Machine Learning Pilot for Electronic Phenotyping

Sentinel Protocol

Teresa B. Gibson, PhD,[1] Timothy Burrell, MD,[1] Sarah Bloemers, MPH,[1] Cory Pack,[1] Frank Yoon, PhD,[1] Darren Toh, ScD,[2] Jenna Wong, PhD,[2] Adee Kennedy, MS, MPH,[2] Michael Nguyen, MD,[3] Elande Baro, PhD,[3] Wei Hua, PhD,[3] Sai Dharmarajan, PhD,[3] Rita Ouellet-Hellstrom, PhD,[3] Robert Ball, MD[3]

1. IBM Watson Health, Ann Arbor, MI, 2. Department of Population Medicine, Harvard Pilgrim Health Care Institute and Harvard Medical School, Boston, MA, 3. Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, MD

Version 1.0

April 20, 2020

The Sentinel System is sponsored by the U.S. Food and Drug Administration (FDA) to proactively monitor the safety of FDA-regulated medical products and complements other existing FDA safety surveillance capabilities. The Sentinel System is one piece of FDA's Sentinel Initiative, a long-term, multi-faceted effort to develop a national electronic system. Sentinel Collaborators include Data and Academic Partners that provide access to healthcare data and ongoing scientific, technical, methodological, and organizational expertise. The Sentinel Coordinating Center is funded by the FDA through the Department of Health and Human Services (HHS) Contract number HHSF223201400030I.

Machine Learning Pilot for Electronic Phenotyping

Sentinel Protocol

Table of Contents

History of Modifications

Version	Date	Modification	Author
1.0	April 20, 2020	Original version	IBM Watson Health

I. Background

Claims-based algorithms are used in the Food and Drug Administration (FDA) Sentinel's active surveillance system to identify occurrences of health outcomes of interest (HOIs) and medical product safety assessment activities. Validation of a claims-based HOI algorithm typically involves manual review of information abstracted from medical records. Validation of an HOI through medical record abstraction is time-consuming, labor intensive, and costly. To improve on this method, this project investigates the feasibility of using machine learning to develop a claims-based HOI algorithm based on information from a linked claims-electronic medical record (EMR) database. In addition, machine learning methods to develop and validate claims-based signatures (or phenotypes) hold promise because they can (1) detect nonlinear relationships and interactions of features and (2) quickly identify clusters of inputs that may be difficult or impossible to find using traditional clinically informed or linear methods.[1,2]

This project is implemented as part of the FDA Sentinel's HOI 2.0 portfolio aimed at improving detection of HOIs. HOIs requiring clinical testing (e.g., pathology or imaging) often yield unstructured data (e.g., clinical reports or images) and more complex clinical interpretation to confirm the diagnosis. We chose an HOI requiring few laboratory tests to establish the diagnosis to reduce complexity of the evaluation. We reviewed several potential HOIs and selected rhabdomyolysis.

A. Framework

This proof-of-concept project has the potential to accelerate claims-based signature validation by improving the electronic phenotype development and validation process for outcomes detected from standardized information in a linked claims-EMR database. This new approach to algorithm development diverges from the usual gold standard approach with full physician adjudicated chart review of a claims algorithm. This project represents a new approach to algorithm development that leverages laboratory data as a marker of disease. Outputs generated from this project are intended to provide a claims-based phenotype for an HOI using machine learning analysis. Stakeholders, including the FDA, researchers, and clinicians, may benefit from applying this approach to other HOIs and research projects, potentially streamlining HOI validation.

Figure 1 outlines the framework for development and validation of the HOI signature. Specific data sources for the EMR/Electronic Health Record (EHR) and claims data are the IBM® MarketScan® Commercial and MarketScan Medicare Supplemental Databases and the MarketScan Explorys® Claims-EMR Data Set. These data sources are discussed further in the Methods section. EMR data contain the results of laboratory tests recorded using Logical Observation Identifiers Names and Codes (LOINC). LOINC corresponding to the clinical findings of the HOI are identified, the results of these laboratory tests are analyzed, and a determination of a diagnosis of rhabdomyolysis is made based on this information. Non-cases also can be identified in the EMR creating two groups—rhabdomyolysis cases and non-cases.

Candidate predictors or input features to the claims-based algorithm for the HOI include diagnostic information based on the International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) procedures coding using Current Procedural Terminology, Fourth Edition; the Healthcare Common Procedure Coding System (HCPCS); and Place of Service (e.g., office, inpatient hospital, outpatient hospital). Other information available on claims data also can be included as candidate predictors. The claims data elements and HOI indication are fed into machine learning algorithms and the characteristics within the claims data most closely associated with the HOI are selected. The

combination of claims-based characteristics discovered by each machine learning algorithm determines an HOI electronic phenotype. To determine the accuracy of each phenotype, the claims-based electronic phenotype then is compared with the actual EMR determination of rhabdomyolysis diagnosis (the "gold standard" phenotype) using a validation data set that was not used in the training and development of the algorithm. Such a comparison allows evaluation of the accuracy of the claims-based electronic phenotypes produced by the machine learning algorithm.

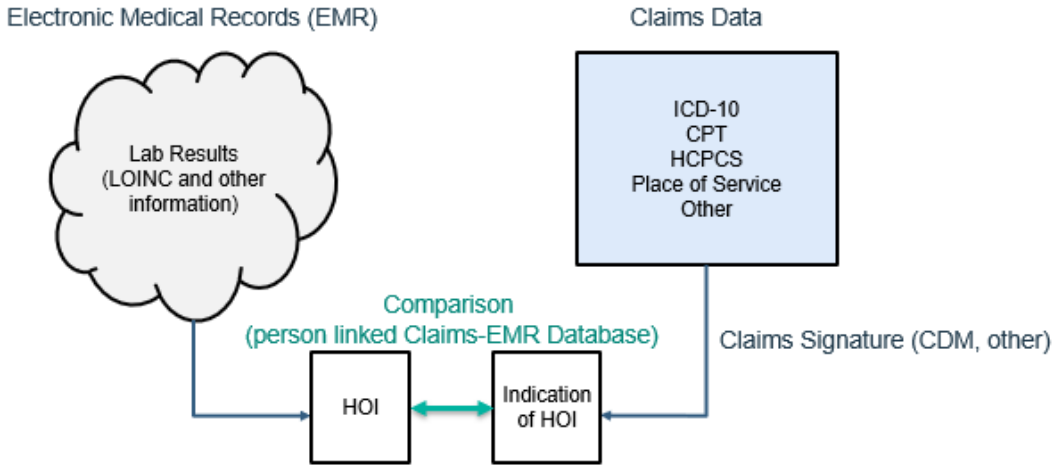


Figure 1. HOI Validation Framework

Abbreviations: CPT, Current Procedural Terminology; HCPCS, Healthcare Common Procedure Coding System; ICD-10, International Classification of Diseases, Tenth Revision; LOINC, Logical Observation Identifiers Names and Codes.

B. Objective

The aim of this project is to use a variety of machine learning classification techniques including logistic regression LASSO (Least Absolute Shrinkage and Selection Operator), support vector machines, a tree-based method, and artificial neural networks, applied to a linked claims-EMR database to demonstrate the feasibility and efficiency of the development and validation of a claims-based HOI algorithm for rhabdomyolysis. This project has the potential to accelerate validation of claims-based signatures by improving the electronic phenotype development and validation process for outcomes detected from standardized information in a linked claims-EMR database.

II. Methods

A. Data Source

Data sources include the MarketScan Commercial and MarketScan Medicare Supplemental Databases and the IBM Explorys EHR Database. The MarketScan Commercial Database provides claims information on employer-sponsored private health insurance provided under a variety of fee-for-service, fully capitated and partially capitated health plans. The MarketScan Medicare data mainly represent claims paid through the Medicare fee for service system. The IBM Explorys EHR Database contains deidentified, longitudinal clinical data across the continuum of care from ambulatory and inpatient to specialty care and post-acute care settings.

The MarketScan Explorys Claims-EMR Data Set links individual patient longitudinal treatment and claims records from the MarketScan Commercial Database and the MarketScan Medicare Supplemental Database to the same individual's records in the IBM Explorys EHR Database. Linking is deterministic at the patient level, and the linked data set is deidentified as required under Health Insurance Portability and Accountability Act provisions.

