Prediction of Neutropenic Events in Chemotherapy Patients: A Machine Learning Approach

Holly Wiberg, BS1; Peter Yu, MD2; Pat Montanaro, MBA, BSN2; Jeff Mather, MS2; Suzi Birz, MSc2; Michelle Schneider, RN2; and Dimitris Bertsimas, PhD1,3

abstract

PURPOSE Severe and febrile neutropenia present serious hazards to patients with cancer undergoing chemotherapy. We seek to develop a machine learning–based neutropenia prediction model that can be used to assess risk at the initiation of a chemotherapy cycle.

MATERIALS AND METHODS We leverage rich electronic medical records (EMRs) data from a large health care system and apply machine learning methods to predict severe and febrile neutropenic events. We outline the data curation process and challenges posed by EMRs data. We explore a range of algorithms with an emphasis on model interpretability and ease of use in a clinical setting.

RESULTS Our final proposed model demonstrates an out-of-sample area under the receiver operating characteristic curve of 0.865 (95% CI, 0.830 to 0.891) in the prediction of neutropenic events on the basis of only 20 clinical features. The model validates known risk factors and offers insight into potential novel clinical indicators and treatment characteristics that elevate risk. It relies on factors that are directly extractable from EMRs, provided a tool can be easily integrated into existing workflows. A cost-based analysis provides insight into optimal risk thresholds and offers a framework for tailoring algorithms to individual hospital needs.

CONCLUSION A better understanding of neutropenic risk on an individual level enables a more informed approach to patient monitoring and treatment decisions.

JCO Clin Cancer Inform 5:904-911. © 2021 by American Society of Clinical Oncology

INTRODUCTION

Severe neutropenia and febrile neutropenia (SN-FN) pose severe risks to patients with cancer receiving chemotherapy. ASCO guidelines recommend the use of granulocyte colony-stimulating factor (G-CSF) prophylaxis when the risk of SN-FN exceeds 20%.1 A model of SN-FN risk for chemotherapy patients over the course of multiple cycles of chemotherapy and adjustable for economic considerations can provide more personalized insights into patient risk throughout their treatment. In this work, we propose a highly accurate and interpretable machine learning (ML) model for predicting neutropenic events using electronic medical record (EMR) data from a large health care system.

Neutropenia risk models have been introduced previously by several independent research groups. Most existing models rely on logistic regression2-6 or ML methods that lack interpretability.7,8 Our proposed model differs from existing studies in several ways: we assess the risk of SN-FN at the initiation of any chemotherapy cycle, not only the patient’s first cycle; we consider a broad set of cancers and drugs, rather than focusing on targeted populations or treatment regimens; and we restrict our data set to discrete EMR fields, allowing for direct integration into oncology workflows without manual data manipulation.

In this work, we present an end-to-end analytical pipeline, from data extraction to model implementation considerations. We develop a neutropenia risk score using optimal feature selection (OFS), a novel approach that trains sparse additive models with strong performance and high usability. The final model provides clinical insight into neutropenic risk, both validating risk factors from existing models and identifying additional clinical indicators.

MATERIALS AND METHODS

Study Population

This retrospective, observational study was carried out at Hartford HealthCare (HHC), an integrated health care system composed of seven acute care hospitals in Connecticut with more than 6,000 analytic cases per annum. The study consists of antineoplastic chemotherapy encounters between May 2016 and October 2019. Leukemia was excluded because often the goal of chemotherapy is to produce profound and prolonged neutropenia through bone marrow suppression; all...
other cancers were included. The full criteria used in querying the database are included in the Data Supplement. Each cycle start date is considered as a separate observation, which implies that a single patient can appear multiple times within the data set.

Data Curation

Data extraction. All data used in this study were curated from HHHC’s EMR (Epic Systems, Verona, WI). Only data directly extractable from the EMR were included. This approach makes it feasible to directly run the model off of the EMR and improves the transferability of our model to other institutions. With these same advantages in mind, unstructured data, such as free-text notes on patient condition, detailed information about regimen adjustments, and imaging results, were not selected for the data set.

