Title

Exploring Surgical Infection Prediction: A Comparative Study of Established Risk Indexes and a Novel Model

Authors

Kjersti Mevik [1,2] M.D., Ph.D., Corresponding author and request for reprint

1 Nordland Hospital, Department of Surgery, 8092 Bodø, Norway. E-mail: kjersti.mevik@nlsh.no, phone: +4797123977

2 University of Calgary, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada

Ashenafi Zebene Woldaregay[3] Ph.D

3 University Hospital of North Norway, SPKI – the Norwegian Centre for Clinical Artificial Intelligence, 9019 Tromsø, Norway

Alexander Ringdal[4] MSc

4 Nordland Hospital Trust, Department of e-health, 8092 Bodø, Norway

Karl Øyvind Mikalsen [3,5] Ph.D.

5 UiT The Arctic University of Norway, Department of Clinical Medicine, 9019 Tromsø, Norway

Yuan Xu[6] M.D., Ph.D.

6 University of Calgary, Departments of Oncology, Community Health Sciences, and Surgery, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada

Keywords: Prediction, surgical site infection, global trigger tool, risk indexes, logistic regression model

Abstract

Background: Surgical site infections are a major health problem that deteriorates the patients' health and increases health care costs. A reliable method to identify patients with modifiable risk of surgical site infection is necessary to reduce the incidence of them but data are limited. Hence the objective is to assess the predictive validity of a logistic regression model compared to risk indexes to identify patients at risk of surgical site infections.

Methods: In this study, we evaluated the predictive validity of a new model which incorporates important predictors based on logistic regression model compared to three state-of-the-art risk indexes to identify high risk patients, recruited from 2016 to 2020 from a medium size hospital in North Norway, prone to surgical site infection.

Results**:** The logistic regression model demonstrated significantly higher scores, defined as high-risk, in 110 patients with surgical site infections than in 110 patients without surgical site infections ($p<0.001$, CI 19-44) compared to risk indexes. The logistic regression model achieved an area under the curve of 80%, which was better than the risk indexes SSIRS (77%), NNIS (59%), and JSS-SSI (52%) for predicting surgical site infections. The logistic regression model identified operating time and admission type as the major predictors of surgical site infections.

Conclusions: The logistic regression model demonstrated better performance in predicting surgical site infections compared to three state-of-the-art risk indexes. The model could be further developed into a decision support tool, by incorporating predictors available prior to surgery, to identify patients with modifiable risk prone to surgical site infection.

1. Introduction

Surgical site infections (SSIs) affect approximately 10% of surgical patients and aggravate the rehabilitation phase with an increased risk of extended hospital stays, readmissions, and postoperative deaths, as well as increasing hospital costs[1–3]. SSIs accounted for 40% of all postoperative complications and 15% of hospital costs identified by the record review method Global Trigger Tool (GTT) in the report “Patient injuries in Norway in 2020”[4]. Worldwide, postoperative death is the third most common cause of death, and one in three postoperative deaths is associated with SSIs[5]. Considering that SSIs have a major impact on morbidity, mortality, and hospital expenses, it is important to identify patients at risk of SSI for early interventions to modify the risks[6]. Several risk factors have been addressed in the literature and included in risk indexes, but few indexes are in clinical use due to suboptimal performance in the risk prediction of SSI or that the risk factors are difficult to modify; thus, a different approach for SSI risk prediction is needed[7,8].

Prediction models should be based on logistic regression (LR) models, which have shown reasonable results and are the preferred method of choice for risk stratification because of their simplicity and interpretability instead of regular risk indexes. But resent reviews shows that artificial intelligence-enabled decision support in surgery is lacking scientific quality[9]. However, to the best of our knowledge, no study has compared risk indexes with LR models. Therefore, the purposes of this study were to i) evaluate the validity of the three state-of-the-art risk indexes to predict the risk of SSI, ii) compare the predictive validity of the indexes to a LR model, and iii) explore which risk factors most significantly increase the risk of SSI. Our goal was to assess and pinpoint the major, and preferably modifiable, predictors for SSI to improve risk stratification.

2. Methods

This study is a prediction and internal validation of a SSI prediction model, using a dataset in which a statistical procedure has been applied to increase the event rate. We have used the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) guidelines for reporting prediction model studies to include all relevant information[10].

2.1. Study population and data

The study included patients who had undergone surgery and were then randomly selected for GTT review at a medium-sized hospital in North Norway between January 1, 2016, and December 31, 2020. The record review method, GTT, is currently considered the most suitable method to measure adverse events due to treatment given in healthcare[11,12]. GTT is a two-step method in which patient records are screened for specific situations (e.g., triggers) that could indicate that an adverse event has occurred, and the screening is for this study performed by the Nordic Clinical Automatic Framework (NCAF from SAS Institute © software), which was developed to identify triggers based on algorithms for both structured and unstructured data[13]. For structured data, the algorithms are built on rule-based algorithms, but for unstructured data, such as "patient fall", text analyses with natural language processing were used. If a trigger is identified, a more exhaustive review of the record is performed manually to verify if the trigger represents an adverse event. The incidents of SSIs for the study were obtained from the GTT review which is more sensitive than using administrative data such as ICD 10 codes (International Classification of Diseases and health problems, version 10)[14].