The MarketScan Explorys Claims-EMR Data Set contains approximately 5 million unique patients represented across the United States, with the highest concentration of population at 41% in the Midwest followed by the South, West, and Northeast regions. It comprises data from a variety of insurers/administrative systems, as well as a variety of provider groups, across the nation.

For each individual in the MarketScan Explorys Claims-EMR Data Set, administrative claims history data are available, which is an essential feature for creating an HOI claims-based signature. Laboratory tests are identified by LOINC code in the EMR and are available for patient services in the inpatient and outpatient setting. Laboratory tests with a categorical result or numeric result value are included within the data set. Qualifying laboratory results include the test result, unit of measurement, and reference points for the test value (e.g., upper limit of normal [ULN]). Data for the study will be accessed for service dates beginning October 1, 2015, to coincide with the introduction of ICD-10-CM. Updated data from the MarketScan Explorys Claims-EMR Data Set are scheduled to be available in the second quarter of 2020, allowing service dates to extend into calendar year 2019.

B. Literature Review of Clinical Indications of Rhabdomyolysis

Rhabdomyolysis is a serious condition that results from the destruction of muscle tissue that can lead to acute kidney injury and death. We conducted a literature review to assess clinical characteristics to inform candidate predictors or input features for the machine learning algorithms. Search specifications included peer-reviewed articles focusing on rhabdomyolysis and associated outcomes with the search covering articles published from 1994 through 2019.

Clinical symptoms for rhabdomyolysis were consistently noted in the literature as muscle pain, muscle weakness, and dark urine color. Rhabdomyolysis incidence was observed to occur most often in adults and rarely in children.[3,4] It also was determined that men experience rhabdomyolysis more often than women.[3,5] Etiology varied within the literature and focused on three causes: traumatic, exertional, and nonexertional (i.e., toxin or medication related).[3-5, 43-61]

A well-know nonexertional cause of rhabdomyolysis are lipid lowering agents such hydroxymethylglutaryl (HMG) CoA reductase inhibitors, also known as statins. These are used for cardiovascular disease prevention and are often discontinued due to side effects such as muscle aches or muscle injury. Medications that inhibit statin metabolism, such as CYP3A4 inhibitors and fibrates (Appendix B – Medications) increase the risk of side effects and of rhabdomyolysis even more in the setting of statin therapy.[6-42]

Diagnosis and clinical determination of the condition focused on several laboratory tests. The principal laboratory evidence for diagnosing rhabdomyolysis is an elevated creatine kinase (CK) serum level. Elevated CK is the most sensitive laboratory test for evaluating muscle injury leading to rhabdomyolysis.[43] Rosenson (2014) provided the following specific measure of CK levels: "Mild as 3-fold greater than upper limit of normal (ULN), Moderate as 10-fold greater ULN, and Severe as 50-fold or more than the ULN."[44] Other researchers have stated that CK levels greater than 5 times the ULN are indicative of rhabdomyolysis.[45] In the setting of statin-induced rhabdomyolysis, CK elevation levels are defined as 10 times the ULN.[46,47,48,49] Additional research articles note that no defined magnitude of

elevation for CK level is available and call for a homogenous definition of CK levels for rhabdomyolysis.[50,51] The timeline for serum CK elevation is approximately 2–12 hours after muscle injury, with peak levels at 24–72 hours after injury. Levels decline back to normal over the following 6–10 days.[52,53]

Myoglobulin levels also were evaluated as potential evidence for causing dark urine. The literature notes different views on utilization of this parameter for diagnosing rhabdomyolysis and discusses the issues of using myoglobin as urine dipstick testing because it is nonspecific and has a short half-life, which limits the sensitivity in testing.[49,54,55] However, in testing for rhabdomyolysis in the intensive care unit, the low cost of testing myoglobulin and its initial pathophysiological role may call for its utilization in assessing for acute kidney injury (AKI) in the setting of rhabdomyolysis.[18] However, it may be more useful as a screening tool rather than a diagnostic test. Additional laboratory testing for rhabdomyolysis included evaluation for increased muscle enzyme levels for lactate dehydrogenase, aspartate aminotransferase, and alanine aminotransferase.[49,51,55]

In addition to diagnostic markers, the literature focuses on outcomes related to rhabdomyolysis. Severity of the condition can range from asymptomatic elevation of CK to severe outcomes such as compartment syndrome, hyperkalemia, and renal failure.[54,56] Renal failure, or AKI, was noted as one of the more severe outcomes of rhabdomyolysis. Individuals under the age of 10 years and with a body weight of less than 50 kg were noted to be at higher risk for potentially fatal outcomes.[5] Authors noted the importance of early identification of causation through laboratory testing of potassium, calcium, and serum creatinine levels.[57,58] To reduce potential negative outcomes associated with rhabdomyolysis, early identification and treatment with renal replacement therapy is essential.[59]

Overall, the literature supports relying on elevated serum CK levels to diagnose rhabdomyolysis of any etiology. To distinguish elevated CK levels from other potential health outcomes, such as myocardial infarction, troponin levels are essential to evaluate as well.[60] Early on, urine myoglobulin presence can be used for disease screening to expedite diagnosis and treatment to improve outcomes. Appropriate diagnosis methods and evaluation for AKI is essential for reducing severe outcomes such as AKI. Surgery without myocardial infarction involvement does not appear to be a confounder because routine postoperative CK level rise does not appear to be clinically significant.[61]

C. Definition of HOI Using EMR Data

The primary laboratory test used to identify rhabdomyolysis is CK. As noted in the literature review, CK is a primary laboratory test for evidence of muscle damage. CK level ULN for this study is defined as 200 units/liter (U/L). The standard unit of measurement of CK will initially be set at U/L and variations of this measurement (international U/L, etc.). Other units of measure such as nanograms per milliliter that can be converted to the U/L threshold may be incorporated in the validation exercise.

Serum creatinine testing is used to assess of the presence of AKI. Creatinine varies by sex as defined in Table 1. Definition of HOI Scenarios, RhabdomyolysisTable 1, with a 1.5 increase over ULN. Troponin testing is added to help distinguish rhabdomyolysis from other potential conditions with elevated CK levels such as myocardial infarction.[60] Because troponin is used to distinguish rhabdomyolysis from other clinical conditions, lack of testing or a negative value will be assessed. Troponin negative results vary by LOINC test as defined in Table 1.

Table 1. Definition of HOI Scenarios, Rhabdomyolysis

LEVEL	CRITERIA	SCENARIO 1	SCENARIO 2
Definite	Must have all of the following: CK elevation > 10–50 × ULN[d] Serum creatinine 1.5 × ULN - Males: 1.5 × ULN = 2.03 mg/dL - Females: 1.5 × ULN = 1.56 mg/dL Negative troponin or no troponin lab present	CK > 10–50 × ULN[d] Serum creatinine 1.5 × ULN Negative troponin or no troponin[a,b,c]	CK > 10–50 × ULN[d] Serum creatinine 1.5 × ULN Negative troponin or no troponin[a,b,c]
	or CK elevation > 50 × ULN[d]	or CK elevation > 50 × ULN[d]	or CK elevation > 50 × ULN[d]
Probable	Must have all of the following: CK elevation > 10–50 × ULN[d] Negative troponin or no troponin lab present	CK > 10 × ULN[d] Negative troponin or no troponin[a,b,c]	CK > 10 × ULN[d] Negative troponin or no troponin[a,b,c]
Possible			CK elevation > 4 × ULN[d]

Abbreviations: CK, creatine kinase; HOI, health outcome of interest; LOINC, Logical Observation Identifiers Names and Codes; ng/mL, nanograms per milliliter; ULN, upper limit of normal.

[a] Negative troponin levels for LOINC 10839-9: value <0.04 ng/mL.

[b] Negative troponin levels for LOINC 6598-7: value <0.01 ng/mL.

[c] Negative troponin levels for LOINC 67151-1: value <15 ng/mL.

[d] CK level ULN for this study is defined as 200 units/liter (U/L)

As outlined in Table 1, two scenarios demonstrate cases for rhabdomyolysis with varying CK, serum creatinine, and troponin levels. Scenario 1 includes cases within the Definite and Probable categories, and Scenario 2 includes Definite, Probable, and Possible cases. Laboratory testing is specified through LOINC as outlined in Appendix A.

