Accented: 24 January 2021

## RESEARCH ARTICLE

## URGICAL ONCOLOGY WILEY

# Predicting post-discharge cancer surgery complications via telemonitoring of patient-reported outcomes and patient-generated health data

Lorenzo A. Rossi PhD<sup>1</sup> Virginia Sun PhD, RN<sup>2,3</sup>

<sup>1</sup>Department of Applied AI & Data Science, City of Hope National Medical Center, Duarte, California, USA

<sup>2</sup>Department of Surgery, City of Hope National Medical Center, Duarte, California, USA

<sup>3</sup>Department of Population Sciences, City of Hope National Medical Center, Duarte, California, USA

#### Correspondence

Lorenzo A. Rossi, PhD, Department of Applied AI & Data Science, City of Hope National Medical Center, 1500 E Duarte Rd, Duarte, CA 91010, USA. Email: Irossi@coh.org

#### Funding information

National Cancer Institute, Grant/Award Number: (P30CA33572); City of Hope's Center for Cancer and Aging

Lorenzo A. Rossi PhD<sup>1</sup> | Laleh G. Melstrom MD<sup>2</sup> | Yuman Fong MD<sup>2</sup> |

## Abstract

Revised: 22 January 2021

**Background and Objectives:** Post-discharge oncologic surgical complications are costly for patients, families, and healthcare systems. The capacity to predict complications and early intervention can improve postoperative outcomes. In this proof-of-concept study, we used a machine learning approach to explore the potential added value of patient-reported outcomes (PROs) and patient-generated health data (PGHD) in predicting post-discharge complications for gastrointestinal (GI) and lung cancer surgery patients.

**Methods:** We formulated post-discharge complication prediction as a binary classification task. Features were extracted from clinical variables, PROs (MD Anderson Symptom Inventory [MDASI]), and PGHD (VivoFit) from a cohort of 52 patients with 134 temporal observation points pre- and post-discharge that were collected from two pilot studies. We trained and evaluated supervised learning classifiers via nested cross-validation.

**Results:** A logistic regression model with  $L_2$  regularization trained with clinical data, PROs and PGHD from wearable pedometers achieved an area under the receiver operating characteristic of 0.74.

**Conclusions:** PROs and PGHDs captured through remote patient telemonitoring approaches have the potential to improve prediction performance for postoperative complications.

#### KEYWORDS

machine learning, patient-generated health data, patient-reported outcomes, supervised learning, wearable computing

## 1 | INTRODUCTION

Many cancer surgeries have a high risk of post-discharge complications that may result in undesirable healthcare resource use such as hospital readmissions. Postoperative complications and readmissions are costly and burdensome for the patients, their families, and the healthcare system. Postoperative complications and readmissions are important surgical outcomes that are often used as quality metrics in surgical oncology. The ability to predict postoperative complications is critical for improving the quality and value of surgical care. The predictive ability can assist with identifying patients that are at higher risk for postoperative morbidities early and allocate healthcare resources appropriately to improve postoperative recovery. Complications may also be caused by events occurring after discharge, and therefore, not observable by prediction approaches based on traditional EHR data. Patient-generated health data (PGHD), including wearable activity

ROSSI ET AL.

trackers and electronic patient-reported outcomes (ePROs) can be feasibly obtained while patients are at home and in their communities. These data, particularly ePROs (symptoms, quality of life), provide the patient's lived experience away from the hospital, and have been shown to improve survival in patients with advanced cancer.<sup>1</sup> Perioperative telemonitoring technologies through wearable and webbased applications may enable real-time remote patient observations and complement traditional clinical, surgical, and disease-specific data to improve the prediction of negative outcomes and prevent undesirable healthcare resource use. In addition, policy and reimbursement barriers related to telehealth and telemonitoring have been temporarily removed to enable their rapid uptake and use during the COVID-19 pandemic. However, PGHDs and ePROs are not routinely captured in surgical oncology. Furthermore, there is a dearth in the current literature on understanding the meaning and potential clinical benefits of PGHD and ePROs in surgical oncology. As for PGHDs, Bae et al.<sup>2</sup> used features from step counts to predict 30-day unplanned readmission for a cohort of 25 postsurgical cancer inpatients. Doryab et al.<sup>3</sup> used step and heart rate data collected from smartwatches to extract bio-behavioural rhythm figures to predict complication risk for a cohort of 52 pancreatic surgery patients (82% cancer). To the best of our knowledge, there is no work using ePROs to predict postsurgical outcomes for oncology patients. In this proof-of-concept study, we used a machine learning approach to explore the potential added value of functional recovery PGHD (as daily steps measured by wristband pedometers) and ePROs (symptoms) for predicting postoperative complications. Specifically, we were interested in the following research questions. First, does functional recovery PGHD (as measured by daily steps) bring added value to postoperative complications predictive models that also include traditional clinical/surgical/ disease-related variables? The second guestion we aimed to address was do ePROs (symptoms) bring added value to postoperative complications predictive models that also include traditional clinical/ surgical/disease-related variables?

