**What do we know about contributing factors for “never events” in operating rooms? A machine learning analysis**

Dana Arad, MSN1, 3, Ariel Rosenfeld, PhD2, Racheli Magnezi, PhD1

1 Department of Management, Health Management Program, Faculty of Sciences, Bar-Ilan University, Israel

2 Department of Information Science, Bar-Ilan University, Israel

3 Patient Safety Division, Ministry of Health, Israel

Corresponding author:

Dana Arad, RN, BA, MSN, ACNP

danaarad@gmail.com

Word Count:

 **Additional material**

Additional file 1:

12697-additional\_file1--4\_EN.docx; Surgical safety standards checklists – checklists used for monitoring surgical safety standards

**Declarations**

**Ethics approval and consent to participate**

The research was approved by the “Helsinki” ethics committee of the Israeli Ministry of Health (MOH). Approval number 1/2020 to trial registration number MOH 032-2019. The need for informed consent was waived by the MOH’s ethics committee.

**Consent for publication**

 Not applicable

**Availability of data and materials**

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

**Competing interests**

To the best of our knowledge, the named authors have no competing interests, financial or otherwise to disclose.

**Funding**

This study was funded by grant #MOHIG 14-2019 from the Medical Research Fund for Health Services–Jerusalem.

**Authors’ contributions**

D.A. performed the data collection and analyzed the observations and root cause analyses for the dataset and possible contributing factors. A.R. interpreted and created the algorithms for machine learning analysis, R.M. made a major contribution to the writing of the manuscript. All authors read and approved the final manuscript.

**Acknowledgments**

The authors would like to thank the Medical Research Fund of the Israel Ministry of Health for supporting this study.

**What do we know about contributing factors for “never events” in operating rooms? A machine learning analysis**

**ABSTRACT**

**Background**

A surgical “Never Event” (NE) is a preventable error. Various factors contribute to the occurrence of wrong site surgery and retained foreign item, but little is known about their quantified risk in relation to a surgery’s characteristics. Our study uses machine learning to reveal risk factors and quantify their risk to improve patient safety and quality of care.

**Methods**

We used data from 9234 observations on safety standards and 101 root-cause analyses from actual Nes. We utilized three random forest supervised machine learning models. Using a standard 10-cross validation technique, we evaluated the models’ metrics, and, through Gini impurity, we measured the impact of factors thereof to the occurrence of the two types of NEs.

**Results**

We identified 24 contributing factors in six surgical departments. Two factors had an impact of >900% in Urology, Orthopedics, and General Surgery; six had an impact of 0–900% in Gynecology, Urology, and Cardiology; and 17 had an impact of <0%.

Factors' combination revealed 15–20 pairs with an increased probability in five departments: Gynecology, 875–1900%; Urology, 1900–2600%; Cardiology, 833–1500%; Orthopedics,1825–4225%; and General Surgery, 2720–13,600%.

Five factors affected the occurrence of wrong site surgery (-60.96 to 503.92%) and five affected retained foreign body (-74.65 to 151.43%), three of them overlapping: two nurses (66.26–87.92%), surgery length <1 hour (85.56–122.91%), and surgery length 1–2 hours (-60.96 to 85.56%).

**Conclusions**

The use of machine learning has enabled us to quantify the potential impact of risk factors for wrong site surgeries and retained foreign items, in relation to a surgery’s characteristics, which in turn suggests safety standards should be tailored accordingly.

Keywords: Never Event, surgery department, machine learning, patient safety

**Trial registration number:** MOH 032-2019

**BACKGROUND**

Adverse medical events can lead to significant morbidity and mortality and increase healthcare expenditures [1]. A Never Event (NE) is an unacceptable adverse event, both preventable and unjustifiable; NEs should be reduced to zero through quality improvement measures [2]. Major NEs in perioperative care include incorrect surgery sites and foreign items retained in patients following surgery [3-4].

The human factors approach recognizes that human error is often the result of a combination of both individual surgeon factors and work system factors [5], which makes human error the main contributing factor to NEs [6]. Human error includes surgeon distraction [7], lack of situational awareness of the surgical team to possible error, and miscommunication among team members [8]. Additionally, institutional factors and working conditions, such as increased workload and clinician pressure, can create a work climate that is not conducive to meeting the standards required to maintain patient safety [9] and effective teamwork [10].

