Visit level suicidality/self-harm phenotyping in bipolar disorder

Praveen Kumar[1*], Anastasiya Nestsiarovich, MD, PhD[2], Aurélien J. Mazurie, PhD[3], Stuart J. Nelson[4,5], Nathaniel G. Hurwitz, MD[6], Christophe G. Lambert, PhD[2,5*]

[1]Department of Computer Science, University of New Mexico (UNM), Albuquerque, NM, USA. [2]Center for Global Health, Department of Internal Medicine, UNM Health Sciences Center, Albuquerque, NM, USA. [3]TwoFoldChange Consulting, Bozeman, MT, USA. [4]Health Sciences Library and Informatics Center, UNM Health Sciences Center, Albuquerque, NM, USA. [5]Division of Translational Informatics, Department of Internal Medicine, UNM Health Sciences Center, Albuquerque, NM, USA. [6]New Mexico Behavioral Health Institute, Las Vegas, NM, USA. ***Corresponding Authors' emails: {pkumar81, cglambert} @unm.edu**

Abstract: Patients suffering from bipolar disorder have an increased risk of attempting self-harm/suicide by numerous means. Although many suicidal patients are hospitalized for diagnosis and treatment, the vast majority of such visits are not documented with suicidality/self-harm diagnostic codes in administrative claims data, which makes studies related to this outcome difficult. In the Truven Health Analytics MarketScan® database, we observed in a cohort of 1.3M bipolar disorder patients that visits containing likely self-inflicted external injuries (suspicious injuries), are rarely accompanied by self-harm diagnostic codes. The fraction of visits with suspicious injuries in which suicidality/self-harm was also coded varies from 1.83-33.7%, depending on the injury code and state. Summary statistics are presented, along with preliminary machine learning approaches to imputing suicidality/self-harm at a visit level to support investigation of this phenotype in time-to-event studies.

Introduction

Bipolar Disorder (BD) is associated with excess mortality/morbidity from suicide and high rates of attempted suicide[1]. Despite the fact that most patients who attempt suicide are treated in a hospital or emergency department setting[2], most suicidal behavior and self-harm is not explicitly coded in administrative claims billing data. The frequency of using ICD9CM diagnostic codes to report suicidal ideation and suicidal attempts in patients with depression was shown to be only 3% and 19% respectively in primary care organizations[3]. Sensitivity of various approaches to detecting suicidality in administrative claims data ranges from 13.8% to 65%, with positive predictive value ranging from 4.0% to 100%[4]. In addition, most approaches to phenotype learning are calculated over a extended observation period, as is done with Aphrodite[5], rather than at a visit level, which is required for time-to-event and other observational study designs. The low frequency of labeled self-harm visits relative to reality makes it challenging to develop a suicidality classifier to label outcomes at a visit-level. Previous work on imputing phenotypes, such as Aphrodite do not have a mechanism for visit-level labeling. We present our initial efforts towards using machine learning to address this challenge, as well as addressing the asymmetry in mislabeling suicidality events.

Methods

- We used the *Truven Health Analytics MarketScan®* administrative claims database to analyze data on 1.3M inpatient and outpatient individuals with at least two diagnoses of bipolar or schizoaffective disorder during the observation period 2003-2015.
- To get a comprehensive picture of the events accompanying each visit, we constructed "meta-visits", defined as a consecutive sequence of visits. Meta-visits:
  - Might include an ER visit, an outpatient visit, and a subsequent psychiatric hospitalization.
  - Contain more information than a single visit, thus, we expect the percentage of correctly reported cases of self-harm to be higher.
- We identified a set of ICD9 and ICD10 procedure (*Table 1*) and diagnostic (*Table 2*) codes "suspicious" for suicide/self-harm based on self-harm methods existing in current international classifications of diseases. We aimed to find what percentage of visits/meta-visits with suspicious codes were accompanied with a diagnosis of suicide/self-harm (ICD9CM E95[0-9]*; ICD10CM X7[1-9]*, X8[0-3]*; SNOMED 59274003, 276853009, 418420002 and descendants).
- **Injury covariate:** a pool of diagnostic and procedure codes was created manually to identify codes consistent with external injury, excluding explicitly coded suicide/self-harm. Three MD raters independently scored 10,000+ codes for probable injury. Discrepancies were resolved by consensus

