HHS Public Access

Author manuscript

Inform Med Unlocked. Author manuscript; available in PMC 2023 November 09.

Published in final edited form as:

Inform Med Unlocked. 2023 ; 42: . doi:10.1016/j.imu.2023.101345.

Association rule mining of real-world data: Uncovering links between race, glycemic control, lipid profiles, and suicide attempts in individuals with diabetes

Ploypun Narindrarangkura[a], **Patricia E. Alafaireet**[b], **Uzma Khan**[c,d], **Min Soon Kim**[b,e,*]

[a]Phramongkutklao College of Medicine, 317 Ratchawithi Rd, Thung Phaya Thai, Ratchathewi, Bangkok, 10400, Thailand

[b]Department of Health Management and Informatics, University of Missouri, 5 Hospital Drive, Columbia, MO, 65212, USA

[c]Cosmopolitan International Diabetes and Endocrinology Center, USA

[d]Department of Medicine, University of Missouri, 5 Hospital Drive, Columbia, MO, 65212, USA

[e]University of Missouri Institute for Data Science and Informatics, 5 Hospital Drive, Columbia, MO, 65212, USA

Abstract

Aims: The increased risk of suicide among individuals with diabetes is a significant public health concern. However, few studies have focused on understanding the relationship between suicide attempts and diabetes. Association rule mining (ARM) is a data mining technique to discover a set of high-risk factors of a given disease. Therefore, this study aimed to utilize ARM to identify a high-risk group of suicide attempts among patients with diabetes using Cerner Real-World Data™ (CRWD).

Methods: The study analyzed a large multicenter electronic health records data of 3,265,041 patients with diabetes from 2010 to 2020. The Least Absolute Shrinkage and Selection Operator regression with ten-fold cross-validation and the Apriori algorithm with ARM were used to uncover groups of high-risk suicide attempts.

Results: Of the 52,217,517 unique patients in the CRWD, 3,266,856 were diagnosed with diabetes. There were 7764 (0.2%) patients with diabetes who had a history of suicide attempts. The study revealed that patients with diabetes who were never married and had average blood glucose levels below 150 mg/dl were more likely to attempt suicide. In contrast, patients with diabetes aged 60 and older who had diabetes for less than five years and A1C levels between 6.5 and 8.9% were less likely to attempt suicide. Risk factors were strongly associated with suicide

*Corresponding author. Department of Health Management and Informatics, University of Missouri, Columbia University of Missouri Institute for Data Science and Informatics, 5 Hospital Drive, Columbia, MO, 65212, USA. kimms@health.missouri.edu (M.S. Kim).

Authors' contributions

All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Narindrarangkura. The first draft of the manuscript was written by Narindrarangkura, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

attempts, including never married, White, blood glucose levels below 150 mg/dl, and LDL levels below 100 mg/dl.

Conclusions: This is the first study utilizing ARM to discover the risk patterns for suicide attempts in individuals with diabetes. ARM showed the potential for knowledge discovery in large multi-center electronic health records data. The results are explainable and could be practically used by providers during outpatient clinic visits. Further studies are needed to validate the results and investigate the cause-and-effect relationship of suicide attempts among individuals with diabetes.

Keywords

Association rule; Data mining; Diabetes; Risk pattern; Suicide attempts

1. Introduction

In 2020, suicide was the fifth leading cause of death for ages 18–65 years in the United States (US) [1]. The National Survey on Drug Use and Health reported that approximately one million American adults, or 0.5% of the US population, had attempted suicide in the past year [2]. In 2019, 37.1 million adults aged 18 and older, or 14.7% of US adults, were diagnosed with diabetes [3]. Mental health disorders, such as depression, anxiety, schizophrenia, and bipolar disorder, occur in people with diabetes more frequently than in the general population [4].

Glycemic control and target goal achievements are challenges in the diabetes population because of the psychological response to glycemic control and sufficient delivery of care [5]. The questions of self-worth, fear of being a burden for family and caretakers, fear of complications, fear of the future, and limitation of job opportunities are additional challenges faced by people with diabetes [6]. Consequently, people with diabetes have a higher risk of attempted suicide than the general population [7]. A meta-analysis in 2017 confirmed that diabetes significantly increased the risk of suicide, and approximately 90,000 people with diabetes commit suicide yearly [8]. However, poor mental health is still a neglected comorbidity for people with diabetes [9]. Standards of Medical Care in Diabetes 2022, the latest guideline for diagnosing and managing diabetes, provides a comprehensive medical evaluation and assessment of the comorbidities [10]. The guideline included the assessment of comorbidities of autoimmune diseases, cancer, cognitive impairment/dementia, nonalcoholic fatty liver disease, hepatitis C infection, pancreatitis, fractures, sensory impairment, low testosterone in men, obstructive sleep apnea, and periodontal disease with diabetes. However, there were no recommendations for mental health comorbidities with diabetes included.

In medicine, Machine Learning (ML) can support providers with diagnosis, treatment, prognosis, and clinical workflow [11]. ML is defined as “the field of study that gives computers the ability to learn without being explicitly programmed [12].” Most studies that used ML applied logistic regression to identify risk factors for suicide attempts in people with diabetes [7,13,14]. Logistic regression is commonly used in medical research because the dependent variable or outcome of research is usually binary (i.e., positive vs.

negative or alive vs. dead). One of the advantages of logistic regression is that the odds ratio (OR) can be conveniently interpreted [15]. Feature selection is an important step of the ML pipeline that improves the model performance [16]. Logistic regression is prone to overfitting when the number of variables increases [17]. To overcome the overfitting issue, the Least Absolute Shrinkage and Selection Operator (LASSO) regression plays a role. LASSO was proposed in 1996 by Tibshirani as a new method for estimation in linear models [18]. LASSO regression is a regularization technique used to reduce overfitting by adding a penalty term (λ) an absolute value of the magnitude of the coefficient, to shrink the coefficients of some features to exactly zero. Therefore, LASSO regression can be used as a feature selection that eliminates redundant or irrelevant variables with minimal loss of information [19]. For suicide attempts related research, Nock et al. studied patients who visited an emergency department with psychiatric problems to develop one- and six-month suicide attempt prediction models. These researchers compared feature selection techniques between random forest and LASSO regression and reported that LASSO regression provided a higher Area Under the Curve (AUC) of the prediction models [20]. To our knowledge, no study uses the ML approach to determine risk factors and predict suicide attempts among people with diabetes.

