

What do we Know about Contributing Factors for “Never Events” in Operating Rooms? A Machine Learning Analysis

Dana Arad (✉ dana.arad@moh.gov.il)

Bar-Ilan University

Ariel Rosenfeld

Bar-Ilan University

Racheli Magnezi

Bar-Ilan University

Research Article

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

Posted Date: October 6th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-915811/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

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 surgery's characteristics. Our study uses machine learning to reveal factors and quantify their risk to improve patient safety and quality of care.

Methods

We used data from 9,234 observations on safety standards and 101 Root-Cause Analysis from actual NEs, and utilized three Random Forest supervised machine learning models. Using a standard 10-cross validation technique, we evaluated the model's metrics, and, through Gini impurity we measured the impact of factors thereof to occurrence of the two types of NEs.

Results

We identified 24 contributing factors in six surgical departments. Two had an impact of >90% in Urology, Orthopedics and General Surgery, six had an impact of 0–90% 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: 1,900:2,600%; Cardiology:833–1,500%; Orthopedics:1,825–4,225%; and General Surgery:2,720–13,600%.

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

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 surgery's characteristics, which in turn suggests tailoring the safety standards accordingly.

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 unjustified, and should be reduced to zero through quality improvement. [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, working conditions, such as increased workload and clinician pressure, 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 aiming to reduce NE occurrence: 1) the WHO Surgical Safety Checklist; [11] and 2) surgical counts of all items used during the surgery. [12]

Yet, partial compliance, unstandardized implementation of these standards, [13] and other possible unknown factors keep the incidence of NEs unchanged. [14] In Israel, the incidence of retained foreign items during surgery is 3.2 in every 100,000 surgeries. [15] The incidence for wrong site procedure is unclear, but is generally estimated as 1 in every 100,000 surgeries.

This study adopts a machine learning (ML) approach [16] to identify currently unknown contributors to NE occurrence. Previous studies leveraging ML methods in healthcare have demonstrated the benefits of analyzing and revealing non-trivial insights from diverse data types when compared to 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 ORs.

Methods

Study Design

We utilized a supervised ML method called Random Forest (RF), [18–19] incorporating the popular 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 needed for properly conducting 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 is 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). It is worthwhile mentioning that RFs have been used extensively in the medical field for clinical risk prediction, [21] among other applications.

Safety Standards used in the OR (surgical safety checklists and surgical counts) were divided into safety verification 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 the 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 9,234 surgeries performed between January 2018 and February 2019 in which no NE occurred in the surgeries observed, and *root cause analyses* (RCA) 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 that included simulations. In each OR, at least two observers passively observed randomly selected surgeries, and 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, 9,234 observations were conducted. Each observation was translated into a 93-feature long vector, representing characteristics of the surgery (Appendix 1). To maintain reliability, entries with greater than 5% discordance among annotators in one OR were discarded (<1%).

Root Cause Analyses (RCA)

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, thus, 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 further on.

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% missing feature values revealed that most were strongly dependent on the NE type. Namely, 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 on which the wrong hand was operated, 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 popular 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 high imbalance of the dataset. Specifically, with over 9,000 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 underrepresented 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 NEs Type A; and *Model 3* for differing between observations and NE Type B. We used a standard 10-cross validation technique to evaluate the 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 of 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 NE occurrence probability 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 compared to the value False.

The 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 data set according to surgical specialty

Observations n=9234							Never Events n=101
Phase	*Pre-procedure	Sign in	Time out	First count	Second count	Third count	
Specialty	(n=1,539) (missing data on 760 cases)	(n=1,504)	(n=1,498)	(n=1,518)	(n=1,501)	(n=1,498)	
Urology	72	156	148	124	118	124	7 (6.93%)
Orthopedics	185	331	324	341	302	326	16 (15.84%)
ENT	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%)

In order to evaluate our models, we adopted the Area Under the Curve (AUC) measure which is especially suited for imbalanced data, as in our case in this study, since it does not have any bias toward models that perform well on the minority of majority classes in the expense of the other. [26] Our three RF models demonstrated good performance, exhibiting an Area Under the Curve (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 presents the top contributing features to the occurrence of NEs (of both types combined) in the six departments along with the associated probability change.

