Surgery 170 (2021) 790–796

Contents lists available at ScienceDirect

Surgery

journal homepage: www.elsevier.com/locate/surg

Innovation

Predicting need for hospital-specific interventional care after surgery using electronic health record data

Davy van de Sande, BSc[a,*], Michel E. van Genderen, MD, PhD[a], C. Verhoef, MD, PhD[b], Jasper van Bommel, MD, PhD[a], Diederik Gommers, MD, PhD[a], Edwin van Unen, Ir.[c], Joost Huiskens, MD, PhD[a,c], D.J. Grünhagen, MD, PhD[b]

[a] *Department of Adult Intensive Care, Erasmus University Medical Center, Rotterdam, The Netherlands*
[b] *Department of Surgical Oncology, Erasmus MC Cancer Institute University Medical Center, Rotterdam, The Netherlands*
[c] *Health Care Analytics, SAS Institute, Huizen, The Netherlands*

A R T I C L E I N F O

Article history:
Accepted 4 May 2021
Available online 3 June 2021

A B S T R A C T

Background: A significant proportion of surgical inpatients is often admitted longer than necessary. Early identification of patients who do not need care that is strictly provided within hospitals would allow timely discharge of patients to a postoperative nursing home for further recovery. We aimed to develop a model to predict whether a patient needs hospital-specific interventional care beyond the second postoperative day.
Methods: This study included all adult patients discharged from surgical care in the surgical oncology department from June 2017 to February 2020. The primary outcome was to predict whether a patient still needs hospital-specific interventional care beyond the second postoperative day. Hospital-specific care was defined as unplanned reoperations, radiological interventions, and intravenous antibiotics administration. Different analytical methods were compared with respect to the area under the receiver-operating characteristics curve, sensitivity, specificity, positive predictive value, and negative predictive value.
Results: Each model was trained on 1,174 episodes. In total, 847 (50.5%) patients required an intervention during postoperative admission. A random forest model performed best with an area under the receiver-operating characteristics curve of 0.88 (95% confidence interval 0.83–0.93), sensitivity of 79.1% (95% confidence interval 0.67–0.92), specificity of 80.0% (0.73–0.87), positive predictive value of 57.6% (0.45–0.70) and negative predictive value of 91.7% (0.87–0.97).
Conclusion: This proof-of-concept study found that a random forest model could successfully predict whether a patient could be safely discharged to a nursing home and does not need hospital care anymore. Such a model could aid hospitals in addressing capacity challenges and improve patient flow, allowing for timely surgical care.

Introduction

Healthcare facilities and clinicians are challenged to maintain balance in efficiently distributing healthcare resources on the one hand, and on the other hand physicians need to effectively treat each patient according to their specific condition. In recent years, hospitals have experienced an increasing demand for medical services such as surgical care. One of the factors involved in providing timely surgical care is the availability of postoperative beds. Nevertheless, hospitals have limited postoperative beds to accommodate inpatients, so there is a need for efficient and creative capacity management without compromising quality of care.

The decision to discharge surgical inpatients is multifactorial with a combination of patient progress in recovery and the clinical expertise of the medical-surgical team. Among surgical patients, delay in discharge is a common and longstanding problem.[1] Several barriers prevent surgical patients from being discharged on a daily basis; these include both clinical as well as nonclinical reasons.[2] As a consequence, discharge is frequently postponed and hospital stay prolonged. Prolonged stay is associated with an increased risk on

* Reprint requests: Davy van de Sande, Erasmus Medical Center, Department of Adult Intensive Care, Room Ne-403, Doctor Molewaterplein 40, 3015 GD Rotterdam, The Netherlands.
E-mail address: d.sande@erasmusmc.nl (D. van de Sande);
Twitter: @ErasmusMC

https://doi.org/10.1016/j.surg.2021.05.005

hospital-acquired infections, poor nutrition levels, and other complications.[3,4] Moreover, lack of optimal patient flow leads to waits, delays, and diversions.[5]

Electronic health record (EHR) systems contain increasing amounts of valuable detailed patient information and are widely used within hospitals. Yet these huge amounts of data, however, do not aid discharge decision-making.[6] After surgery, early identification of patients who do not need care that is strictly provided within hospitals would allow timely discharge of patients to a postoperative nursing home for further recovery. However, in many cases this remains difficult for clinicians to estimate shortly after surgery, and the amount of these data is too large to be processed by clinicians.