One of the problems with ML is interpretability, the degree to which a human can understand the cause of a decision [21]. According to cognitive science, humans use curiosity to understand and facilitate learning, and humans look for explanations to find meaning and manage social interaction [22]. Therefore, issues that raise skepticism about adopting ML in clinical practice may include unfamiliarity with the technology, inability to understand the algorithm, lack of trust, and inability to detect ML failures [23]. On the other hand, Association Rule Mining (ARM) is an explainable rule-based ML technique that has shown great potential in biomedical research, especially in large datasets, for knowledge discovery and for determining associations [24]. The concept of association rules was introduced by Agrawal et al. using a large database of customer transactions to find the frequent IF-THEN patterns [25]; for instance, if milk is purchased, then eggs are likely to be purchased. In the medical field, ARM has become a powerful technique to discover unknown associations, such as disease co-occurrences, drug-disease associations, and symptomatic patterns of disease, from Electronic Health Record (EHR) data [26]. Because few studies have focused on understanding the relationship between suicide attempts and diabetes, ARM for knowledge discovery has a high potential to discover previously unknown relationships between suicide attempts and diabetes [27]. Knowledge discovery is discovering patterns and/or sets of items that frequently occur together, leading to actionable knowledge.

The objective of this study was to discover patterns of risk factors in suicide attempts among people with diabetes. The long-term goal of this study is to develop a clinical decision support system that can be integrated into an EHR in a clinical setting to allow providers to identify patients with DM at high risk for suicide attempts and enable them to provide proper preventive measures during clinic visits. Patients with suicide attempts tend to be recorded in clinical information systems using a coding system, such as the International Classification of Diseases (ICD), better than patients who experienced suicidal ideation, a suicidal behavior with thoughts of engaging in suicide but no actual attempt. This study focused on suicide attempts because it is suicidal behavior with any intent to die, but no death occurred. As such, intervention at this stage will directly impact patient care.

2. Materials and methods

2.1. Study design

This study performed a cross-sectional study to discover patterns of risk factors in suicide attempts among people with diabetes by using EHR data from Cerner Real-World Data (CRWD). CRWD is a big data source consisting of approximately 100 million de-identified patients and more than 1.5 billion encounters from 117 hospitals' EHRs across the United States. CRWD from each hospital is merged into an aggregated data warehouse and processed to reduce duplication of identifiers (e.g., person IDs or encounter IDs) between hospitals. CRWD contains longitudinal data, including allergy, clinical event, condition, lab, demographics, encounter, immunization, measurement, medication, order list, and procedure data [28]. This study was approved by the University of Missouri Health Care Institutional Review Board.

2.2. Study population and data collection

This study retrieved CRWD data on patients with type 1 or type 2 diabetes mellitus (T1DM or T2DM), which were over 95% of diabetes patients, for all age groups from January 2010 to December 2020 using Structured Query Language (SQL) query, a language managing structured data in the database. The time period of data was limited to control the quality of data and avoid issues such as incomplete information and timestamp errors that were often observed in CRWD data documented before 2010. A total of 3,266,856 subjects with diabetes were included, and 7764 (0.2%) had a documented diagnosis of suicide attempts (Fig. 1). The data from CRWD was extracted through the HealtheDataLab™ platform, a cloud-based, high-performance computing service [29]. The data elements were identified by ICD 9th and 10th Revision, Clinical Modification (ICD-9-CM and ICD-10-CM). SQL query was used to help identify data elements that were not coded with ICD-9/10-CM, such as drug names. The final dataset included patients' characteristics, diabetes complications, comorbidities, medication profiles, and laboratory profiles.

2.3. Statistical analysis

Data elements with values missing more than 50% of the time were removed, and those data elements with values missing less than 50% had those missing values replaced by the imputed values generated from Multivariate Imputation by Chained Equations (MICE) [30]. Descriptive analysis was performed to determine the prevalence of suicide attempts among people with diabetes and the distribution of patients' characteristics. In this dataset, only 0.2% of total patients with diabetes had a history of attempted suicide, which caused an imbalanced dataset. The Random Over-Sampling Examples (ROSE), a package for binary imbalanced learning, was utilized on the training dataset by adding synthetic data to balance the class distribution [31]. Data were split into 90% as training and 10% as testing datasets.

After cleaning the dataset, LASSO was used for feature selection to reduce the number of variables in this study (Fig. 1). LASSO regression was performed with ten-fold cross-validation, and the largest λ within one standard error that contributes most to the error to exactly zero was selected [18]. The largest cross-validated λ within one standard error of the minimum binomial deviance was 0.00079, which was used in the final model to remove the redundant or irrelevant variables [32].

Then, an Apriori algorithm [33], a well-known standard approach for ARM, was used to investigate the patterns of risk factors associated with suicide attempts among people with diabetes (Fig. 1). The dataset was transformed into transactions or sets of items format before applying the Apriori algorithm. The Apriori algorithm uses a bottom-up approach to identify frequent items in the database by incrementing one item at a time to make the frequent items larger until the frequent itemset has no item to be derived from the data.

$$\text{Support } (\mathrm{X} \Longrightarrow Y) = \frac{(\text{number of patients with X and Y})}{\text{Total number of patients}} \tag{1}$$

$$\text{Confidence } (\mathrm{X} \Longrightarrow Y) = \frac{(\text{number of patients with X and Y})}{\text{number of patients with X}} \tag{2}$$

$$\text{Lift } (\mathrm{X} \Longrightarrow Y) = \frac{\text{Support } (\mathrm{X} \Longrightarrow \mathrm{Y})}{\text{Support(X)} \times \text{Support(Y)}} \tag{3}$$

The effectiveness of discovered rules is measured by support, confidence, and lift. In this study, support indicates the support of rule $\mathrm{X} \Longrightarrow Y$, which may occur by chance. Confidence indicates the reliability of the rule. Higher confidence means patients who have Y are more likely to have X [34]. Support and confidence were interpreted in percentage. Lift measures the importance of the rule $\mathrm{X} \Longrightarrow Y$. Lift greater than one means X and Y appear more often together than expected. Lift less than one means X and Y appear less often together than expected. Lift near one means X and Y appear almost as often together as expected [35]. The analysis was performed in R (version 4.0.2) using the “arulesCBA” package on HealtheDataLab™ [36]. Mine Class Association Rules [37] was utilized to perform ARM by setting the minimum number of items required in the rule (*minlen*) as 2, the maximum items (*maxlen*) as 10, and the threshold of support (support) and confidence (*confidence*) as 0.2 and 0.8. To determine the impact of well-known risk factors for suicide attempts, we compared association rules before and after removing depression, anxiety, antidepressants, and anti-anxiety medications [38–40]. Patterns or sets of rules were ranked by confidence.

3. Results

3.1. Characteristics of the study population

The patients’ characteristics are shown in Table 1. Of the 52,217,517 unique patients in the CRWD, 3,266,856 were diagnosed with diabetes. We removed less than 5% of patients with abnormal values and two variables with missing values of more than 50%. After data cleaning, 3,265,041 patients were included in this study. There were 7764 (0.2%) patients with diabetes who had a history of suicide attempts. The average age of patients with documented diabetes and suicide attempts was 45, most of whom were between 41 and 60 years old (47.3%). More than half (56.7%) were female, and 52.4% were recorded as never married. Most were white (74.3%) and not Hispanic or Latino (73.9%). Approximately 64% were overweight. More than 90% of study subjects had no history of psychiatric illness in the family, no documented death of family members or divorce, and no documented history of child abuse.