The top 14 contributing features varied significantly across departments, and there was no single feature set which was consistently more informative across all operations in 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 1,540%. There were 10 features that consistently decreased the chance of an NE, including [F]; **Surgeon scans the cavity/fascia before closure during the second count**, which affected five out of six departments, which 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 per department shows a variation among contributing features. For example, in Ophthalmology, the probability was consistently -100% in five features, while in General Surgery, two features that increased the probability of an error varied between 1,168–1,283%: 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 (1,540–1,950%). 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**, -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.

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 1,287% and 1,168%, 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 1,000–2,000%. 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), results show there were dozens of pairs with an effect of 1,900–2,500%, 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 1,900–4,200%, while the effect of a single feature had an <1,950% indication on error, and the rest even lower percentages.

Features Affecting Types A and B

Turning to Models 2 and 3, there is an overlap in three of the top five contributing features to Types A and B errors (Figures 3 and 4): 1) the presence of two nurses during the surgery predicts 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 52% with Type B (Figure 4).

Discussion

Surgical errors are a serious public health problem and uncovering their causes is challenging. [29] In this study, we aimed to uncover contributing factors to NEs by using ML methods to identify heretofore unknown contributors, since ML automatically looks for patterns not seen by classic 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 their widespread implementation. [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 the universal safety standards. [29, 34] Some studies have suggested that counting alone is insufficient, and even when declared correct, there have been items left in the patient, [35–36] mostly in the abdomen and pelvis [35, 37] This may explain our higher probability of Type B error 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, what can be explained by partial compliance with the standards. In shorter surgeries, the staff rushes and skips some phases of the checklists [38] and the complex sets used challenges the counts. [31, 39]

We found that the occurrence of wrong site surgery increases in Ophthalmology during short surgeries and when two nurses are 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 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 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 amount, 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 to the relatively high number of analyzed observations. 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 model and estimate feature importance. Nevertheless, given the above, the feature impact should be considered carefully and validated in future study.

In future study we plan to further expand our data pool with newly obtained observations and NEs as those are accumulated. In another avenue, we explore the use of transfer learning of NEs from other countries which could be used to better inform our model. This avenue could prove valuable in mitigating the imbalanced nature of our data yet may introduce significant 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 will consider additional factors such as those identified in this work. These more specific guidelines may be used adjust risk management programs to improve patient safety.

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. All the methods were performed in accordance with guidelines of the "Helsinki" Ethics Committee of the Israeli Ministry of Health. The approval is based on Public Health Regulations – Clinical Experiments on Humans, 5741-1980, including revisions and amendments (1999) Informed consent was waived by the MOH's "Helsinki" 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.

