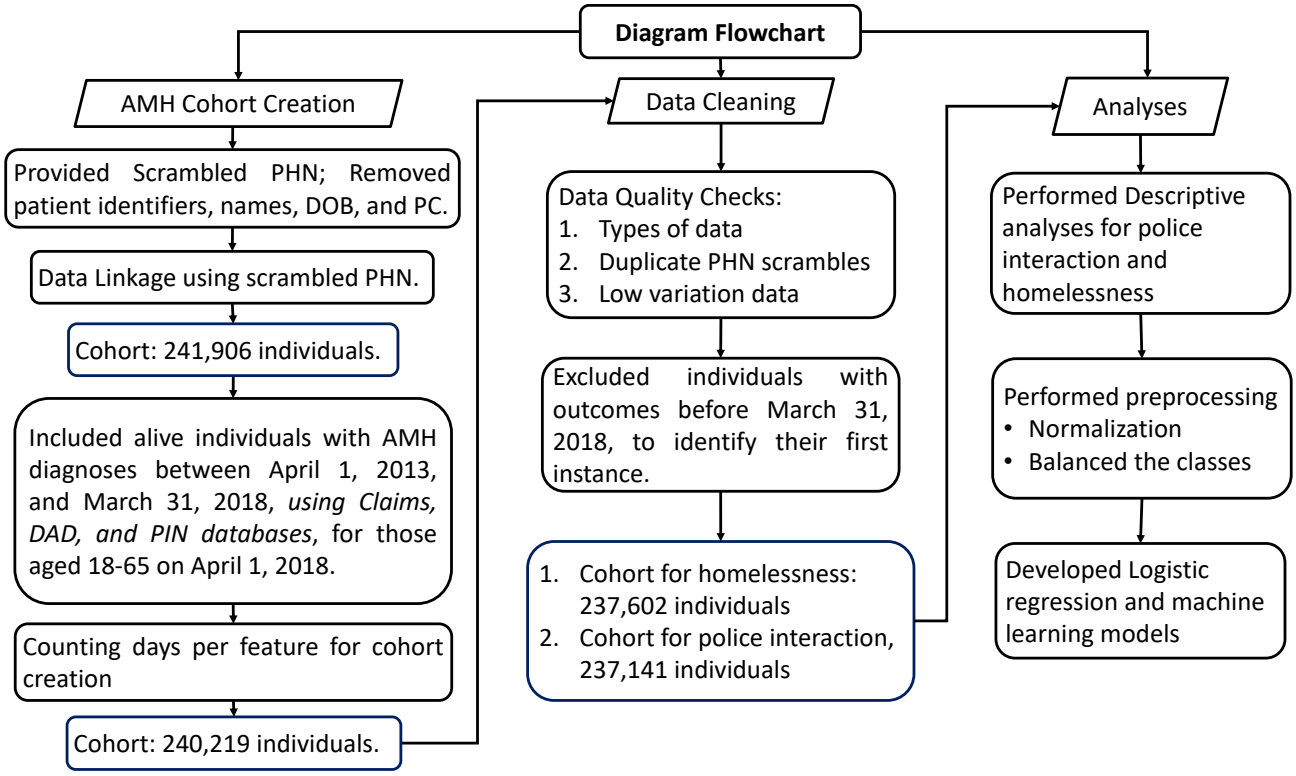
APPENDIX 1: The RECORD checklist of items that were reported in this study.

	RECORD items	Location in manuscript
Title and abstract		
	RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included.	The database was stated in the title and abstract.
	RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract.	The location was stated in the title and abstract.
	RECORD 1.3: If linkage between databases were conducted for the study, this should be clearly stated in the title or abstract.	The linkage was stated in the abstract.
Methods		
Participants	RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided.	In the method of the study, the population selection was explained.
	RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided.	This is not applicable as our population was explained in RECORD 6.1.
	RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	In Appendix 3, we illustrated all the steps to show the steps we took to create the cohort.
Variables	RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	All the codes were reported in Appendix 2.
Data access and cleaning methods	RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population.	All the data sources that we had access to were explained in the data source section of the method.
	RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	Data cleaning was explained in the methodology, and the steps clearly defined in Appendix 3.
Linkage	RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	Person-level linkage was performed, and it is clearly stated in the linkage part in the methodology and illustrated in Appendix 3.
Results		
Participants	RECORD 13.1: Describe in detail the selection of the persons included in the study (i.e., study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	The selection of the persons all described in the results, and it was clearly illustrated in the flowchart in Appendix 3.
Discussion		
Limitations	RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	The limitation of the study is clearly stated in the discussion.
Other Information		
	RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	supplemental information is submitted to the journal. Data access and programming codes are available in the Data Availability section.