Currently, there are two essential international standards aimed at reducing NE occurrence: 1) the World Health Organization (WHO) Surgical Safety Checklist [11]; and 2) surgical counts of all items used during a surgery [12]. However, incomplete compliance, non-standardized implementation of these standards [13], and other possible unknown factors have led to the incidence of NEs remaining unchanged [14]. In Israel, the incidence of retained foreign items during surgery is 3.2 in every 100,000 surgeries [15]. The incidence for an incorrect site procedure is unclear, but is generally estimated to be 1 in every 100,000 surgeries.

For this study, we adopted a machine learning (ML) approach [16] to identify currently unknown contributors to NE occurrence. Previous studies leveraging ML methods in healthcare have demonstrated its advantages in analyzing diverse data types and revealing non-trivial insights when compared with traditional methods [17]. To the best of our knowledge, this is the first study to use ML methods to identify potential contributing factors to the occurrence of NEs in operating rooms (ORs).

**METHODS**

**Study design**

We utilized a supervised ML method know as random forest (RF) [18-19], incorporating the commonly used extra tree classifier [20]. RF is an ensemble learning method that trains multiple “simple” decision tree models and merges them to achieve a more accurate and stable prediction. The use of RF entails several desired elements we needed to properly conduct the analysis for this study. First, RFs are used to rank the importance of features in a natural way. Specifically, the importance of features can be determined by examining to what extent the tree nodes using a feature reduce the impurity (i.e., the uncertainty in classification) across all “trees in the forest.” Second, RFs are known to cope well with imbalanced datasets (as was the case in this study) and avoid overfitting the data. Finally, RFs compared favorably with several other supervised ML algorithms we tested using our data, including popular deep neural networks and support vector machines (SVMs). RFs have been extensively used in the medical field for clinical risk prediction [21] and other applications.

Safety standards used in the OR (surgical safety checklists and surgical counts) were divided into safety verifications at three distinct time periods – pre-procedure, sign in, and time out [11] – and addressed incorrect surgery site errors, which we will define as type A errors. Surgical counts were divided into three separate counts throughout a surgery to address retained foreign body errors, which we will define as type B errors: prior to skin incision; initiation of closure of fascia/cavity; and following skin closure [22]. In addition, we added general features, such as the name of the hospital, length of surgery, patient’s gender and age, surgeon’s specialty, and number of physicians and nurses present during surgery.

**Data collection and annotation**

Data were collected from 29 Israeli hospitals and consisted of two types of data entries: observations of 9234 surgeries performed between January 2018 and February 2019 in which no NEs occurred during the surgeries observed, and root cause analyses (RCAs) of 101 NEs that occurred between January 2016 and February 2020 in the examined hospitals.

Observations

Initiated by the supervisory arm of the Israeli Ministry of Health, passive observations by medical students, physicians, nursing students, or RNs are routinely performed in ORs. Observers for this study underwent an eight-hour long training program that included simulations. In each OR, at least two observers passively observed randomly selected surgeries; they recorded and annotated the surgery process using a pre-defined set of features. Observations were then transferred to a central database and routinely assessed for variability and reliability. Overall, 9234 observations were conducted. Each observation was translated into a 93-feature-long vector, representing characteristics of the surgery (see Additional file 1). To maintain reliability, entries with greater than 5% discordance among annotators in one OR were discarded (<1%).

Root cause analyses (RCAs)

RCAs were performed in response to NEs that occurred between January 2016 and February 2020. Overall, we reported 101 NEs: 49 of type A and 52 of type B. The obtained RCAs were manually annotated by the authors using the same 93-feature-long representation used to characterize the observations. However, unlike the observations, RCAs were performed retrospectively and therefore a significant portion of the features was missing and could not be obtained. Specifically, up to 40% of all other feature values were missing, a challenge we address later.

**Pre-processing and analysis technique**

As some features were non-binary (e.g., patient age, length of surgery), we first discretized them, resulting in 250 binary features. This and subsequent steps were performed using a designated Python 3 program implemented by the authors that uses the standard scikit-learn ML package (https://scikit-learn.org/stable).