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)
5. 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.
6. 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.
7. 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.
8. 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.
9. 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
10. 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.
11. 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)
12. 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.
13. Urbach DR, Govindarajan A, Saskin R, Wilton AS, Baxter NN. Introduction of surgical safety checklists in Ontario, Canada. *N Engl J Med* .2014;370(13):1029-38.
14. 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
15. OECD. Foreign body left in during procedure, 2017 (or nearest year). In: *Quality and outcomes of care*. OECD Publishing: Paris; 2019.
16. Logan-Phelan T. The buzz around learning analytics—enablers and challenges identified through the# VLEIreland Project. *Irish Journal of Technology Enhanced Learning* .2018;3(2):77-85.
17. Doupe P, Faghmous J, Basu S. Machine learning for health services researchers. *Value Health* 2019;22(7):808-15.
18. 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.
19. Shalev-Shwartz S, Ben-David S. *Understanding machine learning: From theory to algorithms*. Cambridge, England: Cambridge University Press; 2014.
20. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Mach learn* .2006;63(1):3-42.
21. Gong J, Simon GE, Liu S. Machine learning discovery of longitudinal patterns of depression and suicidal ideation. *PloS one* .2019;14(9): e0222665.
22. 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.
23. Goldberg JL, Feldman DL. Implementing AORN recommended practices for prevention of retained surgical items. *AORN* .2012;95(2):205-19.
24. 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.
25. Elkan, C. The foundations of cost-sensitive learning. In: *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence*. 2001; 17(1): 973-8.
26. He, H, Ma Y Eds. *Imbalanced learning: Foundations, algorithms, and applications*. New Jersey: John Wiley & Sons; 2013.
27. Hosmer DW, Lemeshow S. *Applied logistic regression*. 2nd ed. New York, NY: John Wiley & Sons; 2000. 160-164.
28. Fernández A, García S, Galar M, Prati RC, Krawczyk B, Herrera F. *Learning from imbalanced data sets*. Vol. 10. Berlin: Springer; 2018.
29. Nembrini S, König IR, Wright MN. The revival of the Gini importance? *Bioinformatics* .2018;34(21):3711-8.
30. Rodziewicz TL, Houseman B, Hipskind JE. *Medical error prevention*. Treasure Island, FL: Stat Pearls Publishing; 2020.
31. Moshtaghi O, Haidar YM, Sahyouni R, et al. Wrong-site surgery in California, 2007-2014. *Otolaryngol Head Neck Surg* .2017;157(1):48-52.
32. 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.
33. 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
34. 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.
35. Stawicki SP, Moffatt-Bruce SD, Ahmed HM, Anderson HL, Baliya TM, Bernescu I, et al. Retained surgical items: A problem yet to be solved. *J Am Coll Surg* .2013; 216(1):15-22.
36. 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.

37. 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.
38. 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.
39. 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.
40. 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.
41. Eleyse 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.
42. Barimani B, Ahangar P, Nandra R, Porter K. The WHO surgical safety checklist: a review of outcomes and implementation strategies. *perioperative care and operating room management.* 2020;21: 100117.

Figures

Feature	Description	Operation Stage						
A	In case of discrepancy in the count of absorbable items, their package is taken out from the OR (verified out loud)	First Count	-2.78	11				
B	Surgery is paused because of discrepancy in third count	Third count	269	1125		1540	1283	
C	Discrepancy in second count	Second Count	269	1150	160	1950	1168	
D	Length of surgery 1-2 hours	General		-100				
E	Length of surgery >4 hours	General		577				
F	Surgeon scans the cavity/fascia before closure	Second Count	-69		-100	-65	-100	-100
G	Length of surgery <1 hour	General			128			
H	Second count is performed before closure of fascia/cavity	Second Count				-87	-81	
I	Procedure's type is compared to the one written patient's file (verified out loud)	Time out				-87		-100
J	Second count is performed before closure of fascia/cavity	Second Count					-81	
K	Closure of fascia/cavity begins after verification of correct count	Second Count	-77		-100			
L	Third count starts after removing all items from surgical field	Third Count			-100			
M	All staff members agreed to the details mentioned during time out (verified out loud)	Time out						-100
N	Small rural hospital	General						-100
O	Anesthesia consent forms are completed and signed (verified out loud)	Time out						-100

Feature impact on prediction $\geq 900\%$

Feature impact on prediction $> 0\%$ and $<900\%$

Feature impact on prediction $< 0\%$ (decrease chance of an error)