Machine learning (ML) can discover patterns in large amounts of complex patient data and, as such, can uncover clinically relevant information.[7] ML models improve through experience rather than act upon preprogrammed rules. Some examples of ML models in medicine consist of models to detect diabetic retinopathy and skin cancer.[8,9]

In the current era, capacity management is important and thus unnecessary hospital bed occupation should be avoided. In such decision-making it is crucial to know whether a patient still needs hospital care in terms of interventions such as reoperations, radiological interventions, or intravenous antibiotics. The aim of this study was therefore to develop and internally validate a model to predict whether a patient still needs hospital-specific interventional care 2 days after surgery or can be safely discharged.

Methods

This study was a single-center retrospective cohort study. All research data were readily available and automatically extracted from the EHR system of the Erasmus MC University Medical Center, Rotterdam, The Netherlands, which is a tertiary hospital. The protocol of this study was approved by the Medical Research Ethics Committee (reference number: MEC-2019–0728) of the Erasmus MC University Medical Center, who waived the need for consent due to the nature of the study. All data were anonymized. The study was prepared in accordance with the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) statement.[10]

Setting and participants

The setting included 1 inpatient medical-surgical unit with 34 floor beds. To be eligible, (1) patients must have undergone surgery in the surgical oncology department; and (2) patients' intra- and postoperative data had to be discoverable from the EHR system. Eligible patients were adults (≥18 years) who were discharged between June 2017 and February 2020.

Outcomes

We defined care that is strictly provided within hospitals (herein referred to as "hospital-specific interventional care") as reoperations, radiological interventions, and intravenous antibiotics administration. Readmissions were not included since the aim of our study was to predict whether a patient can be safely discharged to a postoperative nursing home. We selected the distinct/unique types of medication that were administrated. For all other data items we collected unaggregated values. The primary outcome was determined as the need for hospital-specific interventional care beyond 2 days after surgery (ie, postoperative day 2). As such, for each patient, the outcome was predicted on postoperative day 2. Hospital-specific interventional care was defined as all reoperations, radiological interventions, and administration of intravenous antibiotics.

The secondary outcome was the predicted total number of avoidable bed-days for 2019. Avoidable bed-days per patient were defined as the number of days a patient remained admitted after the model predicted that a patient was no longer in need of hospital-specific interventional care.

Predictors

For model predictors, we used routinely collected data from (1) the preoperative anesthesia assessment, (2) patient demographic information and comorbidities, (3) admission information, (4) multidisciplinary board notes, (5) surgery details, (6) clinical test results, (7) medication administration, and (8) bedside assessments (see Table I). Most data used were structured, including either numerical values (eg, age, clinical test results, and laboratory results) or categorical values chosen out of predetermined text phrases (eg, intoxications and smoking status).

Preprocessing data

Before initiation of data analysis, the data set was deidentified and all identifiable patient data was removed. For each predictor in the data set, the amount of missing data was assessed. Predictors with 50% or more missing were rejected for modeling by default. For regression-based models such as logistic regression and neural networks, missing values were imputed using multiple imputation. For tree-based models, missing values were treated as a separate category for splitting criteria. In addition, we evaluated standardization for scaling of continuous variables.

To make the data understandable for the computer, we had to apply logic to the data. Since a patient could have been accommodated multiple times within the study period, we created unique episodes for every time a patient was admitted for a surgical procedure. The start of an episode was the date of primary surgery, and the end of an episode was the date of discharge in the EHR, with admission number corresponding to the date of admission.

In addition, in order to make hospital-specific interventions identifiable in the data, we used the following approach: hospital-specific interventions were defined as all reoperations, radiological interventions, and administration of intravenous antibiotics that took place after the primary surgical procedure, within 1 episode.

Statistical analysis and model validation

Baseline characteristics and outcomes are presented as count and percentage (%) for categorical variables and as median, interquartile range for continuous variables.

We calculated the sample size following the method proposed by Riley et al.[11] In the absence of information on the prevalence of the prespecified outcome of interest, the prevalence was estimated based on the experience of the medical-surgical team at half of the surgical inpatients (50%). The number of predictors for the prediction model was restricted to a maximum of 25. A minimum sample size of 1,374 patients was required.