D. Classification of HOI

HOI (i.e., rhabdomyolysis) status represents the outcome that will be predicted by the machine learning algorithms in this analysis. The model outcome is expressed as a dichotomous (1/0) outcome using the two scenarios described in Table 1:

- In Scenario 1, the HOI is defined as Rhabdomyolysis = *Yes* if Definite or Probable and *No* otherwise.
- In Scenario 2, the HOI is defined as Rhabdomyolysis = *Yes* if Definite, Probable, or Possible and *No* otherwise.

In both scenarios, noncases (0) are defined as individuals reporting a CK test with a normal result. To develop the claims signature, we use Scenario 1 as the main scenario and Scenario 2 as a second analysis to determine whether the claims signature varies when the definition of rhabdomyolysis is expanded to include the Possible cases.

In the future, if time, resources, and sample sizes allow it may be instructive to develop a claims signature for each level of outcome (i.e., a multicategorical outcome variable). One can use various techniques such as an ordered logit to take advantage of the ordered nature of the outcome (e.g., *Definite* is defined using higher levels of CK than *Probable*). Threshold models with rhabdomyolysis diagnosis as a latent term may be a future refinement to take advantage of the continuous nature of the laboratory results values.

E. Inputs to Machine Learning Models

With clinical input and information from the literature review from Section II.B., we will develop definitions of candidate predictors or input features for the machine learning models. We consider only features that can be derived from claims data such as medications, diagnoses and procedures. Features should be meaningful, not highly correlated with each other, and measurable.

Definitions of candidate predictors or input features will be aligned with definitions of data elements in the Sentinel Common Data Model (SCDM). Analysis of medication is solely designed to assess the predictive performance of the algorithm and not to explain the causal role of medications. The largest predictors will be examined to analyze the role of medications, cast in the context of predictive purpose. We may also include additional information not present in the SCDM to test the explanatory or discriminatory power of the additional information associated with inclusion. These include laboratory tests and provider specialty not in the SCDM.

Initial input features are based on categories derived from the SCDM. Table 2 includes categories containing input features or candidate predictors. Appendix B contains further details on the predictors.

Table 2. Predictor Categories from the Sentinel Common Data Model

Administrative Data
Demographic (Age at Index Date, Sex)
Encounter Data
Encounter Type (Inpatient)
Diagnosis
Diagnosis
Diagnosis Code Type (ICD-10-CM, Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT))
Medications (models will be estimated with and without medications)
Medication
Procedure
Procedure
Procedure Code Type (ICD-10-PCS, CPT)
Other
Provider specialty

Abbreviations: CPT, Current Procedural Terminology; ICD-10-CM, International Classification of Diseases, Tenth Revision, Clinical Modification; ICD-10-PCS, International Classification of Diseases, Procedure Coding System; SNOMED CT, Systematized Nomenclature of Medicine – Clinical Terms.

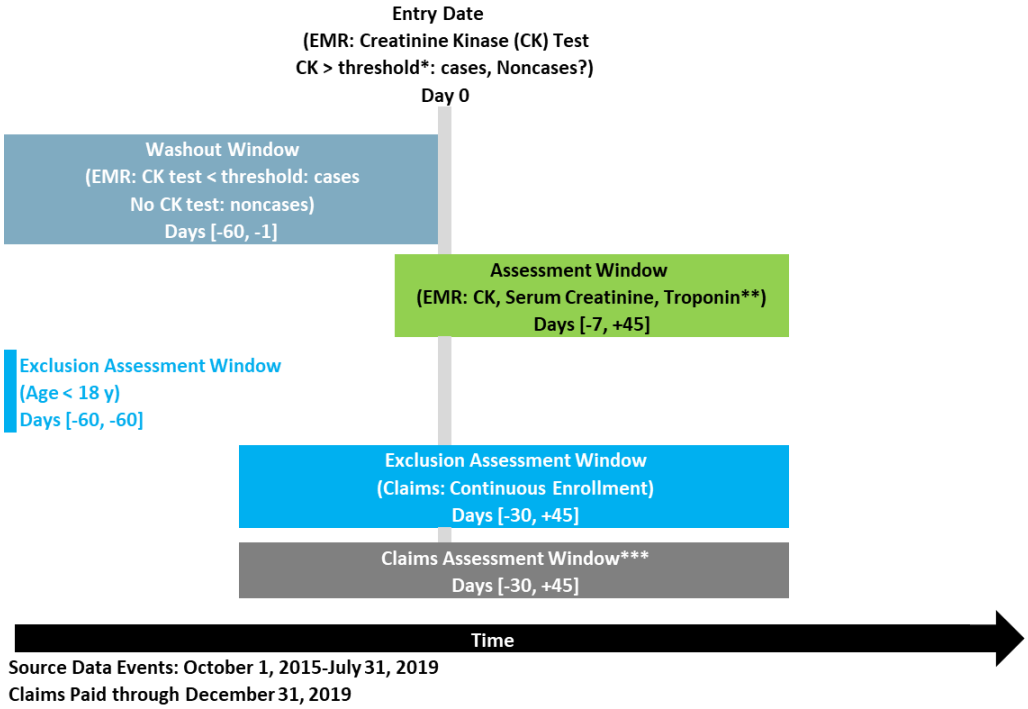
F. Study Design: Construction of Episodes

To construct a claims signature for rhabdomyolysis, instances of rhabdomyolysis are first identified in the EMR laboratory results data and a time window is placed around the rhabdomyolysis EMR events. Concurrent data within this event window are then extracted from the claims database to inform modeling efforts to discover a claims signature for rhabdomyolysis.

The longitudinal database design is displayed in Figure 2 and follows Schneeweiss et al. (2019).[62] First, data are extracted from the Explorys Claims-EMR Data Set with event dates from November 1, 2015, through July 31, 2019, coinciding with the introduction of ICD-10 codes in November 2015. Claims data contain paid dates through December 31, 2019. In each entry listed in the figure, the data source is specified: EMR for the EMR database and Claims for the claims databases (inpatient medical, outpatient medical, outpatient pharmacy, and enrollment data). This study anchors in patient event time, not calendar time.

The figure is read from top to bottom, starting with an episode-entry date based on the existence of a CK test in the EMR data. An attrition table (i.e., patient counts) will be constructed to coincide with the order of instructions in Figure 2. For cases, this date (Day 0) coincides with the first instance of a CK test exceeding 4 x ULN. To determine the anchor date, a washout window is applied. For cases, this must be the first instance of high CK with a washout window of 60 days prior (Days [–60, –1]). After cases are determined based on CK results, noncases are determined and the anchor date for this group is the first instance of a CK test below 4 x ULN. Noncases must not have a CK result reported within the same washout window (Days [–60, –1]).

Figure 2. Database Design and Episode Construction



* CK threshold is 4 x Upper Limit of Normal (ULN), ULN = 200 U/L
** Assess levels of laboratory test results to assign patient to Definite, Probable, Possible, Noncase
*** Gather potential, concurrent indications of rhabdomyolysis from claims data

Abbreviation: EMR, electronic medical record.

EMR laboratory test results starting 1 week before the anchor date and 45 days after the anchor date are examined to determine the rhabdomyolysis level (Table 1, e.g., Definite, Probable) for each patient. All levels of all relevant laboratory tests (CK, serum creatinine, and troponin) are examined in this time window (Days [–7, +45]) to determine the level. Troponin and serum creatinine will be coded as binary (indicating higher/lower than threshold) to assess the level of rhabdomyolysis. Although most occurrences of rhabdomyolysis are resolved in 6 to 10 days,[45,48] we allow for longer episodes by following experience in the EMR for a fixed 45 days following the index date. We look back 7 days to allow for different sites of care to report laboratory testing results in the EMR on dates that are relatively close in time.

After the level is determined, patients who are not adults during the time window (starting 60 days before) are excluded (Day [–60, –60]).

To ensure that complete claims data are available to study, an exclusion is added to ensure that all patients must be continuously enrolled in the claims database for the study time window of 30 days before the anchor date and 45 days after the anchor date (Days [–30, +45]). Finally, all medical and outpatient pharmacy claims are extracted during the study time window with claims data elements such as diagnosis codes, dates of service, procedure codes, place of service, and provider type available to determine the claims signature (Days [–30, +45]).