Figure 1

Top 15 contributing features for the six examined departments

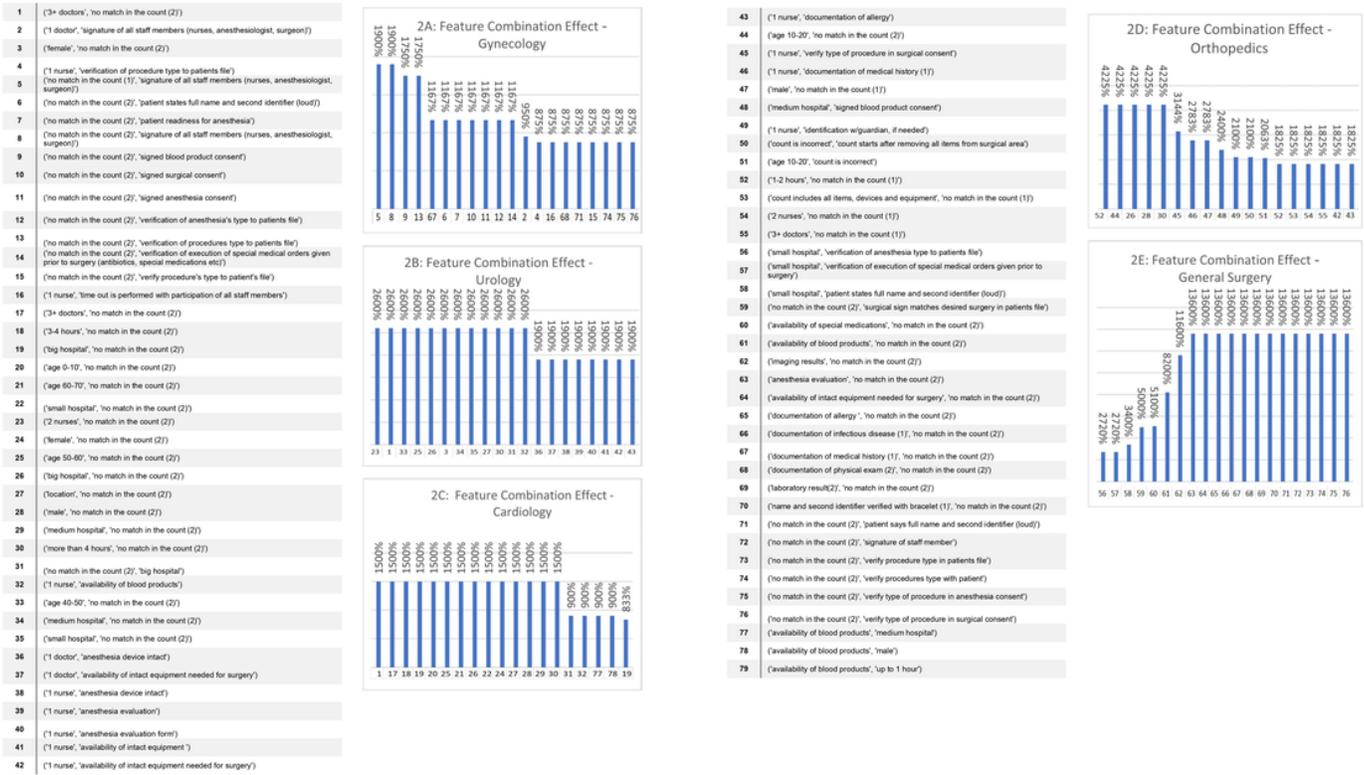


Figure 2

Effect of two features' combination on prediction by surgical departments

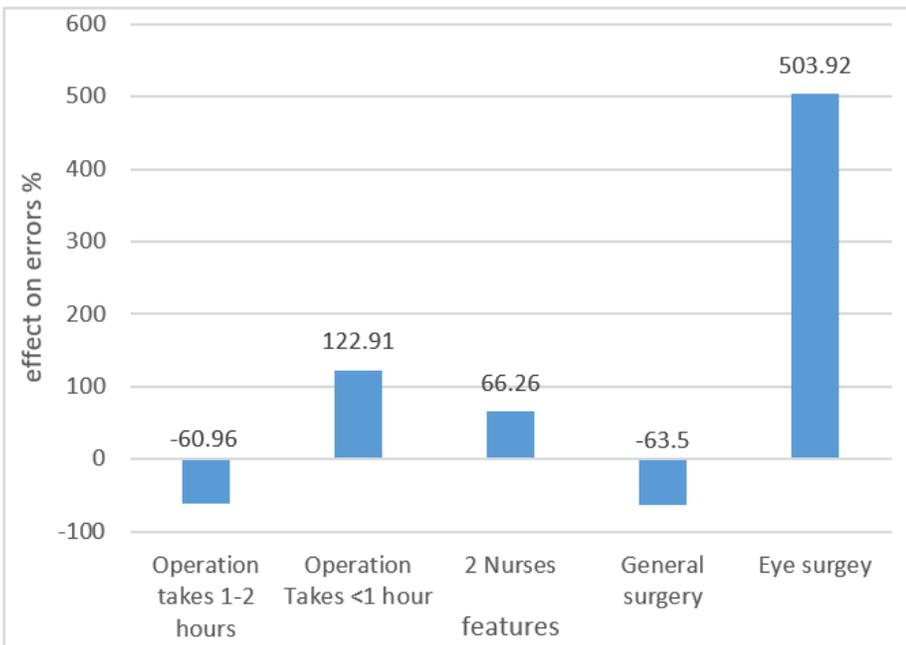