Classification

- We used the XGboost[6] machine learning approach to develop a classification model based on these data. Class '1' was assigned to the visits that were documented as self-harm/suicide and class '0' was assigned to the rest of the visits.
- In the Truven data, a small percentage of visits related to suicide or self-harm have been documented with valid concept codes, which means a large percentage of suicide/self-harm visits were assigned class '0' for the classification. Therefore, we used XGboost's *scale_pos_weight* parameter to control the balance of class '1' and class '0' weights.
- For data with imbalanced class, most of the standard machine learning algorithms tend to be biased towards the majority class. Thus, we used Matthews Correlation Coefficient (MCC) in addition to AUC ROC as a performance metric, the former being widely used for data with class imbalance[7].
- We identified 58 concept ids that were directly related to suicide/self-harm and excluded them as covariates from the model building process.
- Covariates characterizing visits included observations, conditions, procedures, drug ingredients, two manually curated sets of covariates for injuries, as well as BD clinical characteristics, comorbidities, and concomitant drug classes.
- For all possible combinations of covariates (full-factorial analysis), we performed 5-fold cross-validation with 20 repetitions and computed the ROC_AUC score and MCC in each run.
- To know the effect of ancestors on the performance of the classification model, we selected all ancestors of given covariates and added ancestor terms as additional covariates and repeated our experiment to compute ROC_AUC score and MCC.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

Procedure Concept Code	Description
2007958	Gastric lavage
2102818	Exploration of penetrating wound (separate procedure); neck
2107701	Repair blood vessel, direct; neck
2107718	Repair blood vessel with vein graft; neck
2107734	Repair blood vessel with graft other than vein; neck
2108382	Ligation, major artery (eg, post-traumatic, rupture); neck
45889617	Anesthesia for procedures on major vessels of neck
2102818	Exploration of penetrating wound (separate procedure); neck
40757044	Gastric intubation and aspiration(s) therapeutic, necessitating physician's skill (eg, for gastrointestinal hemorrhage), including lavage if performed
2104152	Repair, tendon or muscle, flexor, forearm and/or wrist; primary, single, each tendon or muscle
2104153	Repair, tendon or muscle, flexor, forearm and/or wrist; secondary, single, each tendon or muscle

Table 1: Examples of "suspicious" procedure codes and their descriptions, rated as being likely to involve an injury from an external cause, as opposed to a disease process.

Condition Concept Codes	Description
S61.5*	Open wound of wrist
S65.(00|01|09|10|11|19|8|9)*	Injury of blood vessels at wrist and hand level
S51.8[0,1,3,4]*	Open wound of elbow and forearm
S55*	Injury of blood vessels at forearm level
R09.01*	Asphyxia
T71*	Asphyxiation
W15*	Fall from cliff
Y21*-Y33*	Event of undetermined intent
T51*-T65*	Toxic effects of substances chiefly nonmedicinal as to source
960*-979*	Poisoning By Drugs, Medicinal And Biological Substances
E98[0-9]*	Injury Undetermined Whether Accidentally Or Purposely Inflicted
E850*-E858*	Accidental Poisoning By Drugs, Medicinal Substances, And Biologicals
E860*-E869*	Accidental Poisoning By Other Solid And Liquid Substances, Gases, And Vapors
799.01*	Asphyxia
994.7*	Asphyxiation and strangulation
980*-989*	Toxic Effects Of Substances Chiefly Nonmedicinal As To Source

Table 2: Examples of "suspicious" condition codes and their descriptions, rated as being likely to involve an injury from an external cause, as opposed to a disease process.