3.2. Association rules mining

ARM generated 263 rules from 2,938,538 transactions and 79 items (dummy variables). There were 250 rules for *Suicide attempts=Yes* and 13 rules for *Suicide attempts=No*. The top ten frequent items of transactions were 1) T2DM, 2) LDL <100 mg/dl, 3) diabetes duration <5 years, 4) hypertension, 5) race identified as White, 6) high-density lipoprotein (HDL) of females <50 mg/dl or males <40 mg/dl, 7) BMI ≥ 30 kg/m^2, 8) benzodiazepines 9) A1C 6.5–8.9%, and 10) female gender. After removing depression, anxiety, antidepressants, and anti-anxiety medications, ARM generated 185 rules from 2,938,538 transactions and 69 items (dummy variables). There were 54 rules for *Suicide attempts=Yes* and 131 rules for *Suicide attempts=No*.

3.2.1. Rules for "Suicide attempts = Yes" before removing depression, anxiety, antidepressants, and anti-anxiety medications—Table 2 shows the top five association rules for *"Suicide attempts=Yes"* among people with diabetes that were ranked by confidence. As an example, the first rule, *{anxiety, atypical}* => *{suicide=Yes}*, Support = 0.18, Confidence = 0.90, means "18% of total diabetes patients had a record of anxiety, atypical antipsychotic agents, and suicide attempts occurred together." In addition, "90% of those patients who had anxiety and received atypical antipsychotic agents also had a record of suicide attempts."

3.2.2. Rules for "Suicide attempts = Yes" after removing depression, anxiety, antidepressants, and anti-anxiety medications—Table 2 shows the top five association rules for "*Suicide attempts=Yes*." The first rule, *{never.married, glucose <150* => *{suicide=Yes}*, Support = 0.14, Confidence = 0.72, means "14% of total diabetes patients had never married marital status, average blood glucose <150 mg/dl, and a record of suicide attempts occurred together." In addition, "72% of patients who never married and average blood glucose <150 mg/dl also had a record of suicide attempts."

3.2.3. Rules for "Suicide attempts = No" before removing depression, anxiety, antidepressants, and anti-anxiety medications—Table 3 shows the top five association rules for *"Suicide attempts=No"* among people with diabetes ranked by confidence. As an example, the first rule, *{age >60, diabetes duration <5, A1C 6.5-8.9} => {suicide=No}*, Support = 0.18, Confidence = 0.84, means "18% of total diabetes patients had age 60 and older, diabetes duration less than five years, A1C 6.5–8.9%, and no record of suicide attempts occurred together." In addition, "84% of those patients aged 60 and older, who had diabetes less than five years and A1C 6.5–8.9% also had no record of suicide attempts."

3.2.4. Rules for "Suicide attempts = No" after removing depression, anxiety, antidepressants, and anti-anxiety medications—

Table 3 shows the top five association rules for "*Suicide attempts=No*" among people with diabetes after removing depression, anxiety, antidepressants, and anti-anxiety medications ranked by confidence. As an example, the first rule is *{age >60, diabetes duration* < *5, A1C 6.5-8.9}* => *{suicide=No}*, Support = 0.18, Confidence = 0.84. The result of no record of suicide attempts remained the same as before removed depression, anxiety, antidepressants, and anti-anxiety medications.

3.3. The graph-based visualization of the top five association rules

Fig. 2 shows a graph-based visualization of the top five association rules for *"Suicide attempts=Yes"* and *"Suicide attempts=No"* among people with diabetes. The shading color from blue to red represents confidence from lower to higher, and the bigger circle means greater support. The arrow from an item that points to the same circle is considered in the same pattern; then, the arrow from the circle points to the target variable. For example, Panel (a) plot shows that *anxiety* and *atypical* point to the circle of red color (highest confidence). Then, from the circle, the arrow points to *suicide=Yes*, which means patients who had *anxiety* and received *atypical antipsychotic agents* also had a record of *suicide attempts*.

4. Discussion

To our knowledge, this is the first study that utilized Association Rule Mining (ARM) to discover unknown associations between suicide attempts and diabetes using a large clinical dataset. Furthermore, this study performed feature selection using LASSO regression, which helped to reduce the number of factors included in the analysis. This study indicated the top five association rules with the highest confidence related to developing suicide attempts among people with diabetes. Within the top five association rules, there were seven unique items: 1) never married, 2) race identified as White, 3) anxiety, 4) depression, 5) atypical antipsychotic agents, 6) benzodiazepines, and 7) LDL <100 mg/dl. After we removed depression, anxiety, antidepressants, and anti-anxiety medications, there were five unique items within the top five association rules: 1) never married, 2) race identified as White, 3) T2DM, 4) glucose <150 mg/dl and 5) LDL <100 mg/dl. As no previous study had applied ARM to identify the risk patterns of suicide attempts among people with diabetes, we explored the potential reasons associated with such attempts.

First, we found that marital status = "*Never married"* was associated with suicide attempts among people with diabetes. This is consistent with findings in the general population, where those who are unmarried or live alone are more prone to attempting suicide compared to those who are married. It might be caused by a lack of social support and feelings of loneliness [41–43].

Second, we found that race = *"White"* was associated with suicide attempts among people with diabetes. The suicide rate among non-Hispanic White people was the second highest in the general population (16.9 per 100,000) [44]. In addition, another study found that non-Hispanic Whites reported suicide attempts more than Hispanics (5.1% vs. 3.2%; $p <$ 0.001) [45]. We suggest a need for further examination of racial disparities in suicide to inform and enhance suicide prevention programs.

Third, type of diabetes = "*T2DM"* was another factor found in the association rules. Löfman et al. identified the relationship between the type of diabetes and suicide attempts [46]. These researchers found that the prevalence of previous suicide attempts among people with T2DM was 1.7-fold (25.5%) compared to people with T1DM (14.8%) ($\chi^2 = 1.18$, $p = 0.277$). Wang et al. conducted a meta-analysis to estimate the risk of suicide in people with diabetes [8]. They reported that the relative risk of suicide associated with T2DM was 1.65 (95% CI: 0.95–2.85).

Fourth, we found that an *average blood glucose* < 150 mg/dl was associated with suicide attempts in people with diabetes. However, mixed results were obtained from previous studies. Bendix et al. discovered that suicide attempters aged 18 or older had higher levels of insulin in cerebrospinal fluid and plasma and lower levels of glucagon in cerebrospinal fluid than the control group, which could explain the lower blood glucose levels seen in suicide attempters, which may explain lower blood glucose levels in suicide attempters [47]. In contrast, Necho et al. reported that individuals with diabetes who had fasting blood glucose levels greater than 130 mg/dl had a higher risk of suicide attempts than controls [48]. Meanwhile, Gómez-Peralta et al. found no significant association between blood glucose levels equal to or greater than 125 mg/dl and suicide attempts in patients with T2DM [7]. Further research is needed to understand better the role of glucose homeostasis in suicide attempts among diabetes patients.