Due to the complexity and heterogeneity of the data, we considered traditional logistic regression models next to more advanced analytical methods: neural networks, random forest, decision trees, and gradient boosting. The decision to use ML techniques is motivated by our assumption that the classification problem (in need versus not in need of hospital-specific interventional care) is associated with both clinical as well as nonclinical variables, which interact nonlinearly.[12] The performance of different ML techniques may vary on different types of data,

Table I
Overview of all collected patient data

Clinical data category	Data item
Patient demographic information	Age
	Sex
Surgery information	Actual duration of surgery
	Anesthesia type used (eg, general, epidural)
	Expected duration of surgery
	Intubation conditions
	Surgical procedure type
	Surgical urgency (emergency versus elective)
Clinical test results	Laboratory values
	Radiologic (eg, computed tomography scans, x-rays)
Admission information	Admission type (eg, clinical)
	Emergency room registration
	Postoperative destination
	Preoperative origin location
Preoperative anesthesia assessment	ASA score
	Chemotherapy history
	Intoxications (eg, smoking, alcohol)
	Transfusion history
	Presence of irregular antibodies
	Prophylaxis policy
Bedside assessments	Cognitive assessments (eg, neurological and delirium risk assessment)
	Fluid intake
	Fluid output
	Vital signs (eg, heart rate, blood pressure, temperature)
Multidisciplinary board notes	Medical history (eg, prior surgery, prior admissions)
	Multidisciplinary board type
	Therapy recommendation
	TNM stage
Medication administration	Dose administrated
	Frequency of administration
	Route of administration (eg, oral, IV)
	Type of medication

ASA, American Society of Anesthesiologists; *IV*, intravenous; *TNM*, tumor nodes metastases.

depending on the characteristics of the predictors and the outcome. Therefore, we applied several ML techniques and compared predictive performances. Model hyperparameters (eg, number of hidden layers and nodes for the neural network), leaf size, and number of trees (for tree-based algorithms) were tuned through an iterative process taking into account the predictive performance and the degree of overfitting. Supervised learning was used for training of the models. The models are considered to be "supervised" since they are trained to predict a target (need for hospital-specific interventional care). These training data are the inputs and outcomes for the training data set (ie, the classification whether a patient will need hospital-specific interventional care or not) and are known in retrospect.

Each model was internally validated by randomly partitioning the total data set into 70% training data, 20% validation data, and 10% test data.[13] The partition was determined arbitrarily and ensured each model was trained and validated before finally testing on an independent data set the model had not seen before.[14] We assessed the discriminative performance of the models by generating the area under the receiver-operating characteristics (AUROC) curve.[14,15] We used each model to predict on the test data set and compared performance with respect to misclassification rate, sensitivity (%), specificity (%), positive predictive value (PPV) (%), and negative predictive value (NPV) (%). The optimal classification threshold was calculated for each model on the validation data set using the Youden's J statistic.[16]

For 2019, the total number of avoidable-bed days saved was calculated by summing all avoidable bed-days saved per patient from all patients who underwent surgical care in 2019. We penalized for misclassification by multiplying the avoidable bed-days by the predicted probability of needing hospital-specific interventional care. All analyses were carried out in SAS Viya version 8.3. For modeling specifically the SAS Visual Data Mining and ML application was used.

Results

The characteristics of patients used to train, validate, and test the prediction model are presented in Table II. In total, 1,677 episodes were included. The median age was 63 (interquartile range: 51–72) years, and 837 (50%) were men. The training set included 1,174 episodes, the validation set 335 episodes, and the test set 168 episodes.

The majority of the surgical procedures' urgency were elective (1.514 [90.3%]), 163 (9.7%) were emergency surgical procedures, and none of the reoperations were planned. An intervention occurred in 847 (50.5%) out of 1,677 episodes. The frequency of hospital-specific interventions is presented in Table III. The majority of interventions were administration of intravenous antibiotics [$n = 588$ (35%)]. Figure 1A demonstrates the distribution of the length of stay after surgery, which is right skewed. Five hundred and eighty (34.6%) patients were admitted for a maximum of 2 days (Fig 1B).

Model performance

Based on predictive performance, logistic regression, gradient boosting, and random forest were explored in further detail and corresponding hyperparamters were optimized (Fig 2). Standardization was used to scale continuous variables for random forest and gradient boosting. Compared to logistic regression and gradient boosting, random forest (hyperparameters: 100 trees, 5 variables per split, 50 bins, and 2 branches) achieved the best predictive performance on the test data set, with a misclassification rate of 0.179 and an AUROC of 0.88 (95% CI 0.83–0.93) (logistic regression 0.89 [0.84–0.94] and gradient boosting 0.84 [0.79–0.89]) (Fig 3). The optimal classification threshold for the random forest was 0.25 with a corresponding Youden's J statistic of 0.62. This means when a patient's predicted probability exceeds 25%, the model will predict a patient to be in need for hospital-specific interventional care. On the test data set the random forest had a sensitivity of 79.1% (95% CI 0.67–0.92), specificity of 80.0% (0.73–0.87), PPV of 57.6% (0.45–0.70), and NPV of 91.7% (0.87–0.97) (Table IV).

