Proceedings of Machine Learning Research 1–10

Predicting Early Inpatient Rehospitalization for Psychiatric Patients with Insurance Claims Data

Sooraj Boominathan, Christopher Wang, Kevin Weng
Massachusetts Institute of Technology, Cambridge, MA

1. Introduction

Rehospitalization after a psychiatric admission is a negative outcome that disrupts the patient's life and leads to increased health care costs. Additionally, early rehospitalization following a discharge may indicate inadequate post-discharge intervention. Improved prediction to identify at-risk patients can help guide targeted interventions and prevent rehospitalization. Our goal is to (1) improve prediction of rehospitalization within 30, 90, and 180 days and (2) understand the diverse health states that underlie admissions for mental disorders and how they might explain rehospitalization.

Our work focuses on improving prediction of rehospitalization using information found in health insurance claims data. We use data from the Truven MarketScan Database (Adamson and Chang (2018)), which contains diagnoses, procedures, and drug codes, as well as demographics and hospital admission data. We train logistic regression and neural network models on simple and learned representations of this data and show that our approach for learning representations yields an improvement in predictive performance, particularly on the 30-day prediction task. We also use LDA to discover subpopulations in our cohort and show that both rehospitalization enrichment and predictive model performance vary across these subpopulations.

We propose an approach for learning representations of admissions data that can improve prediction of rehospitalization outcomes. This improvement will be useful in the face of challenges such as data scarcity. Furthermore, characterizing the subpopulations that are admitted for psychiatric reasons can help identify potential explanations of rehospitalization, which may be used to guide focused interventions.

2. Related Work

There has been previous work on finding good low-dimensional representations of medical concepts. Work by Choi et al. (2016) presented an approach for learning low dimensional embeddings of diagnosis, procedure, and medication codes such that related medical concepts are also close in the learned embedding space. Related work by Choi et al. (2018) proposed an approach for learning low dimensional representations of EHR data that leverages the hierarchical relationship between diagnoses and drugs. Our work focuses on finding useful representations of psychiatric hospital admissions with an approach inspired by the work of Choi et al. (2018), although we do not make use of any hierarchical relationships between the components of an admission.

Previous work has also used topic modeling approaches, such as LDA, to characterize sub-populations related to various medical outcomes. Work by Rumshisky et al. (2016) used LDA to learn topics from electronic health record (EHR) discharge summaries for psychiatric patients. Using these topics as features in predictive models led to improved prediction of rehospitalization. Recent work by Gong et al. (2018) applied LDA to medical codes, instead of narrative text, and showed that this approach can be used to characterize patient subpopulations suffering from opioid abuse. Work by Suresh et al. (2018) presented a framework to discover heterogenous patient subpopulations in the ICU and predict patient mortality outcomes for each subpopulation. We adopt a similar approach to discover subpopulations of patients admitted for psychiatric reasons and analyze the effects of the variance between subpopulations on predictive performance.

3. Methods

We developed models to predict inpatient rehospitalization for patients with mental disorders using information in health insurance claims data. Our models used diagnosis, procedure, and medication codes as input features, along with demographics and auxiliary information about a patient's admission history, such as admission duration and number of previous admissions.

We first developed predictive models by training 'off-the-shelf' logistic regression and feedforward neural network models with naive representations of these features. We then used unsupervised methods to learn more effective representations of single admissions, and used these learned admission representations for training the same types of models. Finally, we used topic modeling algorithms to characterize the subpopulations of patients within our cohort for which the models exhibited particularly good and poor predictive performance.

All of our code is public and open-source on GitHub [1].

3.1. Prediction Problem

We first formally define the prediction problem examined in this paper. Given an inpatient admission A^t with a mental disorder diagnosis related group (DRG), we predict a binary label corresponding to whether the patient was rehospitalized within a fixed time window from the discharge date of A^t. Any inpatient admission within this time window is counted as a rehospitalization, regardless of the admission DRG. The admission history for a patient $A^1, \ldots, A^{t-1}$ consists of all past inpatient admissions for that individual, regardless of the admission DRG. This history is also provided as input to our models to predict rehospitalization after A^t. For the remainder of this paper, we use the term "current admission" to refer to A^t, the admission after which we are predicting rehospitalization.

3.2. Feature Construction

We used three groups of features in our predictive models: medical codes (diagnoses, procedures, and medications), demographic information, and metadata about a patient's admission history. Section 4 provides a comprehensive overview of the features included in our models. We considered both medical codes associated with the current admission and those from the patient's admission history.