Examination of the 40% of missing feature values revealed that most were strongly dependent on the NE type. Specifically, for type A NEs, features that were assumed to be more related to NEs of type B were not investigated and vice versa. For example, for an NE in which the wrong hand was operated on, there was no indication as to whether the surgeon scanned the surgical cavity for retained surgical items before closure. To mitigate this artifact, we used the iterative data imputation approach [23], where we predicted the value of each missing value while relying on the present features and available examples. Specifically, using the entire dataset, each missing value was estimated using a standard decision-tree regressor.

In addition, balancing steps were taken to cope with the highly imbalanced dataset. Specifically, with more than 9000 observations and only 101 NEs, we adopted a cost-sensitive training approach [24], whereby our model adjusted for prediction mistakes on the minority class (NEs) by an amount proportional to how under-represented it was (here, approximately 90 times under-represented).

We implemented three RF models using our data: model 1 for differing between observations and NEs; model 2 for differing between observations and type A NEs; and model 3 for differing between observations and type B NEs. We used a standard 10-cross validation technique to evaluate each model’s metrics and adopted the standard Gini impurity [25] measure to estimate the importance of features and their combination in our models. Intuitively, Gini impurity captures the “noise” in a set by measuring how often a randomly chosen element from the set would be incorrectly labeled if it were randomly labeled according to the distribution of labels in the set. Feature importance ranking was conducted using the trained RF models, and we reported the change in the probability of NE occurrence given the entire data set. We considered each feature separately and calculated the probability of NE occurrence when that feature assumed the value “True” as opposed to the value “False”.

This study was approved by the University’s and Ministry of Health Ethics Committee (MOH 032-2019).

**RESULTS**

The majority of NEs (62.32%) occurred in six main departments: General Surgery, 19 (18.81%); Gynecology, 17 (16.83%); Orthopedics, 16 (15.84%); Cardiac and Cardiothoracic, 15 (14.85%); Ophthalmology, 8 (7.92%); and Urology, 7 (6.93%) (Table 1). Therefore, our analysis focused on the occurrence of NEs in these six departments.

Table 1: Characteristics of the dataset according to surgical specialty

|  |  |
| --- | --- |
| **Observations n=9234** |  **Never Eventsn=101** |
| PhaseSpecialty | \*Pre-procedure(n=1539)(missing data on 760 cases) | Sign in (n=1504) | Time out(n=1498) | First count(n=1518) | Second count(n=1501) | Third count(n=1498) |  |
| Urology | 72 | 156 | 148 | 124 | 118 | 124 | 7 (6.93%) |
| Orthopedics | 185 | 331 | 324 | 341 | 302 | 326 | 16 (15.84%) |
| Ear, nose, and throat | 64 | 105 | 105 | 99 | 102 | 93 | 3 (2.97%) |
| Gynecology | 63 | 143 | 139 | 149 | 153 | 153 | 17 (16.83%) |
| General surgery | 313 | 537 | 558 | 576 | 623 | 604 | 19 (18.81%) |
| Plastic surgery | 22 | 39 | 37 | 40 | 36 | 42 | 2 (1.98%) |
| Vascular surgery | 18 | 45 | 42 | 45 | 42 | 43 | 5 (4.95%) |
| Neurosurgery | 7 | 25 | 19 | 22 | 19 | 19 | 5 (4.95%) |
| Dermatology | 7 | 16 | 26 | 21 | 22 | 24 | 2 (1.98%) |
| Ophthalmology | 12 | 41 | 34 | 33 | 19 | 18 | 8 (7.92%) |
| Maxillofacial | 3 | 12 | 10 | 8 | 10 | 11 | 2 (1.98%) |
| Cardiac and Cardiothoracic  | 13 | 54 | 56 | 60 | 55 | 41 | 15 (14.85%) |