Last, people with diabetes who had *LDL* < 100 mg/dl were found to be associated with suicide attempts. Segoviano-Mendoza et al. reported that LDL levels were significantly lower in patients with major depressive disorder (MDD) and suicide attempts ($p = 0.013$) [49]. In contrast, Messaoud et al., who studied suicide risk in patients with MDD, reported no significant differences in LDL levels between patients with suicide attempts and those without suicide attempts [50]. Overall, the relationship between LDL and suicide attempts in people with diabetes is not well understood and requires further study.

Among patients with diabetes who did not attempt suicide, the association rule with the highest confidence was identified in those *over 60 years old*, who had a *diabetes duration of fewer than five years, with A1C levels ranging from 6.5% to 8.9%*. Previous studies have shown that *older patients with diabetes* have a lower risk of attempting suicide than younger patients [7]. Additionally, Lee et al. found that patients with diabetes with a duration of 5 years or more and patients with diabetes who had A1C levels of 6.5% or higher had a significantly higher risk of suicidal ideation, while diabetes with a duration of fewer than five years and A1C levels less than 6.5% were not associated with suicidal ideation [42]. Further investigation is necessary to understand the relationship between diabetes duration, A1C levels, and suicide attempts in patients with diabetes.

This study discovered the risk patterns of suicide attempts in patients with diabetes, thereby providing essential insights for various stake-holders. Healthcare providers may consider leveraging the findings to refine diagnostic processes, facilitating the implementation of tailored therapeutic interventions and counseling modalities for those identified as high-risk. Concurrently, policy framers and public health administrators are poised to allocate resources judiciously, orchestrating targeted interventions and propagating educational initiatives focused on the nexus between diabetes and mental well-being. Moreover, familial caregivers are better positioned to discern subtle behavioral or emotional deviations in their kin, prompting timely intervention. For researchers, the contradictory findings highlighted offer fertile ground for deeper exploration and might even pave the way for developing standardized tools for early detection in patients with diabetes. While this study provides a structured framework, the real-world application necessitates a compassionate, personalized, and comprehensive, given the complex interplay of mental health determinants in individuals with diabetes.

This study has several limitations that should be considered. First, this study was cross-sectional and cannot establish a causal relationship between risk factors and suicide attempts in people with diabetes. Second, de-identified patient information limits the confirmation of diagnosis codes, which may result in incomplete, inconsistent, or inaccurate data. However, this study improved data quality by removing outliers and data elements with missing values greater than 50%. Third, ARM might discover too many rules and discover poorly understandable rules; however, the study adjusted support and confidence factors to obtain interesting rules and reduced the size of the rules to improve the comprehensibility of discovered rules [51]. Last, the prevalence of suicide attempts in people with diabetes may have been underestimated because of unidentified suicide attempts in ICD-9/10-CM coding. Despite this, the sample size of 64 was calculated to have sufficient power (>95%) to detect a significant difference between the diabetes patients with and without suicide attempts, based on an estimated population incidence of 7.6% [48] and study group incidence of 0.2% with Alpha of 0.05, Beta of 0.05 and power of 0.95.

5. Conclusion

This is the first study utilizing Association Rules Mining (ARM) to discover the pattern of risk factors for suicide attempts in people with diabetes. Our model showed potential for knowledge discovery from large multicenter electronic health records data. The study confirms the links between T2DM, anxiety, depression, and suicide attempts in individuals with diabetes. New findings associated with suicide attempts from our study included *race identified as White, average blood glucose* < 150 mg/dl, and *LDL* < 100 mg/dl. Additionally, we discovered the association rule for individuals with diabetes who had no record of suicide attempts, which was *{age >60, diabetes duration < 5, and A1C 6.5-8.9%}*. This knowledge can be useful for healthcare providers in identifying patients with a high risk of suicide attempts and taking preventive measures during routine clinic visits. Further research is necessary to validate these findings and explore the causal relationship between suicide attempts and diabetes.

Funding statement

This work was supported by the Agency for Healthcare Research and Quality [R21HS028032] and the National Institute of Diabetes and Digestive and Kidney Disorders [P30DK092950].

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: The authors declared no potential conflicts of interest with respect to the research, authorship,

and/or publication of this article: Min Soon Kim reports financial support was provided by Agency for Healthcare Research and Quality and the National Institute of Diabetes and Digestive and Kidney Disorders.

References

[1]. Centers for Disease Control and Prevention. WISQARS (Web-based injury statistics query and reporting system). https://www.cdc.gov/injury/wisqars/index.html. [Accessed 17 September 2022].

[2]. Substance Abuse and Mental Health Services Administration. National Survey of drug use and health (NSDUH) releases. U.S. Department of Health and Human Services; 2021; 2020.

[3]. Centers for Disease Control and Prevention. National diabetes statistics report website. 2022. https://www.cdc.gov/diabetes/data/statistics-report/index.html. [Accessed 4 October 2022].

[4]. Robinson DJ, Coons M, Haensel H, Vallis M, Yale J-F. Diabetes and mental health. Can J Diabetes 2018;42:S130–41. 10.1016/j.jcjd.2017.10.031. [PubMed: 29650085]

[5]. Tareen RS, Tareen K. Psychosocial aspects of diabetes management: dilemma of diabetes distress. Transl Pediatr 2017;6:383–96. 10.21037/tp.2017.10.04. [PubMed: 29184819]

[6]. Barnard-Kelly KD, Naranjo D, Majidi S, Akturk HK, Breton M, Courtet P, et al. Suicide and self-inflicted injury in diabetes: a balancing act. J Diabetes Sci Technol 2020;14:1010–6. 10.1177/1932296819891136. [PubMed: 31801353]

[7]. Gómez-Peralta T, González-Castro T, Fresan A, Tovilla-Zárate C, Juárez-Rojop I, Villar-Soto M, et al. Risk factors and prevalence of suicide attempt in patients with type 2 diabetes in the Mexican population. IJERPH 2018;15:1198. 10.3390/ijerphl5061198. [PubMed: 29880751]

[8]. Wang B, An X, Shi X, Zhang J. Suicide risk in patients with diabetes: a systematic review and meta-analysis. Eur J Endocrinol 2017;177:R169–81. 10.1530/EJE-16-0952. [PubMed: 28512134]

[9]. The Lancet Diabetes & Endocrinology. Poor mental health in diabetes: still a neglected comorbidity. Lancet Diabetes Endocrinol 2015;3:393. 10.1016/S2213-8587(15)00144-8. [PubMed: 26003751]

[10]. American Diabetes Association Professional Practice Committee. 4. Comprehensive medical evaluation and assessment of comorbidities: standards of medical care in diabetes—2022. Diabetes Care 2021;45:S46–59. 10.2337/dc22-S004.

[11]. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med 2019;380:1347–58. 10.1056/NEJMral814259. [PubMed: 30943338]

[12]. Samuel AL. Some studies in machine learning using the game of checkers. IBM J Res Dev 1959;3:210–29. 10.1147/rd.33.0210.