Results

- There was a 3.33-fold higher number of suicide/self-harm E-codes among visits with versus without "suspicious" diagnoses (4.9% versus 1.5%) (Table 3).
- The number of self-harm diagnoses among visits/meta-visits with suspicious codes varied depending on the type of injury and US state (1.83-33.7% for visits and 2.07-35.07% for meta-visits) (Figures 2, 3), which may be explained by existing regional differences in electronic health recording[8].
- In the "suspicious" subset of visits we observed that the majority of states reported suicide/self-harm in <6% of cases and only two states reported them in >10% of cases (Figure 1).
- Surprisingly, state mandates to code for injury E-codes (which includes suicide codes) do not appear to increase the fraction of probable suicides captured in administrative claims data (Figure 1).
- We found that observations and our hand-curated covariates contributed most to model classification performance, with procedures providing no benefits.
- With both hand-curated and observation covariates, we achieved high classifier performance with an area under the curve of 0.95 and MCC ranging from 0.26-0.29 across the 20 cross-validated runs.
- When the least significant covariates were used with ancestors, the performance of the classification model improved significantly, whereas addition of ancestors to the most significant covariates did not improve the performance.
- Using the "metavisit" concept allowed capture of slightly more suicide/self-harm E-codes. The average increase in number of detected E-codes was 1.5 % for diagnoses (maximum 4.37%) (Figure 2) and 0.5% for procedures (maximum 2.13%) (Figure 3).

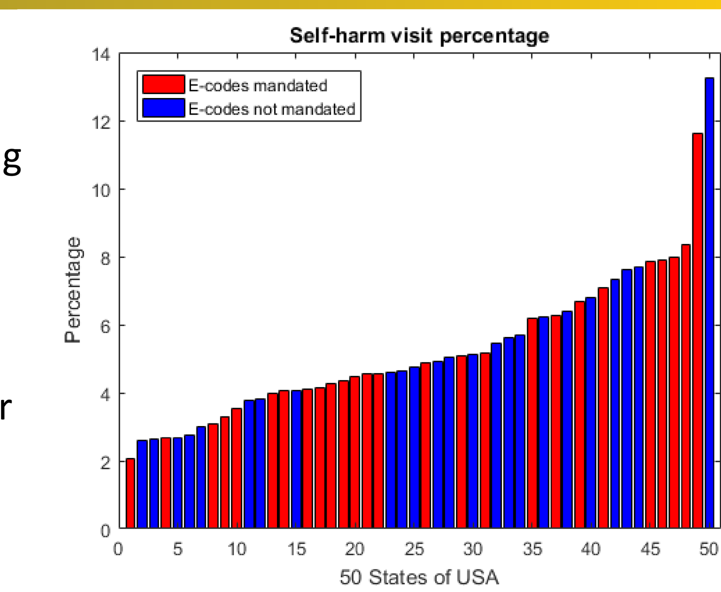

Figure 1: Fraction of visits with suspicious injuries coded as suicidal by US State. There is little evidence that state mandates to code injures increase the capture of suicide/self-harm in administrative claims data.

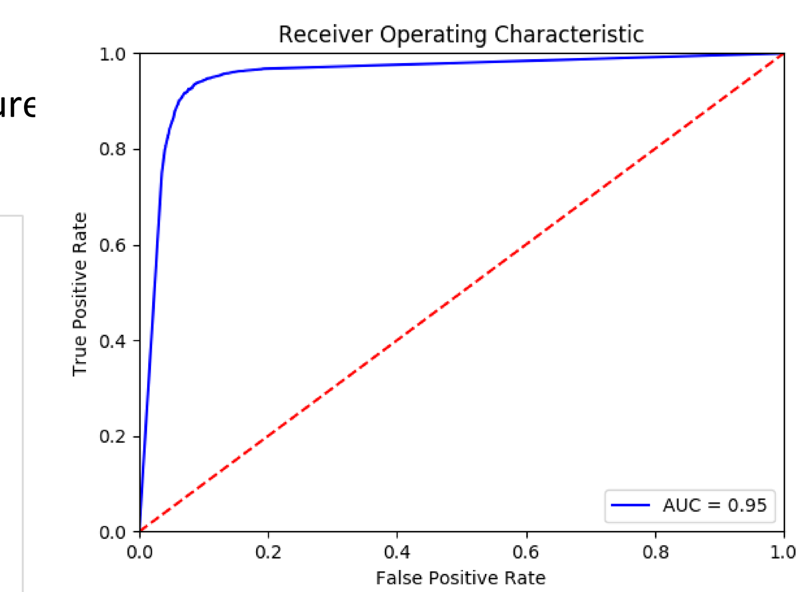

Figure 4: ROC_AUC when all covariates as well as their ancestors were selected.