Clinical features. The outcome of interest was the occurrence of either SN or FN within 4 weeks (28 days) of a chemotherapy encounter.9 We defined SN as absolute neutrophil count below 500 cells/µL and FN as absolute neutrophil count < 1,000 cells/µL, accompanied by a fever (> 101°F).9 We curated discrete EMR data elements reflecting the patient’s demographics, medical features, and cancer treatment information. By incorporating both static and temporal components, we captured how these factors change over time. Table 1 lists the data elements and sources. In total, each encounter was represented by 107 distinct clinical features. Details on the individual data elements are included in the Data Supplement.

Although the data we selected from the EMR are more highly structured than free-text notes, there is still significant variability in data capture that hinders the creation of a unified data set. For example, a single clinical entity such as temperature might be captured in different ways on the basis of hospital department and equipment or laboratory results may be a mix of numeric and text data (10, 15, < 5, > 100). We standardized our data elements through close collaboration with clinical experts, creating data mappings and common vocabularies in the process. These can serve as artifacts for future projects to facilitate meaningful information extraction.

Data imputation. Missing data present a challenge in developing a comprehensive feature space, particularly for vital and laboratory features. Many measurements are not recorded at all visits, resulting in missing data for some encounters. Any features that were missing in more than 40% of encounters were fully excluded. For the remaining features, imputation was performed using MedImpute,10 a novel method that estimates missing values on the basis of the known values of similar observations. The imputation balances the known values from proximal encounters of the same patient with data from other patients; this is formalized through an optimization algorithm.

ML Methods

The prediction of whether an encounter will be followed by SN-FN is a binary classification problem. Table 2 outlines the methods considered, which offer various levels of interpretability and complexity. The models were trained using a common training set of 80% of the data. Five-fold cross-validation was used on the training set to tune the relevant internal parameters for each method. Details on the algorithm implementations and tuned parameters are included in the Data Supplement. The final models were then evaluated on the remaining out-of-sample data (20%). The training and testing data were split by patient, meaning that no patient can have encounters in both the training and testing set, to prevent bias in model evaluation. We additionally report results on a temporal split of the data in the Data Supplement. In both cases, the training and testing sets were imputed independently.

Model Evaluation

When evaluating candidate models, we considered two primary criteria: quantitative performance and model interpretability.
**Quantitative performance.** A model must provide accurate predictions to be useful in a clinical setting. Binary classification models output a probability of a positive response; in this setting, a positive response is defined as the occurrence of a neutropenic event. We used out-of-sample area under the receiver operating characteristic curve (AUC) as the primary performance metric. Given the low incidence of SN-FN, we also considered the average precision for each model. Average precision provides a threshold-independent measure of the precision-recall curve, such as AUC for the receiver operating characteristic curve, that is particularly useful when the outcomes are highly imbalanced. We report these metrics for our final models along with bootstrapped 95% confidence intervals.

Although the model returns a probability of SN-FN, in practice, a probability threshold is often used to label high-risk patients. For a fixed threshold \( \tau \), all encounters with probabilities greater than \( \tau \) would be categorized as having a high risk of an SN-FN event. This is useful in making the outcomes of the predictive tool actionable; for example, the threshold could be used to determine when to surface EMR alerts. For a fixed threshold, we can assess the number of false negatives (\( FN_\tau \)) and false positives (\( FP_\tau \)) incurred by the model. Lower thresholds predict more neutropenic events: this increases sensitivity at the expense of specificity. The opposite is true as the threshold increases.

The desirable threshold is user-determined and driven by the relative costs of mistaken positive and negative events. To determine an optimal risk cutoff threshold, we must quantify the cost of a false positive (\( CFN_\tau \)) and false negative (\( CFP_\tau \)). From a financial perspective, the negative consequence of a false positive is unnecessary intervention and for a false negative, hospitalization. For this estimate, we assume that the intervention for high-risk patients would be G-CSF administration. For a given threshold \( \tau \), the cost incurred is then given as follows:

\[
C_\tau = C_{FP_\tau} + C_{CFN_\tau}
\]

The optimal threshold minimizes this cost.