[13]. Chung JH, Moon K, Kim DH, Min J-W, Kim TH, Hwang H-J. Suicidal ideation and suicide attempts among diabetes mellitus: the korea national health and nutrition examination Survey (KNHANES IV, V) from 2007 to 2012. J Psychosom Res 2014;77:457–61. 10.1016/j.jpsychores.2014.08.008. [PubMed: 25258359]

[14]. Roy A, Roy M, Janal M. Suicide attempts and ideation in African-American type 1 diabetic patients. Psychiatr Res 2010;179:53–6. 10.1016/j.psychres.2010.06.004.

[15]. Schober P, Vetter TR. Logistic regression in medical research. Anesth Analg 2021;132:365–6. 10.1213/ANE.0000000000005247. [PubMed: 33449558]

[16]. Brank J, Mladenić D, Grobelnik M, Liu H, Mladenić D, Flach PA, et al. Feature selection. In: Sammut C, Webb GI, editors. Encyclopedia of machine learning. Boston, MA: Springer US; 2011. p. 402–6. 10.1007/978-0-387-30164-8_306.

[17]. Sperandei S. Understanding logistic regression analysis. Biochem Med 2014;12–8.10.11613/BM.2014.003.

[18]. Tibshirani R Regression shrinkage and selection via the lasso. J Roy Stat Soc B 1996;58:267–88.

[19]. Muthukrishnan R, Rohini R. LASSO: a feature selection technique in predictive modeling for machine learning. In: 2016 IEEE international conference on advances in computer applications. ICACA); 2016. p. 18–20. 10.1109/ICACA.2016.7887916.

[20]. Nock MK, Millner AJ, Ross EL, Kennedy CJ, Al-Suwaidi M, Barak-Corren Y, et al. Prediction of suicide attempts using clinician assessment, patient self-report, and electronic

health records. JAMA Netw Open 2022;5:e2144373. 10.1001/jamanetworkopen.2021.44373. [PubMed: 35084483]

[21]. Miller T. Explanation in artificial intelligence: insights from the social sciences. 2017. 10.48550/ARXIV.1706.07269.

[22]. Molnar C Interpretable machine learning: a guide for making black box models explainable. second ed. Munich, Germany: Christoph Molnar; 2022.

[23]. Barragán-Montero A, Bibal A, Dastarac MH, Draguet C, Valdés G, Nguyen D, et al. Towards a safe and efficient clinical implementation of machine learning in radiation oncology by exploring model interpretability, explainability and data-model dependency. Phys Med Biol 2022;67:11TR01. 10.1088/1361-6560/ac678a.

[24]. Abeysinghe R, Cui L. Query-constraint-based mining of association rules for exploratory analysis of clinical datasets in the National Sleep Research Resource. BMC Med Inf Decis Making 2018;18:58. 10.1186/s12911-018-0633-7.

[25]. Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases. SIGMOD Rec 1993;22:207–16. 10.1145/170036.170072.

[26]. Kost R, Littenberg B, Chen ES. Exploring generalized association rule mining for disease Co-occurrences. AMIA Annu Symp Proc 2012;2012:1284–93. [PubMed: 23304407]

[27]. Loshin DChapter 17 - knowledge discovery and data mining for predictive analytics. In: Loshin D, editor. Business intelligence, second ed. Morgan Kaufmann; 2013. p. 271–86. 10.1016/B978-0-12-385889-4.00017-X.

[28]. Ehwerhemuepha L, Carlson K, Moog R, Bondurant B, Akridge C, Moreno T, et al. Cerner real-world data (CRWD) - a de-identified multicenter electronic health records database. Data Brief 2022;42:108120. 10.1016/j.dib.2022.108120. [PubMed: 35434225]

[29]. Ehwerhemuepha L, Gasperino G, Bischoff N, Taraman S, Chang A, Feaster W. HealtheDataLab - a cloud computing solution for data science and advanced analytics in healthcare with application to predicting multi-center pediatric readmissions. BMC Med Inf Decis Making 2020;20:115. 10.1186/s12911-020-01153-7.

[30]. Buuren S van, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained Equations in R. J Stat Software 2011;45. 10.18637/jss.v045.i03.

[31]. Lunardon N, Menardi G, Torelli N. ROSE: random over-sampling examples. 2021.

[32]. Narindrarangkura P, Alafaireet PE, Khan U, Kim MS. Predicting suicide attempts among people with diabetes using a large multicenter electronic health records dataset. Int J Psychiatr Med 2023:009121742311624. 10.1177/00912174231162477.

[33]. Agrawal R, Srikant R. Fast algorithms for mining association rules in large databases. In: Proceedings of the 20th international conference on very large data bases. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 1994. p. 487–99.

[34]. Tan P-N, Steinbach M, Karpatne A, Kumar V. Introduction to data mining. second ed. NY NY: Pearson; 2019.

[35]. Lift in an association rule. https://prod.ibmdocs-production-dal-6099123ce774e592a519d7c33db8265e-0000.us-south.containers.appdomain.cloud/docs/en/db2/10.5?topic=SSEPGG_10.5.0/com.ibm.im.model.doc/c_lift_in_an_association_rule.html. [Accessed 17 September 2022].

[36]. Hahsler M, Grün B, Hornik K. Arules - a computational environment for mining association rules and frequent item sets. J Stat Software 2005;14. 10.18637/jss.v014.i15.

[37]. Hahsler MR. Mine class association rules. R documentation. 2020. https://search.r-project.org/CRAN/refmans/arulesCBA/html/mineCARs.html. [Accessed 29 December 2022].

[38]. Cato V, Holländare F, Nordenskjöld A, Sellin T. Association between benzodiazepines and suicide risk: a matched case-control study. BMC Psychiatr 2019;19:317. 10.1186/s12888-019-2312-3.

[39]. Hawton K, Casañas i Comabella C, Haw C, Saunders K. Risk factors for suicide in individuals with depression: a systematic review. J Affect Disord 2013;147:17–28. 10.1016/j.jad.2013.01.004. [PubMed: 23411024]

[40]. Sareen J, Cox BJ, Afifi TO, de Graaf R, Asmundson GJG, ten Have M, et al. Anxiety disorders and risk for suicidal ideation and suicide attempts: a population-based longitudinal study of adults. Arch Gen Psychiatr 2005;62:1249. 10.1001/archpsyc.62.11.1249. [PubMed: 16275812]

[41]. Calati R, Ferrari C, Brittner M, Oasi O, Olié E, Carvalho AF, et al. Suicidal thoughts and behaviors and social isolation: a narrative review of the literature. J Affect Disord 2019;245:653–67. 10.1016/j.jad.2018.ll.022. [PubMed: 30445391]

[42]. Lee H-Y, Hahm M-I, Lee SG. Risk of suicidal ideation in diabetes varies by diabetes regimen, diabetes duration, and HbA1c level. J Psychosom Res 2014;76:275–9. 10.1016/j.jpsychores.2014.02.003. [PubMed: 24630176]

[43]. Øien-Ødegaard C, Hauge LJ, Reneflot A. Marital status, educational attainment, and suicide risk: a Norwegian register-based population study. Popul Health Metrics 2021;19:33. 10.1186/s12963-021-00263-2.