To ensure that the random split of the data set did not erroneously influence the results, we further stabilized the analysis. We used the total data set and additionally validated the random forest using 20-fold cross validation. On average, the random forest had an AUROC of 0.96 (95% CI 0.93–0.98), sensitivity of 97.0% (95% CI 0.96–0.98), specificity of 83.8% (0.82–0.85), PPV of 67.4% (0.66–0.68), and NPV of 98.8% (0.98–0.99) over the 20 folds of cross-validation.

In order to make the model more accessible to the medical-surgical team, we constructed a nomogram representing the relative importance of predictors included in the prediction model in a graphical format (Fig 4). The final model used 17 different predictors. Length of time in the operating room turned out to be very predictive compared to the other predictors (Fig 4). We carried out a sensitivity analysis, which is graphically presented in Supplemental Figure S1.

In 2019, patients in 565 out of 976 episodes remained admitted after 2 days. The random forest classified patients in 116 out of 565 episodes as not in need of hospital-specific interventional care. After penalizing for misclassification, the total predicted avoidable bed-days were 187 days for 2019.

Table II
Characteristics of the training, validation and test data set

Characteristic	Patients, no. (%) Training	Validation	Test
No. of unique episodes	1,174 (70)	335 (20)	168 (10)
Sex			
Male	594 (50.6)	165 (49.3)	78 (46.4)
Female	580 (49.4)	170 (50.7)	90 (53.6)
Surgical urgency			
Elective	1,065 (90.7)	298 (89)	151 (89.9)
Emergency	109 (9.3)	37 (11)	17 (10.1)
Surgical type			
Adrenal gland	21 (1.8)	3 (0.9)	2 (1.2)
Breast	132 (11.2)	40 (11.9)	21 (12.5)
Colon and rectum	128 (10.9)	31 (9.3)	23 (13.7)
Diagnostic laparoscopy	37 (3.2)	14 (4.2)	7 (4.2)
Esophagus	86 (7.3)	16 (4.8)	5 (3)
Hernia	15 (1.3)	6 (1.8)	3 (1.8)
Liver	70 (6)	26 (7.8)	10 (6)
Lymph node dissection	99 (8.4)	28 (8.4)	10 (6)
Melanoma	27 (2.3)	4 (1.2)	2 (1.2)
Ostomy	33 (2.8)	8 (2.4)	1 (0.6)
Sarcoma	81 (6.9)	21 (6.3)	18 (10.7)
Stomach	62 (5.3)	19 (5.8)	8 (4.8)
Thyroid gland	66 (5.6)	20 (6)	12 (7.1)
Others	317 (27)	99 (29.6)	46 (27.4)

Table III
Frequency of hospital-specific interventions

Intravenous antibiotics	Reoperations	Radiological interventions	Number of patients
Yes	Yes	Yes	26
		No	55
	No	Yes	135
		No	588
No	Yes	No	4
	No	Yes	39
		No	830

Reoperations were all unplanned. There were no prespecified planned reoperations in our data set. Eight hundred and forty-seven patients (50.5%) required at least 1 intervention; 4 only a reoperation, 39 only a radiological intervention, and 588 only intravenous antibiotics.

Discussion

This retrospective cohort study examined the concept of constructing a model to predict whether a patient who had undergone surgery would need hospital-specific interventional care during admission. In studying this concept, a random forest model demonstrated strong discrimination with an AUROC of 0.88 (0.83–0.93), sensitivity of 79.1% (0.67–0.92), and specificity of 80.0% (0.73–0.87). Using the prediction model on the patients in our institution, the total number of predicted avoidable bed-days was 187 days in 2019.