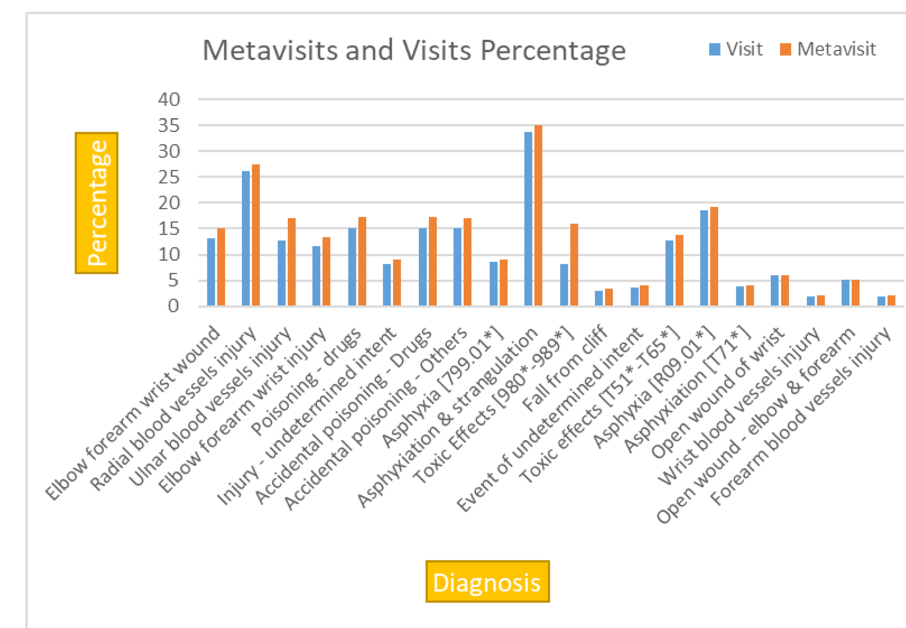

Figure 2: Percentage of visits and metavisits with suspicious diagnosis codes that were documented as self-harm/suicide visits.

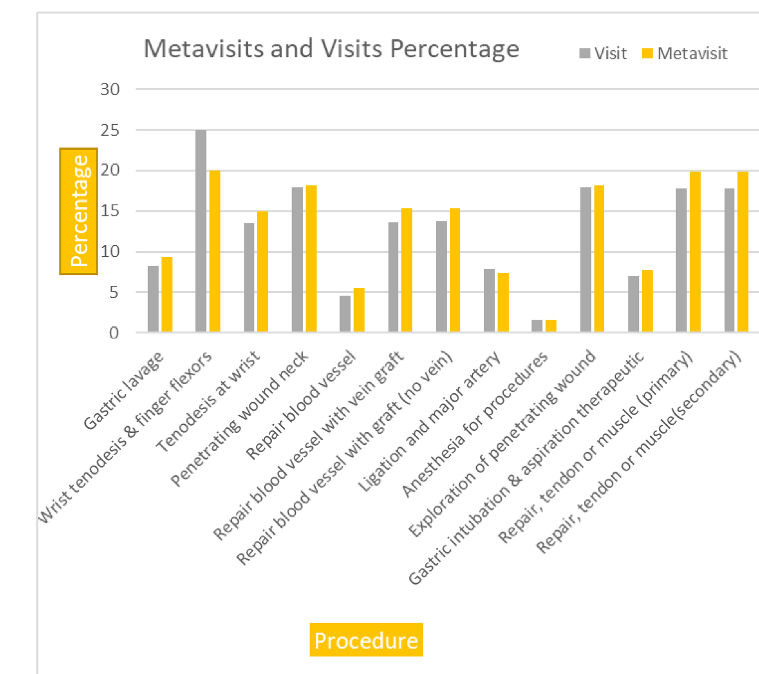

Figure 3: Percentage of visits and metavisits with suspicious procedure codes that were documented as self-harm/suicide visits.

	With "Suspicious" Diagnosis Codes	Without "Suspicious" Diagnosis Codes
Coded suicide/self-harm	14,451 (0.92%)	18,658 (1.2%)
Not coded suicide/self-harm	281,752 (18.0%)	1,249,648 (79.9%)

Table 3: Counts of unique visits with/without "suspicious" diagnosis codes that are/aren't coded as suicide/self-harm. 4.9% of visits with "suspicious" codes also were coded with suicide/self-harm vs. 1.5% of visits without "suspicious" codes.

