34%, whereas the coded annual incidence was 0.28%.
- Out of 83,113 meta-visits coded for self-harm, 79,882 (96.11%) had class "1" probability >0.5 and 62,929 (75.71%) had Class "1" probability ≥0.95.
- Out of 185,234 covariates used to build the "Full-data-model", only 2,205 (1.19%) had relative gain >0.
- Poisoning, Mood disorder, Traumatic injury, External injury, and Mental disorder were the top five covariates with the highest gain scores.
- Adding "Ancestor" concepts had a negligible effect on the model performance, except for the "Procedures"concepts, where the AUC increased from 0.800 to 0.828 after adding the ICD-10 PCS ancestor terms.
- The agreement between ML and the "Gold standard" varied from 54% to 100%, whereas inter-expert agreement varied from 50% to 98%.
- The fraction of coded self-harm was higher in younger individuals versus older ones, and in females versus males.
- Patients with multiple comorbid MMI diagnoses had two-fold higher self-harm incidence.
- All but two US states (Montana and Utah) coded less than 10% of the imputed self-harm.

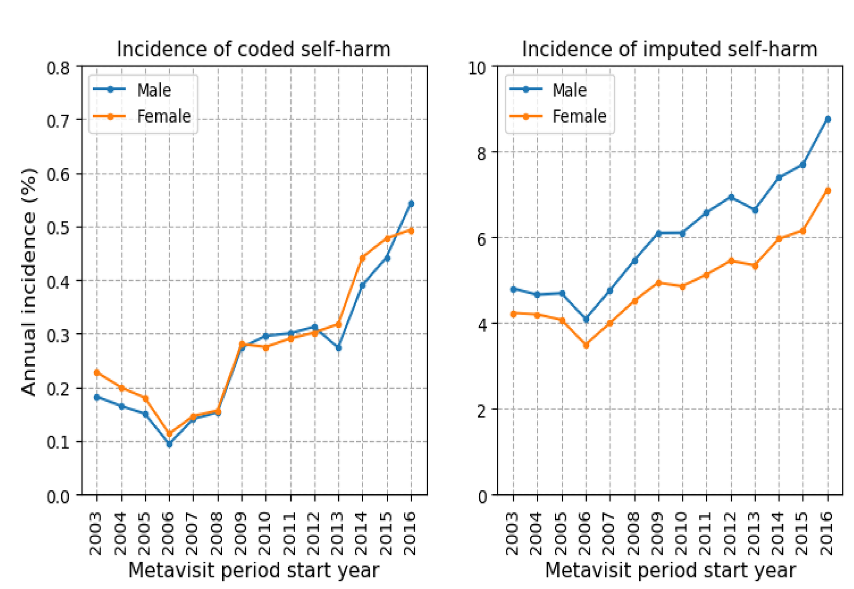

Figure 2. The annual coded vs. imputed incidence of meta-visits with self-harm in patients with MMI by sex and year. The left graph shows the annual percentage incidence of coded self-harm for males (blue line) and females (orange line); the right graph shows the annual percentage incidence of imputed self-harm for males (blue line) and females (orange line). *Self-harm rates have generally risen since 2006, with male self-harm undercoded more than female self-harm.*

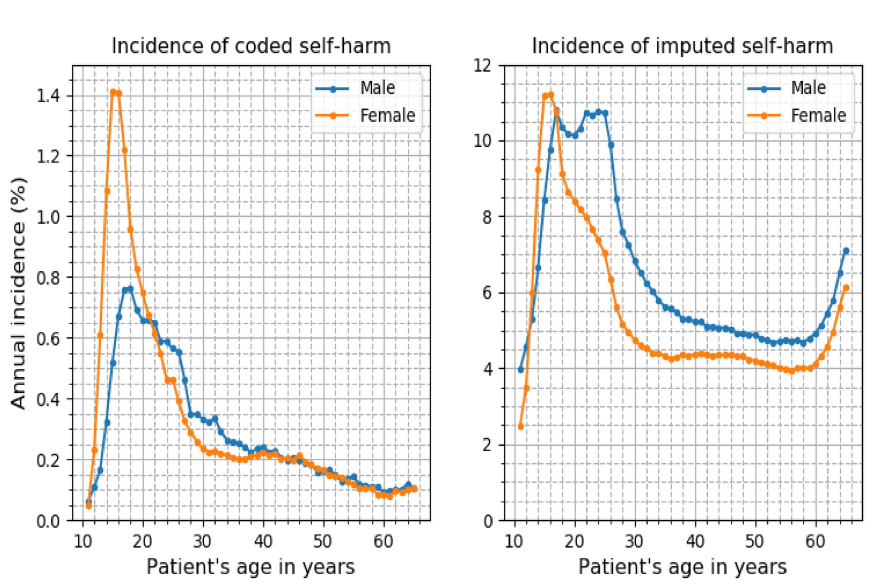

Figure 3. The annual coded vs. imputed incidence of meta-visits with self-harm in patients with MMI by age and sex. The left graph shows the annual percentage incidence of coded self-harm in males (blue line) and females (orange line). The right graph shows the annual percentage incidence of imputed self-harm in males (blue line) and females (orange line). *Risk peaks earlier for females than males (age 15 vs. 17), but males have a longer period of high risk. Male self-harm goes undercoded more than female self-harm.*