Surgery was defined as a procedure that included a skin incision performed in the operating theatre. The included patients were admitted for at least 24 hours and were over 18 years of age and extracted randomly from the seven main departments in pools of ten and ten patients every second week from the discharged patient lists. The study was performed in accordance with the Helsinki Declaration of 1975 and approved by the data protection official in the Nordland Hospital Trust and by the Norwegian Regional Ethics Committee (ref 2021/343618). The committee waived the requirement for obtaining written informed consent.

SSI was defined as a wound infection that led to antibiotic treatment, prolonged hospital stay, immediate intervention, or death. Seventeen unique risk factors (see Table 2) included in the risk indexes, in addition to age and length of stay (LOS), were identified in the patient records by manually reviewing the records and included as variables in the study. Patients with incomplete or missing surgical data were excluded from the study.

2.2. Risk indexes NNIS, SSIRS, and JSS-SSI

We choose three state-of-the art risk indexes; NNIS, SSIRS and JSS-SSI, to compare their performances to a novel logistic regression model. The rationale behind the selection of these three indexes was mainly due to their extensive evaluations in literature.

The National Nosocomial Infections Surveillance (NNIS) was developed from the first risk index for predicting SSI in the 1980s named as the Efficacy of Nosocomial Infection Control (SENIC)[15,16]. The NNIS uses three variables: American Society of Anesthesiology (ASA) score ≥ 3, contaminated wound, and above-average surgical time. Instead of the stringent cut-off time of 2 hours used in SENIC, the NNIS includes a procedure-specific surgical time based on the 75th percentile of the operating time of the individual procedure being performed. The tool stratifies patients into low (one feature present), moderate (two features present), or high (all three features present) risk of SSI[16].

Surgical Site Infection Risk Score (SSIRS) was developed in 2013, available via a web-based calculator, where the patient and procedure data are plotted, and the score is calculated based on these data[17]. SSIRS includes 13 variables where five of them are binary variables, answered with yes or no if the variable is present and the rest are categorical variables. The risk score ranges from 0 to 1 which equals 0-100%[17].

The third index JSS-SSI, a Canadian risk scoring tool from 2018, uses five variables: elevated care at discharge (add 42 points), surgery time > 3 hours (add 19 points), admission before surgery (add 24 points), general, gynecologic, ear-nose-throat, thoracic, or urologic surgery (add 22 points), and contaminated or infected wound (add 19 points). The score ranges from 0 to 100 points where the low risk of SSI ranges from 0 to 42 points, moderate risk ranges from 43 to 58 points, and the high risk of SSI ranges from 59 to 100 points[18,19].

2.3. Development of a new SSI prediction model based on logistic regression model

The LR model was developed based on the variables used in all three risk indexes. Categorical variables were encoded using one-hot encoding, and numerical variables were rescaled using a Min-Max scaler. Any missing values were imputed with simple fill, where the mean value is used for numerical variables, and the most frequent value is used for categorical variables. The variables incorporated several overlapping variables such as ASA, contaminated wounds, and operating time. To this end, pairwise correlation and variance inflation factors were used to estimate the degree of correlation and variance inflation caused by the variable. Consequently, the variables that exhibited significant correlations, with a threshold of 0.7, were removed from the variable set. This results in excluding ASA $\geq$ 3, above-average operating time, and contaminated wounds but including ASA and wound type (clean, contaminated, or infected). Operating time > 3 hours was also excluded as we included hours of operating time. It is worth noting that while removing the variables, we followed an iterative approach, where we closely monitored model performance through repeated training and evaluation with and without these variables, and here we report the best-performing model. Univariate logistic regression was used to evaluate the association between each variable and SSI. Furthermore, these variables were fitted to a multivariable logistic regression, and variables that were no longer significant ($p < 0.01$) were removed using backward selection. Body Mass Index (BMI) and age were included irrespective of the significance levels, and the final set of variables used in the model development is depicted in Table 1. During the entire variable selection, medical expert recommendations were incorporated in the process.

Table 1: Predictor variables included in the logistic regression model **(N= numerical, B= binary, C= categorical)**

Category	Variable
Pre-procedure patient characteristics	Age (N)
	BMI (N)
	ASA (C)
Procedure-related	Type of admission (C)
	Length of stay (N)
	Operating time (N)
	Admitted before surgery (B)
	High-risk surgery (B)

2.4. Statistical analyses

The dataset initially consisted of an unbalanced number of patients with and without SSI from the set of patients who underwent surgery and were reviewed by the GTT. In the dataset of 2210 patients, 110 patients were identified as having SSI, and 2100 patients were identified as having no SSI. In classification task, class imbalance poses significant challenges, where the number of observations belonging to one class significantly outnumbers the observations in another class. Resampling techniques are used to mitigate the class imbalance, which could be over-sampling or under-sampling. For our specific case, propensity scoring was used to match the number of samples with SSI with that of without SSI based on age and length of stay. The propensity scoring method was used due to its popularity and widespread use in medical literature[20–22]. The analysis was conducted in R, using the MatchIt package[23,24]. No imputation of missing data was necessary since the set of data used for modelling was complete with no missing values.