[44]. Centers for Disease Control and Prevention. Disparities in suicide. https://www.cdc.gov/suicide/facts/disparities-in-suicide.html. [Accessed 3 December 2022].

[45]. Sorenson SB, Golding JM. Suicide ideation and attempts in Hispanics and non-hispanic whites: demographic and psychiatric disorder issues. Suicide Life-Threatening Behav 1988;18:205–18. 10.1111/j.1943-278X.1988.tb00157.x.

[46]. Löfman S, Hakko H, Mainio A, Timonen M, Räsänen P. Characteristics of suicide among diabetes patients: a population based study of suicide victims in Northern Finland J Psychosom Res 2012;73:268–71. 10.1016/j.jpsychores.2012.08.002. [PubMed: 22980531]

[47]. Bendix M, Uvnäs-Moberg K, Petersson M, Kaldo V, Åsberg M, Jokinen J. Insulin and glucagon in plasma and cerebrospinal fluid in suicide attempters and healthy controls. Psychoneuroendocrinology 2017;81:1–7. 10.1016/j.psyneuen.2017.03.019. [PubMed: 28391069]

[48]. Necho M, Mekonnen S, Haile K, Birkie M, Belete A. Suicidal plan, attempt, and associated factors among patients with diabetes in Felegehiwot referral hospital, Bahirdar, Ethiopia: cross-sectional study. BMC Psychiatr 2019;19:258. 10.1186/s12888-019-2253-x.

[49]. Segoviano-Mendoza M, Cárdenas-de la Cruz M, Salas-Pacheco J, Vázquez-Alaniz F, La Llave-León O, Castellanos-Juárez F, et al. Hypocholesterolemia is an independent risk factor for depression disorder and suicide attempt in Northern Mexican population. BMC Psychiatr 2018;18:7. 10.1186/s12888-018-1596-z.

[50]. Messaoud A, Mensi R, Mrad A, Mhalla A, Azizi I, Amemou B, et al. Is low total cholesterol levels associated with suicide attempt in depressive patients? Ann Gen Psychiatr 2017;16:20. 10.1186/s12991-017-0144-4.

[51]. García E, Romero C, Ventura S, Calders T. Drawbacks and solutions of applying association rule mining in learning management systems. CEUR Workshop Proceedings 2007;305:13–22.

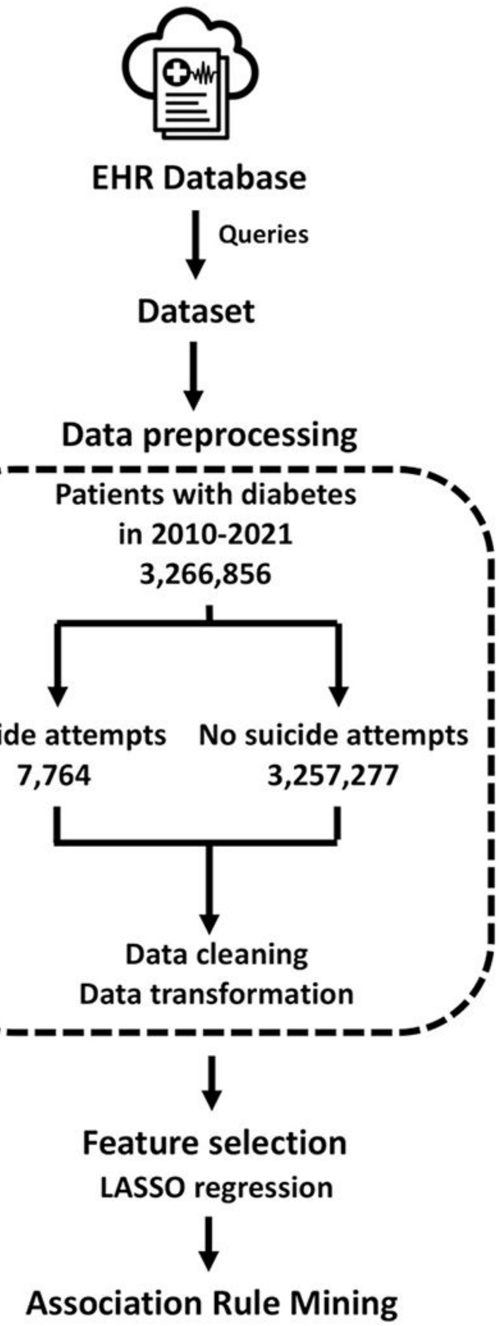

Fig. 1.

Diagram of study workflow. Structured Query Language or SQL was used to query information from the database to create a dataset. Then, data preprocessing was performed to prepare the dataset for analysis.