We represented patient demographic information using one-hot vectors. We standardized the values of the scalar variables in our feature space (i.e, number of prior admissions) and concatenated them with the other two feature groups. We used two simple representations of the medical codes in a patient's data.

Bag-of-Words Representation We first represented each admission as a bag-of-words over medical codes. The feature vector for an admission has 1s in the dimensions representing a code associated with that admission and 0s in all other dimensions. We incorporated medical codes associated with prior admissions for this patient into the same feature vector; thus, this feature representation did not distinguish between codes associated with the current admission and codes associated with past admissions. Any diagnoses or procedures that a patient received multiple times were counted only once.

Pre-Trained Embedding Representation We next represented the codes for an admission using pre-trained embeddings from Choi et al. (2016), which were learned from a large insurance claims dataset. These embeddings capture meaningful structure in the data, so that clinically-related diagnoses, procedures, and medications are close to one another in the embedding space. We averaged all diagnosis embeddings for a given admission to obtain the overall "diagnosis embedding" for an admission. We used the same operation to combine the procedures and medications in an admission. We concatenated the averaged embeddings for the three code types to obtain a single "admission embedding". This procedure is displayed in Figure 1(a). We performed this step separately for each admission in a patient's history to obtain individual admission representations. We averaged all admissions representations in a patient's history to obtain a single "history embedding" and concatenated this with the embedding of the current admission.

1. https://github.com/czwmit/mlhc_final_project

3.3. Learning Admission Representations

The representations in the previous section contain most information about a patient's admission history, but they represent the data in a flattened manner that may not capture complex interactions among the various components of an admission. We used unsupervised learning techniques to learn low-dimensional embeddings of individual admissions that attempt to model some of these interactions to improve the performance of our predictive models.

To do this, we developed an admission autoencoder architecture which takes as input all diagnosis codes $\{d_i\}$ and procedure codes $\{p_i\}$ for a given admission A^t, along with the patient's previous admission count (n) and current admission length (l). For simplicity, we did not provide medication codes as an input to our architecture. The diagnosis and procedure codes are represented using the same pre-trained embeddings discussed in Section 3.2. The encoder uses these inputs to calculate an embedding o_a for an admission a as follows:

$$v_j = \sigma\Big(W_d r_d(d_j)\Big) \otimes \sum_{i=1}^{p_{max}} r_p(p_i) \tag{1}$$

$$e_a = \sigma\left(\sum_{j=1}^{d_{max}} W_p v_j\right) + \sum_{j=1}^{d_{max}} v_j \tag{2}$$

$$o_a = \tanh\Big(W_a e_a\Big) \tag{3}$$

where r_d, r_p are the embedding functions for diagnoses and procedures, respectively, and p_{max}, d_{max} are the maximum number of diagnoses/procedures for any given admission. We also concatenated n and l to e_a, which is not shown in Eq. (2) above. Figure 1(b) contains a detailed diagram of our proposed architecture.

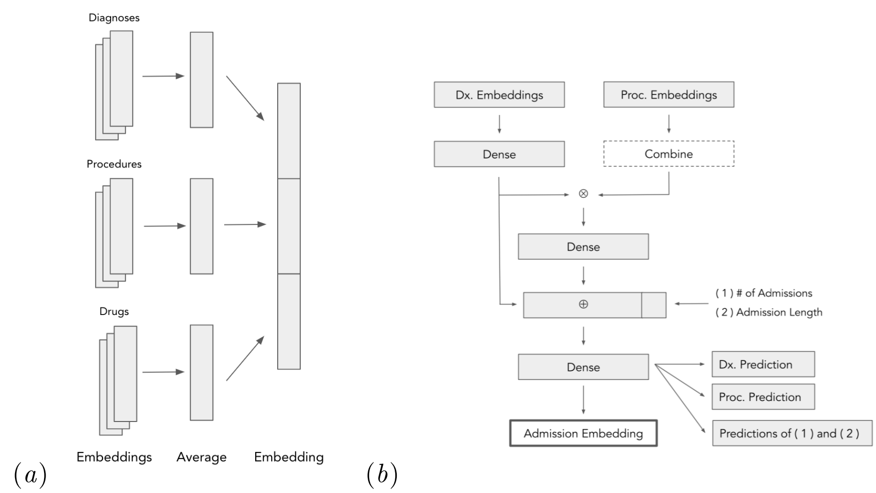

Figure 1: (a) Procedure for calculating averaged pre-trained embedding representation. (b) Autoencoder architecture for learning admission representations

