



Published in final edited form as:

J Biomed Inform. 2017 November ; 75 Suppl: S149–S159. doi:10.1016/j.jbi.2017.08.009.

Exploring Associations of Clinical and Social Parameters with Violent Behaviors among Psychiatric Patients

Hong-Jie Dai, PhD^{1,2,*}, Emily Chia-Yu Su, PhD³, Mohy Uddin, PhD⁴, Jitendra Jonnagaddala^{5,6,*}, Chi-Shin Wu, MD⁷, and Shabbir Syed-Abdul, MD, PhD^{3,*}

¹Department of Computer Science and Information Engineering, National Taitung University, Taitung, Taiwan

²Interdisciplinary Program of Green and Information Technology, National Taitung University, Taitung, Taiwan

³Graduate Institute of Biomedical Informatics, Taipei Medical University, Taipei, Taiwan

⁴King Abdullah International Medical Research Center, King Saud bin Abdulaziz University for Health Sciences, Publication Office, King Abdulaziz Medical City, Ministry of National Guard Health Affairs, Riyadh, Kingdom of Saudi Arabia

⁵School of Public Health and Community Medicine, UNSW Sydney, Australia

⁶Prince of Wales Clinical School, UNSW Sydney, Australia

⁷Department of Psychiatry, National Taiwan University Hospital and College of Medicine, National Taiwan University, Taipei, Taiwan

Abstract

Evidence has revealed interesting associations of clinical and social parameters with violent behaviors of patients with psychiatric disorders. Men are more violent preceding and during hospitalization, whereas women are more violent than men throughout the 3 days following a hospital admission. It has also been proven that mental disorders may be a consistent risk factor for the occurrence of violence. In order to better understand violent behaviors of patients with psychiatric disorders, it is important to investigate both the clinical symptoms and psychosocial factors that accompany violence in these patients. In this study, we utilized a dataset released by the Partners Healthcare and Neuropsychiatric Genome-scale and RDoC Individualized Domains project of Harvard Medical School to develop a unique text mining pipeline that processes unstructured clinical data in order to recognize clinical and social parameters such as age, gender, history of alcohol use, violent behaviors, etc., and explored the associations between these parameters and violent behaviors of patients with psychiatric disorders. The aim of our work was to demonstrate the feasibility of mining factors that are strongly associated with violent behaviors

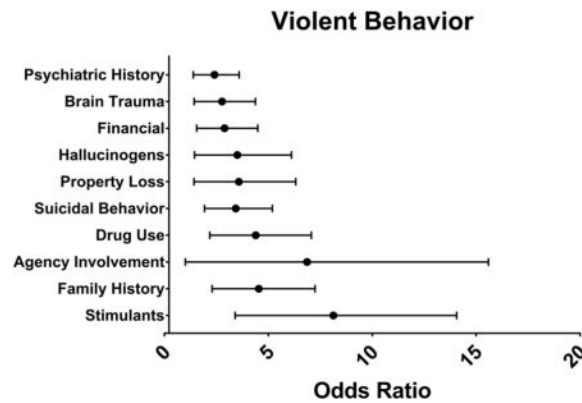
*Corresponding authors.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

There is no conflict of interest regarding the publication of this article.

among psychiatric patients from unstructured psychiatric evaluation records using clinical text mining. Experiment results showed that stimulants, followed by a family history of violent behavior, suicidal behaviors, and financial stress were strongly associated with violent behaviors. Key aspects explicated in this paper include employing our text mining pipeline to extract clinical and social factors linked with violent behaviors, generating association rules to uncover possible associations between these factors and violent behaviors, and lastly the ranking of top rules associated with violent behaviors using statistical analysis and interpretation.

Graphical abstract



1. Introduction

The prevalence of violent behaviors has gained media, business, and political attention lately. In various societies around the world, violence and mental illnesses have reduced the quality of life and raised health concerns of people [1]. Hence, it has recently become a major area of research in the field of mental health. Due to the health consequences of crime and violence, analytical and operational approaches developed by public health practitioners, researchers, and communities are now being used by criminal researchers and practitioners to control criminal violence. These include preventing incidents of violence rather than dealing with their consequences, identifying risk factors that increase the violence, and designing programs to reduce the violence effects [1].

The World Health Organization (WHO) has declared violence to be a major public health problem in 1996, defining it as: “the intentional use of physical force or power, threatened or actual, against oneself, another person, or against a group or community, that either results in or has a high likelihood of resulting in injury, death, psychological harm, maldevelopment or deprivation” [2]. This definition encompasses threat, intimidation, neglect, and abuse along with acts of self-harm and suicidal behaviors [3]. The first World Report on Violence and Health of WHO divided violence into three broad categories: self-inflicted, interpersonal, and collective [2]. Each category was subdivided to show specific types of violence, settings of violence, and the nature of violent acts [4]. Violence cannot be attributed to a single factor, as the etiology of violence in psychological patients reflects various contributing factors. Through the analysis of an ecological model, which can also be considered as a framework for violence prevention, the four key level factors resulting in violence described

by the WHO report include biological and personal factors (e.g., demographics, personality disorders, and a history of violent behaviors), close relationship factors (e.g., family and friends), community contextual factors (e.g., educational environments, work places, and neighborhoods), and broad societal factors (e.g., the criminal justice system, social welfare system, cultural attributes, and the role of media) [2].

Factors such as stressful life situations and brain damage may increase risks of developing mental health problems. A vast amount of research has been conducted in order to identify, assess, and evaluate various types of factors involved in violent behaviors among patients with psychiatric disorders. For example, one study reported that the rate of violent reoffending was lower during periods in which individuals were dispensed antipsychotics, psychostimulants, and drugs for addictive disorders [5]. Use of substances, such as alcohol and drugs, has been linked with a direct increase in risks of violent incidents in psychiatric patients [3, 6, 7]. Diehl *et al.* [8] showed that violent crime was strongly associated with various sexual behaviors and the severity of substance dependence [8], which are clinical manifestations of positive valence systems [9]. Several studies concluded that patients with psychiatric disorders commit serious crimes more frequently compared to normal patients without psychiatric disorders [3, 10–14]. In addition, some studies specifically focused on reviewing the social context and associations of interlinked variables, such as empathy deficits, aggression, and violent behaviors, among children and adolescents [15–18]. A number of studies and reviews have been conducted to assess the incidence, prevalence, and risk factors associated with violent behaviors in psychiatric patients. However, in order to better comprehend the violent behaviors of patients with psychiatric disorders, it is important to investigate the clinical symptoms and psychosocial factors that are strongly correlated to violent behaviors.