**Interpretability.** The methods considered vary in their inherent interpretability. Linear models and single decision trees explicitly tie clinical inputs to the resultant predictions. Ensemble methods, such as random forests and gradient boosted machines, aggregate many individual models, which limits their interpretability. Models that lack clinical interpretation make it difficult to justify predictions and assess their validity, hindering clinician trust.\(^{19}\)

---

**TABLE 1. Overview of Clinical Features and Sources**

<table>
<thead>
<tr>
<th>Category</th>
<th>Source</th>
<th>Sample Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographics</td>
<td>Patient data</td>
<td>Age, sex</td>
</tr>
<tr>
<td>Comorbidities</td>
<td>Problem list</td>
<td>Cerebrovascular disease, diabetes, hypertension</td>
</tr>
<tr>
<td>Other procedures</td>
<td>Procedure charges</td>
<td>Indicator of concurrent radiation</td>
</tr>
<tr>
<td>Treatment information</td>
<td>Cancer patient history</td>
<td>Cancer site, treatment intent (eg, curative and maintenance)</td>
</tr>
<tr>
<td>Drugs administered</td>
<td>Medication charges</td>
<td>Chemotherapy drugs (individual drugs and combinations), indicator of G-CSF administration</td>
</tr>
<tr>
<td>Vitals measurements</td>
<td>Flow sheets</td>
<td>BMI, pulse, systolic blood pressure</td>
</tr>
<tr>
<td>Laboratory measurements</td>
<td>Laboratory order results</td>
<td>Complete blood count results</td>
</tr>
</tbody>
</table>

Abbreviations: BMI, body mass index; G-CSF, granulocyte colony-stimulating factor.

**TABLE 2. Overview of Machine Learning Methods Used for Binary Classification**

<table>
<thead>
<tr>
<th>Method</th>
<th>Reference</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR</td>
<td>11</td>
<td>Fits an additive model. Regularization is used to limit model complexity and control overfitting. This serves as a benchmark traditional statistical method</td>
</tr>
<tr>
<td>OFS</td>
<td>12</td>
<td>Fits an additive model using a sparse set of features. For example, OFS(_{50}), with a maximum sparsity of 50, would fit a model with at most 50 features having nonzero coefficients. In this study, three maximum sparsity parameters, 20, 30, and 50, are considered</td>
</tr>
<tr>
<td>CART</td>
<td>13</td>
<td>Partitions data using a single decision tree. The tree comprises binary feature splits, and each leaf yields a predicted risk probability</td>
</tr>
<tr>
<td>OCT</td>
<td>14, 15</td>
<td>Constructs a single decision tree, as with CART. In contrast to traditional greedy tree-based models such as CART, OCT uses an optimization framework when fitting the tree, which generally demonstrates superior performance</td>
</tr>
<tr>
<td>RF</td>
<td>16</td>
<td>Fits many decision trees, each using a subset of features and data, forming an ensemble of models. Final predictions aggregate the votes of the individual trees</td>
</tr>
<tr>
<td>XGB</td>
<td>17, 18</td>
<td>Another ensemble approach that trains many decision trees but uses a weighting scheme to better account for errors in individual learners</td>
</tr>
</tbody>
</table>

Abbreviations: CART, classification and regression trees; LR, logistic regression; OCT, optimal classification trees; OFS, optimal feature selection; OFS\(_{50}\), OFS with 50 features; RF, random forests; XGB, gradient boosted machines.
An additional aspect of interpretability is model sparsity, namely the number of features used to generate a prediction. For example, a decision tree with six splits would use a maximum of six features to output a prediction. Although the input feature space contains 107 unique features, it is desirable for the final model to rely on a subset of these features for clinical interpretability.

RESULTS

Study Population

The final population consists of 17,513 encounters across 2,806 patients. Four hundred forty-nine (2.6%) encounters had a neutropenic event within 28 days of the encounter. Four hundred twenty-one encounters had SN and 77 had FN; outcomes that met the criteria for both SN and FN count as a single neutropic event. The most common cancers observed in the data are breast, lung, colon, and rectal or anal cancer and multiple myeloma, which comprise more than 60% of the encounters. Descriptive summaries of all features are included in the Data Supplement.