Table 2: Characteristics of patients and surgery in the dataset

|  |  |  |
| --- | --- | --- |
| **Characteristic** | **Observations** | **Never Events** |
| **Average age** | 50.8 years (SD 20.4) | 46 |
| **Gender** | Male (n=388 (49.8%)), Female (n=391 (50.2%))  | Male (n=46 (45.5%)) Female n=55 (54.5%) |
| **Length of surgery** | Up to 1 hour: 2124 (23%)1–2 hours: 4340 (47%)3–4 hours: 2031 (22%)Over 4 hours: 739 (8%) | Length of surgery:Up to 1 hour: 54 (53.5%)1–2 hours: 13 (12.9%)3–4 hours: 17 (16.8%)Over 4 hours: 17 (16.8%) |

To evaluate our models, we adopted the area under the curve (AUC) measure, which is especially suited for imbalanced data, as was the case in this study, as it does not have any bias toward models that perform well on the minority of majority classes at the expense of the other [26]. Our three RF models each demonstrated good performance, exhibiting an AUC between 0.81 and 0.85. Generally, AUC scores between 0.8 to 0.9 are considered excellent [27]. AUC is interpreted as the probability that our model will rank a randomly chosen positive instance higher than a randomly chosen negative one [28]. As such, our models can be considered relatively strong and accurate, despite their limitations.

**Feature importance**

Figure 1 shows the most common contributing features to the occurrence of NEs (of both types combined) in the six departments, along with the associated probability change.

Figure 1

The top 14 contributing features varied significantly across departments, and there was no single feature set that was consistently more informative across all operations for predicting NEs. For example, feature [C], “Discrepancy in second count”, varied significantly across departments (160% to 1,950%). Feature [B], “Surgery is paused because of discrepancy in third count”,appeared in four of the six departments, and the associated probability change varied dramatically as well, between 269% and 1540%. There were 10 features that consistently decreased the chance of an NE occurring, including [F] “Surgeon scans the cavity/fascia before closure during the second count”, which affected five out of six departments and was consistent in its probability change, between 65%–100%. Features [I], [J], [ K], [L], [M], and [N] decreased the chances of NEs between 2%–100% in three departments. Three features, [A] “Discrepancy in absorbing materials”, [E] “Surgery time >4 hours”, and [G] “Surgery time <1 hour”appeared just once across departments, with a medium impact on NE occurrence.

Analysis of the results by department shows variation among the contributing features. For example, in Ophthalmology, the probability was consistently -100% for five features, while in General Surgery, two features that increased the probability of an error varied between 1168–1283%: features [B] “Surgery is paused because of discrepancy in third count” and[C] “Discrepancy in second count”. In Orthopedics, those same two features, [B] and [C], increased the probability of error (1540–1950%). Three features decreased the probability of error: [F] “Surgeon scans the cavity/fascia before closure”; [H] “Second count is performed before closure of fascia/cavity”; and (I) “Procedure type is compared to the one written in patient’s file”,by -65 to -87%.

**Effects of feature combinations**

In the following analysis (Figure 2), we examine the effects of paired features, i.e., features that occur together in the data. It is important to note that when considering feature combinations, their occurrence is expected to be very low, especially in the NEs class. As such, the estimated effects are likely to be very high, yet their confidence is significantly low.

Figure 2

Interestingly, in General Surgery, there were 14 feature combinations that caused a probability change of 13,600% (Figure 2A). In comparison, the single feature analysis (Figure 1) revealed a probability change of 1287% and 1168%, surprisingly by two features that were not part of the 14 feature combinations identified here.

In Figure 2A (Gynecology), the effect of every feature combination is associated with a probability change of 1000–2000%. In the single feature analysis (Table 2), the effect of two of the features separately was <900%, and the rest lagged behind with <150%. In Urology (Figure 2B), the results showed there were dozens of pairs with an effect of 1900–2500%, while the effect of a single feature had <1150% effect on error. In General Surgery (Figure 2E), the accumulated effect of two features together showed a dozen pairs with an effect of 1900–4200%, while the effect of a single feature had a <1950% indication on error, and the rest showed even lower percentages.