Initial psychiatric evaluation records contain a variety of psychiatric assessments that can be employed to extract clinical symptoms and psychosocial risk factors of patients. Unfortunately, most of the information is concealed in the form of unstructured electronic health records (EHRs) [19]. Few studies have explored the clinical and social aspects among psychiatric patients using unstructured EHRs. To address this knowledge gap, we developed a novel medical text mining pipeline using a dataset released by the Partners Healthcare and Neuropsychiatric Genome-scale and RDoC Individualized Domains (N-GRID) project of Harvard Medical School in clinical natural language processing [20] to demonstrate the feasibility of observing the correspondence of clinical and social factors that play a part in violent behaviors among psychiatric patients. First, clinical and social factors linked with violent behaviors were extracted using our text mining pipeline. Subsequently, association rule mining was conducted to locate their possible links with violent behaviors. Finally, the top rules associated with violent behaviors were presented with relative interpretations.

2. Methods

Data

A dataset released by the Centers of Excellence in Genomic Science (CEGS) N-GRID 2016 shared task was used to explore the associations. The dataset contains 1000 de-identified initial psychiatric evaluation records collected from the Partners Healthcare and the N-GRID

project on a per-patient basis [21]. Figure 1 displays an excerpt of an initial psychiatric evaluation record.

As shown in Figure 1, these records include a variety of psychiatric assessments based on different dimensions, such as lifestyle and social behaviors, to identify a patient's psychiatric signs and symptoms. Many lines are structured as a question-answer pair in which the answer may either be short such as 'Yes' or 'No', or longer texts with additional information. For instance, the patient depicted in the example has financial stress, a history of violent behavior, suicidal behaviors, etc. The physician uses a couple of sentences to describe his/her history of alcohol use. Among the released records, some questions are more likely to appear than the others, and their order of appearance in a record is arbitrary.

Text Mining Pipeline

In order to understand the associations between violent behaviors and other factors, we constructed a text mining pipeline to recognize the factors of interest in all of the records. The pipeline integrates an entity recognition system developed in our previous works [22, 23] and systems developed for recognizing protected health information (PHI) and classifying symptom severity of a patient in the CEGS N-GRID 2016 shared task [24]. Figure 2 illustrates the overall text mining pipeline.

Preprocessing

First, all records were tokenized and split into sentences by [nttmuClinical.NET](#) [23]. Each token of a sentence was then processed by the Hunspell checker to correct spelling errors.

Medical Entity Extraction

Various medical entities were extracted using rules and machine learning models established in our previous works [22, 23, 25, 26]. The entities recognized by our pipeline include age, gender, lab values such as glycated hemoglobin, glucose, cholesterol, blood pressure, body mass index, and other heart disease-related risk factors. Appendix A outlines all types of entities extracted by our pipeline. For medications prescribed in the records, we specifically added a rule-based method to distinguish medication names that exist before their corresponding dosages. For instance, the medication "modafinil" can be derived from the phrase "modafinil 200mg QD". A total of 187 identified medications are listed in Appendix B.

Violent Behavior Classification

Recognizing records with violent behavior is an essential step of our study. The accuracy of the classification results is critical to ensure that the outcome of the subsequent investigations on associations can be interpreted properly. Therefore, an iterative approach involving manual validation was adapted.

First, existing questions in the dataset presented in Figure 1 were extracted by regular expression patterns and summarized into a list containing a total of 84 questions. The third and fourth author with clinical background and related work experience manually reviewed the list, and selected 49 questions to form the final list shown in Appendix C. These

questions were the basis of the psychiatric factors considered in this study. Questions specifically related to violent behaviors are shown in Table 1.

Answers to each question listed in Table 1 and Appendix C can be numeric values, nominal values such as gender, binary answers such as ‘yes’/‘no’, or narrative descriptions like “Used to punch walls all the time” for “History (Hx) of Violent Behavior”. We generated a rule-based method to extract the answer to each question. For a given question heading, one of the rules was implemented to extract all of the tokens between the recognized heading and a new line or another question heading that follows as the corresponding answer. The developed rules were used to extract psychiatric factors listed in Appendix C, including violent behaviors.

Furthermore, two keywords (*violent* and *violence*) along with negative word detection (such as *denied* and *no*) were specifically used to determine if a patient had violent behavior or had undergone family violence in case the question did not present in the record. Violence records identified this way were manually reviewed to confirm whether the patient truly had violent behaviors or not. The process described ran iteratively until all records were annotated for violent behaviors.

Psychiatric Risk Factor Extraction

Psychiatric risk factors such as assessment of substance abuse and psychiatric illnesses were extracted with a similar rule-based approach used for Appendix C. In addition, a positive valence severity score was assigned to each patient’s record by our symptom severity classifier based on deep convolutional neural networks [24] and listed as one of the parameters considered for association analysis. The classifier categorized each record into an ordinal scale ranging from 0~3 as listed in Table 2.

Missing Violent Behavior Status

As described in the previous subsection, for each answer type we developed a rule-based method to extract the answer and manually determine the answer category if the extracted answer was a narrative description. Although the output from the developed method was reliable, a lot of missing data were still found. Missing data is defined as no data values extracted for the considered parameter, or the extracted value for the parameter was not informative. For example, with the question-answer rules, 77 records contain missing data for violent behaviors when we only depend on the questions listed in Table 1 for classifying violent behaviors. The violence records identified in this manner were manually reviewed because some patients may have suffered from domestic violence, but did not actually exhibit violent behaviors themselves.

“Once or twice slapped his wife (about four years ago). Spanking kids for discipline. No other violence, weapon use”.

In the example illustrated above, we categorized “slapped his wife” as violence because it was a direct physical aggressive behavior. Nevertheless, “spanking kids for discipline” was uncertain since physical punishment might be allowed in some cultural contexts.

Data Representation for the Association Analysis

For each psychiatric evaluation record, a vector consisting of 244 parameters including violent behaviors, medications, and all factors mentioned was generated. Table 3 shows a snippet of the vectors created from the dataset. Each row represents one evaluation record, and each column indicates an extracted parameter. A question mark “?” was used to indicate missing values.