We trained the network by using the admission embedding o_a to predict all input characteristics of the admission. Given D total diagnoses and P total procedures, the loss function for a single admission is:

$$L(A^t, \hat{A}^t) = \sum_{i=1}^{D} CE(d_i, \hat{d}_i) + \sum_{i=1}^{P} CE(p_i, \hat{p}_i) + (n - \hat{n})^2 + (l - \hat{l})^2$$

where $CE(\cdot, \cdot)$ is the binary cross-entropy loss function, and $\hat{d}_i, \hat{p}_i, \hat{n}, \hat{l}$ are the autoencoder's predictions for the features of this admission.

We train the autoencoder to learn representations using all admissions in the training cohort. We use the model to encode all admissions in both our training and test cohort, and use this low-dimensional representation when training our predictive models to predict rehospitalization.

3.4. Topic Modeling

We used Latent Dirichlet Allocation (LDA) to model the patient subpopulations in our cohort in an interpretable manner. LDA models a generative process in which each document is generated from a distribution over latent topics. LDA learns these topics in an unsupervised manner. For our setting, we modeled each inpatient admission and associated history as a "document" and model all medical codes as the "words" of this document. We used a binary bag-of-words representation for these medical codes. Note that this representation does not indicate whether a given diagnoses/procedure appeared multiple times in a patient's history. We applied the standard LDA algorithm to this dataset, selecting the number of topics to provide distinct, interpretable topics related to patient health states.

We analyzed the learned topics by calculating the "rehospitalization enrichment" for each topic, which roughly represents the probability of rehospitalization associated with the topic. It is defined as:

$$\text{Enrichment(topic } k) = \frac{\sum_{i=1}^{n} p_i^k y_i}{\sum_{i=1}^{n} p_i^k}$$

where p_i^k is the proportion of the k^{th} topic in the i^{th} admission, and y_i is the binary rehospitalization label for the i^{th} admission. In order to determine the characteristics of inpatients for whom we are able to successfully predict rehospitalization, we also analyzed the predictive performance of our models on each of the learned topics. We partitioned the admissions by assigning each one to the topic with highest weight in its topic distribution and analyzed the predictive performance of our models on each group.

4. Data / Experiment Setup

We used data from the Truven MarketScan Database, a nationwide private health insurance claims database containing de-identified enrollment history, pharmacy and medical claims of employees and their dependents (Adamson and Chang (2018)). The data contains date-stamped records of medications, diagnoses and procedures undergone by a patient during their time enrolled in coverage.

4.1. Cohort Selection

We considered all individuals with a mood order diagnosis code (ICD code=296.xx) anytime between 2011-14. This includes diagnoses such as bipolar disorder, schizophrenia, and depression. Our cohort consists of inpatient admissions for these patients during this time period with a mental disorder diagnosis group (DRG = 885) that ended in a discharge to home (DSTATUS = 1). Our training set consisted of admissions from 2011-13, and the test set consisted of admissions in 2014. We removed any patient overlap between the train and test set. Statistics for this cohort are shown in Table 1.

	Overall	**Train (2011-13)**	**Test (2014)**
Admissions	15,830	12,768	3,062
Patients	12,033	9,417	2,616
Age (y, mean)	31.3	31.4	31.1
% Male	40.5%	40.6%	40.1%
30-day Rehosp. Rate	12.7%	13.2%	10.8%
90-day Rehosp. Rate	21.3%	22.2%	17.4%
180-day Rehosp. Rate	27.5%	28.8%	22.3%

Table 1: Cohort statistics

4.2. Feature Set Description

As stated in Section 3, we used three groups of features in our models: medical codes (ICD, CPT, NDC), demographics, and admission history metadata. We excluded any medical codes that occurred 10 times or fewer across all admissions in the train set to reduce the feature space. Table 2 lists all of the features considered. Our set of "core features", which are included in all models, consists of all diagnosis/procedure codes for the current admission, along with demographics, past admission count, and admission duration. Our processed feature set consists of 824 ICD codes, 442 CPT codes, and 1867 NDC codes.