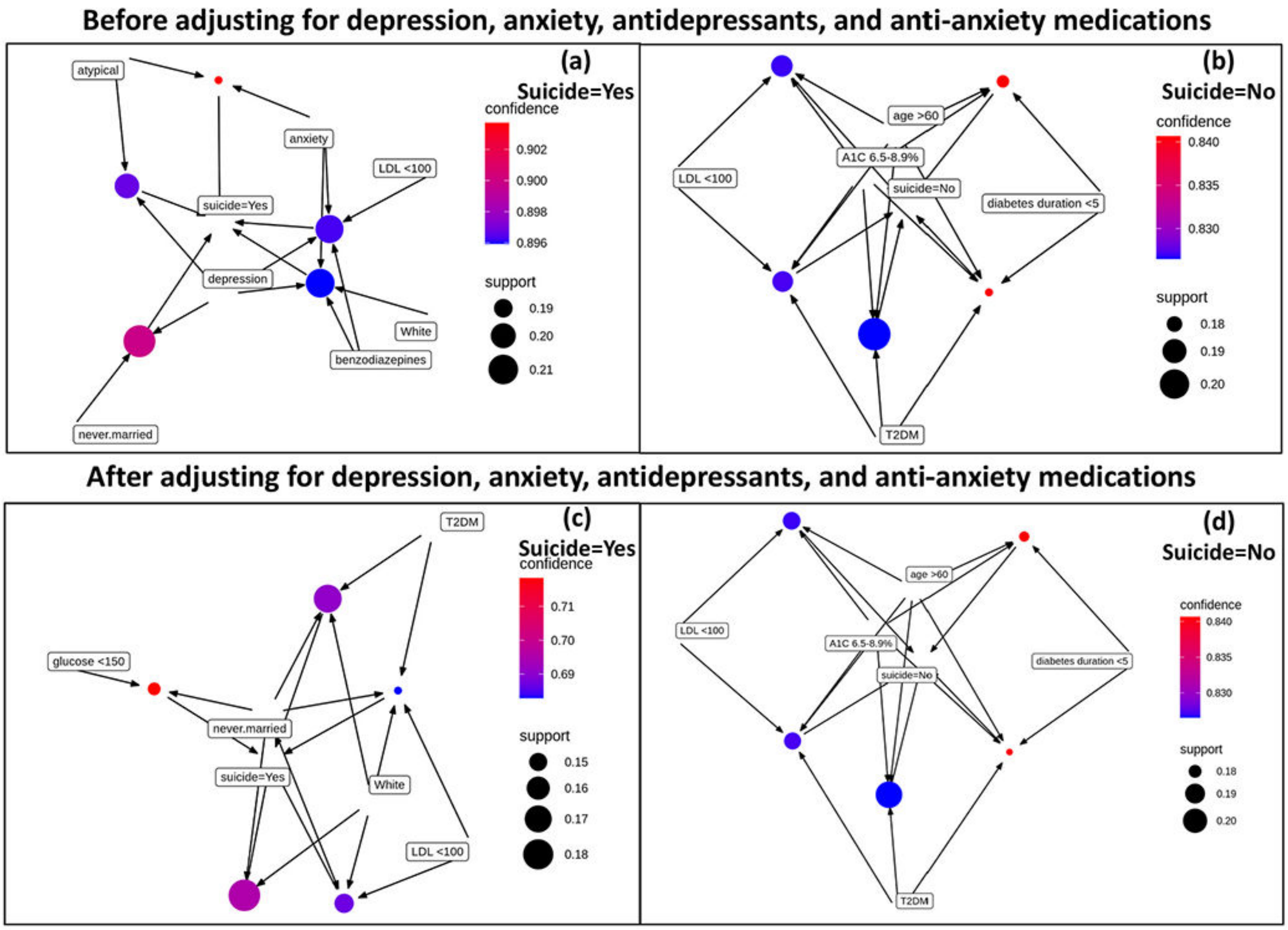

Fig. 2.

The graph-based visualization of the top five association rules. Each rule is represented by a circle, where the shade of blue to red indicates the confidence level. A red color indicates higher confidence, while blue represents lower confidence. Additionally, the size of the circle reflects the level of support for the rule. Panels **(a)** and **(b)** present the top five association rules for "Suicide attempts = Yes" and "Suicide attempts = No" without any adjustments. Panels **(c)** and **(d)** show the top five association rules for "Suicide attempts = Yes" and "Suicide attempts = No" after adjusting for variables such as depression, anxiety, antidepressants, and anti-anxiety medications. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 1

Patient characteristics extracted from Cemer Real-World Data.

	DM			
	Suicide attempts (N = 7764)		No suicide attempts (N = 3,257,277)	
	N	%	n	%
Demographics				
Age (years)	45.1 ± 16.0 (8-88)	62.3 ± 15.9 (0-89)		
≤20	656	8.5	59,905	1.8
21-40	2208	28.4	258,077	7.9
41-60	3674	47.3	1,009,974	31.0
>60	1226	15.8	1,928,753	59.3
Gender				
Female	4402	56.7	1,634,420	50.2
Male	3347	43.1	1,615,831	49.6
Others	7	0.1	3052	0.1
Unknown	5	0.1	3589	0.1
Marital status				
Never Married	4032	52.4	902,970	28.3
Married	1664	21.6	1,468,839	46.0
Divorced	927	12	252,412	7.9
Widowed	460	6	367,412	11.5
Separated	268	3.5	48,280	1.5
Others	30	0.4	19,468	0.6
Unknown	315	4.1	134,370	4.2
Race				
White	5620	74.3	2,143,593	67.2
Black or African American	929	12.3	499,394	15.7
American Indian or Alaska Native	214	2.8	38,446	1.2
Asian	97	1.3	76,544	2.4
Hispanic or Latino	74	1	103,160	3.2
Native Hawaiian or Other Pacific Islander	29	0.4	13,718	0.4
Others	429	5.7	195,950	6.1
Unknown/refused	170	2.2	119,535	3.8
Ethnicities				
Not Hispanic or Latino	4250	73.9	1,831,662	74.6
Hispanic or Latino	1162	20.2	450,827	18.4
Others	93	1.6	23,289	1.0
Unknown/refused	249	4.3	153,131	6.0
BMI (kg/m2)				
<18.5	85	1.2	41,969	1.5
18.5–24.9	901	12.3	368,728	12.6
25–29.9	1656	22.7	731,729	25.0

	DM			
	Suicide attempts (N = 7764)		No suicide attempts (N = 3,257,277)	
	N	%	n	%
≥30	4657	63.8	1,781,737	60.9
Having a family history of psychiatric illness				
No	7084	91.2	3,243,865	99.6
Yes	680	8.8	13,412	0.4
Having the death of family members or divorce				
No	7532	97	3,247,504	99.7
Yes	232	3	9773	0.3
Having a history of childhood abuse				
No	7419	95.6	3,253,788	99.9
Yes	345	4.4	3489	0.1
Having a Family history of diabetes				
No	6705	86.4	3,048,179	93.6
Yes	1059	13.6	209,098	6.4
Diabetes characteristics				
Types of Diabetes				
Type 1	363	4.7	58,193	1.8
Type 2	7334	95.3	3,189,016	98.2
Diabetes duration (years)				
<5	5898	76.0	2,833,887	87.0
≥5	1866	24.0	422,567	13.0
Complications				
Number of complications				
<2	6438	23.7	2,788,994	85.6
≥2	1326	59.1	468,283	14.4
Diabetes complications				
Amyotrophy	2	0.0	1169	0.0
Arthropathy	13	0.2	7943	0.2
Foot ulcer	25	0.3	52,104	1.6
Hyperglycemia	887	11.4	446,270	13.7
Hyperosmolarity	24	0.3	14,340	0.4
Hypoglycemia	117	1.5	81,583	2.5
Ketoacidosis	302	3.9	61,540	1.9
Nephropathy	274	3.5	285,714	8.8
Neuropathy	1060	13.7	270,630	8.3
Coma	10	0.1	816	0.0
Peripheral circulatory disorders	77	1.0	105,897	3.3
Retinopathy	125	1.6	82,644	2.5
Skin complications	21	0.3	32,482	1.0
Unspecified	4323	55.7	942,939	28.9
Comorbidity				

	DM			
	Suicide attempts (N = 7764)		No suicide attempts (N = 3,257,277)	
	N	%	n	%
Number of comorbidities				
<5	5997	77.2	3,074,661	94.4
≥5	1767	22.8	182,616	5.6
Comorbidities				
Anxiety	5568	71.7	536,226	16.5
Cerebral infarction	536	6.9	210,100	6.5
Cognitive impairment	284	3.7	84,179	2.6
Depression	6354	81.8	532,848	16.4
Eating disorder	144	1.9	7897	0.2
Erectile dysfunction	98	1.3	82,186	5.1
Hypertension	5639	72.6	2,316,649	71.1
Hypothyroidism	1619	20.9	508,204	15.6
Ischemic heart disease	1794	23.1	785,227	24.1
Myocardial infarction	704	9.1	229,751	7.1
Obesity	3259	42.0	918,232	28.2
Medication profiles				
Number of prescribed medications in the past year				
<10	3648	47.0	1,248,164	46.1
10-50	2993	38.5	1,327,095	49.0
≥50	1123	14.5	130,812	4.8
Number of prescribed diabetes medications				
<2	6391	82.3	2,032,229	75.1
≥2	1373	17.7	673,842	24.9
Diabetes medications				
Insulin	3700	47.7	1,166,349	43.1
Biguanides	3194	41.1	1,248,492	46.1
Sulfonylureas	1147	14.8	575,843	21.3
Thiazolidinedione	187	2.4	110,411	4.1
DPP-4 inhibitor	460	5.9	250,310	9.2
SGLT2 inhibitor	229	2.9	143,865	5.3
GLP-1 receptor agonist	284	3.7	160,071	5.9
Antidiabetic combinations	167	2.2	116,090	4.3
Number of prescribed antidepressants				
<2	4632	59.7	3,013,035	92.5
≥2	3132	40.3	244,242	7.5
Antidepressants				
SSRI	3789	48.8	521,874	16.0
SNRI	1957	25.2	216,305	6.6
Tricyclic	562	7.2	99,192	3.0
MAOI	10	0.1	733	0.0