An attrition table with sample sizes will be produced. For sample size considerations, including statistical power of validation testing, see Section III.D.

III. Analysis Approach

A. Univariate and Bivariate Statistics

Univariate statistics are calculated for outcomes and features in the analytic file, including mean, median, percentile, standard deviation, skewness, and kurtosis for continuous variables. Histograms also are created for continuous variables. For categorical variables, frequency distributions of values are created. Of importance are features that have no variation, which should be removed before model construction. We also will count the number of episodes created per individual.

Bivariate statistics also are created. Pairwise correlations can identify very highly correlated features (e.g., those with a correlation of 0.95 or greater) that provide little independent information and may be combined or reduced before model construction.

In addition, contingency tables between rhabdomyolysis and each feature illustrate the unadjusted association between each feature and laboratory evidence of rhabdomyolysis. To assess differences in feature distributions between rhabdomyolysis cases and noncases, *t*-tests are performed for continuous features and chi-squared tests are performed for categorical features; these can be quickly implemented by regression modeling commands in statistical software.

In some scenarios, features with no predictive value are dropped. However, to maintain consistency across methods, we analyze all candidate features and predictors through the univariate and bivariate statistics.

B. Model Approaches

1. Candidate Models

Developing a claims-based signature for rhabdomyolysis is a classification problem in which the class label for episode *i*, y_i, is either 0 (noncase) or 1 (rhabdomyolysis case). Multiple episodes from the same patient are deduplicated when they contain the same outcomes and features, specifically clinical conditions, because they would provide redundant information for training the models. The data set can be denoted by D = {($\mathbf{x_1}$, y_1), ($\mathbf{x_2}$, y_2), …, ($\mathbf{x}_n$, y_n)}, where $\mathbf{x}_i$ are *m*-dimensional vectors of input variables or features. A classification algorithm generally attempts to model to the relationship between *y* and **x**—describing the relationship by a probability distribution $P(\mathbf{x}, y)$, which characterizes the joint probability of class membership and features. In implementation, the class label is based on a threshold of probability (e.g., $Pr(y_i/\mathbf{x}_i) \geq 0.50$ is defined as a case). The thresholds can be calibrated to a model performance measure, such as prediction error or similar diagnostic for misclassification (e.g., sensitivity or specificity). To determine the optimal thresholding rule, we apply cross-validation in the training set to determine thresholds that optimize model performance.

The most common machine learning classification models for dichotomous outcomes are logistic regression, decision trees (e.g., classification trees and random forests), support vector machines, and artificial neural networks (see Table 3).[63] For this collection of models, we will tune models in the training dataset to optimize their predictive performance on the scaled Brier score (see Section C.3).

2. Ensemble Methods

To leverage the predictive performance of all models, we apply SuperLearner, an ensemble-based method that weights model predictions from multiple candidate models in the ensemble and then yields a composited prediction that typically outperforms the prediction from any single model. The SuperLearner algorithm uses *k*-fold cross-validation to determine optimal weights that minimize prediction error.[64] To implement the ensemble, we incorporate the individual models (Table 3), which will have been tuned separately in the training set. The Super Learner is accordingly tuned on the same diagnostic on which the candidate models were separately tuned, that is, the scaled Brier score.

Table 3. Description of Candidate Models

MODEL	DESCRIPTION
Logistic regression	Logistic regression is characterized by a parameter/outcome relationship $P(y/\mathbf{x}) = f(\mathbf{x}, \beta)$, where *f* is the logistic function. Model parameters (β) are estimated by maximum likelihood estimation. Logistic regression typically is less complex than other methods and results in parameter estimates of main effects unless interaction terms are explicitly entered in the model. In comparison to other classification models, logistic regression presents a greater degree of interpretability, that is, to characterize associations between outcome and features.
LASSO logistic regression	LASSO regression applies optimization and shrinkage techniques to a regression model. The aim is to select the smallest number of predictors with the best performance; as such, feature selection occurs automatically as part of the model fitting process.

MODEL	DESCRIPTION
Classification tree	CART methods implement recursive partitioning to identify subclasses in which study units are homogeneous in the target variable (i.e., outcome), according to a prespecified objective function. Variants of CART conduct statistical tests at each node, such as conditional inference trees. Akin to logistic regression, a classification tree can offer interpretability, by showing how the dataset is recursively partitioned to form feature subclasses that are homogeneous in the outcome.
Random forest	This method is characterized by aggregation of predictions made by multiple decision trees of varying depth. Each tree is grown on a separate bootstrap resample of the training data; at each split the three chooses among a randomly selected subset of candidate features to help de-correlate the trees in the forest. This process also induces feature selection.
Support vector machines	This method maps the data onto a multidimensional "kernel space" within the feature space and uses optimization techniques to determine the location and coordinates of multidimensional hyperplanes that best separate observations from different outcome classes in Euclidean space.
Artificial neural networks	Neural networks are characterized by a semiparametric or nonparametric probability model (i.e., *f* is not known). The output is a nonlinear function of features. As an extension of the conventional regression framework which adopts only an input and output layer, a neural network contains one or more "hidden" layers between those layers so that the network can automatically model more complex, nonlinear relationships. A "deep model" has at least two hidden layers.

Abbreviation: CART, Classification and Regression Trees; LASSO, Least Absolute Shrinkage and Selection Operator.

We will enter all input features into each model and estimate models as is feasible. We will produce a list of features by importance as produced in each model. For the logistic model, this will be based on the coefficient absolute value. For the support vector machine with a linear kernel, feature importance can be determined by the relative coordinates (weights) of the vector orthogonal to the hyperplane. For artificial neural networks, the relative importance of a feature can be characterized by the sum of weighted connections between relevant nodes; additionally, the artificial neural network and its weighted connections can be visualized in an interpretation diagram, which shows connections between layers that define feature importance.

C. Model Training

1. Data Splitting

Generally, low prevalence of an HOI in claims data poses challenges for standard data-splitting approaches—that is, into training and testing sets and for cross-validation. Data splitting based on stratified sampling can ensure that subsets contain exactly the same proportion of rhabdomyolysis cases and noncases—that is, class balancing. Specifically, we apply random sampling within each stratum defined by the class

First, we will split the study data into training and testing sets in a random 80/20 split. Model training is based on the training set, an 80-percent sample of the study data; we will tune the model hyperparameters (described below) through k-fold cross-validation. Specifically, we use k-fold cross-validation in the training set to prevent or minimize the risk of overfitting when fitting our models. We will explore 5- and 10-fold cross-validation; the folds will be randomly sampled from the training set with class balancing on cases and noncases.

As the "hold out" sample, the testing set is used to assess the model's generalization error, that is, performance on an unseen 20-percent sample that was not used for training. Using an independent testing set in this way, we will evaluate whether we have indeed overfit our final models to the training set. In particular, reflecting a real-world test setting, it will not be balanced on rhabdomyolysis classes (i.e., it will not necessarily have the same proportion of cases/noncases as the training set).

2. Hyperparameters

Most machine learning models require tuning of hyperparameters. Hyperparameters represent higher-level concepts whose values are specified prior to the model-fitting process and affect the structure and complexity (and thus potential performance) of machine learning models.

Table 4. Hyperparameters

METHOD	R PACKAGE/PROCEDURE	HYPERPARAMETERS
Logistic regression	glm	N/A
LASSO logistic regression	glmnet	λ—regularization term
Classification tree; random forest	randomForest	Mtry—number of variables randomly samples as candidates at each split Ntree—number of trees to grow
Support vector machines	kernlab	C—regularization term Polynomial degree—kernel specification Sigma—radial basis function kernel Scale Tau (for regularization)
Artificial neural networks	nnet	Size—number of hidden units Decay—weight decay

Abbreviation: LASSO, Least Absolute Shrinkage and Selection Operator; N/A, not applicable.

We describe the main hyperparameters for each machine learning method in Table 4. To identify the optimal value for each hyperparameter (i.e., "tune" the hyperparameters), we will implement a grid-search procedure (described later)—an approach commonly used to perform hyperparameter tuning.