Medical Codes	ICD (diagnoses), CPT (procedures), NDC (medications)
Demographics	Sex, Age Group, Geographic Region, Admission Type, Plan Type, Relation to Employee
Admission Metadata	Admission length, Number of Past Admissions, Days Since Last Admission

Table 2: Features used in our models

4.3. Experiment Details

We trained logistic regression (LR) models and simple feedforward NNs and utilized a 80/20 train/validation stratified split for hyperparameter tuning. We chose the optimal hyperparameters as the set which gave the highest mean validation AUC across 10 such splits. Due to the amount of required training time, we only used one such train/validation split when training NNs.

For our LR models, we optimized over 20 different hyperparameter settings: L1/L2 regularization and 10 different regularization strengths ($10^{-4} \leq C \leq 1$). Our feedforward NNs had two hidden layers with 64 and 32 hidden nodes and ReLU activation. We used the Adam optimizer with a learning rate of 0.0001 and trained the network for 100 epochs. We considered three different dropout rate settings between hidden layers: 0.2, 0.5, and 0.8.

5. Results

5.1. Rehospitalization Prediction

We first trained LR models using the bag-of-words representation of procedures and diagnoses. We also evaluated the impact of including admission history and medications in the feature set. Table 3 contains the test AUCs for this set of experiments.

Task	**Core Features**	**Core + History**	**Core+History/Meds**
30 Day	0.619	0.621	0.621
90 Day	0.623	0.629	0.629
180 Day	0.634	0.638	0.640

Table 3: Test AUCs for LR models trained on bag-of-words feature representation

We find that incorporating a patient's past admission history and medications in a naive manner slightly improves performance for the 90 and 180 day prediction tasks, and has almost no effect for the 30 day prediction task. We also observe that 30-day rehospitalization prediction is more difficult than 90 and 180-day prediction.

We next train logistic regression models on the averaged embedding representations of medical codes described in Section 3.2, preserving the remainder of the features from the previous set of experiments. We also train logistic regression models on the admission representations learned using our autoencoder architecture and compare the performance of these two embedding-based representations. Results of these experiments are shown in Table 4.

Task	**Averaged**	**Avg.+ History**	**Learned**
30 Day	0.623	0.615	0.648
90 Day	0.628	0.627	0.640
180 Day	0.634	0.633	0.644

Table 4: Test AUC - LR models trained on averaged pre-trained embeddings and learned admission embedding representations.

We found that LR models trained on averaged embeddings do not yield significant improvement over the bag-of-words representation on any of the prediction tasks. Again, incorporating information about diagnoses/procedures in past admissions does not significantly improve predictive performance either. However, the LR models trained on the learned embeddings displayed significant improvement over the baseline models for the 30 and 90 day prediction tasks. The improvement was particularly significant for the 30 day prediction task.

We also trained feedforward neural network models using the baseline bag-of-words representation and the learned admission embeddings (ignoring history and medications for the reasons outlined earlier). We found that significantly less regularization is required on the networks trained with the learned embeddings to achieve good validation set performance (and consequently, good test generalization) relative to the networks trained with the bag-of-words representations. Table 5 compares the test performance of a 2-layer feedforward neural net with two different dropout rates (D) between fully connected layers on the 30 day rehospitalization task. Even with a dropout rate of 0.5, the neural network trained on the bag-of-words representation overfits severely and produces very poor validation/test performance. When trained on the learned admission embeddings instead, the same network architecture produced significantly better results.

Task	**BoW, D=.2**	**BoW, D=.5**	**Learned, D=.2**	**Learned, D=.5**
Train AUC	0.966	0.910	0.661	0.656
Validation AUC	0.580	0.607	0.647	0.642
Test AUC	0.529	0.555	0.648	0.653

Table 5: Test AUCs - feedforward NN models trained on bag-of-words and learned admission embedding representations.

These results also show that training standard feedforward NN models do not provide a significant improvement relative to logistic regression models. However, it may be possible to achieve superior performance by further tuning the architecture of the neural network and considering deeper networks.

5.2. Model Interpretation

We examined the coefficients in our logistic regression models trained on the bag-of-words representation to understand the most positive and negative predictors of rehospitalization. Table 6 contains the most important predictors for the 30 day task; the most significant features for the other two tasks were similar.

Feature	Coefficient	Feature	Coefficient
Num. of Visits	0.140	Employee	-0.027
Bipolar Disorder	0.033	Age: 35-44	-0.013
Dependent	0.023	Major depressive disorder	-0.013
Unspecified Psychosis	0.022	Region: South	-0.010
Schizoaffective disorder	0.021	Chest radiology exam	-0.010

Table 6: Most positive (left) and negative (right) predictors in LR model for 30-day rehospitalization.