	DM			
	Suicide attempts (N = 7764)		**No suicide attempts (N = 3,257,277)**	
	N	**%**	**n**	**%**
Atypical agents	3430	44.2	322,057	9.9
Number of prescribed anti-anxiety medications				
<2	5068	65.3	3,032,556	93.1
≥2	2696	34.7	224,721	6.9
Anti-anxiety medications				
Azaperone	783	10.1	59,281	1.8
Benzodiazepines	5973	76.9	1,548,727	47.5
Antihistamines	2311	29.8	146,222	4.5
Anticonvulsants	646	8.3	124,431	3.8
Laboratory profiles				
Average A1C (%)	7.5 ± 2.2(3.4–20.2)	7.6 ± 1.9(1.0–28.0)		
<5.7	1208	21.0	201,468	9.8
5.7–6.4	1371	23.9	434,782	21.1
6.5–8.9	1835	32.0	1,028,128	50.0
≥9	1327	23.1	393,709	19.1
Average random glucose (mg/dl)	154.2 ± 54.9(60.0–524.9)		163.9 ± 60.5(1.0–1984.0)	
<150	4225	56.9	1,418,077	49.7
151-200	1637	22.1	825,139	28.9
201-300	1458	19.6	516,553	18.1
301-400	95	1.3	78,616	2.8
≥401	11	0.1	16,535	0.6
Average cholesterol (mg/dl)	171.0 ± 43.7(57.0–441.4)		165.7 ± 44.7(1.0–4000.0)	
<200	3898	78.4	1,392,887	81.2
≥200	1077	21.6	322,727	18.8
Average HDL (mg/dl)	43.3 ± 15.4(2.7–226.0)		43.1 ± 15.8 (1.0–3291.0)	
Female≥50, Male≥40 (Desirable)	1734	36.8	681,424	40.7
Female<50, Male<40 (At risk)	2974	63.2	993,276	59.3
Average LDL (mg/dl)	85.7 ± 34.6(3.0–306.0)		84.5 ± 35.6 (1.0–3923.0)	
<100	3308	68.8	1,175,499	70.3
100-129	1009	21.0	326,986	19.5
130-159	356	7.4	122,647	7.3
160-189	101	2.1	34,753	2.1
≥190	37	0.8	12,977	0.8
Average TG (mg/dl)	189.5 ± 178.9 (2.0–3939.5)		168.9 ± 166.6 (1.0–29237.0)	
<150	2567	50.9	996,667	57.3
150-199	954	18.9	372,035	21.4
200-499	1347	26.7	334,311	19.2

	DM			
	Suicide attempts (N = 7764)		No suicide attempts (N = 3,257,277)	
	N	%	n	%
≥500	173	3.4	36,631	2.1

Table 2

The top five association rules for *"Suicide attempts=Yes"* among people with diabetes before and after being adjusted with depression, anxiety, and antipsychotic medications ranking by confidence.

	rules		support	confidence	coverage	lift	count
Before adjusting for depression, anxiety, and antipsychotic medications							
1	{anxiety, atypical}	=> {suicide = Yes}	0.184	0.904	0.203	1.808	539314
2	{depression, never.married}	=> {suicide = Yes}	0.216	0.900	0.240	1.800	634455
3	{depression, atypical}	=> {suicide = Yes}	0.199	0.897	0.222	1.794	585738
4	{anxiety, depression, benzodiazepines, LDL<100}	=> {suicide = Yes}	0.206	0.896	0.230	1.793	606540
5	{anxiety, depression, benzodiazepines, White}	=> {suicide = Yes}	0.208	0.896	0.232	1.792	611033
After adjusting for depression, anxiety, and antipsychotic medications							
1	{never.married, glucose <150}	=> {suicide = Yes}	0.144	0.718	0.200	1.437	422865
2	{never.married, White}	=> {suicide = Yes}	0.185	0.695	0.266	1.391	544150
3	{never.married, White, T2DM}	=> {suicide = Yes}	0.173	0.691	0.250	1.382	507574
4	{never.married, White, LDL <100}	=> {suicide = Yes}	0.152	0.687	0.221	1.374	446656
5	{never.married, White, T2DM, LDL <100}	=> {suicide = Yes}	0.142	0.683	0.208	1.366	417421

Table 3

The top five association rules for *“Suicide attempts=No”* among people with diabetes before and after being adjusted with depression, anxiety, and antipsychotic medications ranking by confidence.

	rules		support	confidence	coverage	lift	count
Before adjusting for depression, anxiety, and antipsychotic medications							
1	{age >60, diabetes duration<5, A1C 6.5–8.9%}	=> {suicide = No}	0.178	0.841	0.212	1.681	523339
2	{age >60, T2DM, diabetes duration<5, A1C 6.5–8.9%}	=> {suicide = No}	0.177	0.840	0.211	1.681	520511
3	{age >60, T2DM, A1C 6.5–8.9%, LDL <100}	=> {suicide = No}	0.185	0.827	0.224	1.654	543988
4	{age >60, A1C 6.5–8.9%, LDL <100}	=> {suicide = No}	0.186	0.827	0.225	1.654	546910
5	{age >60, T2DM, A1C 6.5–8.9%}	=> {suicide = No}	0.206	0.827	0.250	1.653	606652
After adjusting for depression, anxiety, and antipsychotic medications							
1	{age >60, diabetes duration<5, A1C 6.5–8.9%}	=> {suicide = No}	0.178	0.841	0.212	1.681	523339
2	{age >60, T2DM, diabetes duration<5, A1C 6.5–8.9%}	=> {suicide = No}	0.177	0.840	0.211	1.681	520511
3	{age >60, T2DM, A1C 6.5–8.9%, LDL <100}	=> {suicide = No}	0.185	0.827	0.224	1.654	543988
4	{age >60, A1C 6.5–8.9%, LDL <100}	=> {suicide = No}	0.186	0.827	0.225	1.654	546910
5	{age >60, T2DM, A1C 6.5–8.9%}	=> {suicide = No}	0.206	0.827	0.250	1.653	606652