3. Diagnostics

To tune the machine learning models (during training) and assess the performance of the final models (during testing), a loss (or objective) function must be chosen over which model performance will be evaluated. For rhabdomyolysis classification, accuracy is defined by the predicted class from the models, which generally yield predictions on the probability scale. To better account for the probability scale in assessing model performance for hyperparameter tuning (including threshold probabilities) in the training set, we propose a scaled Brier score as the primary performance diagnostic.[65] To test the performance of the final model, we will apply bootstrapping procedures. Specifically, we will calculate confidence intervals on the estimated positive predicted value (PPV); these confidence intervals capture statistical uncertainty induced by the testing set selection.

The final model, or ensemble, will be evaluated in the testing set, using an expanded list of diagnostic measures (Table 5). The result of the claims-based signature (indicating rhabdomyolysis or no rhabdomyolysis) is compared with the rhabdomyolysis test results (case/noncase). The same set of

performance diagnostics are used between Scenarios 1 and 2 to assess generalization error on the testing set; specifically, we focus on positive predictive value for model performance. Sensitivity analysis to determine the suitability of Scenarios 1 and 2 is based on the false negative and false positive rates (which are the additive inverses of sensitivity and specificity, specifically).

Table 5. Statistic Measure Definitions

MEASURE/STATISTIC	DEFINITION
Prediction error	Proportion of patients whose claims-based signature (positive or negative for rhabdomyolysis) does not agree with clinical determination (from lab test values)
Positive predictive value	Proportion of patients with a claims-based signature indicating rhabdomyolysis who have rhabdomyolysis
Negative predictive value	Proportion of patients with a claims-based signature indicating that the case is not rhabdomyolysis who do not have rhabdomyolysis
Sensitivity	Proportion of patients with rhabdomyolysis who have a claims-based signature indicating rhabdomyolysis
Specificity	Proportion of patients without rhabdomyolysis who have a claims-based signature indicating that the case is not rhabdomyolysis

4. Hyperparameter Tuning

We adopt grid search to optimize hyperparameter values and probability thresholds. In particular, the scaled Brier score and classification accuracy will be estimated over a multidimensional grid of hyperparameter values through cross-validation. Specifically, the training set first will be split into k-folds (using class balancing techniques, described above), and then each model will be run through cross-validation using each unique combination of values defined by the model's hyperparameter grid and probability threshold range. We will determine the hyperparameter values and probability thresholds that optimize the predictive performance of the individual models and eventual ensemble.

D. Model Testing and Validation

Over the testing set, we will calculate the performance diagnostics in the expanded list of Table 6. Splitting testing from training ensures independence between data that were used to estimate the candidate models and those that will be used to assess its performance. To understand how model performance is affected by hyperparameter values, we will analyze the association between diagnostics and the candidate values (i.e., specified by the hyperparameter grid). Although cross-validation can mitigate some concerns about overfitting, the final model's performance on the testing set will reveal whether these concerns remain.

1. Predictions in Testing Set

Models are trained on 80 percent of the target cohort from the MarketScan Explorys Claims-EMR Data Set. Using the final model, we will calculate predicted class probabilities in the 20-percent testing set, which have been set aside for this purpose. Each episode in the testing set is assigned to an HOI class—that is, rhabdomyolysis case or noncase—using the threshold probability. These predicted classes are compared against the true class determined by laboratory test values under each scenario.

2. Overfitting

One concern in model training is overfitting. In large part, we can preempt this concern by ensuring that final model performance is assessed on the holdout testing set and not on the training data that are

used to build the model. Overfitting also can be addressed by regularization methods that aim to stabilize model parameter estimates at the expense of some bias. To this end, we adopt LASSO regression in the final ensemble (Table 4).

3. Statistical Inference

To test the performance of the final model, we will apply bootstrapping procedures. Specifically, we will calculate confidence intervals on the estimated PPV; these confidence intervals capture statistical uncertainty induced by the testing set selection using the 80/20 random split. Statistical power to assess model performance on the estimated PPV is an inherent property of the given sample size (the 20-percent testing set), the underlying signal (true positive cases among estimated cases), and the significance level (control of Type I error). We will calculate the statistical power to detect model performance above a threshold PPV value using conventional tests of binomial proportions.

Table 6. Positive Predictive Value Calculation

CLAIMS ALGORITHM	ELECTRONIC HEALTH RECORDS CASES		ROW TOTAL
	POSITIVE	NEGATIVE	
Positive	y^+	y^-	n^+_{test}
Negative	z^+	z^-	n^-_{test}
Column total	n^+_{ref}	n^-_{ref}	n

Table 6 reflects the underlying data to calculate the estimated PPV from the predicted and true classes across episodes in the testing set. Specifically, estimated PPV is calculated as follows:

$$\widehat{PPV} = \frac{y^+}{y^+ + y^-}$$

We obtain a bootstrap sample of the testing set using random sampling with replacement, whereby the probability of selection is the inverse of the testing set size (i.e., if there are N episodes in the testing set, the probability is 1/N). Five hundred (B=500) bootstrap samples will be created. In each bootstrap sample, we will apply the final model to determine the predicted class for each episode and compare the prediction with the true class from the EMR laboratory values. A positive rhabdomyolysis case that is assigned to positive predicted HOI class will contribute to the count y^+, whereas those with a negative predicted class will contribute to z^+. Rhabdomyolysis noncases assigned to a negative predicted class will contribute to z^-, whereas those assigned to a positive predicted class will contribute to y^-. The PPV will be estimated in each bootstrap sample; over the 500 bootstrap iterations, the 2.5th and 97.5th percentile values of the estimated PPVs will be used to calculate the 95-percent confidence interval around the estimated PPV in the testing set.

IV. Reporting of Results

Results are reported in tables and figures. This section includes sample table shells (Tables 7–10) and figure descriptions for reporting.

A. Table Shells

Table 7. Sample Table for Bivariate Analysis, Scenario X (X=1 or 2)

VARIABLE	RHABDOMYOLYSIS	NO RHABDOMYOLYSIS	p-VALUE
Number			
Female, percent			
Age, years, average (std)			
Age categories, years, percent			
18–34			
35–44			
45–54			
55–64			
65–74			
75-84			
85+			
ICD-10-CM diagnosis of rhabdomyolysis, percent			
Symptom of muscle pain, percent			
Dialysis procedures, percent			
Additional variables created in file build			

Table 8. Model Comparisons, Performance Metrics

MODEL*	PPV	NPV	SENSITIVITY	SPECIFICITY	AUC
Logistic regression					
LASSO logistic regression					
Random forest					
Support vector machines					
Artificial neural networks					
Super Learner					

Abbreviations: AUC, area under the curve; LASSO, Least Absolute Shrinkage and Selection Operator; NPV, negative predictive value; PPV, positive predictive value.

*Models will be estimated as is feasible.

Table 9. Sample Output of Model Estimation, by Machine Learning Model

PARAMETER ESTIMATE	S.E.	Z-VALUE	Pr(>|Z|)

Table 10. Feature Importance by Machine Learning Model, Example

MACHINE LEARNING MODEL (E.G., LASSO LOGISTIC REGRESSION)	FEATURES, RANKED BY IMPORTANCE

Abbreviation: LASSO, Least Absolute Shrinkage and Selection Operator.

B. Figure Descriptions

Figures—Model Calibration Plots for Each Machine Learning Method

Model calibration plots divide predicted probabilities into 10 bins (e.g., 0–0.1) and plot the mean actual probability of rhabdomyolysis for each bin on the y-axis. The 45-degree line depicts perfect prediction.

Figures—Receiver Operating Characteristic Curves for Each Machine Learning Method

V. Next Steps and Timeline

After the protocol, next steps include execution of the machine learning analysis. Manuscript and project completion are set for July 2020.

VI. References

1. James G, Witten D, Hastie T, Tibshirani R. Introduction to Statistical Learning With Applications in R. New York, NY: Springer; 2013.
2. Parikh RB, Manz C, Chivers C, et al. Machine learning approaches to predict 6-month mortality among patients with cancer. *JAMA Netw Open*. 2019;2(10):e1915997.
3. Rawson ES, Clarkson PM, Tarnopolsky MA. Perspectives on exertional rhabdomyolysis. *Sports Med*. 2017;47(Suppl 1):33-49.
4. Alaygut D, Torun Bayram M, Kasap B, et al. Rhabdomyolysis with different etiologies in childhood. *World J Clin Pediatr*. 2017;6(4):161-168.
5. Ohima Y. Characteristics of drug-associated rhabdomyolysis: analysis of 8,610 cases reported to the U.S. Food and Drug Administration. *Intern. Med.* 2011;50(8):845-853.
6. Tobert JA. Efficacy and long-term adverse effect pattern of lovastatin. *Am J Cardio*l. 1988; 62:28J.
7. Pierce LR, Wysowski DK, Gross TP. Myopathy and rhabdomyolysis associated with lovastatin-gemfibrozil combination therapy. *JAMA.* 1990; 264:71.
8. Grundy SM. HMG-CoA reductase inhibitors for treatment of hypercholesterolemia. *N Engl J Med.* 1988; 319:24.