**Features affecting types A and B**

Turning to Models 2 and 3, there was an overlap in three of the top five contributing features to type A and B errors (Figures 3 and 4): 1) the presence of two nurses during the surgery predicted a greater occurrence of type A (66%) and type B (88%); 2) an operation <1 hour had a greater occurrence of type A (122%) and type B (87%); and 3) when the operation lasted between one to two hours, both types A and B were less frequent, decreasing by 60% and 74%, respectively. The surgical department that was most affected regarding the occurrence of type A Nes was Ophthalmology, with a prevalence of 504%, while General Surgery was associated with a decrease of 63% in type A (Figure 3). For type B, the two remaining features were staff driven; the feature “more than three physicians” was associated with an increased prevalence of type B (151%), while “two physicians” was associated with a decreased prevalence of Type B, by 52% (Figure 4).

Figure 3: Features affecting incorrect surgery site (Type A)

Figure 4: Features affecting retained foreign item during surgery (type B)

**DISCUSSION**

Surgical errors are a serious public health problem and uncovering their causes is challenging [29]. In this study, we aimed to uncover factors that contribute to NEs by using ML methods to identify heretofore unknown contributors, as ML can automatically look for patterns not seen when using traditional methods [18, 30].

Despite the widespread use of the surgical safety checklist and strict surgical counts, the prevalence of NEs has not decreased significantly since the widespread implementation of these checks [31-32]. The human factor, and not system error, has been identified as the main contributing factor to NEs [31,33]. For example, in one study using an analysis and classification system, 628 human factors were divided into four categories that influenced NEs: preconditions for action, unsafe actions, oversight and supervisory factors, and organization influences [6]. Additional studies have identified lack of communication and lack of empirical evidence as barriers to the implementation of universal safety standards [29,34]. Some studies have suggested that surgical counts alone are insufficient; even when declared to be correct, items have been left in patients [35-36], mostly in the abdomen and pelvis [35, 37]. This may explain our finding of a higher probability of type B errors in General Surgery and Urology, which involve those regions.

We further analyzed paired contributing factors representing the relative risk in the OR’s complex work environment, when the graded risk increased compared to single feature analysis. For example, in Orthopedics, discrepancy in the count in combination with a surgery length of 1–2 hours increased the chances for an NE, which can be explained by partial compliance with the standards. In shorter surgeries, staff may rush and skip some phases of the checklists [38] and the complex sets used challenges the counts [31,39].

We found that the occurrence of incorrect surgery site increased in Ophthalmology during short surgeries and when two nurses were present. Its occurrence decreased in general surgery. This increased risk in could be due to the difficulty of performing a time out because the surgeons have antiseptic hands and cannot review charts, or perhaps because doing so is not made a priority [40]. The decrease in general surgery could be explained by better implementation of the time out process in that specialty [41-42].

One of the main factors contributing to the occurrence of NEs is a lack of communication among participating members in the surgery [33], which may explain our findings that the number of staff had an increasing/decreasing effect on NE occurrence.

We recognize that the current study is limited by the quantity, quality, and diversity of the data used. In the context of this work, our samples come from two distinct sources: prospective observations and retrospective investigations of Nes, where the latter consists of a small number of NEs compared with the relatively high number of observations analyzed. We believe that these limitations are inherent to the problem at hand, as performing prospective analyses of NEs is virtually impossible due to their infrequency, and the number of NEs is nominally small. To mitigate some of these concerns, we have used grounded statistical techniques that allowed us to train adequate models and estimate feature importance. Nevertheless, given the above, the feature impact should be carefully considered and validated in future studies.

In the future, we plan to further expand our data pool with newly obtained observations and NEs as they are accumulated. In other work, we will explore the use of transfer learning about NEs from other countries, which could be used to better inform our model. This approach could prove valuable in mitigating the imbalanced nature of our data, although it could introduce considerable biases due to the variety of data sources.

**Conclusion**

Our results suggest that the existing “one size fits all” safety approach currently in place may significantly benefit from tailored adjustments that consider additional factors such as those identified in this work. These more specific guidelines may be used to adjust risk management programs and improve patient safety.