Finally, data in Table 3 were represented in a binary format in Table 4 for our association analysis. Continuous attributes such as age were discretized into a finite number of intervals by grouping adjacent values of a continuous attribute. Categorical binary attributes, such as gender, were transformed by creating a new item for each distinct attribute-value pair.

Association Rule Mining with Odds Ratio (OR)-based Pruning

Since our goal was to explore the associations of clinical and social variables with violent behaviors among psychiatric patients, we utilized the Apriori algorithm for association rule mining [27] to uncover relationships concealed within the extracted data. Associations related to violent behaviors can be represented in rules of the form $X \rightarrow \text{Violent Behavior}$, where $X \subseteq F$, $F = \{p_1, p_2, \dots, p_n\}$ is a set of variables extracted by our text mining pipeline as shown in Table 4. Each parameter in F is termed an item, and a collection of zero or more items is termed an itemset.

One advantage of using association rule mining is that it can produce all possible associations. By applying the algorithm, $2^n - 1$ contingency tables could be generated, where n is the number of variables under analysis (-1 accounts for the combination where no variable is selected as a violence parameter). The contingency tables can provide a domain-independent data-driven approach to objectively evaluate the quality of associations. Figure 3 displays a two-way contingency table for a pair of parameters, “financial stress” and “violent behavior”. We use the notation “+” and “-” to indicate records with and without the respective parameter. Each entry in the table denotes a frequency count. For example, a is the number of records with both violent behavior and a family history of violent behavior, b is the number of records with violent behavior without suffering violence from family members, c is the number of records with family history of violent behavior but does not exhibit violent behavior, and d is the number of records with neither violent behavior nor a family history of violent behavior. One particular usage of contingency tables is to calculate the odds ratio (OR). In statistics, the odds of an event reflect the likelihood that the event will occur. In Figure 3, the odds of violent behavior occurring can be expressed as $a:c$, which can be computed by a/c .

As a result of applying the association rule mining method, we ended up with 1.4×10^{73} possible contingency tables from 243 variables. This value is an upper bound of the number of contingency tables the algorithm needs to evaluate. However, one can usually adjust the minimum support and confidence to avoid evaluating combinations in which the support is too low. “Confidence” represents how often a rule is found to be true, and is defined as the percentage of violent patients among the total number of patients with antecedent criteria.

“Support” denotes how frequently an item appears in the dataset, and is defined as the proportion of records that contain variable set X .

In this study, we set low values for minimum support since the itemsets associated with significant changes in the odds of demonstrating violent behavior may not appear too often in our dataset. Nonetheless, when working with low values of minimum support and confidence, the Apriori algorithm is likely to output an incredibly high number of rules. Therefore, after conducting an association analysis, we calculated the OR to prune the generated associations. ORs are frequently used in clinical research and decision-making because as an effect-size statistic, they provide clear and direct information to clinicians about which treatment approach has the best odds of benefiting a patient. Including the OR in the step of association rule mining can preserve interpretability of the mined rules and also help focus on rules with interesting ORs [28].

Following the notation of Figure 3, for patients with violent behaviors, the odds of the family history of violent behaviors is $a:b$, while the odds for patients without violence is $c:d$. The steps undertaken to calculate the OR and its standard error (SE) and confidence interval (CI) are described as follows:

$$OR = \frac{a \times d}{b \times c}, SE = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}, \text{ and } CI = \exp(\ln(OR) \pm 1.96 \times SE).$$

When a generated rule involves more than one parameter, the contingency table is updated. Figure 4 shows an example for the rule ‘Financial stress’=Yes and ‘Has the patient had episodes of sudden intense anxiety with physical sensations’ = Yes \rightarrow ‘Violent behavior’ = Yes. The group $a+c$ includes subjects who have both financial stress and intense anxiety, while the group $b+d$ comprises every subject who did not have all of the parameters under analysis. Based on the values in Figure 4, the OR for the rule was 7.743 with a CI of 3.46 to 17.33.

All rules can then be pruned by the following two rules proposed by Toti *et al.* [29].

- A rule is pruned if its 95% CI for the OR overlaps the null value (e.g., OR=1), which indicates a lack of statistical significance of the rule and we cannot be confident that it is effective. For the example shown in Figure 4, the association is statistically significant at $\alpha = 0.05$ because its 95% CI does not contain the value 1.0.
- A rule is pruned if one or more of the parameters considered in the rule can be removed without significant changes in the associated OR. To assure a significant OR difference, we demanded that the 95% CI of the rule cannot overlap with that of any of its parent rules. For the rule studied in Figure 4, the rule “Financial stress’=Yes \rightarrow ‘Violent behavior’ = Yes” is one of its parents. The parent rule has a CI of 1.55~4.49 which overlaps with the studied rule. Therefore, the studied rule does not substantially differ and should be pruned.

3. Results

Using our rule-based violent behavior extraction component, a total of 86 patients with violent behaviors and 851 patients without violent behaviors were identified. Among these, there were 47 cases with missing values and 16 cases noted as uncertain, which might be due to incomplete information or conflicting reports between self-reports and previous medical records. Thus, cases as such were excluded in our final analysis. Table 5 presents the characteristics of the cohort used for our analysis.

Table 6 lists the top 12 extracted association rules that imply violence behaviors calculated from random sampling data including 43 negative and 86 positive psychiatric evaluation records. Variables under the Antecedent column include ‘Financial stress’= *Yes*, ‘Psychiatric history of inpatient treatment’= *Yes*, ‘Gender’= *Male*, ‘History of drug use’= *Yes*, ‘Suicidal behavior/history of suicidal behavior’= *Yes*, and ‘History of brain injury’= *Yes*. The rule with the highest confidence (i.e., 0.95) indicated that if a patient with a psychiatric history of inpatient treatment has financial stress, then it is very likely that he or she will have violent behaviors. In addition, combinations of financial stress with the gender of a patient being male or a history of drug use also suggests the patient has a high chance of exhibiting violent behaviors with confidences > 0.9. If we examine association rules with only a single variable, ‘Financial stress’= *Yes*, ‘Gender’= *Male*, and ‘History of brain injury’= *Yes* all achieved a confidence > 0.8. This implies that these variables are highly correlated with violent behaviors, and these findings can be provided as indicators for violent behaviors in clinical practice.

After applying the association rule mining method with OR-based pruning, the algorithm reported 10 rules that fit the defined criteria of minimum support and a significant OR interval. Results in Table 7 show that stimulants (i.e., illicit drugs like amphetamines) were strongly associated with violent behaviors with an OR of 6.9 (95% CI: 3.4~14.06), followed by a family history of violent behavior (OR: 4.08, 95% CI: 2.29~7.25) and suicidal behavior (OR: 3.17, 95% CI: 1.93~5.19).