Model Performance

Table 3 reports the test AUC and average precision for the various models. The methods have out-of-sample AUCs ranging from 0.789 to 0.869 and average precisions ranging from 0.080 to 0.148. Given the significant class imbalance in our data (2.6% positives), the baseline precision is 0.026. Therefore, the best performing models offer a roughly five-fold increase from the baseline precision. The additive models, OFS and logistic regression, demonstrate strong performance. OFS with 20 features (OFS20) is able to achieve the second-best AUC (0.865; 95% CI, 0.830 to 0.891) and the highest average precision (0.148; 95% CI, 0.117 to 0.188) with fewer features than other linear models. Random forests performs comparably but at the price of lower interpretability. Overall, the strength of the linear models suggests that nonlinear feature interactions are not highly significant in this prediction problem.

Of the models considered, OFS20 offers the most insight given its balance of both quantitative performance and model interpretability. This is therefore our proposed final model.

Threshold-Based Analysis

An optimal decision threshold can be found after estimating the costs and probabilities associated with false positives and false negatives. To illustrate threshold selection, we consider an example for patients with non–small-cell lung cancer. On the basis of analysis by Li et al., we estimate the cost of G-CSF administration as $2,580 US dollars (USD) and the cost of a neutropenia-related hospitalization for a patient with non–small-cell lung cancer at $21,822.50 USD, $5,075 USD per day with an average length of 4.3 days. Figure 1 shows the total expected cost incurred across all thresholds $\tau \in [0, 1]$ for the OFS20 model on the test set. The cost-minimizing threshold is $\tau = 0.16$. At this threshold, the model obtains out-of-sample specificity of 95.7% and sensitivity of 42.9%.

To obtain a more general characterization of the trade-offs between false positives and false negatives, we compute the optimal threshold as a function of the ratio between $C_{FN}$ and $C_{FP}$. The optimal threshold is determined purely by the cost ratio. The example above demonstrates the computation for a cost ratio of 8.5 ($C_{FN}/C_{FP} = 21,822.50/2,580$ USD). Figure 2 shows the optimal thresholds for varied cost ratios using the OFS20 model on the test set. As the ratio increases, the optimal threshold decreases; model sensitivity (identification of true positives) becomes more valuable, and so the model flags more patients at high risk. The optimal threshold begins to decrease above a ratio of 5. This implies that when the cost of hospitalization is more than five times as expensive as G-CSF intervention, it is economically advantageous to lower the decision threshold, which allows the model to recover true positives despite the risk of overtreatment of false positives. The optimal threshold stabilizes at 16% for ratios between 5 and 10 and lowers to 8% for higher cost ratios between 15 and 30.

Model Interpretation

The OFS20 coefficients are shown in Table 4. Positive coefficients indicate an increase in risk as the value increases (eg, risk increases with the number of drugs given in recent weeks), whereas negative coefficients represent an inverse relationship (risk decreases as the cumulative infusion count increases). Individual drugs have varied risk impacts. Since a patient can receive multiple drugs in a single encounter, the net contribution of the drugs is determined by the sum of these coefficients.
In this work, we have developed a practical tool for assessing SN-FN risk in patients upon initiation of a chemotherapy cycle. We leveraged discrete EMR data and used state-of-the-art algorithms to synthesize clinical features into a single risk score. The resultant model enables a personalized approach to patient care on the basis of an individual’s cancer characteristics, vitals, laboratory values, and course of treatment.

We compared various ML algorithms but ultimately selected the OFS20 model because of its high interpretability and competitive quantitative performance. Our final model has an out-of-sample AUC of 0.865. This model outperforms the neutropenia risk model (out-of-sample AUC of 0.81) proposed by Lyman et al, while also using a smaller feature set that is directly extractable from the EMR. Cho et al report an AUC of 0.908 in their proposed ML model for FN prediction, although this model addresses a narrower clinical question of neutropenic risk for Korean patients with breast cancer. The OFS20 model coefficients provide insights into the risk contributions of individual features. Risk increases when more drugs are involved in the regimen, which could indicate more aggressive treatments. G-CSF administration leads to lower risk, consistent with its use as an intervention to mitigate SN-FN risk, as other models have found.

Genitourinary comorbidities, which include renal diseases,
increase risk; this aligns with findings of increased risk associated with kidney dysfunction. Higher blood counts (hematocrit and platelets) are associated with lower risk, as is a lower pulse. Finally, to our knowledge, our model is the first to incorporate temporal elements. Risk decreases as patients are further along in treatment, that is, as they have more previous cycles. We also see that the change in clinical features over time, particularly a decrease in weight, implies higher risk. We note that as with any retrospective study using observational data, we cannot establish causation of the observed risk factors or rule out the significance of unobserved factors. Nevertheless, the proposed final model provides highly accurate characterizations of patient risk on the basis of the included features.