Our findings contribute to the limited amount of available literature on prediction models designed to aid clinicians and improve clinical processes. Previous studies examined whether ML could be used to predict length of stay in surgical patients, defined as the total amount of time spent in a hospital.[17,18] Furthermore, several studies developed preoperative risk prediction models able to predict adverse postoperative outcomes and subsequently demonstrated that such models have high predictive value, can be implemented during the consent procedure, and improve patient satisfaction.[19–22] Nonetheless, no other study constructed a prediction model able to predict whether a patient will need hospital-specific interventional care shortly (ie, 2 days) after surgery. Previous models focused solely on the development of postoperative complications instead of focusing on the need for hospital needed care.[23,24] In comparison, the current model takes postoperative complications into account, but in particular the need for hospital-specific interventions. This model is the first of its kind, and we therefore believe that such a model can not only predict postoperative complications, but also could aid clinicians in addressing capacity challenges and, as such, may improve clinical patient flow. Future studies should determine whether such a prediction model could enable clinicians in decision-making to safely discharge a postoperative patient from the hospital. We only used data that originated from the EHR system and that are accessible in most hospital information systems; thus it can be reproduced in other hospitals. However, when applying this model to another hospital it must be taken into account that there may exist differences in the available predictors in the EHR system across institutions. We explicitly chose to use routinely collected data so that no additional information needs to be collected by the medical-surgical team and to ensure that the model can be easily adopted into clinical practice in the future. In addition, the predictors eventually used by the random forest model are part of routine data collection in most institutions. Moreover, the predictors used for the development of the ML model are in line with currently available variables in clinical practice. Therefore, the ML model allows for updates either periodically as new data are acquired or in real time. Updating the model with new available data is an essential step toward establishing a self-learning healthcare system.[6]

In this study, different prediction models based on different techniques were evaluated. We evaluated logistic regression as well as more sophisticated ML techniques. We compared the performance of different models with respect to AUROC, sensitivity, specificity, PPV, NPV, and misclassification rate. In this study, a random forest outperformed the other models on the test data set. In future, even more predictors can be created by combining different data, potentially resulting in better discrimination performance.

In this model, a challenge remains to find the “appropriate” probability threshold (eg, a threshold of 0.9 or 90% can be interpreted as: when the predicted probability for a given patient is >0.9 or 90%, then the model will classify the patient as “not in need for hospital-specific interventional care”) above which a patient is predicted to be not in need of hospital-specific interventional care. This probability threshold can be flexibly adjusted according to different circumstances. In the event of a shortage of beds, a hospital can decide to lower the threshold to improve patient flow and increase bed capacity. However, the impact on the readmission rate must be carefully considered before deciding to lower the threshold. In a scenario where there is no pressure on bed capacity, a hospital can increase the threshold to improve the predictive accuracy. The challenge in determining the appropriate probability threshold is to make a trade-off between the hospital needs (eg, bed capacity and improved patient flow) and the number of readmissions a hospital considers acceptable (in this article we do not elaborate on readmissions since it was not one of our outcomes). Thus, the appropriate threshold really depends on the organization itself.

Of note, our model was developed for the purpose of optimizing patient flow and the discharge process and not replacing decisions of the medical-surgical team. The output was solely intended to aid clinicians and interact with their knowledge not defined in the model. Thus, the final decision to discharge a patient remains within control of the medical-surgical team, taking into account the new information provided by the model.

Although several studies have already shown the potential of ML in medicine, it remains challenging to balance assembling a large, representative, and diverse data set with privacy and regulatory requirements.[6,17,18,25–30] A key issue for many applications of