Text Mining Pipeline Performance

The psychiatric evaluation records of the dataset follow the question-answer template to examine the health and psychiatric illnesses of patients. This information was extracted using rules established on the basis of questions listed in Appendix C. Furthermore, answers extracted from these questions usually conform to a range of values, such as “Yes”, “No”, or “Uncertain”. This feature allows us to manually verify the results that did not fall into the range to ensure the quality of the information extracted. Thus, the output of our pipeline related to violent behaviors and its associated risk factors is reliable based on our spot checks.

The overall F-measure of our entity recognition in heart disease risk factors, such as medications, blood pressure, and glucose, was about 0.83. Detailed evaluation results were reported in our previous work [26]. Likewise, specific methods related to smoking status detection are reported in [25]. For positive valence severity score classification, the

developed convolutional neural network-based classifier achieved a mean absolute error of 0.539 on the test set of the N-GRID 2016 shared task.

4. Discussion

Results of our study revealed that patients with substance dependence like drug use were highly associated with violent behaviors, similar to the results of a study by Diehl *et al.* [8]. Possession of a firearm was documented in the records we analyzed, but results of the association analysis did not detect a significant correlation between firearm possession alone with the odds of violence. Empirical research by Stroebe [30] on the association of firearm possession with suicides and homicides reported that both suicides and homicides reflected intentional behaviors with the goal of killing oneself or another person. In our dataset, all patients who possessed a firearm do not have a record of violent behavior, indicating that this conduct is not a primary cause of violence.

Our results also supported the evidence that long-term psychological effects of victimization and trauma exposures combined with factors like homelessness, adverse social environments, substance use, and treatment non-compliance can result in a remarkable increased risk of violence in patients with severe mental illness, as reported by Swanson *et al.* [13]. Moreover, we discovered that the high OR of a family history of violent behavior demonstrates that people witnessing routine violent experiences in their surroundings or communities for a long time may begin acting violently as a result of learned behaviors/reactions, resembling the findings of Swanson *et al.* [13].

Biological, psychological, and social factors play roles in violence and aggressive behaviors exhibited by psychiatric patients, although individual factors may manifest distinct degrees of influence, respectively [31, 32]. Examples of biological factors include genetics, hormonal mechanisms, and neurochemical interventions. Miczek *et al.* [33] claimed that it is possible that low blood glucose levels (hypoglycemia) may be conducive to aggressive behaviors. However, the glucose level information is rarely observed in our dataset. By contrast, the ‘suicidal behavior/history of suicidal behavior’ association observed in our study illustrates the individual side mentioned by Rueve and Welton [31], which includes biological-environmental interaction, self-destructive impulses, death instincts, de-individuation, perceptions of deprivation and punishment, and feelings of frustration, fear, injustice, and anger.

Our results also showed that a history of brain injury is a factor in violence. Several studies [34, 35] provided strong evidence that the majority of the patients convicted of violent crimes had various psychiatric disorders. The evidence also substantiates an association of aggressiveness with psychological disorders, implying that patients with psychological disorders are more aggressive in different ways compared to the general community [34, 35]. Aggressiveness is strongly dependent on the type of disorder with which it is associated, as there are different disorders such as psychiatric disorders, neurological diseases, and personality disorders [34, 36]. Different studies have shown that psychopathic, psychotic symptoms and related factors relatively contribute to the development of aggressive

behaviors [7]. In the field of suicidal behavior, major depression, impulsiveness, and aggressiveness are some of the main factors associated with suicide [37, 38]. [26].

Finally, although medical entities were mainly extracted using hand crafted rules in this work, another approach worth exploring is to extract the risk factors by employing supervised learning methods using a specific terminology [39]. However, currently there is no specific ontology or terminology for violence related risk factors.

Limitations

The main limitation of our study was the lack of a systematic evaluation of the developed text mining pipeline. For instance, our text mining pipeline included a glucose level recognizer that can identify high/low glucose level mentioned in electronic medical records [23]. However, we observed that this type of information was rarely available in the dataset, so the competence of the recognizer cannot be assessed. In cases in which the violent behavior status was missing, one of the authors manually determined whether the record was positive or negative. Nevertheless, the decision is subjective and may not be consistent with that of an experienced psychiatrist. This variance may affect the quality of the results. Another major limitation is the inconsistent reporting of the violent behavior status in psychiatric evaluation records. For example, a psychiatrist might summarize that a patient had no history of family violence, but the narrative record contradicts that as follows.

“Difficult childhood. Father w/h/o heavy ETOH. Patient endorsed physical and emotional abuse from him, as well as witnessing domestic violence in the home.”

“Father was out of control and verbally, emotionally, threatened physical violence - would try to kill family in the car. PTSD: Does the patient feel”

In some cases, the psychologist reported no violent behavior in the violence question, but he wrote conflicting comments as follows:

“... was aggressive towards his wife as noted in HPI “he usually just vents with the mouth, “ his wife states and notes no other violent behavior ...”

It is also very important to recognize that dates were shifted in this publicly available dataset to protect the privacy of patients. As a result, some of the time constrained entities such as age might not be the true age because of the time shift process. There is no way to verify this, so in this work we assumed that the temporal nature is maintained.

5. Conclusion

This study utilized text mining techniques to recognize and explore different clinical and social parameters involved in violent behaviors. Association mining rules were generated to produce all possible associations between these parameters and violent behaviors of patients with psychiatric disorders using datasets that contained 1000 de-identified initial psychiatric evaluation records. Results showed that stimulants, followed by a family history of violent behavior, suicidal behaviors, and financial stress were strongly associated with violent behaviors. Limitations of this work included the performance of the text mining pipeline, missing violent behavior data, inconsistent records, and its cross-sectional nature, which can

be further improved by using large-scale population and longitudinal data from multiple sources in the future. Also, additional researches should be conducted to understand the causal nature of these associations of various clinical, social, and cultural parameters with violent behavior, as well as links of these parameters to the motivation of violent behaviors in broader and global contexts.