## 2 | MATERIALS AND METHODS

#### 2.1 | Data source

PGHD and ePRO data from two telehealth/telemonitoring pilot studies were accessed for this analysis (n = 52). The first study was a proof-ofconcept pilot to explore the feasibility and acceptability of remote wireless perioperative telemonitoring in patients undergoing surgery for the treatment of gastrointestinal (GI) cancers (n = 20).<sup>4</sup> The second study was a feasibility trial of a telehealth perioperative physical activity intervention in older adults with lung and GI cancer (n = 32).<sup>5</sup> In both studies, patients wore wristband pedometers (VivoFit<sup>TM</sup>) to assess functional recovery (as measured by daily steps) and completed ePROs (MD Anderson Symptom Inventory [MDASI]) to assess symptoms. Telemonitoring for both studies began before surgery and continued up to 4 weeks post-discharge. Daily steps data were captured continuously, while ePROs were collected intermittently (before surgery, at discharge, and at 1, 2, and 4 weeks post-discharge. Surgical outcomes (postoperative complications, hospital readmissions) were captured via electronic health records (EHRs).

Functional recovery PGHD was captured using the VivoFit (Garmin Company). This device is a commercially available tracking wristband pedometer. It continuously monitors daily step progress without requiring battery change for one year. It is also waterproof and can be worn in the shower. VivoFit stores daily steps count right on the device and can be wirelessly synced for free to supported smartphones and computers. A systematic review of 22 studies concluded that the validity and reliability of wearable pedometers were generally high for measuring daily steps.<sup>6</sup>

Symptoms were assessed using the MDASI. This is a brief, wellvalidated measure of 13 common cancer-related symptoms: pain, fatigue, nausea, sleep, distress/anxiety, shortness of breath, memory, appetite, dry mouth, drowsiness, sadness, vomiting, and numbness/ tingling. Each symptom's severity is rated on a 10-point scale. A score of 1–3 indicates mild intensity, and 4–10 indicates moderateto-severe intensity. Patients also rate how much their symptoms interfered with six common functional domains: walking, activity, working/housework, social relationships, enjoyment of life, and mood. The MDASI applies broadly across cancer types and treatments and is easy for patients to complete. A movement of 1.2 points is clinically meaningful and cut-off scores for indication of clinical deficits have been defined by the tool authors. The MDASI has been validated in surgical populations; Cronbach's  $\alpha$  reliability ranges from .82 to .94.<sup>7.8</sup>

Out of the 52 patients enrolled, 47 participants submitted answers MDASI (ePROs) at least once. Forty of the 47 patients wore the VivoFit pedometer at least once. The numbers of male and female patients were 25 and 27, respectively. Age ranged between 22 and 84 years, with a median of 68 years. Complications were recorded for up to 30 days after discharge. The overall post-discharge complication rate for the 52 patients was 21.1%. The length of stay (LOS) ranged between 1 and 13 days with a median of 6 days. In total, 36 and 16 patients underwent surgeries for treatment of GI and lung cancers, respectively.

#### 2.2 | Temporal segmentation of the data

From the standpoint of data collection, there are three distinct types of temporal segments for each patient—T1: a period of 3–14 days preceding admission, days, used to collect baseline PRO and step data; T2: hospital visit for the surgery (median LOS = 6 days); and T3: 2–4 weeks following hospital discharge.