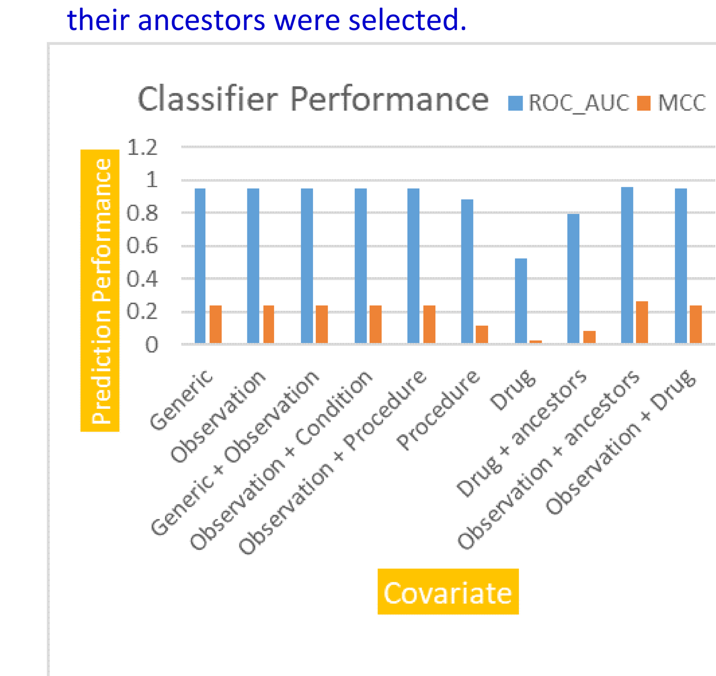

Figure 5: ROC_AUC and MCC values for some of the covariate combinataions.

Conclusions

The results of our study show that most bipolar patients hospitalized for attempted suicide/self-harm do not have associated billing codes for such, with significant regional biases in data collection, which could confound observational studies. Initial machine learning results suggest the lack of a harmonized vocabulary for procedures that harmonizes CPT4, HCPCS, ICD9Proc and ICD10PCS codes prevents the utilization of important procedure information in predicting suicidality. Expert-curated collections of codes improve model performance. The notion of meta-visits, which coincides with Vocabulary Working Group efforts to create a visit_era table, appear to increase the detection of suicidality. Further work remains to develop unbiased classifiers for visit-level suicidality.

Acknowledgments

Research reported in this publication was funded through a Patient-Centered Outcomes Research Institute (PCORI) award (CER-1507-3160). The views in this publication are solely the responsibility of the authors and do not necessarily represent the views of PCORI, its Board of Governors or Methodology Committee.

References

1. Oquendo MA, Carballo JJ, Rajouria N, Currier D, Tin A, Merville J, et al. Are high-lethality suicide attempters with bipolar disorder a distinct phenotype? Arch Suicide Res. 2009;13(3):247–56.
2. Gairin I, House A, Owens D. Attendance at the accident and emergency department in the year before suicide: retrospective study. Br J Psychiatry. 2003 Jul;183:28–33.
3. Anderson HD, Pace WD, Brandt E, Nielsen RD, Allen RR, Libby AM, et al. Monitoring suicidal patients in primary care using electronic health records. J Am Board Fam Med. 2015 Jan;28(1):65–71.
4. Walkup JT, Townsend L, Crystal S, Olfson M. A systematic review of validated methods for identifying suicide or suicidal ideation using administrative or claims data Pharmacoepidemiol Drug Saf. 2012 Jan;21 Suppl 1:174-82.
5. Vibhu A, Tanya P, Juan MB, Veena G, Tiffany IL, Evan PM, Timothy ES, Elsie G, Nigam HS Learning statistical models of phenotypes using noisy labeled training data J Am Med Inform Assoc 2016;23:1166–1173.
6. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: ACM; 2016. p. 785–94. (KDD '16).
7. Sabri Boughorbel, Fethi Jarray, Mohammed El-Anbari. Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. PLoS ONE 12(6): e0177678
8. Abellera J, Annest JL, Conn JM, Kohn M. How states are collecting and using cause of injury data: 2004 update to the 1997 report. Council of State and Territorial Epidemiologists, Data Committee Injury Control and Emergency Health Services Section, American Public Health Association, and State and Territorial Injury Prevention Directors Association. 2005;