Acknowledgments

The study was supported by the following grants: Ministry of Science and Technology (MOST) 103-2221-E-038-014, MOST 103-2221-E-038-016, MOST 104-2221-E-038-013, MOST 104-3011-E-038-001, and MOST 105-2221-E-143-003, Health and Welfare Surcharge of Tobacco Products grant MOHW 104-TDU-B-212-124-001, and Ministry of Education grant TMUTOP103006-6. Two grants made the organization of the CEGS N-GRID 2016 shared task possible: NIH P50 MH106933 (to PI: Isaac Kohane) and NIH 4R13LM011411 (to PI: Ozlem Uzuner).

Appendix A

Entity name	Entity category
Age	Demographics
Gender	Demographics
Body-mass index	Physical exam
Obesity	Physical exam
Waist circumference	Physical exam
Blood pressure values	Physical exam
Irregular pulse	Physical exam
Family history	Family history
Insulin therapy	Medication
Blood pressure treatment	Medication
Lipid-lowering therapy	Medication
Smoking	Social history
Chronic kidney disease	Medical history
Diabetic retinopathy	Medical history
Duration of diabetes	Medical history
Neuropathy	Medical history
Hypertension	Medical history
Diabetes	Medical history
Albumin creatinine ratio	Lab values
Creatinine	Lab values
Creatinine kinase	Lab values
Glycated hemoglobin (HbA1C)	Lab values
Lipoprotein-a	Lab values
Triglyceride	Lab values
Low-density lipoprotein	Lab values
High-density lipoprotein (HDL)	Lab values
Total cholesterol	Lab values

Entity name	Entity category
Total cholesterol-HDL ratio	Lab values

Appendix B

The following table shows medications extracted from the dataset by our text mining pipeline.

abacavir, acetaminophen, acetazolamide, acetylcholine esterase inhibitor, acetylsalicylic acid, acyclovir, advair diskus, advair hfa, albuterol and ipratropium nebulizer, albuterol inhaler, albuterol inhaler hfa, alendronate, allopurinol, alprazolam, altavera, alteplase, amitriptyline, amitriptyline hcl, amlodipine, amoxicillin, anastrozole, ARB, aripiprazole, ascorbic acid, aspirin, aspirin enteric coated, atazanavir, atenolol, ativan, atorvastatin, aviane, azathioprine, azelastine, azithromycin, baclofen, balziva, benzac ac wash, beta blocker, betamethasone valerate, bevacizumab, bisacodyl rectal, budesonide nebulizer susp., buprenorphine, dihydrate, bupropion hcl, buspirone hcl, calcitriol, calcium, calcium carbonate, calcium channel blocker, calcium citrate, carvedilol, cetirizine, cholecalciferol, citalopram, claritin D, clindamycin, clobetasol propionate, clonazepam, clonidine, clopidogrel, codeine sulfate, creon, cyanocobalamin, cyclobenzaprine hcl, darunavir, dexamethasone, dextroamphetamine sulfate sustained release, diazepam, diphenhydramine oral, diuretic, divalproex sodium, docusate sodium, doxycycline hyclate, duloxetine, enalapril maleate, enoxaparin, epinephrine autoinjector, epipen, erythromycin, escitalopram, etodolac, etravirine, exemestane, ferrous sulfate, ferrous sulfate take, flovent hfa, fluconazole, fluoxetine hcl, fluticasone nasal spray, fluticasone propionate, fluticasone propionate nasal spray, fluvoxamine maleate, folic acid, furosemide, gabapentin, glimepiride, glipizide, hydrochlorothiazide, hydrocodone, hydrocortisone, hydroxyzine hcl, ibuprofen, isosorbide dinitrate, itraconazole, junel, kariva, ketoconazole, labetalol hcl, lamotrigine, leuprolide acetate, levetiracetam, levofloxacin, levothyroxine sodium, lidex, lidocaine, lisinopril, lithium carbonate, lithium carbonate controlled release, loratadine, lorazepam, losartan, lovastatin, melatonin, metformin, metformin, methylphenidate extended release, methylphenidate extended release capsule, methylprednisolone, metoprolol succinate extended release, metoprolol tartrate, metrogel, mirtazapine, motrin, multivitamins, naltrexone, naprosyn, naratriptan, nitrate, nortriptyline hcl, nystatin powder, nystatin suspension, olanzapine, omega, omeprazole, oxybutynin chloride, oxycodone, pantoprazole, percocet, pimizide, prazosin hcl, prednisone, prenatal multivitamins, prochlorperazine maleate, progesterone, propranolol hcl, prozac, pt taking, quetiapine, retinoic acid, risperidone, selenium sulfide, sennosides, seroquel, sertraline, sildenafil, simvastatin, sinemet, sprintec, statin, sumatriptan, takin, thiamine hcl, thiazolidinedione, topiramate, tramadol, trazodone, tretinoin, triamcinolone, trivora, unithroid, venlafaxine, yaz, zolpidem tartrate

Appendix C

The following table shows the final list of questions. We further categorized the questions for a better understanding to develop the rules required to extract the information.

Violence assessment
Aggressive Thoughts Family History of Violent Behavior Past verbal, emotional, physical, sexual abuse Hx of Violent Behavior Past Hx of Violence Firearms
Psychosocial stressors
Employment Currently employed Financial Stress Is Patient at risk of losing current housing Pain Treatment Pain Does patient feel safe in current living situation Does patient have any children
Function level assessment

Dressing Grooming Learning Disabilities
Suicidality assessment
Suicidal Thoughts Family History of Suicidal Behavior Self Injurious Behavior Hx of Suicidal Behavior
Assessment of psychiatric illness
Has the patient had periods of time lasting two weeks or longer in which, most of the day on most days, they felt little interest or pleasure in doing things, or they had to push themselves to do things Does patient have chronic high risk Has patient ever had periods of being persistently irritable for several days, or had verbal/physical fights that seemed clearly out of character Has the patient had unusual experiences that are hard to explain Does the patient often have thoughts that make sense to them, but that other people say are strange Has the patient had times when they worried excessively about day to day matters for most of the day, more days than not Has the patient had episodes of sudden intense anxiety with physical sensations such as heart palpitations, trouble breathing, or dizziness that reached a peak very quickly and presented without warning Does the patient have persistent fear triggered by specific objects (phobias) or situations (social anxiety) or by thought of having a panic attack Does the patient have other repetitive, unwanted thoughts or behaviors that are non-functional and difficult to stop (e.g. excessive preoccupation with appearance, hairpulling/skin picking, motor or vocal tics) Does the patient have longstanding problems sustaining their attention in activities that are of mediocre interest to them Does the patient experience trauma related flashbacks or recurrent dreams/nightmares Does the patient feel themselves getting very upset whenever they are reminded of their traumatic experience Has the patient had periods of time during which they were concerned about eating or their weight Does the patient think they have an eating disorder Has it been more than 6 months since the loss of a loved one, and does grief continue to significantly interfere with the patients daily living Hallucinations Delusions Hx of Brain Injury Hx of Inpatient Treatment
Substance abuse assessment
Self Abuse Thoughts Family History of Substance Abuse History of drug use Has patient ever had a period of time when he/she felt “up” or “high” without the use of substances How often did you have a drink containing alcohol in the past year How many drinks containing alcohol did you have on a typical day when you were drinking in the past year How often did you have six or more drinks on one occasion in the past year Audit-C Total Score Hallucinogens Marijuana Cocaine Sedative-Hypnotics Stimulants
Unknown
Is there a need for one Other Agency Involvement