As expected, we see that the previous number of inpatient admissions strongly predicts the likelihood of rehospitalization. While not shown in Table 6, we also find that longer inpatient admissions are predictive of higher rehospitalization risk. We find that the presence of bipolar disorders and schizoaffective disorders are positive predictors of rehospitalization, while middle-age and depression are negatively predictive.

5.3. Subpopulation Characterization

We used LDA with 10 topics to model the different subpopulations within our cohort. We calculated the rehospitalization enrichment (defined in Section 3.4) for 30 day rehospitalization associated with each topic. For each topic, we also calculated the AUC of logistic regression models evaluated on the admissions associated with each topic (see Section 3.4). The logistic regression models used for testing had been trained on the learned admission representations. The results of these analyses are shown in Figure 2. Table 7 contains a description of selected topics, with the most important diagnoses/procedures associated with these topics.

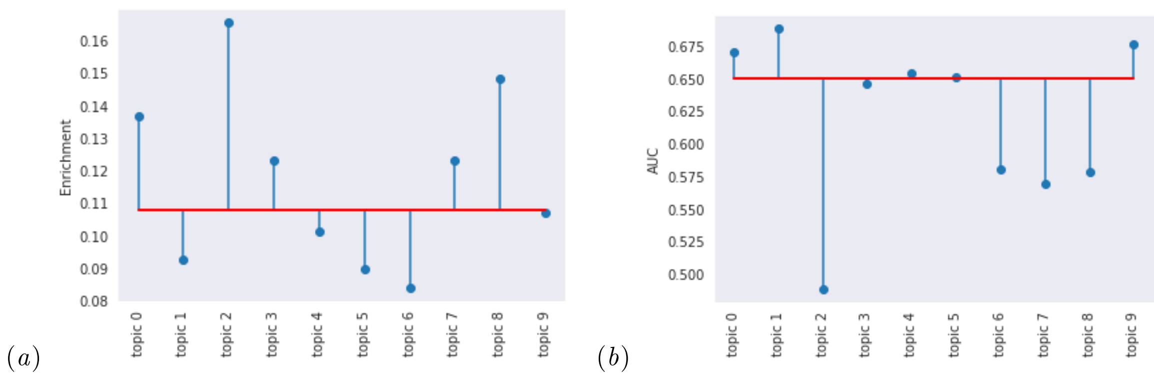

Figure 2: Left: Enrichment of each topic compared to the baseline readmission rate across the entire cohort (red). Right: Test AUC of each topic compared to the baseline AUC on the entire cohort Section 4.

Topic #	**Top Words in Distribution**
0	ECG report, Hypertension, Ages 55-64, Chest x-ray, Chest pain, Diabetes mellitus
2	Abdominal pain, Abdominal Imaging, Chronic pain, Female, Nausea, Bipolar disorder
6	Female, Region: South, Suicidal ideation, Major depressive disorder (recurrent, severe), Major depressive disorder (single episode), Ages 35-44, Ages 18-34

Table 7: Top words in selected topics learned via LDA. Words in each topic are ordered by weight in the topic's word distribution. For complete topics, see Appendix A

We find that there is significant variability in rehospitalization enrichment among the various topics. The topic associated with bipolar disorder and abdominal pain exhibits high enrichment, while the topic associated with depression/suicidal ideation exhibits low enrichment. We also find that our predictive models exhibit different performance between these different subpopulations. We perform particularly poorly on the subpopulation corresponding to bipolar disorder.

6. Discussion

In this work, we developed predictive models for predicting rehospitalization after inpatient admissions for psychiatric reasons. We demonstrated the usefulness of learned admission representations for improving predictive performance and characterized various subpopulations for whom our model exhibited good and poor performance. We found that predicting rehospitalization for psychiatric patients from insurance claims data alone is difficult; there is a lot of valuable information in clinical notes that would have helped considerably, such as more specific descriptions of a patient's behavior and mental state. We also found that predicting rehospitalization for mental disorder-related admissions is more difficult than predicting across all admissions, possibly due to the increased instability of these conditions.

There are several strengths to our analysis. The dataset was compiled from a nationwide population, making it more likely that our results will generalize successfully to other datasets. We also split our data into training and test sets by time, more closely emulating the data that would be available to a model deployed in the real world.