Figure 3

Features affecting the wrong site surgery (Type A)

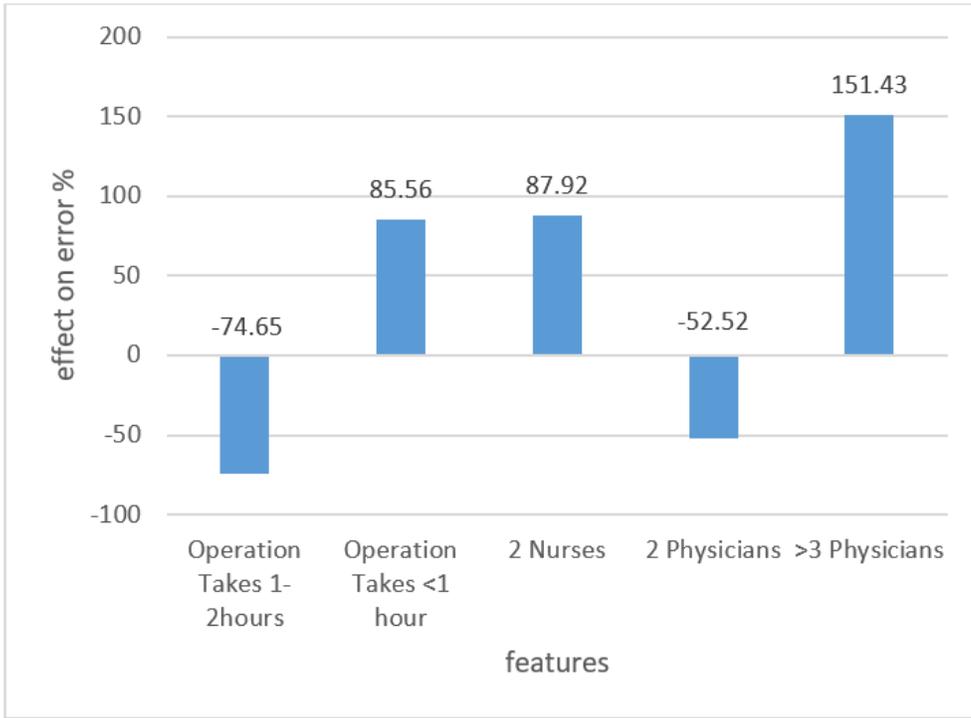


Figure 4

Features affecting retained foreign item during surgery (Type B)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [appendix1.docx](#)