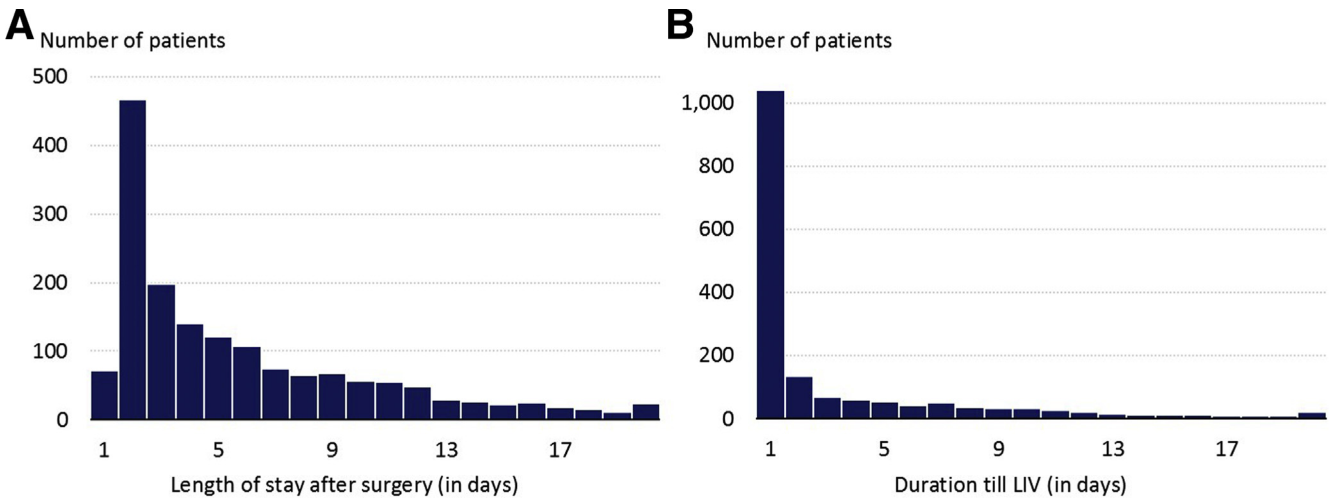

Fig 1. (A) Distribution of the length of stay after surgery in days. (B) Distribution of the number of days between surgery and the last hospital-specific intervention that took place. *LIV*, last hospital-specific intervention.

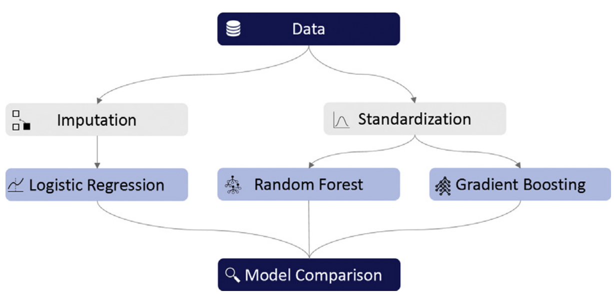

Fig 2. Overview of the prediction model development process.

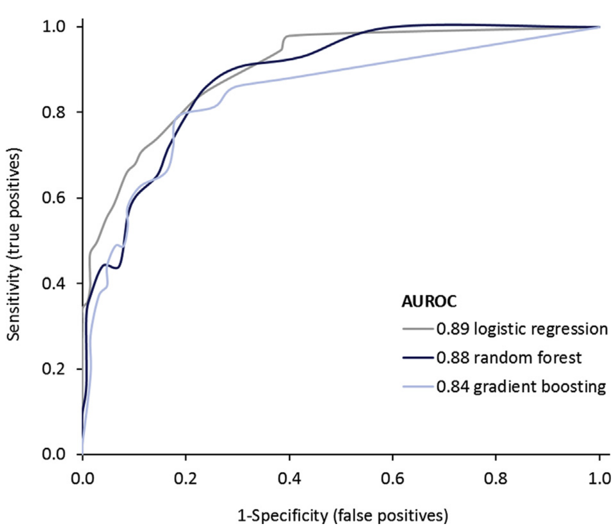

Fig 3. Receiver-operating characteristic (ROC) curves. Discriminative performance of random forest, logistic regression, and gradient boosting on the test data set ($n = 168$ patients). *AUROC*, area under the receiver-operating characteristic curve.

ML is the percentage of initiatives actually engaging in effective scaling practices. This is as low as 8% in several sectors such as banking, economics, healthcare, and insurance.[31] As such, not many ML models have contributed meaningfully to clinical care.[32] Thus, there is a need for clinical trials in which the performance of ML models is evaluated in real-world clinical practices, in order to close the gap between theoretical proof-of-concept studies and daily clinical practice.[30] In this study, we developed and internally validated the prediction model. In future studies, we will address the external validity of the currently presented model. Furthermore, we will evaluate feasibility and reliability in clinical practice to eventually determine the impact on clinical decision-making.

Limitations

Although our model was able to accurately predict the need for hospital-specific interventional care, our findings must be considered in light of several limitations. First, our model was developed and internally validated based on data from a single tertiary medical center in The Netherland, therefore generalizability may be limited. A strength of our study, however, is that our model building is transparent and assessment can easily be applied to and replicated in other healthcare facilities. Future studies should address the external validity and clinical feasibility of our model.

Second, we attempted to create a diverse and at the same time highly representative data set; treatment of patients in the surgical oncology department has changed over time. To allow for minimal change in treatment over time, only patients who underwent surgery within a timeframe of 2.5 years before the start of the study were eligible. Moreover, some data relied on manual input of medical care workers and are hence prone to human error.

Third, the input for our model was limited to those variables entered into the EHR. If data were not entered in the EHR (ie, verbal discussions regarding discharge), they could not be included in our model. Another issue is that we were not able to include the main diagnosis for which a patient underwent surgical care in the prediction model. This is due to the fact that we could not link the correct diagnosis from the EHR to the surgical procedure. We believe that the main diagnosis is important information that may improve the predictive power of the model even further.

