

RESEARCH

Open Access



Factors contributing to preventing operating room “never events”: a machine learning analysis

Dana Arad^{1,3*}, Ariel Rosenfeld² and Racheli Magnezi¹

Abstract

Background A surgical “Never Event” is a preventable error occurring immediately before, during or immediately following surgery. Various factors contribute to the occurrence of major Never Events, but little is known about their quantified risk in relation to a surgery’s characteristics. Our study uses machine learning to reveal and quantify risk factors with the goal of improving patient safety and quality of care.

Methods We used data from 9,234 observations on safety standards and 101 root-cause analyses from actual, major “Never Events” including wrong site surgery and retained foreign item, and three random forest supervised machine learning models to identify risk factors. Using a standard 10-cross validation technique, we evaluated the models’ metrics, measuring their impact on the occurrence of the two types of Never Events through Gini impurity.

Results We identified 24 contributing factors in six surgical departments: two 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%. Combining factors 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 wrong site surgery’s occurrence (-60.96 to 503.92%) and five affected retained foreign body (-74.65 to 151.43%): two nurses (66.26–87.92%), surgery length < 1 h (85.56–122.91%), and surgery length 1–2 h (-60.96 to 85.56%).

Conclusions Using machine learning, we could quantify the risk factors’ potential impact on wrong site surgeries and retained foreign items in relation to a surgery’s characteristics, suggesting that safety standards should be adjusted to surgery’s characteristics based on risk assessment in each operating room. .

Trial registration number MOH 032-2019.

Keywords Never event, Surgery department, Machine learning, Patient safety

*Correspondence:

Dana Arad
danaarad@gmail.com

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

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

³Patient Safety Division, Ministry of Health, Ramat Gan, Israel



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Adverse medical events are preventable, unjustifiable errors that can lead to significant morbidity and mortality and increase healthcare expenditures [1]. They are considered to be entirely preventable with the implementation of quality improvement measures [2]. Major Never Events 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 individual surgeon factors together with work system factors [5], meaning human error is the main contributing factor to Never Events [6]. Human error includes surgeon distraction [7], the surgical team's lack of situational awareness to possible error, and miscommunication among team members [8]. Additionally, institutional factors and working conditions, including increased workload and clinician pressure, can create a work climate un conducive to meeting the standards required to maintain patient safety [9] and effective teamwork [10].

Currently, two essential international standards aim to reduce Never Event 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 meant that the incidence of Never Events has remained unchanged [14]. In Israel, the incidence of retained foreign items during surgery is 3.2 in every 100,000 surgeries [15]. The incidence of wrong site surgeries is unclear but is generally estimated to be 1 in every 100,000 surgeries in Israel.

For this study, we adopted a machine learning approach [16] to identify currently unknown contributors to Never Event occurrence. Previous studies leveraging machine learning methods in health care have demonstrated its advantages in analyzing diverse data types and revealing non-trivial insights compared with traditional methods [17]. To the best of our knowledge, this is the first study using machine learning methods to identify potential contributing factors to the occurrence of Never Events in operating rooms (ORs).

Methods

Study design

We utilized a supervised machine learning method known as random forest [18, 19], incorporating the commonly used extra tree classifier [20]. Random Forest 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 random forest entails several desired elements needed to properly conduct this study's analysis. First, random forests

are used to rank the importance of features in a natural way, determining their importance by examining to what extent the tree nodes using a feature reduce the impurity (i.e., uncertainty in classification) across all "trees in the forest." Second, random forests can cope well with imbalanced datasets (as was the case in this study) and avoid overfitting the data. Finally, random forests compared favorably with several other supervised machine learning algorithms we tested using our data, including popular deep neural networks and support vector machines (SVMs). Random forests have been extensively used in the medical field for clinical risk prediction [21] and other applications.

Safety standards used in the operating room (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 define as type A errors. Surgical counts were divided into three separate counts throughout a surgery to address retained foreign body errors, which we 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, including the hospital's name, length of surgery, patient 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 Never Events occurred during the surgeries observed, and root cause analyses (RCAs) of 101 Never Events that occurred between January 2016 and February 2020 in the examined hospitals.