# **References**

1. Kjellberg J, Wolf RT, Kruse M, Rasmussen SR, Vestergaard J, Nielsen KJ, Rasmussen K. Costs associated with adverse events among acute patients. BMC Health Serv Res. 2017;17(1):1-7.
2. Robert MC, Choi CJ, Shapiro FE, Urman RD, Melki S. Avoidance of serious medical errors in refractive surgery using a custom preoperative checklist. J Cataract Refract Surg. 2015;41(10):2171-8.
3. Provisional publication of Never Events reported as occurring between 1 February and 31 March 2018. London, England: National Health Service, April 27, 2018. (https://improvement.nhs.uk/documents/2613/Never\_Events\_1\_February\_to\_31\_March\_2018\_.pdf)
4. Provisional publication of Never Events reported as occurring between 1 April

2018 and 31 January 2019. London, England: National Health Service, February 27, 2019. (<https://improvement.nhs.uk/documents/4872/Provisional_publication_-__NE_1_April_to_31_Jan_FINAL.pdf>)

1. El Bardissi AW, Sundt TM. Human factors and operating room safety. Surg Clin North Am. 2012;92(1):21-35. doi: 10.1016/j.suc.2011.11.007.
2. Thiels CA, Lal TM, Nienow JM, Pasupathy KS, Blocker RC, Aho JM et al. Surgical never events and contributing human factors. Surgery. 2015;158(2):515-21. doi: 10.1016/j.surg.2015.03.053.
3. Jung JJ, Jüni P, Lebovic G, Grantcharov T. First year analysis of the Operating Room Black Box Study. Ann Surg. 2020;271(1):122-7.
4. Singer SJ, Molina G, Li Z, Jiang W, Nurudeen S, Kite JG et al. Relationship between operating room teamwork, contextual factors, and safety checklist performance. J Am Coll Surg. 2016;223(4):568-80.
5. Göras C, Unbeck M, Nilsson U*,* Ehrenberg A*.* Interprofessional team assessments of the patient safety climate in Swedish operating rooms: A cross-sectional survey. BMJ Open.2017;7:e015607. doi: 10.1136/bmjopen-2016-015607
6. Paige JT, Garbee DD, Bonanno LS, Kerdolff KE. Qualitative analysis of effective teamwork in the operating room (OR). J Surge Ed. 2021;78(3):967-79.
7. Surgical Safety Checklist. The World Health Organization, January, 2009. (https://apps.who.int/iris/bitstream/handle/10665/44186/9789241598590\_eng\_Checklist.pdf;jsessionid=1908B5C90ED0DC4F1362F25B6DE63AEA?sequence)
8. Stawicki SP, Evans DC, Cipolla J, Seamon MJ, Lukaszczyk JJ, Prosciak MP et al. Retained surgical foreign bodies: a comprehensive review of risks and preventive strategies. Scand J Surg. 2009;98(1):8-17.
9. Urbach DR, Govindarajan A, Saskin R, Wilton AS, Baxter NN. Introduction of surgical safety checklists in Ontario, Canada. N Engl J Med. 2014;13(370):1029-38.
10. Moppett IK, Moppett SH. Surgical caseload and the risk of surgical Never Events in England. Anaesthesia. 2016;71(1):17-30. doi: 10.1111/anae.13290
11. OECD. Foreign body left in during procedure, 2017 (or nearest year). In: Quality and outcomes of care. OECD Publishing: Paris; 2019.
12. Logan-Phelan T. The buzz around learning analytics–enablers and challenges identified through the# VLEIreland Project. Ir J Technol Enhanc Learn. 2018;3(2):77-85.
13. Doupe P, Faghmous J, Basu S. Machine learning for health services researchers. Value Health2019;22(7):808-15.
14. Alhusseini MI, Abuzaid F, Rogers AJ, et al. Machine learning to classify intracardiac electrical patterns during atrial fibrillation: machine learning of atrial fibrillation. Circ Arrhythm Electrophysiol. 2020;13(8): e008160.
15. Shalev-Shwartz S, Ben-David S. Understanding machine learning: From theory to algorithms. Cambridge, England: Cambridge University Press; 2014.
16. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. Mach Learn. 2006;63(1):3-42.
17. Gong J, Simon GE, Liu S. Machine learning discovery of longitudinal patterns of depression and suicidal ideation. PloS One*.* 2019;14(9): e0222665.