The training and testing set of the LR model was split at a 70/30 ratio, where the model was trained on 70% and tested with 30% of the data. A grid search with 5-fold cross-validation was employed to estimate the best hyperparameters of the model, and the area under the receiver operating curve (AUC ROC) was used for model ranking. The gride parameters consist of inverse regularization strength (C), penalty (L1 Lasso regression and, L2 Rigid regression), solver (liblinear, 'lbfgs', and 'saga'), and maximum number of iterations. The original dataset was bootstrapped repeatedly resulting in multiple separate sets of training and test data, which was used to fine-tune the parameters on training set via gride search, refit the model with whole training set, and evaluate the model on holdout test set. Bootstrapping is a resampling technique, which randomly samples the original dataset with replacement to create multiple new sets of training and holdout test data. Feature selection and model developments were conducted using Statsmodels and Scikit-learn library, a machine-learning environment in Python[25,26].

For the three risk index, each patient was assigned a risk score based on the sum of the weights of each variable. To circumvent difference in the scores, we established the cut-off values defining the no-risk of SSI and the high-risk of SSI for all three risk indexes by estimating the optimal cut-off threshold value from the ROC curve. Afterwards, the three risk indexes and LR model were compared based on AUC ROC, specificity, sensitivity, positive predictive value (PPV), and negative predictive value (NPV)[27–29]. These metrics are the most widely used performance indicators, and combining these metrics offers a multifaceted evaluation of a model's performance, ensuring a comprehensive understanding of its strengths and weaknesses[30]. T-tests were also performed to evaluate the mean difference between the risk scores in the groups with and without SSI. The average performance of the indexes and the LR model, along with their associated confidence intervals, were estimated using bootstrapping techniques. We performed t-tests at a 5% significance level to test whether the risk scores and the AUC ROC were statistically different. A 5% significance level is chosen to maintain consistency and facilitating further comparison across similar studies predicting post operative infection. We also reported Brier Score for the three risk indexes and the LR model, where low Brier Score signifies better model calibration. The Brier Score is measured by computing the mean squared difference between the predicted probability and the actual outcomes and can be written mathematically as $BS = \frac{1}{N}\sum_{i=1}^{N} (p_i - y_i)^2$. A well-calibrated model, as reflected by a lower Brier Score, is crucial for providing accurate probability risk estimates. Feature importance analysis was also performed to study the association between each variable and the incidence of SSI. The analysis was conducted by computing the percentage contribution of the magnitude of each feature from the logistic regression coefficients. In this way, one can get a crude estimation of the weight of each feature in predicting SSI event[23].

3. Results

110 patients with SSI were identified and 110 patients from the non-SSI patients population were matched. The median age was respectively 63 and 59 years. Median LOS for SSI patients were 9 days and for non-SSI patients 5 days. Further descriptive variables in table 2.

Table 2: Descriptive statistics and variables for patients with and without SSI.

Variables	Patients with SSI (n=110)	Patients without SSI (n=110)
LOS, mean	13 days (range 1-63)	10 days (range 1-98)
Age, mean	60 years (range 19-94)	56 years (range 18-87)
ASA, mean	2.3 (range 1-4)	2.0 (range 1-4)
BMI, mean	29 (range 19-47)	27 (range 15-46)
Operating time, mean	128 minutes (range 9-666)	82 minutes (range 6-333)
ASA ≥ 3 (n)	42 (38 %)	25 (23%)
Contaminated wound (n)	11 (10 %)	12 (11%)
Above-average operating time (n)	29 (27 %)	12 (11 %)
Smoker (n)	23 (21 %)	24 (22 %)
Peripheral Vascular disease (n)	13 (12 %)	6 (5 %)
Metastatic cancer (n)	5 (5 %)	7 (6 %)
Steroid used (n)	7 (6 %)	5 (5 %)
Sepsis (n)	4 (4 %)	4 (4 %)
Type of admissions		
Outpatient (n)	16 (15 %)	1 (1 %)
In-patient non-emergency (n)	54 (49 %)	35 (31 %)
In-patient emergency (n)	40 (36 %)	74 (68 %)
Wound type		
Clean (n)	89 (81 %)	98 (89 %)
Clean/contaminated (n)	14 (13 %)	8 (7 %)
Contaminated/Infected (n)	7 (6 %)	4 (4 %)
General anesthetics (n)	62 (56 %)	58 (53 %)
Additional procedure performed (n)	8 (7 %)	8 (7 %)
Discharged to facility (n)	23 (21 %)	24 (22 %)
Operating time > 3 hours (n)	23 (21 %)	10 (9 %)
Admitted before surgery (n)	10 (9 %)	30 (27 %)
High risk surgery (n)	44 (40 %)	28 (25 %)

3.1. Comparative Performance

T-test statistics demonstrated that the risk scores of the LR model, the NNIS and the SSIRS were significantly higher in patients with SSI than in those without SSI ($p<0.001$, CI 19-44, p=0.004, 95% CI 3-17, p<0.001, 95% CI 7-14), whereas the risk score from the JSS-SSI was not significantly different between the groups (p=0.324, 95% CI-2-7). The models Brier Score were LR (0.1830), JSS-SSI (0.3776), SSIRS (0.3815) and NNIS (0.3468). The comparative performance of the three risk indexes and the LR model based on their respective AUC ROC are shown in Figure 1.