In conclusion, this proof-of-concept study found that a random forest model could predict whether a patient could be safely discharged to a nursing home after the second postoperative day. Such a model can potentially be used to aid hospitals in addressing capacity challenges and to improve patient flow by reducing the number of avoidable bed-days. A future prospective study should

Table IV
Model performance results

Model	AUROC (95% CI)	Sensitivity, % (95% CI)	Specificty, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	Optimal threshold*
Random forest	0.88 (0.83–0.93)	79.1% (0.67–0.92)	80.0% (0.73–0.87)	57.6% (0.45–0.70)	91.7% (0.87–0.97)	0.25
Logistic regression	0.89 (0.84–0.94)	78.9% (0.66–0.91)	81.5% (0.75–0.88)	55.6% (0.42–0.69)	92.9% (0.88–0.98)	0.20
Gradient boosting	0.84 (0.79–0.89)	79.1% (0.67–0.92)	81.6% (0.75–0.88)	59.6% (0.47–0.72)	91.8% (0.87–0.97)	0.20

AUROC, area under the receiver-operating characteristics; *CI*, confidence interval; *NPV*, negative predictive value; *PPV*, positive predictive value.
* The optimal classification threshold was calculated on the validation dataset using the Youden's J statistic.

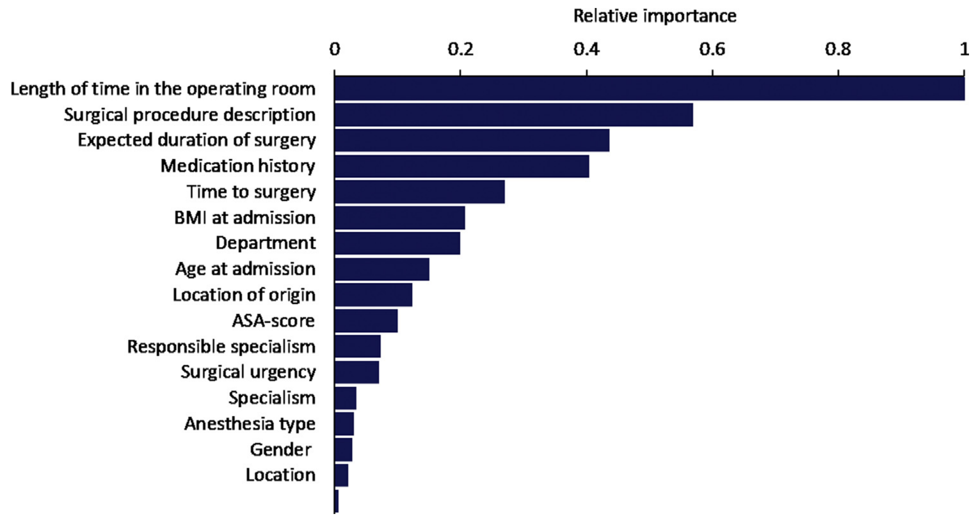

Fig 4. Nomogram representing the relative importance of predictors used by the random forest model. *ASA*, American Society of Anesthesiologists; *BMI*, body mass index.

evaluate the feasibility and reliability of such a model in clinical practice.

Funding/Support

The authors have no funding to declare.

Conflict of interest/Disclosures

DG has received speakers fees and travel expenses from Dräger, GE Healthcare (medical advisory board 2009–12), Maquet, and Novalung (medical advisory board 2015–2018). All other authors declare no competing interests.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [https://doi.org/10.1016/j.surg.2021.05.005].