Patients were asked to submit ePROs at the end of periods T1 and T2 and the end of each week after discharge (T3). Similarly, patients were asked to wear the VivoFit pedometers throughout the above periods. For consistency across the pilots, we considered observations up to the second week after discharge, treating each week as a distinct observation. Hence, after the exclusion of the periods before admission, there were up to three temporal observations for each patient: hospital visit, first, and the second week after discharge. We pointed out the partial compliance of patients with ePROs and pedometers. We obtained 99 observations with at least partial compliance to VivoFit pedometers (63 post-discharge weeks) and 129 observations with compliance to MDASI questionnaires (85 post-discharge weeks). We included observations associated with hospital visits of patients noncompliant to MDASI questionnaires or VivoFit pedometers. We removed post-discharge observations encompassing or following readmission or complication events to avoid leakage.

#### 2.3 | Problem formulation

We considered the periods up to admission only to extract baselines about the patients (e.g., median daily steps before surgery). In this case, we have between one and three temporal segments marked by discharge and/or the completion of the MDASI ePRO questionnaire. Our goal was to predict unplanned readmission and serious complications within 30 days from discharge at discharge and/or at each time of PRO the data available for the associated segment. Therefore, for each patient with at least partially compliant ePRO submissions, we performed predictions at up to three distinct times. We framed the complication prediction task as a binary classification problem (complication, yes or no), where a score between 0 and 1 is produced. A score close to 1 means a high risk of complication/ unplanned readmission.

#### 2.4 | Features

We selected sets of variables shared by both pilots based on clinical domain knowledge and with the goal to evaluate the predictive potential for ePRO and PGHD in the form of steps collected via wearable devices. From the variables, we extracted features to train machine learning models. The small sample size limits the number of variables that can be considered. For this reason, we did not include immediate discharge labs, although potentially predictive of the risk of readmission in larger cohorts. The features used are listed in Table 1. They were computed for the temporal period of the hospital visit and for Weeks 1 and 2 after discharge (dsc): for example, min (daily steps) = the minimum number of daily steps for Week 1 after discharge. The nonlinear transformation of the body mass index (BMI) is motivated by the use of a linear model. The MDASI features range from 0 to 10, according to the level of self-assessed discomfort. "Pain vs. baseline" is the difference between the pain levels reported at, for example, discharge and before admission.

For the purpose of feature extraction, most of the variables did not need processing (e.g., the answers to the MDASI questionnaires consist of scores between 0 and 10). MDASI PRO variables for drowsiness, numbness, and vomiting were excluded based on domain expert assessments. We defined the feature pain with respect to the baseline as pain(T1) – pain(Tk), where T1 is the observation period before admission Journal of

Clinical charts	MDASI ePROs	Pedometer PGHD
ASA score	Pain vs. baseline	Total steps
Number complications before discharge	Pain	Min (daily steps)
Length of stay	Fatigue	Max (daily steps)
Zubrod score	Nausea	Median (daily steps)
(BMI-21.7) <sup>2</sup>	Sleep	Total steps versus baseline
Weeks since discharge	Distress	Linear trend (daily steps)
	Breath	Fraction of days with pedomter
	Memory	
	Appetite	
	Dry mouth	
	Sadness	
	MDASI	
	compliance	

*Note*: The MDASI features range from 0 to 10, according to the level of self-assessed discomfort.

Abbreviations: ASA, American Society of Anesthesiologists; BMI, body mass index; ePRO, electronic patient-reported outcome; MDASI, MD Anderson Symptom Inventory; PGHD, patient-generated health data.

(to collect patient's baseline data) and Tk is either the hospital visit or one of the post-discharge weeks. We introduced an indicator feature to express compliance for MDASI questionnaires—MDASI compliance.

We performed a quadratic transformation on the BMI and extracted its square root to express it as the distance from a normal range. We imputed missing values only for the American Society of Anesthesiologists (ASA) physical status score with 3, the value the majority class. When considering observation periods, after discharge, we used the ASA score assessed during the hospital visit. We excluded age, because the largest of the pilots (n = 32) focused exclusively on elderly patients.

The VivoFit devices recorded the daily steps. From the daily steps over an observation period, we extracted minimum, maximum, median counts of steps, standard deviation of steps, slope and intercept of the linear interpolation, differences with respect to baseline, and the fraction of days the patients wore pedometer during an observation period. The latter also played as an indicator feature for observation periods without pedometers.