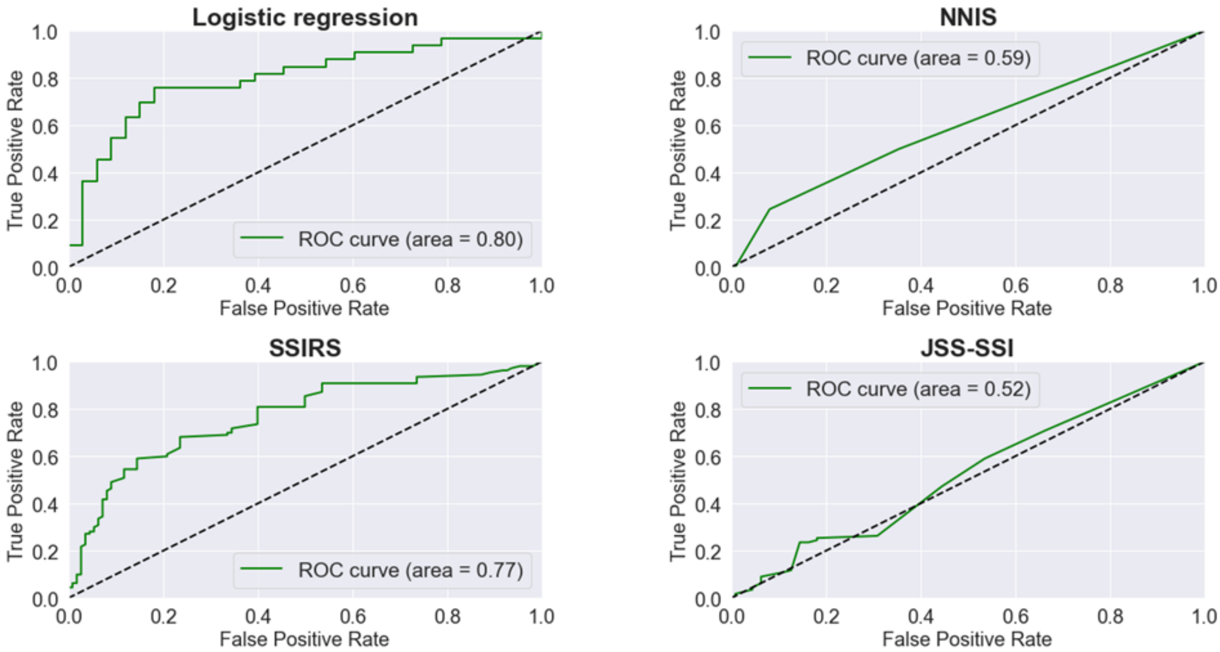

Figure 1: Area under the receiving operating curve

3.2. Major predictors

Variable importance is an indicator of which factors are the major predictors of SSI and can be useful in interpreting the model, as well as quantifying the degree of risk for SSI in response to each variable. The ranking of the variables in the LR model is shown in Figure 2.

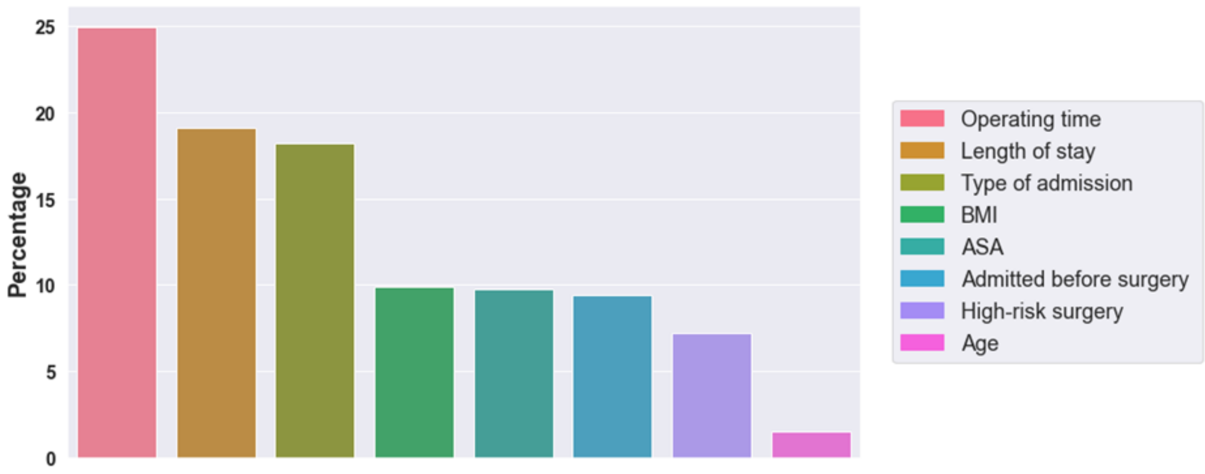

Figure 2: Major predictors Variable importance for predicting SSI. The X-axis depicts the feature index and ordering, and the Y-axis depicts the percentage of coefficients magnitude used by the logistic regression.