References

1. Council, NR. Understanding and Preventing Violence, Volume 4: Consequences and Control. Reiss, AJ., Jr, Roth, JA., editors. Washington, DC: The National Academies Press; 1994. p. 408
2. Krug EG, et al. The world report on violence and health. The Lancet. 2002; 360(9339):1083–1088.
3. Varshney M, et al. Violence and mental illness: what is the true story? Journal of Epidemiology and Community Health. 2016; 70(3):223–225. [PubMed: 26320232]
4. Murphy SN, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). J Am Med Inform Assoc. 2010; 17(2):124–30. [PubMed: 20190053]
5. Chang Z, et al. Association between prescription of major psychotropic medications and violent reoffending after prison release. JAMA. 2016; 316(17):1798–1807. [PubMed: 27802545]

6. Steadman HJ, et al. Violence by people discharged from acute psychiatric inpatient facilities and by others in the same neighborhoods. *Arch Gen Psychiatry*. 1998; 55(5):393–401. [PubMed: 9596041]
7. Volavka J. Aggression in Psychoses. *Advances in Psychiatry*. 2014; 2014:20.
8. Diehl A, et al. Criminality and Sexual Behaviours in Substance Dependents Seeking Treatment. *J Psychoactive Drugs*. 2016; 48(2):124–34. [PubMed: 27163711]
9. Yip SW, Potenza MN. Application of Research Domain Criteria to childhood and adolescent impulsive and addictive disorders: Implications for treatment. *Clinical Psychology Review*. 2016
10. Hornsveld RHJ, Nijman HLI. Evaluation of a cognitive-behavioral program for chronically psychotic forensic inpatients. *International Journal of Law and Psychiatry*. 2005; 28(3):246–254. [PubMed: 15950282]
11. Elbogen EB, Johnson SC. The intricate link between violence and mental disorder: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Arch Gen Psychiatry*. 2009; 66(2):152–61. [PubMed: 19188537]
12. Hodgins S. Violent behaviour among people with schizophrenia: a framework for investigations of causes, and effective treatment, and prevention. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2008; 363(1503):2505–2518.
13. Swanson JW, et al. The Social—Environmental Context of Violent Behavior in Persons Treated for Severe Mental Illness. *American Journal of Public Health*. 2002; 92(9):1523–1531. [PubMed: 12197987]
14. Arango C. Violence in schizophrenia. *Dialogues in Clinical Neuroscience*. 2000; 2(4):392–393. [PubMed: 22033704]
15. Findlay LC, Girardi A, Coplan RJ. Links between empathy, social behavior, and social understanding in early childhood. *Early Childhood Research Quarterly*. 2006; 21(3):347–359.
16. Lovett BJ, Sheffield RA. Affective empathy deficits in aggressive children and adolescents: A critical review. *Clinical Psychology Review*. 2007; 27(1):1–13. [PubMed: 16697094]
17. Castro, MdL, Cunha, SSd, Souza, DPOd. Violence behavior and factors associated among students of Central-West Brazil. *Revista de Saúde Pública*. 2011; 45:1054–1061. [PubMed: 21953077]
18. Silva, RJdS, Soares, NMM., Cabral de Oliveira, AC. Factors Associated with Violent Behavior among Adolescents in Northeastern Brazil. *The Scientific World Journal*. 2014; 2014:863918. [PubMed: 25548796]
19. Jonnagaddala, J., et al. Mining electronic health records to guide and support good clinical decision support systems. In: Moon, J., Galea, MP., editors. *Improving Health Management through Clinical Decision Support Systems*. IGI-Global; 2015.
20. Stubbs A, Filannino M, Uzuner Ö. De-identification of psychiatric intake records: Overview of 2016 CEGS N-GRID Shared Tasks Track 1. *Journal of Biomedical Informatics*. 2017
21. Uzuner O, et al. 2016 CEGS NGRID Shared Tasks and Workshop on Challenges in Natural Language Processing for Clinical Data. 2016
22. Dai HJ, et al. Coreference resolution of medical concepts in discharge summaries by exploiting contextual information. *J Am Med Inform Assoc*. 2012
23. Chang NW, et al. A context-aware approach for progression tracking of medical concepts in electronic medical records. *J of Biomedical Informatics*. 2015; 58(S):S150–S157. [PubMed: 26432355]
24. Hsieh YL, et al. Deep Convolutional Neural Network Document Model for Classifying Symptom Severity in an Research Domain Criteria domain. *Proceedings of the 2016 CEGS NGRID SharedTasks and Workshop on Challenges in Natural Language Processing for Clinical Data*. 2016
25. Jonnagaddala J, et al. A preliminary study on automatic identification of patient smoking status in unstructured electronic health records *ACL-IJCNLP*. 2015; 2015:147–151.
26. Jonnagaddala J, et al. Identification and progression of heart disease risk factors in diabetic patients from longitudinal electronic health records. *BioMed Research International*. 2015
27. Agrawal, R., Imieli ski, T., Swami, A. *Proceedings of the 1993 ACM SIGMOD international conference on Management of data*. New York: ACM; 1993. Mining association rules between sets of items in large databases.