## 2.5 | Training and evaluation

As mentioned, for each patient, we considered hospital visit, Weeks 1 and 2 post-discharge as separate observations (n = 134), to predict complications within 30 days from discharge. Temporal segments encompassing or following complications were removed to avoid leakage. We trained logistic regression classifiers via nested

-WILEY

WILEY-SURGICAL ONCOLO

cross-validation with  $L_1$  or  $L_2$  regularization, using the scikit-learn Python library.<sup>9</sup> Regularization is used to reduce the sensitivity of the coefficients of the model, and therefore, to prevent learning from noise in the data. The feature values were standardized to zero mean and unit variance to facilitate convergence during training. We tuned the model with randomized hyperparameter search via 4-fold patient-cross-validation (inner loop). With this kind of group crossvalidation, the observations from a patient are always in a fold. The tuning parameters were the regularization coefficient and the type of regularization ( $L_1$  or  $L_2$ ). We evaluated the model via leave-onepatient-out cross-validation (outer loop). In both cross-validation loops, all the feature vectors associated with the same patient were either in the same fold, to avoid having segments of the same patient split between training and testing. Cross-validation is a procedure to train and test machine learning models when the number of observations is limited: Every observation is in turn used for both training and testing. The number of folks, K, is the number of groups the observations are split into. We considered the following subsets of data: clinical data (n = 52), clinical + MDASI data (n = 134), clinical + steps data (n = 111), and clinical + steps + MDASI data (n = 134).

We analyzed the contributions of the features toward the prediction score via the Python shap package.<sup>10</sup> The SHapley Additive exPlanations (SHAP) are a measure of the contribution of each feature toward a prediction and are obtained via a game theory approach.

## 3 | RESULTS

We evaluated the performance of the models with the receiver operating characteristic (ROC) and its underlying area (AUROC). The ROC shows the true positive rate of the model as a function of the false positive rate for the possible decision thresholds. The AUROC can be interpreted as the probability that an observation in the positive class (i.e., readmission/complication) has a risk score higher than an observation in the negative class. An AUROC = 0.5 means random guessing, while AUROC = 1 means 100% correct predictions. We also computed the area under the precision-recall curve (AUPRC). The PRcurve shows the precision-recall trade-offs. Precision and recall are also indicated as positive predictive value and sensitivity, respectively. The AUPRC depends on the quality of the model and the prevalence.

The model combining clinical, MDASI PROs, and step features (n = 134) yielded a 0.74 AUROC. When daily step data were removed, the AUROC was 0.70 (n = 134). The AUROC was 0.72 for the model with daily step data, but without ePRO symptoms (n = 111). For the baseline model, solely based on clinical features and observation until discharge, the AUROC was 0.62 (n = 52). The performances are summarized in Table 2. The ROCs of each model are plotted in Figure 1. Table 3 shows the stratified performance of the model for different prediction times: at discharge, 1, and 2 weeks after discharge.

The performances for the models using ePROs and/or PGHDSs are close (AUROC: 0.70–0.74), while the AUROC of the model based only on clinical factors collected before discharge is 0.62. The former models benefit from a larger number of training samples because they include

**TABLE 2** Counts of observation and features, complication rates, and performance metrics for the models

	N	No. of features	Complication rate (%)	AUROC	AUPRC
Baseline	52	5	21.1	0.62	0.43
MDASI	134	18	14.2	0.70	0.27
Steps	111	13	14.4	0.72	0.28
MDASI + steps	134	25	14.2	0.74	0.28

Abbreviations: AUROC, area under the receiver operating characteristic; AUPRC, area under the precision-recall curve; MDASI, MD Anderson Symptom Inventory.

observations from the data collected in the first two weeks after discharge from telemonitoring via MDASI ePROs and VivoFit pedometers. The relatively high value of the AUPRC for the baseline model depends on the higher prevalence. Due to the different values of prevalence for the above models, we chose AUROC for performance comparisons.

