Methods

Classification approaches to detecting suicidality in administrative claims data ranges from 13.8% to 65%, with positive predictive value ranging from 4.0% to 100% [1]. In addition, many suicidal patients are hospitalized for diagnosis and treatment, the vast majority of such visits are not documented with suicidality/self-harm. 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 [2]. Despite the fact that most patients who attempt suicide are treated in a hospital or emergency department setting, suicide is not explicitly coded in administrative claims billing data. The frequency of using ICD9CM diagnostic codes to report suicidal ideation and suicide 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]. Thus, various approaches to phenotype learning are calculated over a extended observation period, as shown with Aphrodite, rather than at a visit level, which is required for time-to-event and other observational study designs. The 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 extended our initial efforts towards using machine learning to address this challenge, as well as addressing the asymmetry in mislabeling suicidality events.

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 "suspicous" 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 suspcious codes were documented with a diagnosis of suicide/self-harm (ICD9CM 690.0*; ICD10CM Y71.8*-Y83.3*; 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. These codes independently scored 10,000+ codes for probable injury. Discrepancies were resolved by consensus.

Classification: We used the XGboost 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 such visits were assigned class ‘0’ for the classification. Therefore, we used XGboost’s scale_pos_weight parameter to control the class 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 [5]. We identified 78 concept ids that were directly related to suicide/self-harm and excluded them from the vocabulary for procedures that harmonizes CPT4, HCPCS, ICD9Proc and ICD10PCS codes prevents the utilization of important procedure information in predicting suicidability. Expert-curated classifications 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.

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 suicidability. Expert-curated classifications 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-1507-3168]. 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.