28. Szumilas M. Explaining Odds Ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*. 2010; 19(3):227–229. [PubMed: 20842279]
29. Toti G, et al. Analysis of correlation between pediatric asthma exacerbation and exposure to pollutant mixtures with association rule mining. *Artificial Intelligence in Medicine*. 2016; 74:44–52. [PubMed: 27964802]
30. Stroebe W. Firearm possession and violent death: A critical review. *Aggression and violent behavior*. 2013; 18(6):709–721.
31. Rueve ME, Welton RS. Violence and Mental Illness. *Psychiatry (Edgmont)*. 2008; 5(5):34–48.
32. Mendes DD, et al. Study review of biological, social and environmental factors associated with aggressive behavior. *Revista Brasileira de Psiquiatria*. 2009; 31:S77–S85. [PubMed: 19967203]
33. Miczek KA, et al. An overview of biological influences on violent behavior. *Understanding and preventing violence*. 1994; 2:1–20.
34. Haller J, Kruk MR. Normal and abnormal aggression: human disorders and novel laboratory models. *Neuroscience & Biobehavioral Reviews*. 2006; 30(3):292–303. [PubMed: 16483889]
35. Kobes MHB, Nijman HHLI, Bulten EBH. Assessing Aggressive Behavior in Forensic Psychiatric Patients: Validity and Clinical Utility of Combining Two Instruments. *Archives of Psychiatric Nursing*. 2012; 26(6):487–494. [PubMed: 23164405]
36. Roaldset JO, et al. A multifaceted model for risk assessment of violent behaviour in acutely admitted psychiatric patients. *Psychiatry Res*. 2012; 200(2–3):773–8. [PubMed: 22609226]
37. Chachamovich E, et al. Which are the recent clinical findings regarding the association between depression and suicide? *Rev Bras Psiquiatr*. 2009; 31(Suppl 1):S18–25. [PubMed: 19565147]
38. Lejoyeux M, et al. An Investigation of Factors Increasing the Risk of Aggressive Behavior among Schizophrenic Inpatients. *Frontiers in Psychiatry*. 2013; 4:97. [PubMed: 24027539]
39. Jonnagaddala J, et al. Improving the dictionary lookup approach for disease normalization using enhanced dictionary and query expansion. *Database*. 2016; 2016:baw112. [PubMed: 27504009]

- Text mining can be used to explore parameters associated with violent behaviors in unstructured clinical notes.
- Mental disorders are a significant risk factor for the violent behavior among the patients.
- Stimulants and suicidal tendency were also strongly associated with the patients' violent behavior.

Subject: Patient Initial Visit Note -Identifying Information Date of Service:
 09/13/13CPT Code: 101: No medical services
 Interactive complexity: + 280
 Age:
 24Sex: Male
 Referral Source:
 Self-referredChief Complaint / HPI Chief Complaint (Patients own words)
 "Not being able to live my life fully."History of Present Illness and Precipitating Events
 Patient reported a long history of Hx of Suicidal Behavior: Yes
 If Yes, comment on Timing, Lethality, Impulsivity, Comorbid Intoxication or Psychosis:
 Patient reported that this past week he experienced a fleeting thought about death. He denied ...
 ...
 Patient reported engaging in skin picking, ... Violent Behavior Hx of Violent Behavior: Yes
 If Yes, comment on Timing, Lethality, Impulsivity, Weapon Use, Comorbid Intoxication or ...
 2Longitudinal Alcohol use History:
 Patient noted drinking excessively from 2023-2029. He denied ... He went to hospital treatment
 program for alcohol abuse.Drug Use: How many times in the past year has the patient used an illegal
 drug or used a prescription medication for nonmedical reasons:
 None reportedHistory of drug use: Yes
 ...
 CoffeeSmoker Status: Smoker
 Family Psych History: Family History of Suicidal Behavior:
 Patient reported that his paternal great grandfather killed himself through affixation.Family History of
 Violent Behavior:
 Patient reported that his brother was violent towards others in the family. ... Family History of
 Psychiatric Illness/Hospitalization:
 Patient reported that his sister possibly has Bipolar; grandfather had depression (deceased); mother has
 ...Family History of Substance Abuse: Yes
 Father and mother have a history of alcohol problems.-Medical History Date of Last Phys. ...
 ... FLOODHx of Brain Injury: Yes
 ...
 Financial Stress: Yes
 Living Situation Current Living Situation (Including type of dwelling and who patient lives with):
 Living alone in an apartmentFirearms: None
 Does patient feel safe in current living situation: Yes
 ...
 Patient has a history of suicidal ideation with plan. She denied ever attempting to take her life, but
 noted coming close to it in the past. ...
 ...

Figure 1.
 An example psychiatric evaluation record (edited).

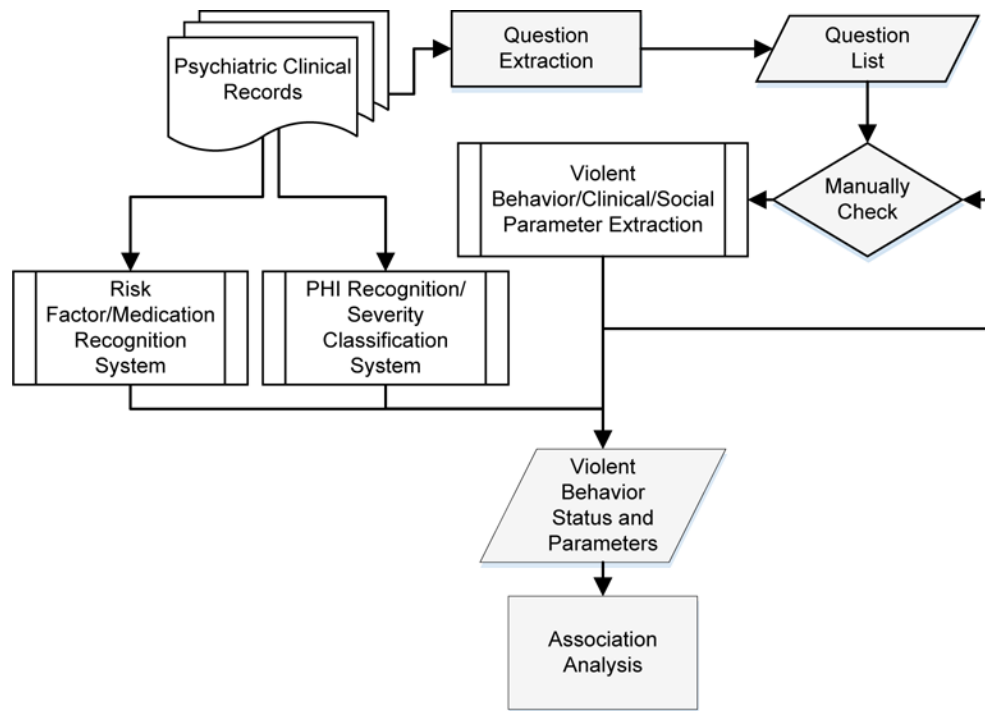


Figure 2.
Flowchart of our study.