9. Keogh A, Macdonald P, Kaan A, et al. Efficacy and safety of pravastatin vs simvastatin after cardiac transplantation. *J Heart Lung Transplant.* 2000; 19:529.
10. Yee HS, Fong NT. Atorvastatin in the treatment of primary hypercholesterolemia and mixed dyslipidemias. *Ann Pharmacother.* 1998; 32:1030.
11. Patel DN, Pagani FD, Koelling TM, et al. Safety and efficacy of atorvastatin in heart transplant recipients. *J Heart Lung Transplant.* 2002; 21:204.
12. Ayanian JZ, Fuchs CS, Stone RM. Lovastatin and rhabdomyolysis. *Ann Intern Med.* 1988; 109:682.
13. Lees RS, Lees AM. Rhabdomyolysis from the coadministration of lovastatin and the antifungal agent itraconazole. *N Engl J Med*. 1995; 333:664.
14. Knopp RH. Drug treatment of lipid disorders. *N Engl J Med.* 1999; 341:498.
15. Fichtenbaum CJ, Gerber JG, Rosenkranz SL, et al. Pharmacokinetic interactions between protease inhibitors and statins in HIV seronegative volunteers: ACTG Study A5047. *AIDS*. 2002; 16:569.
16. Rosenson RS, Colantonio LD, Burkholder GA, et al. Trends in Utilization of Statin Therapy and Contraindicated Statin Use in HIV--Infected Adults Treated With Antiretroviral Therapy From 2007 Through 2015. *J Am Heart Assoc.* 2018; 7:e010345.
17. Baker SK, Goodwin S, Sur M, Tarnopolsky MA. Cytoskeletal myotoxicity from simvastatin and colchicine. *Muscle Nerve.* 2004; 30:799.
18. Penzak SR, Chuck SK. Management of protease inhibitor-associated hyperlipidemia. *Am J Cardiovasc Drugs.* 2002; 2:91.
19. van der Lee M, Sankatsing R, Schippers E, et al. Pharmacokinetics and pharmacodynamics of combined use of lopinavir/ritonavir and rosuvastatin in HIV-infected patients. *Antivir Ther.* 2007; 12:1127.
20. NIH DailyMed. Livalo (pitavastatin) US prescribing information. http://dailymed.nlm.nih.gov/dailymed/lookup.cfm?setid=b3c41a2f-da77-4465-8af6-038006a7ea58
21. FDA Drug Safety Communication: Interactions between certain HIV or hepatitis C drugs and cholesterol-lowering statin drugs can increase the risk of muscle injury http://www.fda.gov/Drugs/DrugSafety/ucm293877.htm
22. Wiggins BS, Lamprecht DG Jr, Page RL 2nd, Saseen JJ. Recommendations for Managing Drug-Drug Interactions with Statins and HIV Medications. *Am J Cardiovasc Drugs.* 2017; 17:375.
23. FDA.www.fda.gov/cder/drug/InfoSheets/HCP/simvastatin_amiodaroneHCP.htm
24. Reddy P, Ellington D, Zhu Y, et al. Serum concentrations and clinical effects of atorvastatin in patients taking grapefruit juice daily. *Br J Clin Pharmacol.* 2011; 72:434.
25. Holdaas H, Hagen E, Asberg A, et al. Evaluation of the pharmacokinetic interaction between fluvastatin XL and cyclosporine in renal transplant recipients. *Int J Clin Pharmacol Ther*. 2006; 44:163.
26. Kantola T, Backman JT, Niemi M, et al. Effect of fluconazole on plasma fluvastatin and pravastatin concentrations. *Eur J Clin Pharmacol.* 2000; 56:225.

27. NIH DailyMed. Crestor (rosuvastatin) US prescribing information (revision 12/2012) http://dailymed.nlm.nih.gov/dailymed/lookup.cfm?setid=bb0f3b5e-4bc6-41c9-66b9-6257e2513512
28. Holdaas H, Hartmann A, Stenstrøm J, et al. Effect of fluvastatin for safely lowering atherogenic lipids in renal transplant patients receiving cyclosporine. *Am J Cardiol.* 1995; 76:102A.
29. Glueck CJ, Oakes N, Speirs J, et al. Gemfibrozil-lovastatin therapy for primary hyperlipoproteinemias. *Am J Cardiol.* 1992; 70:1.
30. Duell PB, Connor WE, Illingworth DR. Rhabdomyolysis after taking atorvastatin with gemfibrozil. *Am J Cardiol*. 1998; 81:368.
31. Alayli G, Cengiz K, Cantürk F, et al. Acute myopathy in a patient with concomitant use of pravastatin and colchicine. *Ann Pharmacother.* 2005; 39:1358.
32. Kotanko P, Kirisits W, Skrabal F. Rhabdomyolysis and acute renal graft impairment in a patient treated with simvastatin, tacrolimus, and fusidic acid. *Nephron*. 2002; 90:234.
33. Wenisch C, Krause R, Fladerer P, et al. Acute rhabdomyolysis after atorvastatin and fusidic acid therapy. *Am J Med.* 2000; 109:78.
34. Yuen SL, McGarity B. Rhabdomyolysis secondary to interaction of fusidic acid and simvastatin. *Med J Aust*.2003; 179:172.
35. Magee CN, Medani SA, Leavey SF, et al. Severe rhabdomyolysis as a consequence of the interaction of fusidic acid and atorvastatin. *Am J Kidney Dis*. 2010; 56:e11.
36. O'Mahony C, Campbell VL, Al-Khayatt MS, Brull DJ. Rhabdomyolysis with atorvastatin and fusidic acid. *Postgrad Med J*. 2008; 84:325.
37. Burtenshaw AJ, Sellors G, Downing R. Presumed interaction of fusidic acid with simvastatin. *Anaesthesia*. 2008; 63:656.
38. Kearney S, Carr AS, McConville J, et al. Rhabdomyolysis after co-prescription of statin and fusidic acid. *BMJ*. 2012; 345:e6562.
39. Karas RH, Kashyap ML, Knopp RH, et al. Long-term safety and efficacy of a combination of niacin extended release and simvastatin in patients with dyslipidemia: the OCEANS study. *Am J Cardiovasc Drugs*. 2008; 8:69.
40. FDA. FDA Drug Safety Communication: Important safety label changes to cholesterol-lowering statin drugs. https://www.fda.gov/drugs/drug-safety-and-availability/fda-drug-safety-communication-important-safety-label-changes-cholesterol-lowering-statin-drugs
41. FDC Reports. "The Pink Sheet". 2002; 64(vol 23):5.
42. NIH DailyMed. Zocor (simvastatin) US prescribing information (revision 10/2012). http://dailymed.nlm.nih.gov/dailymed/lookup.cfm?setid=fdbfe194-b845-42c5-bb87-a48118bc72e7
43. Huerta-Alardín AL, Varon J, Marik PE. Bench-to-bedside review: rhabdomyolysis—an overview for clinicians. *Crit Care*. 2004;9(2):158-169.
44. Rosenson RS, Baker SK, Jacobson TA, Kopecky SL, Parker BA. An assessment by the Statin Muscle Task Force: 2014 update. *J Clin Lipidol*. 2014;8(3, Suppl):S58-71.

45. Knafl EG, Hughes JA, Dimeski G, Eley R. Rhabdomyolysis: patterns, circumstances, and outcomes of patients presenting to the emergency department. *Ochsner J*. 2018;18(3):215-221.

46. Alfirevic A, Neely D, Armitage J, et al. Phenotype standardization for statin-induced myotoxicity. *Nature*. 2014;96(4):470-476.