18. Wongvibulsin S, Wu KC, Zeger SL. Clinical risk prediction with random forests for survival, longitudinal, and multivariate (RF-SLAM) data analysis. BMC Med Res Methodol. 2020;20(1):1-4.
19. Goldberg JL, Feldman DL. Implementing AORN recommended practices for prevention of retained surgical items. AORN. 2012;95(2):205-19.
20. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG et al. Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. BMJ. 2009; 338: b2393.
21. Elkan, C. The foundations of cost-sensitive learning. In: Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence. 2001; 17(1): 973-8.
22. He, H, Ma Y Eds. Imbalanced learning: Foundations, algorithms, and applications. New Jersey: John Wiley & Sons; 2013.‏
23. Hosmer DW, Lemeshow S. Applied logistic regression. 2nd ed. New York, NY: John Wiley & Sons; 2000. 160-164.
24. Fernández A, García S, Galar M, Prati RC, Krawczyk B, Herrera F. Learning from imbalanced data sets. Vol. 10. Berlin: Springer; 2018.
25. Nembrini S, König IR, Wright MN. The revival of the Gini importance? Bioinformatics. 2018;34(21):3711-8.
26. Rodziewicz TL, Houseman B, Hipskind JE. Medical error prevention. Treasure Island, FL: Stat Pearls Publishing; 2020.
27. Moshtaghi O, Haidar YM, Sahyouni R, et al. Wrong-site surgery in California, 2007-2014. Otolaryngol Head Neck Surg. 2017;157(1):48-52.
28. Steelman VM, Shaw C, Shine L, Hardy-Fairbanks AJ. Unintentionally retained foreign objects: a descriptive study of 308 sentinel events and contributing factors. Jt Comm J Qual Patient Saf. 2019; 45(4):249-58.
29. Koleva SI. A literature review exploring common factors contributing to Never Events in surgery. J Perioper Pract. 2020;30(9):256-64. doi:10.1177/1750458919886182
30. Hempel S, Maggard-Gibbons M, Nguyen DK, Dawes AJ, Miake-Lye I, Beroes JM et al. Wrong-site surgery, retained surgical items, and surgical fires: A systematic review of surgical never events. JAMA Surg. 2015;150(8):796-805.
31. Stawicki SP, Moffatt-Bruce SD, Ahmed HM, Anderson HL, Balija TM, Bernescu I, et al. Retained surgical items: A problem yet to be solved. J Am Coll Surg. 2013; 216(1):15-22.
32. Freitas PS, Silveira RC, Clark AM, Galvão CM. Surgical count process for prevention of retained surgical items: An integrative review. J Clin Nurs. 2016;25(13-14):1835-47.
33. Gadelkareem RA. Experience of a tertiary-level urology center in the clinical urological events of rare and very rare incidence. I. Surgical never events: 2. Intracorporeally-retained urological surgical items. Curr Urol. 2017;11(3):151-6.
34. Mahmood T, Mylopoulos M, Bagli D, Damignani R, Haji FA. A mixed methods study of challenges in the implementation and use of the surgical safety checklist. Surgery. 2019;165(4):832-7.
35. Tofte JN, Caldwell LS. Detection of retained foreign objects in upper extremity surgical procedures with incisions of two centimeters or smaller. Iowa Orthop J. 2017; 37:189.
36. Yoo TK, Oh E, Kim HK, Ryu IH, Lee IS, Kim J. Deep learning-based smart speaker to confirm surgical sites for cataract surgeries: A pilot study. PloS One 2020;15(4): e0231322.
37. Elsey EJ, West J, Griffiths G, Humes DJ. Time out of general surgery specialty training in the UK: a national database study. J Surg Educ. 2019;76(1):55-64.
38. Barimani B, Ahangar P, Nandra R, Porter K. The WHO surgical safety checklist: a review of outcomes and implementation strategies. Perioper care Oper Room Manag. 2020;21: 100117.

**Figure titles**

Figure 1: Top 15 contributing features for the six examined departments

Figure 2: Effect of two features’ combination on prediction by surgical departments

Figure 3: Features affecting the wrong surgery site (type A)

Figure 4: Features affecting retained foreign items during surgery (type B)