Data	200 randomly selected meta-visits				
Classifier	Full-data-model	Clinician 1	Clinician 2	Clinician 3	Gold Standard
Full-data-model	1.00	0.81	0.80	0.78	0.84
Clinician 1	0.81	1.00	0.77	0.88	0.88
Clinician 2	0.80	0.77	1.00	0.76	0.86
Clinician 3	0.79	0.88	0.76	1.00	0.87
Gold Standard	0.84	0.88	0.86	0.87	1.00

Table 3. The pairwise agreement between the XGboost "Full-data-model", "Gold standard", and three clinicians regarding the presence of self-harm (with >0.5 probability) in 200 selected meta-visits of patients with MMI. *Inter-rater agreement was comparable between machine learning, individual clinicians, and the "Gold standard".*

Results

XGboost model	Validation method	Dataset	Accuracy	MCC	AUC-ROC
Full-data-model	5-fold cross-validation repeated 10 times.	Full dataset with 20M meta-visits.	$0.960 \pm 4 \times 10^{-3}$	$0.297 \pm 2 \times 10^{-4}$	$0.990 \pm 4 \times 10^{-4}$
Per-person-model	5-fold cross-validation.	Full dataset subset of 6M meta-visits with one random meta-visit per person.	0.966	0.334	0.991
Validation-model	5-fold cross-validation on the training set.	70% random meta-visits from the full dataset.	0.964	0.298	0.991
	Testing on the validation set.	Remaining 30% of meta-visits from the full dataset.	0.963	0.296	0.990
Balanced-data-model	5-fold cross-validation repeated 100 times.	Balanced dataset with 166K meta-visits.	$0.964 \pm 2 \times 10^{-4}$	$0.928 \pm 4 \times 10^{-4}$	$0.991 \pm 4 \times 10^{-4}$
Mislabeled-data-model	5-fold cross-validation. Original labels of meta-visits were used for assessing performance.	Half of the class "1" meta-visits mislabeled in the balanced dataset.	0.962	0.924	0.989
		Half of the class "0" meta-visits mislabeled in the balanced dataset.	0.963	0.926	0.991
Mislabeled-full-data-model		Half of the class "1" meta-visits mislabeled in the full dataset.	0.974	0.347	0.991
Coding-bias-model	5-fold cross-validation.	All meta-visits from the full dataset with class '1.' Probability threshold ≥0.95.	0.679	0.306	0.738
Full-factorial-models	5-fold cross-validation.	Balanced dataset. Only "Condition" covariates.	0.957	0.914	0.988
		Balanced dataset. Only "Hand-curated" covariates.	0.927	0.853	0.977
		Balanced dataset. Only "Billing Code Position" covariates.	0.788	0.577	0.875
		Balanced dataset. Only "Observation" covariates.	0.775	0.562	0.813
		Balanced dataset. Only "Procedure" covariates.	0.708	0.440	0.800
		Balanced dataset. Only "Measurement" covariates.	0.589	0.245	0.594
		Balanced dataset. Only "Drug" covariates.	0.550	0.192	0.586
		Balanced dataset. Only "Device" covariates.	0.516	0.099	0.514

Table 1. Classification results for different XGboost-based classification models on different subsets of meta-visits of patients with MMI. The results for the "Full-data-model" and the "Balanced-data-model" are shown with 80% and 90% confidence intervals, respectively. MCC - Matthews correlation coefficient. AUC-ROC -Receiver Operating Characteristic Area Under the Curve.

Machine learning model/ performance	XGboost balanced-data-model with optimized parameters	XGboost balanced-data-model with default parameters	Logistic Regression	Random Forest	Decision Tree	LinearSVC
Accuracy	$0.964 \pm 2 \times 10^{-4}$	$0.961 \pm 2 \times 10^{-4}$	$0.963 \pm 3 \times 10^{-4}$	$0.946 \pm 1 \times 10^{-3}$	$0.947 \pm 7 \times 10^{-4}$	$0.959 \pm 3 \times 10^{-4}$
MCC	$0.928 \pm 4 \times 10^{-4}$	$0.922 \pm 4 \times 10^{-4}$	$0.926 \pm 6 \times 10^{-4}$	$0.892 \pm 3 \times 10^{-3}$	$0.896 \pm 1 \times 10^{-3}$	$0.919 \pm 7 \times 10^{-4}$
AUC-ROC	$0.991 \pm 4 \times 10^{-4}$	$0.990 \pm 2 \times 10^{-4}$	$0.990 \pm 1 \times 10^{-4}$	$0.982 \pm 6 \times 10^{-4}$	$0.948 \pm 7 \times 10^{-4}$	$0.988 \pm 1 \times 10^{-4}$