Post-predictive analysis (figure 3) of the major predictors, that is operating time, and LOS, attempted to demonstrate the cut-off values used in deciding the prediction class based on the two major predictors. The LR model considers median values of 100 minutes and 8 days as a positive class, resulting in false positives, and a median value of 88.5 minutes and 6.5 days as a negative class, resulting in false negatives.

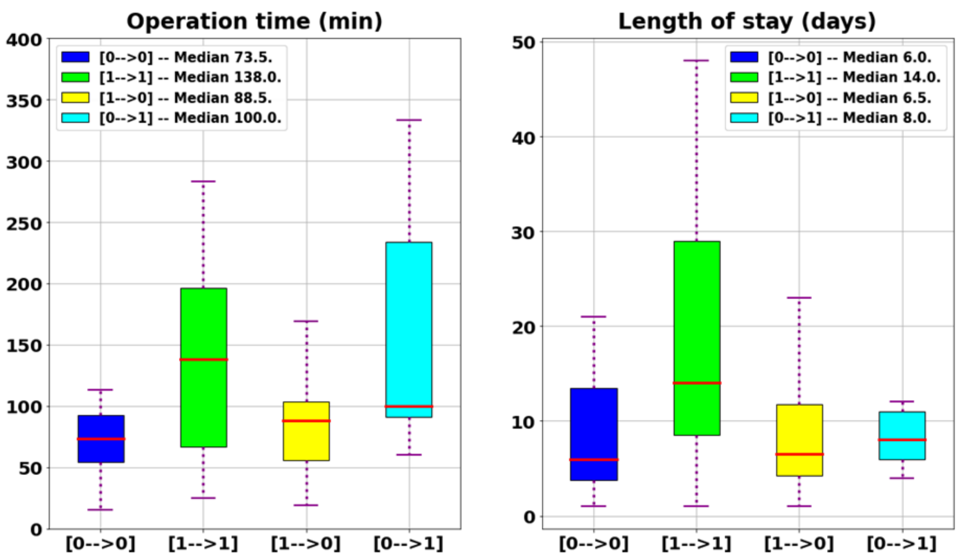

Figure 3: Post predictive analysis. The two predictors, i.e., operating time and length of stay, where the Y-axis depicts the magnitude, and the X-axis depicts each predicted class in relation

to the original class (1: with SSI, and 0: without SSI). For example, [1-->0] depicts a patient with SSI predicted to be without SSI.

4. Discussion

We evaluated three state-of-the-art risk indexes' predictive validity and compared their results to an LR model. The key finding of our study is that the results varied according to the model being used. We found that the LR model and two of the risk indexes identified significantly higher risk scores in patients with SSI than in those without SSI, with the SSIRS and the LR model demonstrating the best results. The PPV was highest in the NNIS; however, the index had a low NPV, which would result in a low ability to correctly identify patients at risk of SSI. The LR model was the only model that achieved reasonable results for all validity properties and AUC ROC.

Due to the consequences of SSI reducing them should be of high priority[31]. The incidence of SSI differs among hospitals and in our study, including all types of surgeries, we found that the overall incidence of SSI was approximately 5%[32]. As our study population was extracted from our GTT database, which consist of only 10 % of the total patient population, some limitations can explain the low incidence of SSI. We consider the use of record review to ascertain patients with surgical site infections as a major strength of the study, compared to other studies that rely on administrative data to identify surgical site infections which is notoriously difficult.

Identifying the factors that contribute most to predicting the risk of SSI is important when developing a risk prediction model. The risk models we tested included both pre-and post-surgery predictors. However, the major predictors were the post-surgery factors. This limits the utility of these models as the clinicians are not able to intervene upon these predictors before surgery. An ideal model would be calculable early enough within a patient's perioperative course, even when surgery is being contemplated, to allow time to intervene. However, the net benefit with decision curve analysis is another way of evaluating the clinical benefit of a risk prediction model[33]. In this study we have only evaluated the accuracy of the models and have not perform a decision curve analysis which can tell us whether using a model to aid clinical decision-making would improve outcomes for our patients[34]. Among all the predictors from the models, we found that operating time was the major predictor of SSI, which is in accordance with literature[35]. The next major predictor we found was LOS, which could be LOS to surgery or LOS to readmission due to treatment of the SSI. As we were not able to conclude if the LOS was related to the main surgery admission or readmission in our data, we do not know if this is a true-positive risk factor. Post-predictive analyses indicate that those who had both predicted and verified SSI had the longest LOS, with a median of 14 days, which was more than twice that of the other groups. These results support the theory that prolonged postoperative LOS is associated with SSI, whereas preoperative LOS has a low association with SSI[36]. ASA was also a major predictor as higher ASA increases the risk of SSI. There are some limitations of ASA as a predictor, as they are highly subjective due to the ASA score is set by the anesthesiologist, with notoriously poor inter-rater reliability as it requires direct clinician assessment, and it cannot be calculated automatically from the electronic health record.