References

1. Strang IW, Boddy FA, Jennett B. Patients in acute surgical wards: a survey in Glasgow. *Br Med J*. 1977;1:545–548.
2. Safavi KC, Khaniyev T, Copenhaver M, et al. Development and validation of a machine learning model to aid discharge processes for inpatient surgical care. *JAMA Netw Open*. 2019;2, e1917221.
3. Caccialanza R, Klersy C, Cereda E, et al. Nutritional parameters associated with prolonged hospital stay among ambulatory adult patients. *CMAJ*. 2010;182:1843–1849.
4. Lucas R, Farley H, Twanmoh J, et al. Emergency department patient flow: the influence of hospital census variables on emergency department length of stay. *Acad Emerg Med*. 2009;16:597–602.
5. Optimizing Patient Flow: Moving Patients Smoothly Through Acute Care Settings. IHI Innovation Series white paper. Boston: Institute for Healthcare Improvement; 2003. Available from: www.IHI.org.
6. Krumholz HM. Big data and new knowledge in medicine: the thinking, training, and tools needed for a learning health system. *Health Aff (Millwood)*. 2014;33:1163–1170.
7. Johnson AE, Ghassemi MM, Nemati S, Niehaus KE, Clifton DA, Clifford GD. Machine learning and decision support in critical care. *Proc IEEE Inst Electr Electron Eng*. 2016;104:444–466.
8. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316:2402–2410.
9. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542:115–118.
10. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Br J Surg*. 2015;102:148–158.
11. Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ*. 2020;368:m441.
12. Breiman L. Random forests. *Mach Learn*. 2001;45:5–32.
13. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning, Data Mining, Inference and Prediction*. Corrected 12th printing ed. Stanford: Springer; 2017.
14. Andrew PB. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*. 1997;30:1145–1159.
15. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143:29–36.
16. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3:32–35.
17. Danilov G, Kotik K, Shifrin M, Strunina U, Pronkina T, Potapov A. Prediction of postoperative hospital stay with deep learning based on 101 654 operative reports in neurosurgery. *Stud Health Technol Inform*. 2019;258:125–129.
18. Daghistani TA, Elshawi R, Sakr S, Ahmed AM, Al-Thwayee A, Al-Mallah MH. Predictors of in-hospital length of stay among cardiac patients: a machine learning approach. *Int J Cardiol*. 2019;288:140–147.
19. Meguid RA, Bronsert MR, Juarez-Colunga E, Hammermeister KE, Henderson WG. Surgical Risk Preoperative Assessment System (SURPAS) I: parsimonious, clinically meaningful groups of postoperative complications by factor analysis. *Ann Surg*. 2016;263:1042–1048.
20. Meguid RA, Bronsert MR, Juarez-Colunga E, Hammermeister KE, Henderson WG. Surgical Risk Preoperative Assessment System (SURPAS) II: parsimonious risk models for postoperative adverse outcomes addressing need for laboratory variables and surgeon specialty-specific models. *Ann Surg*. 2016;264:10–22.
21. Meguid RA, Bronsert MR, Juarez-Colunga E, Hammermeister KE, Henderson WG. Surgical Risk Preoperative Assessment System (SURPAS): III:

accurate preoperative prediction of 8 adverse outcomes using 8 predictor variables. *Ann Surg*. 2016;264:23–31.
22. Wiesen BM, Bronsert MR, Aasen DM, et al. Use of Surgical Risk Preoperative Assessment System (SURPAS) and patient satisfaction during informed consent for surgery. *J Am Coll Surgeons*. 2020;230:1025.
23. Hicks CW, Bronsert M, Hammermeister KE, et al. Operative variables are better predictors of postdischarge infections and unplanned readmissions in vascular surgery patients than patient characteristics. *J Vascular Surg*. 2017;65:1130.
24. Corey KM, Kashyap S, Lorenzi E, et al. 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.
25. Barnes S, Hamrock E, Toerper M, Siddiqui S, Levin S. Real-time prediction of inpatient length of stay for discharge prioritization. *J Am Med Inform Assoc*. 2016;23:e2–e10.
26. Levin SR, Harley ET, Fackler JC, et al. Real-time forecasting of pediatric intensive care unit length of stay using computerized provider orders. *Crit Care Med*. 2012;40:3058–3064.
27. Wasfy JH, Kennedy KF, Masoudi FA, et al. Predicting length of stay and the need for postacute care after acute myocardial infarction to improve healthcare efficiency. *Circ Cardiovasc Qual Outcomes*. 2018;11, e004635.
28. Ramkumar PN, Karnuta JM, Navarro SM, et al. Deep learning preoperatively predicts value metrics for primary total knee arthroplasty: development and validation of an artificial neural network model. *J Arthroplasty*. 2019;34: 2220–2227 e1.
29. Carter EM, Potts HW. Predicting length of stay from an electronic patient record system: a primary total knee replacement example. *BMC Med Inform Decis Mak*. 2014;14:26.
30. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med*. 2019;380:1347–1358.
31. Fleming O, Fountaine T, Henke N, Saleh T. *Ten red flags signaling your analytics program will fail*. Sydney: McKinsey Analytics; 2018. Available from: https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/ten-red-flags-signaling-your-analytics-program-will-fail.
32. Artificial intelligence in health care: within touching distance. *Lancet*. 2017;390:2739.