Table 2. Classification results for 5 different ML algorithms on the balanced dataset of meta-visits of patients with MMI, using five-fold-cross-validation with 100 repetitions, reported with 90% confidence intervals. MCC - Matthews correlation coefficient. AUC-ROC -Receiver Operating Characteristic Area Under the Curve. *XGBoost was selected for its high classification performance and fast runtime on large datasets.*

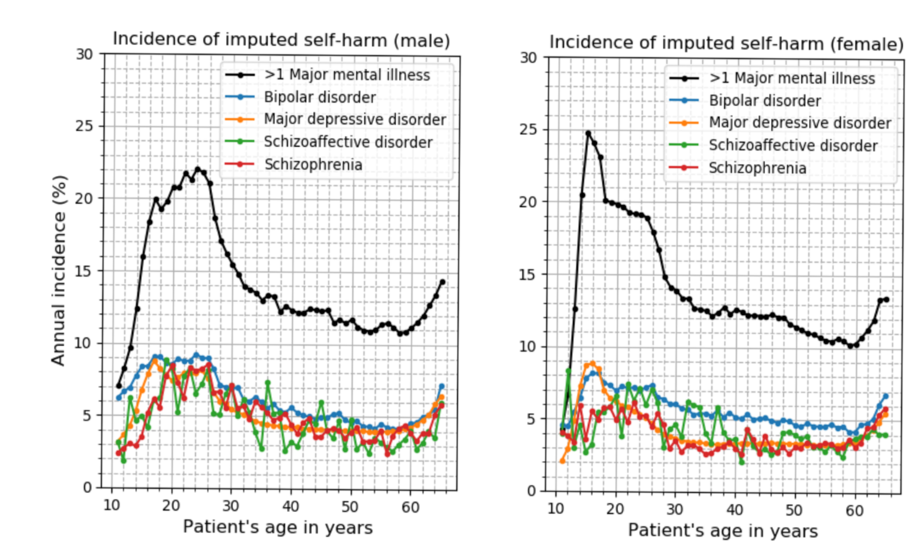

Figure 4. The annual incidence of meta-visits with imputed self-harm by sex and category of major mental illness. The left graph shows the data for males and the right graph for females. *Having 2 or more MMI diagnoses dramatically increases self-harm risk.*

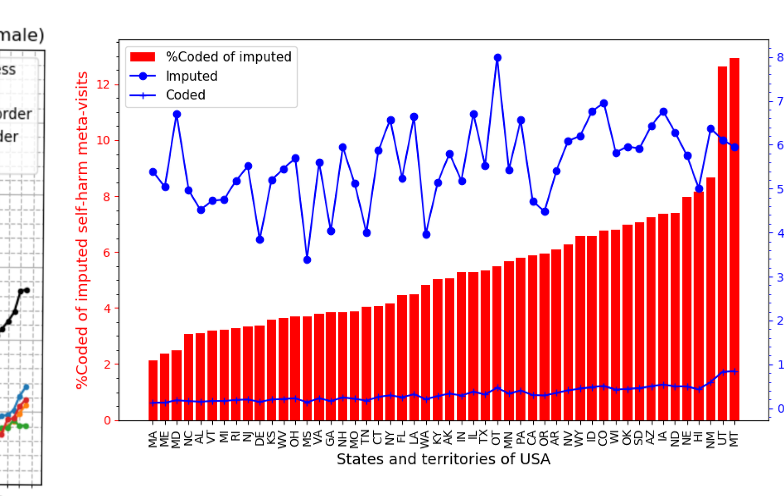

Figure 5. The annual incidence of meta-visits with self-harm and coding percentage in patients with MMI residing in different states and territories of the USA. The blue plots show the annual percentage incidence of imputed self-harm (blue dots) and coded self-harm (blue hatches). The red bars show the fraction of coded self-harm events among the imputed events. 'OT' - "others", which includes DC, Puerto Rico, and other US territories. *While coded rates vary widely across states, self-harm imputation suggests the actual rates are quite comparable.*

Conclusions

- Only about 1 in 19 self-harm events were reported in US claims data among patients with MMI.
- ML methods effectively infer unreported self-harm events.
- The underreporting of self-harm varied by sex and age, suggesting potential coding bias.
- Self-harm undercoding steadily increased with age.
- Self-harm incidence varied considerably 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.