Figure 2 shows the global feature importance for the model using clinical, MDASI, and step data. Figure 3 displays the impact on the prediction score based on the values of the features and it is obtained from the SHAP explanations associated with the prediction of the trained model. The blue and red colors are associated, respectively, with the low and high values of a feature. Dots to the left and right of the vertical axis indicate low- and high-risk scores, respectively. The figure confirms several clinical intuitions. The two most predictive features are *the number of complications before discharge* and *LOS*, with high values associated with a high risk of readmission/complication. Moreover, a low (i.e., 0) functional



**FIGURE 1** Receiver operating characteristic curves for the evaluated models. The one trained with clinical, MDASI ePRO, and step PGHD performs best (AUROC = 0.74). AUROC, area under the receiver operating characteristic; ePRO, electronic patient-reported outcome; FPR, false positive rate; MDASI, MD Anderson Symptom Inventory; TPR, true positive rate [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 3 The area under the received operating characteristic at different observation points for the model based on clinical. MDASI, and step data

Week since dsc.	Ν	AUROC
0	49	0.72
1	44	0.70
2	41	0.69

Abbreviations: AUROC, area under the receiver operating characteristic; dsc., discharge; MDASI, MD Anderson Symptom Inventory.

(Zubrod) score is associated with a low risk of readmission/complications. We also observed that the risk of complication decreases as time progresses (measured via weeks since discharge).

The most predictive among PGHD-related features were the total steps versus the baseline and the linear trend of the daily steps (i.e., the slope of the linear interpolation). The plot suggests that a large decrement of physical activity with respect to the baseline period and/ or a sharp increase in the physical activity levels (measured via the slope of the linear interpolation of the daily step counts) are associated with a higher risk of complications. Similarly, Bae et al.,<sup>2</sup> who used Fitbit data to predict readmission for 25 postsurgical cancer patients, observed that more intense activity by the readmitted group could be risky behavior and deserves more exploration.

In regard to the MDASI ePROs, high levels of reported discomfort for pain (with respect to the baseline as well as absolute values) sleep and appetite imply a higher risk of post-discharge complications. Overall, MDASI ePROs seems less predictive than wearable pedometer PGHD. As a matter of fact, the model based on PGHD features outperformed slightly the one based on ePRO features (AUROC: 0.72 vs. 0.70). Both shared highly predictive clinical features.

Figure 4 shows the top features for an observation of a readmitted patient: 897 steps/day slope, 13 day LOS, and two complications before discharge increased the predicted risk score, while a Zubrod score of 0 was the major factor decreasing the risk score.

Post hoc t-tests showed significant differences over the observations for a number of complications before discharge: no complications ( $\mu = 0.16$ ,  $\sigma = 0.36$ ), complications ( $\mu = 0.6$ ,  $\sigma = 0.8$ ) with t (134) = 2.7, p = .015; LOS: no complications ( $\mu$  = 5.6,  $\sigma$  = 3.0), complications ( $\mu = 8.1$ ,  $\sigma = 3.3$ ) with t(134) = 3.1, p = .005; Zubrod score: no complications ( $\mu = 0.79$ ,  $\sigma = 0.41$ ), complications ( $\mu = 0.95$ ,  $\sigma = 0.23$ ) with *t*(134) = 2.4, *p* = .021.

## DISCUSSION

We proposed and evaluated an approach that uses clinical, ePROs, and wearable PGHDs to predict 30-day post-discharge complications and unplanned readmission for a cohort of 52 GI and lung surgical oncology



for the logistic regression model using clinical, MDASI ePRO, and step PGHD. ASA, American Society of Anesthesiologists; dsc., discharge; ePRO, electronic patient-reported outcome; LOS, length of stay; MDASI, MD Anderson Symptom Inventory; PGHD, patient-generated health data; SHAP, SHapley Additive exPlanations [Color figure can be viewed at wileyonlinelibrary.com]

1350



**FIGURE 3** The contributions of the features based on their values for the model using clinical, MDASI ePRO, and step PGHD. ASA, American Society of Anesthesiologists; dsc., discharge; ePRO, electronic patient-reported outcome; LOS, length of stay; MDASI, MD Anderson Symptom Inventory; PGHD, patient-generated health data; SHAP, SHapley Additive exPlanations [Color figure can be viewed at wileyonlinelibrary.com]

patients. As ePROs and PGHDs were monitored also after discharge to the best of the patients' compliance, our approach updated predictions outside hospital settings, increasing the ability to predict complications. To date, there is limited work on perioperative complication prediction models based on the aforementioned types of data.