		Family Hx of violent behavior	
		+	-
Violent behavior	+	a	b
	-	c	d

Figure 3. Two-way contingency table for parameters “Family Hx of violent behavior” and “violent behavior”.

		+FS, +Q	-FS, -Q
Violent behavior	+	<i>a</i> (24)	<i>b</i> (9)
	-	<i>c</i> (83)	<i>d</i> (241)

Figure 4.

Two-way contingency table for the rule “‘Financial stress’=Yes and ‘Has the patient had episodes of sudden intense anxiety with physical sensations’ = Yes → ‘Violent behavior’ = Yes”. The number in parentheses indicates the value observed in our dataset. FS, Financial Stress; Q, Has the patient had episodes of sudden intense anxiety with physical sensations.

Table 1

Questions for violence assessment.

<ul style="list-style-type: none">• Aggressive Thoughts• Family History of Violent Behavior• Past verbal, emotional, physical, sexual abuse• Hx of Violent Behavior• Past Hx of Violence
--

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Positive valence severity scores

Ordinal scale	Value	Description
Absent	0	No symptoms observed.
Mild	1	Some symptoms present but not a focus of treatment.
Moderate	2	Symptoms present and were a focus of treatment but did not require hospitalization or equivalent.
Severe	3	Symptoms present and required hospitalization, an emergency room visit, or otherwise.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Extracted parameters from initial psychiatric evaluation records.
 VB, violent behavior; Hx, history; SI, suicidal ideation; PVS, positive valence score.

Note ID	Age	Gender	VB	Family Hx of VB	Audit-C Score	Suicidal thoughts	Delusions	PVS	Q: Does patient feel safe in current situation?
0016	47	Female	Yes	Yes	2	Passive SI	No	2	Yes
0059	?	Male	Yes	No	?	No	No	2	?
0247	43	Female	No	No	1	Plan	Persecutory	1	Yes
...
1000	31	Female	No	No	1	?	?	0	Yes
...

Table 4

Example of a binary representation of the extracted data VB, violent behavior.

Note ID	Age < 20	Age ∈ [20, 30)	Age ∈ [30, 40)	...	Gender = Male	Gender = Female	VB = Yes	VB = No	...	Audit-C = 0	Audit-C = 1	...
0016	0	0	0	...	0	1	1	0	...	0	0	...
0059	?	?	?	...	1	0	1	0	...	?	?	...
0247	0	0	0	...	0	1	0	1	...	0	1	...
...
1000	0	0	1	...	0	1	0	1	...	0	1	...

Table 5

Characteristics of the cohort used for analysis. For numeric data, values are expressed as the mean (standard deviation). For categorical data, values are expressed in percentages.

		Violent behaviors		
		No (851)	Yes (86)	
Numeric data	Age	43.14 (17.22)	35.16 (14.70)	
Categorical data	Stimulants	Yes	3.53	19.77
		No	38.19	32.56
	Family history of violent behavior	Yes	0.47	3.49
		Periodic violence	6.11	16.28
		Persistent violence	2.59	9.30
		Homicide	0.24	0
		No	74.38	52.33
	History of drug use	Yes	24.68	38.37
		No	54.52	22.09
	History of suicidal behavior	Yes	14.81	33.72
		No	83.55	56.98
	Is patient at risk of losing current housing	Yes	4.23	12.79
		No	74.85	67.44
	Hallucinogens	Yes	5.41	16.28
		No	36.31	31.40
	Financial stress	Yes	26.79	51.16
		No	43.36	25.58
	History of brain injury	Yes	12.34	23.26
		No	57.34	51.16
	History of psychiatric inpatient treatment	Yes	24.56	39.53
No		69.92	53.49	

Table 6

Extracted association rules that imply violent behaviors

Rule	Antecedent	Consequence	Confidence [#]
1	'Financial stress' = <i>Yes</i> and 'Psychiatric history of inpatient treatment' = <i>Yes</i>	Violent behavior	0.95
2	'Gender' = <i>Male</i> and 'Financial stress' = <i>Yes</i>	Violent behavior	0.92
3	'History of drug use' = <i>Yes</i> and 'Financial stress' = <i>Yes</i>	Violent behavior	0.91
4	'Gender' = <i>Male</i> and 'Psychiatric history of inpatient treatment' = <i>Yes</i>	Violent behavior	0.86
5	'Financial stress' = <i>Yes</i>	Violent behavior	0.83
6	'Gender' = <i>Male</i>	Violent behavior	0.83
7	'Gender' = <i>Male</i> and 'Suicidal behavior/history of suicidal behavior' = <i>Yes</i>	Violent behavior	0.83
8	'Gender' = <i>Male</i> and 'History of drug use' = <i>Yes</i>	Violent behavior	0.80
9	'History of brain injury' = <i>Yes</i>	Violent behavior	0.80
10	'Psychiatric history of inpatient treatment' = <i>Yes</i>	Violent behavior	0.79
11	'History of drug use' = <i>Yes</i>	Violent behavior	0.77
12	'Suicidal behavior/history of suicidal behavior' = <i>Yes</i>	Violent behavior	0.76

[#]Confidence is defined as the percentage of violent patients among the total number of patients with the antecedent criteria.

Table 7

The top 10 rules ranked by the odds ratio (OR).

Violent behavior (a history of violent behavior)	OR	95% Confidence interval	
		Lower	Upper
Use of stimulants	6.92 ^{***}	3.40	14.06
Family history of violent behavior	4.08 ^{***}	2.29	7.25
Other agency involvement	3.98 [*]	1.01	15.59
History of drug use	3.93 ^{***}	2.18	7.07
Suicidal behavior/history of suicidal behavior	3.17 ^{***}	1.93	5.19
Patient at risk of losing current housing	3.0 [*]	1.42	6.32
Use of hallucinogens	2.97 [*]	1.45	6.11
Financial stress	2.64 ^{**}	1.55	4.49
History of brain injury	2.5 ^{**}	1.43	4.38
Psychiatric history of inpatient treatment	2.23 ^{**}	1.39	3.60

^{*} $p < 0.01$;

^{**} $p < 0.001$;

^{***} $p < 0.0001$.