It is also informative to observe the features that were not selected in the model. Although the feature space included cancer site and drug combination, rather than just individual drugs, neither of these features were selected in the final model. Certain clinical elements identified in other models, such as age, also do not appear in our model. This suggests that these factors are less significant in determining a patient’s risk or that their risk contribution can be explained through other observed clinical characteristics. We note that the commitment to the exclusive use of structured EMR data requires the omission of other potentially relevant data elements, such as relative dose intensity or qualitative assessments of patient health, in exchange for portability and reproducible results.

Our proposed approach to determining an optimal cutoff threshold for flagging high-risk patients can be adapted to inform reasonable site-specific cutoffs for new populations and cost estimates. ASCO and National Comprehensive Cancer Network guidelines recommend primary use of G-CSF at a threshold risk level of 20%. National Comprehensive Cancer Network guidelines recommend consideration of G-CSF depending upon patient risk assessment for intermediate risk levels between 10% and 20%. Previously published models are specific to a chemotherapy regimen, a cancer type, or the first cycle of treatment and thus lack generalizability. Our model finds that the cost analysis supports a threshold risk level of 8%-16%. Our framework allows for tuning the performance specifications of the predictive model relative to the economic costs of treatment inherent to a health care delivery system and can be used to guide payer reimbursement policy.

Our threshold modeling provides a framework for determining an appropriate risk threshold that can be extended to incorporate other factors. We did not attempt to model the positive economic benefits of true positives and negatives. An economic model of clinical decision support

---

**TABLE 4.** Optimal Feature Selection Coefficients With Sparsity of 20 Features

<table>
<thead>
<tr>
<th>Category</th>
<th>Feature</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td></td>
<td>-2.460</td>
</tr>
<tr>
<td>Treatment</td>
<td>Antineoplastic drug count (previous 3 weeks)</td>
<td>0.139</td>
</tr>
<tr>
<td></td>
<td>Cumulative number of chemo cycles</td>
<td>-0.021</td>
</tr>
<tr>
<td></td>
<td>Filgrastim (G-CSF) administered? (1 = yes)</td>
<td>-0.142</td>
</tr>
<tr>
<td>Laboratory results or v��als</td>
<td>Hematocrit (%)</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td>Platelet count (τ/μL)</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>Pulse</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>Relative change in weight (lbs) from previous cycle</td>
<td>-3.065</td>
</tr>
<tr>
<td>Comorbidities</td>
<td>Diseases of the genitourinary system? (1 = yes)</td>
<td>0.147</td>
</tr>
<tr>
<td>Drugs</td>
<td>Atezolizumab? (1 = yes)</td>
<td>-0.696</td>
</tr>
<tr>
<td></td>
<td>Carboplatin? (1 = yes)</td>
<td>0.369</td>
</tr>
<tr>
<td></td>
<td>Cisplatin? (1 = yes)</td>
<td>0.374</td>
</tr>
<tr>
<td></td>
<td>Cyclophosphamide? (1 = yes)</td>
<td>0.787</td>
</tr>
<tr>
<td></td>
<td>Dacarbazine? (1 = yes)</td>
<td>0.469</td>
</tr>
<tr>
<td></td>
<td>Docetaxel? (1 = yes)</td>
<td>0.485</td>
</tr>
<tr>
<td></td>
<td>Doxorubicin? (1 = yes)</td>
<td>0.839</td>
</tr>
<tr>
<td></td>
<td>Etoposide? (1 = yes)</td>
<td>1.353</td>
</tr>
<tr>
<td></td>
<td>Pertuzumab? (1 = yes)</td>
<td>-0.321</td>
</tr>
<tr>
<td></td>
<td>Trastuzumab? (1 = yes)</td>
<td>-0.217</td>
</tr>
<tr>
<td></td>
<td>Vinblastine? (1 = yes)</td>
<td>0.469</td>
</tr>
<tr>
<td></td>
<td>Vinorelbine? (1 = yes)</td>
<td>0.791</td>
</tr>
</tbody>
</table>