Furthermore, the type of admission, either inpatient or outpatient, seems to play a role in predicting SSI. However, in our study, 15% of the patients with SSI underwent surgery as outpatients. Among the non-SSI patients, only one underwent surgery as an outpatient. The higher rate of SSI in surgeries performed as outpatients could be explained by the short follow-up time for outpatients. They did not receive the same wound care as the inpatients. However, inpatients undergo more complicated procedures, and their state of health necessitates often hospital admission. These factors should increase the risk of SSI by themselves; therefore, we often consider outpatient surgeries with less risk of SSI, as these patients have less complicated surgeries and are healthier than those admitted. For inpatients, we did not find a difference between the groups regarding planned admission or acute admissions.

Our findings are in accordance with previous literature; for example, a review performed by Korol et al. in 2013 demonstrated that identifying the correct factor for modification to reduce the incidence of SSI is difficult, which can explain why SSIs are difficult to combat[37]. The literature demonstrates that there is a major gap between studies such as this one and the implementation of such models in clinical use. Most of the studies also lack reporting, thus failing them from being used elsewhere. A reporting checklist has therefore been developed in order to increase the reproducibility[38]. The prediction models are therefore mostly used for infection surveillance[39]. The predictive models should not necessarily be targeting the highest risk group, but the group with the most modifiable risk. As the next step, we believe large scale external validation of the model should be performed. Further evaluation and comparison of various sampling strategies on model performance should be performed including synthetic minority over-sampling technique, and over- sampling using propensity scores[20–22].

5. Conclusion

This study demonstrated that patient and procedural factors in an LR model can be used to predict the risk of SSI better than the state-of-art risk indexes. However, a more precise predicting model for clinical decision support tool might include modifiable predictors we might not yet have exposed. For this purpose, more advanced machine learning technics should be applied.

Authors contributions: KM, KØM and AZW designed the study. AR and KM collected the data. YX and AZW analyzed the data. KM drafted the manuscript. All authors interpretated the data, revised and approved the manuscript.

Acknowledgments: We would thank Laila Bjølgerud for making the graphical abstract.

Conflict of interest disclosures: The authors have no related conflicts of interest to declare.

Summary points:

What was already known:

- Studies reporting risk indexes to be used for prediction of surgical site infections demonstrates low sensitivity.
- Several indexes have been developed but few are in clinical use for predicting surgical site infections.

What does this study add to our knowledge：

- State-of-the-art risk indexes are mainly for surveillance, not prediction as they lack modifiable predictors.
- We found major predictors, such as operation time and length of stay, which increase the risk of surgical site infections

References

[1] C.D. Owens, K. Stoessel, Surgical site infections: epidemiology, microbiology and prevention, Journal of Hospital Infection 70 (2008) 3–10. https://doi.org/10.1016/S0195-6701(08)60017-1.

[2] M. Monahan, S. Jowett, T. Pinkney, P. Brocklehurst, D.G. Morton, Z. Abdali, T.E. Roberts, Surgical site infection and costs in low- and middle-income countries: A systematic review of the economic burden, PLoS One 15 (2020) e0232960. https://doi.org/10.1371/JOURNAL.PONE.0232960.

[3] D. Curcio, A. Cane, F. Fernández, J. Correa, Surgical site infection in elective clean and clean-contaminated surgeries in developing countries, International Journal of Infectious Diseases 80 (2019) 34–45. https://doi.org/10.1016/J.IJID.2018.12.013.

[4] Pasientskader i Norge 2022 – Målt med Global Trigger Tool, 2023. https://kudos.dfo.no/documents/46863/files/30149.pdf (accessed January 15, 2024).

[5] Global Guidelines for the Prevention of Surgical Site Infection. Geneva: World Health Organization; 2018. PMID: 30689333., (2018). https://www.ncbi.nlm.nih.gov/books/NBK536404/ (accessed February 15, 2023).

[6] R.E. Al Mamlook, L.J. Wells, R. Sawyer, Machine-Learning Models for Predicting Surgical Site Infections using Patient Pre-Operative Risk and Surgical Procedure Factors, Am J Infect Control (2022). https://doi.org/10.1016/J.AJIC.2022.08.013.

[7] A. Adibi, M. Sadatsafavi, J.P.A. Ioannidis, Validation and Utility Testing of Clinical Prediction Models: Time to Change the Approach, JAMA 324 (2020) 235–236. https://doi.org/10.1001/JAMA.2020.1230.

[8] S.T. Le, D. Corbin, L.C. Myers, P. Kipnis, B. Cohn, V.X. Liu, Development and Validation of an Electronic Health Record-based Score for Triage to Perioperative Medicine, Ann Surg 277 (2023). https://doi.org/10.1097/SLA.0000000000005284.