APPENDIX 2: Data Sources and the Codes in Details

Data Source	Codes	Details
1. DAD	ICD-10-CA for comorbidities (79)	Type: DXCODE1 - DXCODE 25
	ICD-10-CA for substance use disorder	Any hospitalization: F10.X—F19.X, F55.X, F63.X
	ICD-10-CA for mood disorder	F30.X, F31.X, F32.X, F33.X, F34.X, F38.X, F39.X, F53.0
	ICD-10-CA for anxiety disorder	F40.X, F41.X, F42.X, F43.X, F48.8, F48.9
	ICD-10-CA for psychotic disorder	F06.0, F06.1, F06.2, F06.0-2, F20.X, F22.X, F23.X, F24.X, F25.X, F26.X, F27.X, F28.X, F29.X, F22-F29, F53.1
	ICD-10-CA for cognitive disorders	F00X, F01.X, F02.X, F03.X G30.X
	ICD-10-CA for Other psychiatric disorders	F06-F99
	ICD-10-CA for Deliberate self-harm	X60-X84, Y10-Y19, Y28 when DX10CODE1 was not equal to F06-F99
	ICD-10-CA for homelessness	Z590, Z591
	ICD-10-CA for police interaction	Y350-Y357, Z650-Z653
	MPSERV for the number of hospitalizations to a designated psychiatric facility	64
	MPSERV number hospitalizations to non-psychiatric facility	Not equal to 64
2. Claims	ICD-9-CM for comorbidities	HLTH_DX_ICD9x_CODE_1, HLTH_DX_ICD9x_CODE_2 HLTH_DX_ICD9x_CODE_3
	ICD-9-CM for substance use disorder	2 or more physician claims in 2 years with diagnosis: 291, 292,303, 304, 305
	ICD-9-CM for mood disorder	2 physicians claim at least 30 days apart within 2 years with one or more of the diagnoses codes: 296, 311
	ICD-9-CM for anxiety disorder	2 physicians claim at least 30 days apart within 2 years with one or more of the diagnoses codes: 300, 308, 309
	ICD-9-CM for psychotic disorder	2 physicians claim at least 30 days apart within 2 years with one or more of the diagnoses codes: 295, 297, 298
	ICD-9-CM for cognitive disorders	3 physicians claim at least 30 days apart within 2 years with one or more of the diagnoses codes: 290, 331
	ICD-9-CM for Other psychiatric disorders	2 physicians claim at least 30 days apart within 2 years with one or more of the diagnoses codes: 290 – 319
	ICD-9-CM for homelessness	V600, V601
	ICD-9-CM for police interaction	e97.0-e97.6, e97.8, v62.5, z65.0-z65.3, E970-E976, E978, V625
	PERS_CAPB_PRVD_SPEC_AD for Health Service Utilization	GP: family physician visits NEUR: visits by neurologists INMD: visits by general internal medicine PSYC: visits by psychiatrists
3. PIN	DINS codes for cognitive disorders	Donepezil, Galantamine, Rivastigmine or Memantine
2- NACRS	ICD-10-CA for comorbidities (79)	Type: DXCODE1 - DXCODE 25
	ICD-10-CA for a fall-related injury	W01, W10, W11, W12, W13, W06-W08, W09.01-W09.05, W09.08, W09.09, W14-W17, W00, W02.00-W02.05, W02.08, W03, W04, W05.00-W05.04, W05.08, W05.09, W18, W19
	ICD-10-CA for substance use disorder	Any hospitalization: F10.X—F19.X, F55.X, F63.X
	ICD-10-CA for mood disorder	F30.X, F31.X, F32.X, F33.X, F34.X, F38.X, F39.X, F53.0
	ICD-10-CA for anxiety disorder	F40.X, F41.X, F42.X, F43.X, F48.8, F48.9
	ICD-10-CA for psychotic disorder	F06.0, F06.1, F06.2, F06.0-2, F20.X, F22.X, F23.X, F24.X, F25.X, F26.X, F27.X, F28.X, F29.X, F22-F29, F53.1
	ICD-10-CA for cognitive disorders	F00X, F01.X, F02.X, F03.X G30.X
	ICD-10-CA for Other psychiatric disorders	F06-F99
	ICD-10-CA for Deliberate self-harm	X60-X84, Y10-Y19, Y28 when DX10CODE1 was not equal to F06-F99
	ICD-10-CA for homelessness	Z590, Z591
	ICD-10-CA for police interaction	Y350-Y357, Z650-Z653

APPENDIX 3: The diagram flowchart of the study as described in the methodology. Abbreviations: AMH, addiction or mental health; PHN, personal health number; DOB, date of birth; PC, postal code; Claims, practitioner claims; DAD, discharge abstract database; PIN, pharmaceutical information network.



APPENDIX 4: The hyperparameters for each ML model. Abbreviations: Num Exam; number of examinations for homelessness and police interaction (2*), RF; Random Forests, XGBoost; extreme gradient boosting.

Model	Optimum hyperparameters	Experimented hyperparameters	Num Exam
RF	bootstrap= 'TRUE' max_depth= 10 max_features= 'log2 min_samples_leaf= 8 min_samples_split= 10 n_estimators= 1000 criterion= 'entropy'	bootstrap= ('True') max_depth= (10, 20, 30, 40, 50, 60, 70, 80, 90, 100) max_features= ('sqrt', 'log2') min_samples_leaf= (1, 2, 4, 8) min_samples_split= (2, 5, 10) n_estimators= (400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000) criterion= ('gini', 'entropy')	2* 4,320
XGBoost	subsample=0.9 n_estimators=500 max_depth=3 learning_rate=0.01 colsample_bytree=0.7 colsample_bylevel=0.9	subsample= (0.7, 0.8, 0.9) n_estimators= (100, 500, 1000) max_depth= (3, 5, 6, 10, 15, 20) learning_rate= (0.01, 0.1, 0.2, 0.3, 0.4) colsample_bytree= (0.4, 0.5, 0.6, 0.7, 0.8, 0.9) colsample_bylevel= (0.4, 0.5, 0.6, 0.7, 0.8, 0.9)	2* 9,720