Abbreviation: G-CSF, granulocyte colony-stimulating factor.
should ideally minimize unnecessary costs while also maximizing health care benefits. Additionally, although our analysis defines cost as financial costs incurred by either hospitalization or unnecessary intervention, there are also nonfinancial health economic costs that cannot be measured by this model. We remain cognizant of the burden of false positives, which could lead to alarm fatigue, while also recognizing that false negatives associated with hospitalization carry a quality-of-life cost, which in economic terms are disutilities of care. The analysis can be modified to capture additional financial and quality-of-life costs.

A central goal of this study was to create a frictionless point-of-care tool for assessing neutropenic risk while patients are undergoing treatment. This motivated the creation of a feature space using only discrete data elements. Although individual health systems have distinct ways of recording patient data, all the features included in the model should be available as structured data within the EMR. After establishing a mapping of a hospital’s data elements to our feature space, the model can be integrated into a new EMR system to provide real-time insights in clinical encounters. Our selection of a model that relies on a relatively small subset of clinical features reduces the burden of creating such a data mapping; only 20 features need to be extracted from the EMR to calculate the risk score. These considerations lower the barrier to model validation and adoption at external sites. The ultimate test of any risk prediction model is its performance on external populations, and we hope to continue this work through prospective validation both within HHC and at external sites.

In conclusion, this work presents the development of a neutropenia risk prediction tool, from data curation to practical implementation considerations. This tool offers the potential to improve patient care, providing personalized insights for chemotherapy patients that enable more informed treatment planning and care decisions.

**AFFILIATIONS**

1 Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA
2 Hartford HealthCare, Hartford, CT
3 Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA

**CORRESPONDING AUTHOR**

Dimitris Bertsimas, PhD, Sloan School of Management, Massachusetts Institute of Technology, 100 Main St, Building E62-560, Cambridge, MA 02142; e-mail: dbertsim@mit.edu.

**DISCLAIMER**

Any opinion, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

**PRIOR PRESENTATION**


**SUPPORT**

Supported by a grant from Hartford HealthCare. H.W. is also supported by the National Science Foundation Graduate Research Fellowship under Grant No. 174530.

**DATA SHARING STATEMENT**

The data underlying this article cannot be shared publicly as they include sensitive patient information from which it is difficult to guarantee de-identification. As a result, there is a possibility of deductive disclosure of participants and therefore full data access through a public repository is not permitted by the institutional review board of Hartford HealthCare. The data will be shared on reasonable request to the corresponding author.

**AUTHOR CONTRIBUTIONS**

Conception and design: Holly Wiberg, Peter Yu, Pat Montanaro, Jeff Mather, Dimitris Bertsimas

Administrative support: Peter Yu, Pat Montanaro
Provision of study materials or patients: Peter Yu
Collection and assembly of data: Peter Yu, Pat Montanaro, Jeff Mather, Suzi Birz, Michelle Schneider
Data analysis and interpretation: Holly Wiberg, Peter Yu, Pat Montanaro, Dimitris Bertsimas
Manuscript writing: All authors
Final approval of manuscript: All authors
Accountable for all aspects of the work: All authors

**AUTHORS’ DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST**

The following represents disclosure information provided by the authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO’s conflict of interest policy, please refer to www.ascopubs.org/rcw or ascopubs.org/cci/author-center.

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians (Open Payments).

Peter Yu
This author is an Associate Editor of *JCO Clinical Cancer Informatics*. Journal policy recused the author from having any role in the peer review of this manuscript.

No other potential conflicts of interest were reported.

**ACKNOWLEDGMENT**

The authors would like to acknowledge the contributions of many collaborators at Hartford HealthCare who provided expertise on both the data elements and clinical interpretation. These include, listed alphabetically, Marissa Flugrad, Scott Gould, Nitya Joseph, Patty Kaehrle, Judy Keating, Mona Kraynak, Ron M’Sadoques, Nick McNulty, Ewy Mulawka, Amy Noack, Barry Stein, Mary Lou Vasquezza, and Qun Yu. We also thank Srishti Gupta for her contributions to the data processing and analysis.
REFERENCES