Our work does have some limitations. For one, insurance claims are not a complete or true record of a patient's true health state. There are likely spurious diagnosis codes throughout the data that do not actually represent the patient's health. We also do not account for cases where a patient disenrolled from this insurance program, but may have still been rehospitalized. We also do not have mortality data, so we cannot distnguish between patients who died after leaving the hospital and those who were never rehospitalized.

There are several avenues for future exploration. In this work, we developed predictive models for rehospitalization, but did not analyze any causal relationships in the data. In the future, we could attempt to learn interventions that would actually prevent rehospitalization after admissions for psychiatric reasons. We also did not investigate more effective ways of modeling the history of patient admissions; one might be able to obtain improved predictive accuracy by incorporating that information in a better way. Further investigation is also required to better understand why the learned admission representations provide better generalization on future admissions data.

7. Individual Contributions

Sooraj extracted features from the raw data for the bag-of-words representation and developed the workflow for training/testing logistic regression models. He developed the architecture for learning admission embeddings. **Chris** worked on topic modeling and interpreting the results of LDA as well as the analysis of these subpopulations. He also worked on multi-task learning predictions for different subpopulations, which we decided not to pursue to completion due to time constraints. **Kevin** queried the relevant data from the database and created the cohort. He identified and extracted relevant embeddings from Choi et al.

(2016) and developed the average embedding representation of medical codes. He also worked on a RNN architecture that we decided not to further pursue.

References

David M Adamson and Stella Chang. Health research data for the real world: the marketscan databases. *Thomson Healthcare*, 2018.

Edward Choi, Cao Xiao, Walter Stewart, and Jimeng Sun. Mime: Multilevel medical embedding of electronic health records for predictive healthcare. In *Advances in Neural Information Processing Systems*, pages 4547–4557, 2018.

Youngduck Choi, Chill Yi-I Chiu, and David Sontag. Learning low-dimensional representations of medical concepts. *AMIA Summits on Translational Science Proceedings*, 2016:41, 2016.

Jen J Gong, Abigail Z Jacobs, Toby E Stuart, and Mathijs de Vaan. Discovering heterogeneous subpopulations for fine-grained analysis of opioid use and opioid use disorders. *arXiv preprint arXiv:1811.04344*, 2018.

A Rumshisky, M Ghassemi, T Naumann, P Szolovits, VM Castro, TH McCoy, and RH Perlis. Predicting early psychiatric readmission with natural language processing of narrative discharge summaries. *Translational psychiatry*, 6(10):e921, 2016.

Harini Suresh, Jen J Gong, and John V Guttag. Learning tasks for multitask learning: Heterogenous patient populations in the icu. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 802–810. ACM, 2018.

Appendix A.

Topic #	Top Words in Distribution
0	ECG report, Hypertension, Ages 55-64, Chest x-ray, Chest pain, Diabetes mellitus
1	Ages 0-17, Unspecified episodic mood disorder, Female, Male, Suicidal ideation, Region: North Central, Attention deficit disorder with hyperactivity
2	Abdominal pain, Abdominal Imaging, Chronic pain, Female, Nausea, Bipolar disorder
3	Male, Other and unspecified alcohol dependence, Tobacco use disorder, Alcohol abuse (unspecified), Ages 18-34, Bipolar disorder
4	Complete cbc w/auto diff wbc, Comprehensive metabolic panel, Assay of ethanol, Assay thyroid stim hormone, Drug screen (single), Assay of acetaminophen
5	Female, Suicidal ideation, Region: North Central, Major depressive affective disorder (recurrent episode, severe, without mention of psychotic behavior), Major depressive affective disorder (single episode), Emergency dept visit
6	Female, Region: South, Suicidal ideation, Major depressive disorder (recurrent, severe), Major depressive disorder (single episode), Ages 35-44, Ages 18-34
7	Complete cbc w/auto diff wbc, Comprehensive metabolic panel, Assay of ethanol, Basic metabolic panel, Assay of acetaminophen, Assay of salicylate
8	Ages 18-34, Unspecified psychosis, Male, Bipolar disorder (unspecified), Emergency dept visit, Tobacco use disorder, Cannabis abuse (unspecified), Suicidal ideation, Schizoaffective disorder, unspecified
9	Poisoning by unspecified drug or medicinal substance, Electrocardiogram report, Critical care, first hour, Ct head/brain w/o dye, Emergency dept visit, Female