47. El-Abdellati E, Eyselbergs M, Sirimsi H, et al. An observational study on rhabdomyolysis in the intensive care unit. Exploring its risk factors and main complication: acute kidney injury. *Ann Intensive Care.* 2013;3(1):8.

48. Floyd JS, Heckbert SR, Weiss NS, Carrell DS, Psaty BM. Use of administrative data to estimate the incidence of statin-related rhabdomyolysis. *JAMA*. 2012;307(15):1580-1582.

49. Keltz E, Khan FY, Mann G. Rhabdomyolysis: the role of diagnostic and prognostic factors. *Muscles Ligaments Tendons J*. 2014;3(4):303-312.

50. Chavez LO, Leon M, Einav S, Varon J. Beyond muscle destruction: a systematic review of rhabdomyolysis for clinical practice. *Crit Care*. 2016;20(1):135.

51. Khan FY. Rhabdomyolysis: a review of the literature. *Neth J Med.* 2009;67:272-283.

52. Nance JR, Mammen AL. Diagnostic evaluation of rhabdomyolysis. *Muscle Nerve*. 2015;51(6):793-810.

53. Huerta-Alardín AL, Varon J, Marik PE. Bench-to-bedside review: rhabdomyolysis—an overview for clinicians. *Crit Care*. 2004;9(2):158-169.

54. Torres PA, Helmstetter JA, Kaye AM, Kaye AD. Rhabdomyolysis: pathogenesis, diagnosis, and treatment. *Ochsner J*. 2015;15(1):58-69.

55. Ruan XC, Low LL, Kwan YH. Association of aspartate aminotransferase in statin-induced rhabdomyolysis. *J Prim Healthcare*. 2017;9(4):316-320.

56. Fernandes PM, Davenport RJ. How to do it: investigate exertional rhabdomyolysis (or not). *Pract Neurol*. 2019;19(1):43-48.

57. Mishra A, Dave N. Acute renal failure due to rhabdomyolysis following a seizure. *J Family Med Prim Care*. 2013;2(1):86-87.

58. Petejova N, Martinek A. Acute kidney injury due to rhabdomyolysis and renal replacement therapy: a critical review. *Crit Care*. 2014;18(3):224.

59. Veenstra J, Smit WM, Krediet RT, Arisz L. Relationship between elevated creatine phosphokinase and the clinical spectrum of rhabdomyolysis. *Nephrol Dial Transplant*. 1994;9(6):637-641.

60. Mythili S, Malathi N. Diagnostic markers of acute myocardial infarction. *Biomed Rep*. 2015;3(6):743-748.

61. Laurence AS. Serum myoglobin and creatine kinase following surgery. *Br J Anaesth.* 2000;84(6):763-766.

62. Schneeweiss S, Rassen JA, Brown JS, et al. Graphical depiction of longitudinal study designs in health care databases. Ann Intern Med. 2019;170(6):398-407.

63. Lustgarten JL, Gopalakrishnan V, Grover, H, Visweswaran S. Improving classification performance with discretization on biomedical datasets. *AMIA Annu Symp Proc.* 2008;2008:445-449.

64. van der Laan MJ, Polley EC, Hubbard AE. Super learner. *Stat Appl Genet Mol Biol*. 2007;6;Article 25.

65. Wong J, Manderson T, Abrahamowicz M, Buckeridge DL, Tamblyn R. Can hyperparameter tuning improve the performance of a super learner? A case study. *Epidemiology* 2019;30(4):521-531.

Appendix A. LOINC Reference

LABORATORY TEST	LOINC CODE	LOINC CODE DEFINITION
Creatine kinase	50757-4	Creatine kinase.total/Creatine kinase.MB [Enzymatic activity ratio] in Blood
Creatine kinase	2157-6	Creatine kinase [Enzymatic activity/volume] in Serum or Plasma
Serum creatinine	2160-0	Creatinine [Mass/volume] in Serum or Plasma
Troponin	10839-9	Troponin I.cardiac [Mass/volume] in Serum or Plasma
Troponin	6598-7	Troponin T.cardiac [Mass/volume] in Serum or Plasma
Troponin	67151-1	Troponin T.cardiac [Mass/volume] in Serum or Plasma by High sensitivity method

Abbreviation: LOINC, Logical Observation Identifiers Names and Codes.

Appendix B. Common Data Model Predictors

ADMINISTRATIVE	
Demographic	
Age	
Sex	
Male, female	
Encounter	
Encounter type	
Inpatient	
Emergency Department	
Diagnosis	
Description	**ICD-10 Code**
Acidosis	ICD-10 E87.2
Acute kidney injury	ICD-10 N17.9
Asthenia	ICD-10 R53.1
Fever	ICD-10 R50.9
Hyperkalemia	ICD-10 E87.5
Hyperphosphatemia	ICD-10 E83.39
Hypocalcemia	ICD-10 E83.51
Hypovolemia	ICD-10 E86.1
Malignant hyperthermia due to anesthesia	ICD-10 T88.3XXX
Malignant neuroleptic syndrome	ICD-10 G21.0
Muscle weakness (generalized)	ICD-10 M62.81
Myalgia	ICD-10 M79.1X
Myoglobinuria	ICD-10 R82.1
Nausea	ICD-10 R11.XX
Rhabdomyolysis	ICD- 10 M62.82
Tachycardia	ICD-10 R00.0
Unspecified adverse effect of drug or medicament	ICD-10 T88.7XXX
Vomiting	ICD-10 R11.XX
Procedures	
Description	CPT Code
Creatine kinase (CK), (CPK); total	CPT 82550
Creatine kinase (CK), (CPK); isoenzymes	CPT 82552
Creatine kinase (CK), (CPK); isoforms	CPT82554
Calcium; total	CPT 82310
Calcium; ionized	CPT 82330
Creatine kinase (CK), (CPK); MB fraction only	CPT 82553

Creatinine; blood	CPT 82565
Myoglobin	CPT 83874
Phosphorus inorganic (phosphate);	CPT 84100
Potassium; serum, plasma or whole blood	CPT 84132
Transferase; aspartate amino (AST) (SGOT)	CPT 84450
Transferase; alanine amino (ALT) (SGPT)	CPT 84460
Urea nitrogen; quantitative	CPT 84520
Uric acid; blood	CPT 84550
Urinalysis, by dip stick or tablet reagent for bilirubin, glucose, hemoglobin, ketones, leukocytes, nitrite, pH, protein, specific gravity, urobilinogen, any number of these constituents; non-automated, with microscopy	CPT 81000
Urinalysis, by dip stick or tablet reagent for bilirubin, glucose, hemoglobin, ketones, leukocytes, nitrite, pH, protein, specific gravity, urobilinogen, any number of these constituents; automated, with microscopy	CPT 81001
Urinalysis, by dip stick or tablet reagent for bilirubin, glucose, hemoglobin, ketones, leukocytes, nitrite, pH, protein, specific gravity, urobilinogen, any number of these constituents; non-automated, without microscopy	CPT 81002
Urinalysis, by dip stick or tablet reagent for bilirubin, glucose, hemoglobin, ketones, leukocytes, nitrite, pH, protein, specific gravity, urobilinogen, any number of these constituents; automated, without microscopy	CPT 81003
Hemodialysis	CPT 90935–90999
Intravenous infusion, for therapy, prophylaxis, or diagnosis (specify substance or drug); initial, up to 1 hour	CPT 96365
Intravenous infusion, for therapy, prophylaxis, or diagnosis (specify substance or drug); each additional hour (List separately in addition to code for primary procedure)	CPT 96366
Critical care, evaluation and management	CPT 99291, 99292
Office or other outpatient visit, new patient	CPT 99201 - 99205
Office or other outpatient visit, established	CPT 99211 - 99215
Medications[1]	
amiodarone	
aprepitant	
atazanavir	
ceritinib	
cimetidine	
clarithromycin	

[1] Medications listed for completeness, not all medications are marketed in the US.