[9] T.J. Loftus, M.S. Altieri, J.A. Balch, K.L. Abbott, J. Choi, J.S. Marwaha, D.A. Hashimoto, G.A. Brat, Y. Raftopoulos, H.L. Evans, G.P. Jackson, D.S. Walsh, C.J. Tignanelli, Artificial Intelligence-enabled Decision Support in Surgery: State-of-the-art and Future Directions, Ann Surg 278 (2023). https://doi.org/10.1097/SLA.0000000000005853.

[10] G.S. Collins, P. Dhiman, C.L. Andaur Navarro, J. Ma, L. Hooft, J.B. Reitsma, P. Logullo, A.L. Beam, L. Peng, B. Van Calster, M. van Smeden, R.D. Riley, K.G.M. Moons, Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence, BMJ Open 11 (2021). https://doi.org/10.1136/bmjopen-2020-048008.

[11] J.M. Naessens, T.J. O'Byrne, M.G. Johnson, M.B. Vansuch, C.M. McGlone, J.M. Huddleston, Measuring hospital adverse events: assessing inter-rater reliability and trigger performance of the Global Trigger Tool., International Journal for Quality in Health Care : Journal of the International Society for Quality in Health Care / ISQua 22 (2010) 266–74. https://doi.org/10.1093/intqhc/mzq026.

[12] F. Griffin, R. Resar, IHI Global Trigger Tool for measuring adverse events (Second Edition), 2009. https://app.ihi.org/webex/gtt/ihiglobaltriggertoolwhitepaper2009.pdf (accessed November 25, 2014).

[13] K. Mevik, T.E. Hansen, E.C. Deilkås, A.M. Ringdal, B. Vonen, Is a modified Global Trigger Tool method using automatic trigger identification valid when measuring adverse events?A comparison of review methods using automatic and manual trigger identification, International Journal for Quality in Health Care 31 (2019) 535–540. https://doi.org/10.1093/INTQHC/MZY210.

[14] A. Storesund, A.S. Haugen, M. Hjortås, M.W. Nortvedt, H. Flaatten, G.E. Eide, M.A. Boermeester, N. Sevdalis, E. Søfteland, Accuracy of surgical complication rate estimation using ICD-10 codes, BJS 106 (2019) 236–244. https://doi.org/10.1002/bjs.10985.

[15] T.G. Emori, D.H. Culver, T.C. Horan, W.R. Jarvis, J.W. White, D.R. Olson, S. Banerjee, J.R. Edwards, W.J. Martone, R.P. Gaynes, J.M. Hughes, National nosocomial infections surveillance system (NNIS): Description of surveillance methods, Am J Infect Control 19 (1991) 19–35. https://doi.org/10.1016/0196-6553(91)90157-8.

[16] R.W. Haley, D.H. Culver, W.M. Morgan, J.W. White, T.G. Emori, T.M. Hooton, Identifying patients at high risk of surgical wound infection: A simple multivariate index of patient susceptibility and wound contamination, Am J Epidemiol 121 (1985) 206–215. https://doi.org/10.1093/OXFORDJOURNALS.AJE.A113991.

[17] C. van Walraven, R. Musselman, The Surgical Site Infection Risk Score (SSIRS): A Model to Predict the Risk of Surgical Site Infections, PLoS One 8 (2013). https://doi.org/10.1371/journal.pone.0067167.

[18] A. Karellis, Updating and external validation of a surgical site infection risk-index tool, 2018. https://escholarship.mcgill.ca/concern/theses/bc386m90f?locale=en (accessed August 18, 2021).

[19] Karellis A, Vaillancourt J, J. Sampalis, Development of Risk Scoring Tool to Predict Surgical Site Infections, SOJ Surgery 2 (2015) 1–10. https://doi.org/10.15226/2376-4570/2/2/00116.

[20] W.A. Rivera, A. Goel, J.P. Kincaid, OUPS: A combined approach using smote and propensity score matching, in: Proceedings - 2014 13th International Conference on Machine Learning and Applications, ICMLA 2014, 2014. https://doi.org/10.1109/ICMLA.2014.106.

[21] W. Rivera, A. Goel, J.P. Kincaid, Advances in algorithms for re-sampling class-imbalanced educational data sets, in: Artificial Intelligence: Concepts, Methodologies, Tools, and Applications, 2016. https://doi.org/10.4018/978-1-5225-1759-7.ch040.

[22] W.A. Rivera, A. Goel, J.P. Kincaid, Blending propensity score matching and synthetic minority over-sampling technique for imbalanced classification, Proceedings of the 2014 Winter Simulation Conference (2014).

[23] The R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing Vienna Austria 0 (2014). https://www.semanticscholar.org/paper/R%3A-A-language-and-environment-for-statistical-Team/659408b243cec55de8d0a3bc51b81173007aa89b (accessed January 15, 2024).