The prediction of unplanned hospital readmission is a classic task in machine learning for health care. With this goal, Rajikomar et al.<sup>11</sup> proposed a deep learning approach for tasks such a readmission prediction, outperforming a logistic regression baseline by about 1%. Liu et al.<sup>12</sup> applied an artificial neural network to features from embeddings of diagnosis codes and the EHR. There are few studies focused on the prediction of post-discharge complications and unplanned readmissions for surgical patients. For instance, Kalagara et al.<sup>13</sup> proposed a gradient boosting approach to predict readmission for patients subject to lumbar laminectomy. Bolourani et al.<sup>14</sup> focused on predicting readmission for patients who underwent esophagectomy. Corey et al.<sup>15</sup> evaluated

different classifiers in predicting patients at risk of complications before the time of surgery.

PGHD is typically obtained via wearable devices. A limitation of prior studies on the prediction of postsurgical complications is that the inhospital clinical data used to make predictions are only available until the time of discharge. Critical changes in the patient's condition were not detectable if occurred after discharge. The main advantage to monitor patients remotely via wearables is the capacity to unobtrusively observe proxies of their condition after discharge. There are studies on the application of wearable accelerometers to monitor the recovery of stroke patients.<sup>16</sup> The number of studies using PGHD for predicting outcomes for surgical oncology patients is even more limited.<sup>2,3</sup> Additionally, we could not find any studies using ePROs to predict post-discharge outcomes for surgical patients. To the best of our knowledge, ours is the first work to combine clinical data with both PROs and PGHD to predict postdischarge surgical complications for cancer patients.



**FIGURE 4** Top contributing features to the predicted risk for an observation associated with readmission [Color figure can be viewed at wileyonlinelibrary.com]

In general, it may be challenging to recruit large numbers of patients in studies involving wearable devices. For instance, the number of participants in the aforementioned works using PGHD varied from 24 to 52. Similarly, our study relied on data from 52 patients (for which we considered up to 134 temporal observations). Under these constraints, there were not enough observations for a testing set separated from the training, and we trained/evaluated our models via nested group (patient) cross-validation. Nested cross-validation is a recommended approach to train and evaluate models when the sample size is small.<sup>17</sup>

There was an inherent risk of overfitting due to limited sample size, hence overestimating the predictive power of some features. Consequently, the small number of observations also limited the number of features that the model could have. For this reason, we chose to use a linear classifier (logistic regression), as nonlinear classification methods generally do not perform as well for low numbers of observations relative to the dimensionality of the problem. A larger cohort prospective study should include features from immediate discharge labs (e.g., markers of inflammation such as white blood cell count and C-reactive protein, markers of metabolic in-equilibrium, morning glucose, and creatinine).

Other limitations of our study include the fact that our dataset originated from two pilots not designed for machine learning applications and designed with different inclusion criteria. The first pilot enrolled adults (median age = 55), while the second pilot targeted only elderly patients (age > 65) with a median of 75 years. The first pilot enrolled only GI cancer patients, while the second pilot also included lung cancer patients. Finally, steps data in the first pilot were entered manually with the potential risk of errors. Despite these limitations, we believe the data and approach presented here to be novel and have added value to the available literature.

## 5 | CONCLUSIONS

We trained supervised learning models based only on clinical data collected up to discharge, with ePROs and wearable PGHDs collected during and after visits. Our results suggest that both PGHD and ePROs from patient telemonitoring have the potential to enable predictions to be performed after discharge and to improve overall prediction accuracy. Patient-centered outcomes obtained via wearable and web-based technologies may complement clinical data to predict postoperative complications. They seemed particularly useful to update risk prediction once a patient is discharged from the hospital. From our analysis on the impact of each feature, the findings suggest that a high risk of complications was associated with long lenght of stay, a precipitous increase of physical activity, as measured via linear trend in daily steps, and also high levels of symptom discomfort (i.e., pain, sleep, appetite). Future large-scale studies are needed to definitively establish the added predictive value of patient-centered outcomes in surgical oncology to inform treatment decision-making, postoperative care, and postoperative recovery trajectories.

#### ACKNOWLEDGMENTS

We would like to thank Dr. Chi Wah (Alec) Wong for his valuable tips on the methods. The research reported in this paper was supported by the National Cancer Institute of the National Institutes of Health under award number P30CA33572. It also included research that was supported by the City of Hope Center for Cancer and Aging. The content is solely the responsibility of the authors and does not necessarily represent the official view of the Center for Cancer and Aging, the National Cancer Institute, or the National Institutes of Health.

#### CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

#### AUTHOR CONTRIBUTIONS

Lorenzo A. Rossi designed and implemented the machine learning study. Virginia Sun, Laleh G. Melstrom, and Yuman Fong designed the original pilot studies and collected the data. Lorenzo A. Rossi and Virginia Sun drafted the manuscript. All authors contributed to data interpretation. Virginia Sun, Laleh G. Melstrom, and Yuman Fong provided clinical context and interpretation. All authors revised and approved the final manuscript.

#### DATA AVAILABILITY STATEMENT

The data associated with the current study is not publicly available as the data are not legally certified as being deidentified. Summary data may be available by contacting the corresponding author.

#### ORCID

Lorenzo A. Rossi D https://orcid.org/0000-0002-0689-9024 Laleh G. Melstrom D https://orcid.org/0000-0001-6279-7981 Yuman Fong D https://orcid.org/0000-0002-8934-9959

#### REFERENCES

- Basch E, Deal AM, Kris MG, et al. Symptom monitoring with patientreported outcomes during routine cancer treatment: a randomized controlled trial. J Clin Oncol. 2016;34:557-565.
- Bae S, Dey A, Low C. Using passively collected sedentary behavior to predict hospital readmission. Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2016:616-621.
- Doryab A, Dey A, Kao G, Low C. Modeling biobehavioral rhythms with passive sensing in the wild: a case study to predict readmission risk after pancreatic surgery. Proc ACM Interact Mob Wearable Ubiquitous Technol. 2019;3:1-21.
- Sun V, Dumitra S, Ruel N, et al. Wireless monitoring program of patient-centered outcomes and recovery before and after major abdominal cancer surgery. JAMA Surg. 2017;152:852-859.
- Lafaro KJ, Raz DJ, Kim JY, et al. Pilot study of a telehealth perioperative physical activity intervention for older adults with cancer and their caregivers. *Support Care Cancer.* 2020;28:3867-3876.
- Kooiman TJ, Dontje ML, Sprenger SR, Krijnen WP, van der Schans CP, de Groot M. Reliability and validity of ten consumer activity trackers. BMC Sports Sci Med Rehabil. 2015;7:24.
- Cleeland CS, Mendoza TR, Wang XS, et al. Assessing symptom distress in cancer patients: the M.D. Anderson Symptom Inventory. *Cancer.* 2000;89:1634-1646.

WILEY-SURGICAL ONCO

- Cleeland CS, Wang XS, Shi Q, et al. Automated symptom alerts reduce postoperative symptom severity after cancer surgery: a randomized controlled clinical trial. J Clin Oncol. 2011;29:994-1000.
- 9. Pedregosa F, Varoquaux G, Gramfort, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825-2830.
- Scott M, Lundberg, Su-In L. A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems, 2017:4765-4774. Curran Associates Inc
- Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning with electronic health records. NPJ Digit Med. 2018;1:18.
- Liu W, Stansbury C, Singh K, et al. Predicting 30-day hospital readmissions using artificial neural networks with medical code embedding. *PLOS One*. 2020;15:e0221606.
- Kalagara S, Eltorai AEM, Durand WM, DePasse JM, Daniels AH. Machine learning modeling for predicting hospital readmission following lumbar laminectomy. J Neurosurg Spine. 2018;30: 344-352.
- 14. Bolourani S, Tayebi MA, Diao L, et al. Using machine learning to predict early readmission following esophagectomy [published online ahead of print May 29, 2020]. *Journal of Thoracic and Cardiovascular Surgery*. 2020.

- Corey K, Sehj E, Lagoo-Deenadayalan S, Heller K, Whelan K, Balu S. Development and validation of machine learning models to identify high-risk surgical patients using automatically curated electronic health record data (Pythia): a retrospective, single-site study. PLOS Med. 2018;15:e1002701.
- Halloran S, Tang L, Guan Y, Shi J, Eyre J. Remote monitoring of stroke patients' rehabilitation using wearable accelerometers. *Proceedings of* the 23rd International Symposium on Wearable Computers. 2019:72-77.
- 17. Raschka S. Model evaluation, model selection, and algorithm selection in machine learning. *arXiv preprint*. 2018;1811:12808.

How to cite this article: Rossi LA, Melstrom LG, Fong Y, Sun V. Predicting post-discharge cancer surgery complications via telemonitoring of patient-reported outcomes and patient-generated health data. *J Surg Oncol.* 2021;123:1345-1352. https://doi.org/10.1002/jso.26413