cobicistat and cobicistat-containing coformulations
colchicine
conivaptan
crizotinib
cyclosporine
darunavir
diltiazem
duvelisib
dronedarone
erythromycin
fedratinib
fluconazole
fosamprenavir
fosaprepitant
fusidic acid
gemfibrozil
idelalisib
imatinib
indinavir
itraconazole
isavuconazole (isavuconazonium sulfate)
ketoconazole
lefamulin
letermovir
lopinavir
mifepristone
nefazodone
nelfinavir
netupitant
niacin
nilotinib
ombitasvir-paritaprevir-ritonavir
ombitasvir-paritaprevir-ritonavir plus dasabuvir
posaconazole
ribociclib
ritonavir and ritonavir-containing coformulations
saquinavir
schisandra

telithromycin		
verapamil		
voriconazole		
CLINICAL		
Laboratory Results		
Laboratory Test	**LOINC**	**Result**
Creatine Kinase, Total	2157-6	Abnormally high
Creatine Kinase MM		
CK-MM	15049-0	Abnormally high
CK-MM	15049-0	Abnormally high
Urine Color		
Urine-Color	5778-6	Abnormal
Color Ur Auto	50553-7	Abnormal
Urine, Protein		
Urine Protein	20454-5	Abnormally high
Prot Ur-mCnc	2888-6	Abnormally high
Prot Ur Strip. Auto-mCnc	50561-0	Abnormally high
Prot Ur Strip-mCnc	5804-0	Abnormally high
Prot Ur Ql strip. Auto	57735-3	Abnormally high
Urine, Microscopic Examination	12235-8	Abnormal
Urine, Blood Dip		
RBC # Ur Strip	20409-9	Abnormally high
Occult Blood	5794-3	Abnormally high
Hgb Ur Ql Strip. Auto	57751-0	Abnormally high
Urine, RBC Micro		
RBC UrnS Ql Micro	32776-7	Normal
RBC # UrnS HPF	5808-1	Normal
Calcium	17861-6	Abnormally low
Calcium, ionized		
Calcium, Ionized, Serum	17864-0	Abnormally low
Ca-I SerPl-mCnc	17863-2	Abnormally low
Ca-I SerPl-sCnc	1995-0	Abnormally low
Ca-I SerPl ISE-sCnc	12180-6	Abnormally low
Ca-I SerPl Calc-sCnc	13959-2	Abnormally low
Creatine Kinase, MB		
Creatine Kinase (CK), MB	13969-1	Normal
CK MB CFr SerPl Elph	12187-1	Normal
CK MB CFr SerPl Calc	12189-7	Normal

Creatinine	2160-0	Abnormally high
Myoglobin, Serum	2639-3	Abnormally high
Phosphorous	2777-1	Abnormally high
Potassium	2823-3	Abnormally high
Aspartate aminotransferase (AST) (serum glutamic-oxaloacetic transaminase [SGOT])		
AST SerPl w P-5'-P-cCnc	30239-8	Abnormally high
AST (SGOT)	1920-8	Abnormally high
Alanine aminotransferase (ALT) (serum glutamic pyruvic transaminase [SGPT])		
ALT SerPl w P-5'-P-cCnc	1743-4	Abnormally high
ALT SerPl w/o P-5'-P-cCnc	1744-2	Abnormally high
ALT (SGPT)	1742-6	Abnormally high
Blood Urea Nitrogen (BUN)		
BUN SerPl-sCnc	14937-7	Abnormally high
BUN	3094-0	Abnormally high
Uric Acid	3084-1	Abnormally high
OTHER NOT INCLUDED IN SENTINEL COMMON DATA MODEL		
Provider Type		
Nephrologist		
Hospitalist		
Critical Care/Intensive Care		
Pulmonologist		
Emergency Medicine		

Abbreviations: CK, creatine kinase; CPT, Current Procedural Terminology; ICD-10, International Classification of Diseases, Tenth Revision; LOINC, Logical Observation Identifiers Names and Codes; RBC, red blood cell.

Appendix C. Machine Learning Definitions

This appendix provides definitions of key machine learning and statistical terms used throughout the protocol.

TERM	DEFINITION
Bootstrap clustering	A method of unsupervised machine learning method. This common method groups data together into similar subgroups.[a]
Decision tree	Decision tree is a machine learning method that follows a tree-structured classification scheme. Nodes represent the input variables, and the leaves correspond to decision outcomes. Decision tree is one of the earliest and most prominent machine learning methods that has been widely applied for classification purposes. The decisions resulting from its specific architecture allow for adequate reasoning, which makes it an appealing technique.[c]
Feature	Features are the input variables to a machine learning model. For example, when a model is being developed to predict stroke risk, a feature would be a patient's height or weight. Features can be processed before they are entered into a model—for example, height and weight can be combined into a body mass index.[a]
k-fold cross-validation	A method to minimize overfitting when making modeling decisions. This technique uses multiple splits within the development data set to reduce the effects of randomness of the split. For example, if k=2, the development set is split evenly into A and B. Two models are developed: one trained using A and tuned on B and one trained using B and tuned on A. The cross-validated evaluation is usually the average of the two performance estimates using A and B. This approach attempts to use the training set to estimate the generalization error. An independent validation set should be used to evaluate the performance of the final model trained on the entire development set. A leave-one-out cross-validation occurs when k is the total number of data points in the data set.[a]
LASSO	Least Absolute Shrinkage and Selection Operator (LASSO) is a type of penalization regression helps to prevent parameters from becoming too large (shrinkage) and thus overfitting. LASSO has the advantage of incorporating feature selection into the model fitting process, which is helpful in determining the most important input features.[a]
Logistic regression	Logistic regression is a classical statistical method often used in classification models for a dichotomous outcome (HOI/No HOI). For each risk factor, logistic regression determines the relationship between parameters, which are numerical values and binary clinical outcomes such as the presence or absence of a disease entity.[a]
Machine learning	Machine learning refers to the process of developing systems with the ability to learn from and make predictions using data. For example, a machine learning model can process an input (such as a retinal fundus photograph) and produce an output (such as the classification of the image showing that proliferative diabetic retinopathy is present).[a]
Positive predictive value	Positive predictive value is the probability that subjects with a positive model prediction truly have the condition of interest.[e]

TERM	DEFINITION
Predictors	Predictors are variables that are studied for their potential value in explaining the distribution of the outcome in the population. Predictors can include any information that precedes the outcome of interest in time and is believed to predict the outcome of interest. Examples include demographic variables, clinical history, physical examination findings, type and severity of disease, comorbid conditions, and laboratory or imaging results.[b]
Negative predictive value	Negative predictive value is the probability that subjects with a negative model prediction do not have the condition of interest.[e]
Sensitivity	Sensitivity is a model prediction's probability of correctly identifying, solely from among people who are known to have a condition, all those who do indeed have that condition and, at the same time, not categorizing other people as not having the condition when in fact they do have it (i.e., avoiding false negatives).[d]
Separating hyperplanes (support vector machines)	Support vector machines is a form of machine learning that recognizes patterns within data sets to build classifiers. It aims to create a decision boundary between two classes that enables the prediction of labels from one or more feature vectors. This decision boundary, known as the *hyperplane,* is orientated in such a way that it is as far as possible from the closest data points from each of the classes. These closest points are called *support vectors*.[e]
Specificity	Specificity is a model prediction's probability of correctly identifying, solely from among people who are known not to have a condition, all those who do not have that condition (i.e., identifying true negatives), and, at the same time, not categorizing some people as having the condition when in fact they do not have it (i.e., avoiding false positives). [d]
Supervised learning	Supervised learning is a type of machine learning in which labeled data (case/noncase assignment from the EMR in this study) are used for machine learning model development.[a]
Training data set	A training data set uses machine learning to learn parameters. It is the learning model to best match the model output with the reference.[a]
Validation data set	The validation data set is a subset of the development set that is used to tune the hyperparameters of a model. In medical research, a model must be validated using a data set that is completely independent of the training set.[a]

Abbreviation: EMR, electronic medical record; HOI, health outcome of interest.

[a] Liu Y, Chen P, Krause J, Peng L. How to read articles that use machine learning users' guides to the medical literature. *JAMA*. 2019;322(18):1806-1816.

[b] Shipe ME, Deppen SA, Farjah F, Grogan EL. Developing prediction models for clinical use using logistic regression: an overview. *J Thorac Dis*. 2019;11(Suppl 4):S574-S584.

[c] Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J*. 2014;13:8-17.

[d] Trevethan R. Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice. *Front Public Health*. 2017;5:307.

[e] Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W. Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genomics Proteomics*. 2018;15(1):41-51.