[24] D.E. Ho, K. Imai, G. King, E.A. Stuart, MatchIt: Nonparametric Preprocessing for Parametric Causal Inference, J Stat Softw 42 (2011) 1–28. https://doi.org/10.18637/JSS.V042.I08.

[25] S. Seabold, J. Perktold, Statsmodels: Econometric and Statistical Modeling with Python, Undefined (2010) 92–96. https://doi.org/10.25080/MAJORA-92BF1922-011.

[26] F. Pedregosa, V. Michel, M. Blondel, P. Prettenhofer, R. Weiss, J. Vanderplas, D. Cournapeau, G. Varoquaux, A. Gramfort, B. Thirion, O. Grisel, V. Dubourg, A. Passos, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: Machine learning in Python, Jmlr.Org 12 (2011) 2825–2830. https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf?ref=https://githubhelp.com (accessed September 12, 2022).

[27] N.Q.K. Le, D.T. Do, T.T.D. Nguyen, Q.A. Le, A sequence-based prediction of Kruppel-like factors proteins using XGBoost and optimized features, Gene 787 (2021). https://doi.org/10.1016/j.gene.2021.145643.

[28] N.Q.K. Le, Y.Y. Ou, Incorporating efficient radial basis function networks and significant amino acid pairs for predicting GTP binding sites in transport proteins, BMC Bioinformatics 17 (2016). https://doi.org/10.1186/s12859-016-1369-y.

[29] E.W. Steyerberg, A.J. Vickers, N.R. Cook, T. Gerds, M. Gonen, N. Obuchowski, M.J. Pencina, M.W. Kattan, Assessing the performance of prediction models: A framework for traditional and novel measures, Epidemiology 21 (2010). https://doi.org/10.1097/EDE.0b013e3181c30fb2.

[30] T.F. Monaghan, S.N. Rahman, C.W. Agudelo, A.J. Wein, J.M. Lazar, K. Everaert, R.R. Dmochowski, Foundational statistical principles in medical research: Sensitivity, specificity, positive predictive value, and negative predictive value, Medicina (Lithuania) 57 (2021). https://doi.org/10.3390/medicina57050503.

[31] R. Gaynes, C. Richards, J. Edwards, T.G. Emori, T. Horan, J. Alonso-Echanove, S. Fridkin, R. Lawton, G. Peavy, J. Tolson, Feeding back surveillance data to prevent hospital-acquired infections., Emerg Infect Dis 7 (2001) 295. https://doi.org/10.3201/EID0702.010230.

[32] B.M. Gillespie, E. Harbeck, M. Rattray, R. Liang, R. Walker, S. Latimer, L. Thalib, A. Erichsen Andersson, B. Griffin, R. Ware, W. Chaboyer, Worldwide incidence of surgical site infections in general surgical patients: A systematic review and meta-analysis of 488,594 patients, International Journal of Surgery 95 (2021). https://doi.org/10.1016/j.ijsu.2021.106136.

[33] A.J. Vickers, B. Van Calster, E.W. Steyerberg, Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests, BMJ (Online) 352 (2016). https://doi.org/10.1136/bmj.i6.

[34] A.J. Vickers, F. Holland, Decision curve analysis to evaluate the clinical benefit of prediction models, The Spine Journal 21 (2021) 1643–1648. https://doi.org/10.1016/j.spinee.2021.02.024.

[35] H. Cheng, B.P.H. Chen, I.M. Soleas, N.C. Ferko, C.G. Cameron, P. Hinoul, Prolonged Operative Duration Increases Risk of Surgical Site Infections: A Systematic Review, Surg Infect (Larchmt) 18 (2017) 722–735. https://doi.org/10.1089/SUR.2017.089.

[36] E. Mujagic, W.R. Marti, M. Coslovsky, S.D. Soysal, R. Mechera, M. von Strauss, J. Zeindler, F. Saxer, A. Mueller, C.A. Fux, C. Kindler, L. Gurke, W.P. Weber, Associations of Hospital Length of Stay with Surgical Site Infections, World J Surg 42 (2018) 3888–3896. https://doi.org/10.1007/S00268-018-4733-4.

[37] E. Korol, K. Johnston, N. Waser, F. Sifakis, H.S. Jafri, M. Lo, M.H. Kyaw, A Systematic Review of Risk Factors Associated with Surgical Site Infections among Surgical Patients, PLoS One 8 (2013) e83743. https://doi.org/10.1371/JOURNAL.PONE.0083743.

[38] F. Cabitza, A. Campagner, The need to separate the wheat from the chaff in medical informatics: Introducing a comprehensive checklist for the (self)-assessment of medical AI studies, Int J Med Inform 153 (2021). https://doi.org/10.1016/j.ijmedinf.2021.104510.

[39] J.R. Bergquist, C.A. Thiels, D.A. Etzioni, E.B. Habermann, R.R. Cima, Failure of Colorectal Surgical Site Infection Predictive Models Applied to an Independent Dataset: Do They Add Value or Just Confusion?, J Am Coll Surg 222 (2016) 431–438. https://doi.org/10.1016/J.JAMCOLLSURG.2015.12.034.