Observations

Passive observations by medical students, physicians, nursing students, or registered nurses are routinely performed in ORs under the Israel Ministry of Health's supervision. 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, recording and annotating the surgery process using a pre-defined set of features. Observations were then transferred to a central database and were run to assess 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 (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 Never Events occurring between January 2016 and February 2020. We reported 101 Never Events, including 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. Unlike the observations, RCAs were performed retrospectively; 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 machine learning package (<https://scikit-learn.org/stable>).

Examination of the 40% of missing feature values revealed that most were strongly dependent on the Never Event type. Specifically, for type A Never Events, features that were assumed to be more related to Never Events of type B were not investigated and vice versa. For example, for an Never Event 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 pre-closure. To mitigate this artifact, we used the iterative data imputation approach [23], predicting 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 9,000 observations and only 101 Never Events, we adopted a cost-sensitive training approach [24], adjusting our model for prediction mistakes on the minority class (Never Events) by an amount proportional to how under-represented it was (here, approximately 90 times under-represented).

We implemented three random forests models using our data: model 1 to distinguish between observations and Never Events; model 2 for distinguishing between observations and type A Never Events; and model 3 to distinguish between observations and type B Never Events. 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 labels' distribution in the set. We conducted feature importance ranking using the trained random forest models and reported the change in the probability of Never Event occurrence given the entire data set. We considered each feature separately and calculated the probability of Never Event occurrence when that feature assumed the value "True" rather than "False."

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

Results

The majority of Never Events (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

Table 1 Characteristics of the dataset according to surgical specialty

Observations n = 9234							Never Events n = 101
Phase Specialty	*Pre-procedure (n = 1,539) (missing data on 760 cases)	Sign in (n = 1,504)	Time out (n = 1,498)	First count (n = 1,518)	Second count (n = 1,501)	Third count (n = 1,498)	
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 surgeries in the dataset

Characteristic	Observations	Never Events
Gender	Male (n = 388 (49.8%)),	Male (n = 46 (45.5%))
	Female (n = 391 (50.2%))	Female n = 55 (54.5%)
Length of surgery	Up to 1 h: 2124 (23%)	Length of surgery:
	1–2 h: 4340 (47%)	Up to 1 h: 54 (53.5%)
	3–4 h: 2031 (22%)	1–2 h: 13 (12.9%)
	Over 4 h: 739 (8%)	3–4 h: 17 (16.8%)
		Over 4 h: 17 (16.8%)

analysis focused on Never Events’ occurrence in these six departments.

To evaluate our models, we adopted the area under the curve (AUC) measure. This measure 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 random forest models each demonstrated good performance, exhibiting an AUC between 0.81 and 0.85. Generally, AUC scores between 0.8 and 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 Never Events (of both types combined) in the six departments, along with the associated probability change.

The top 14 contributing features varied significantly across departments, and no single feature set was consistently more informative across all operations for predicting Never Events. 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, between 269% and 1,540%. There were 10 features that consistently decreased the chance of a Never Event 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 and 100%. Features [I], [J], [K], [L], [M], and [N] decreased the

Feature	Description	Operation Stage	Ophthalmology	General Surgery	Orthopedics	Cardiology	Urology	Gynecology
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

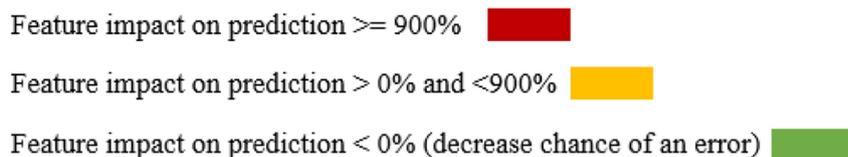


Fig. 1 Top 15 contributing features for the six examined departments

chances of Never Events between 2 and 100% in three departments. Three features, [A] “Discrepancy in absorbing materials,” [E] “Surgery time >4 hours,” and [G] “Surgery time <1 hour” appeared once across departments, with a medium impact on Never Event 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 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,” by -65 to -87%.

Effects of feature combinations

In the following analysis (Fig. 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 Never Events 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% (Fig. 2A). In comparison, the single feature analysis (Fig. 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 Fig. 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 (Fig. 2B), the results showed there were dozens of pairs with an effect of 1,900–2,500%, while the effect of a single feature had <1,150% effect on error. In General Surgery (Fig. 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 a <1,950% effect 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 (Figs. 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 h 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

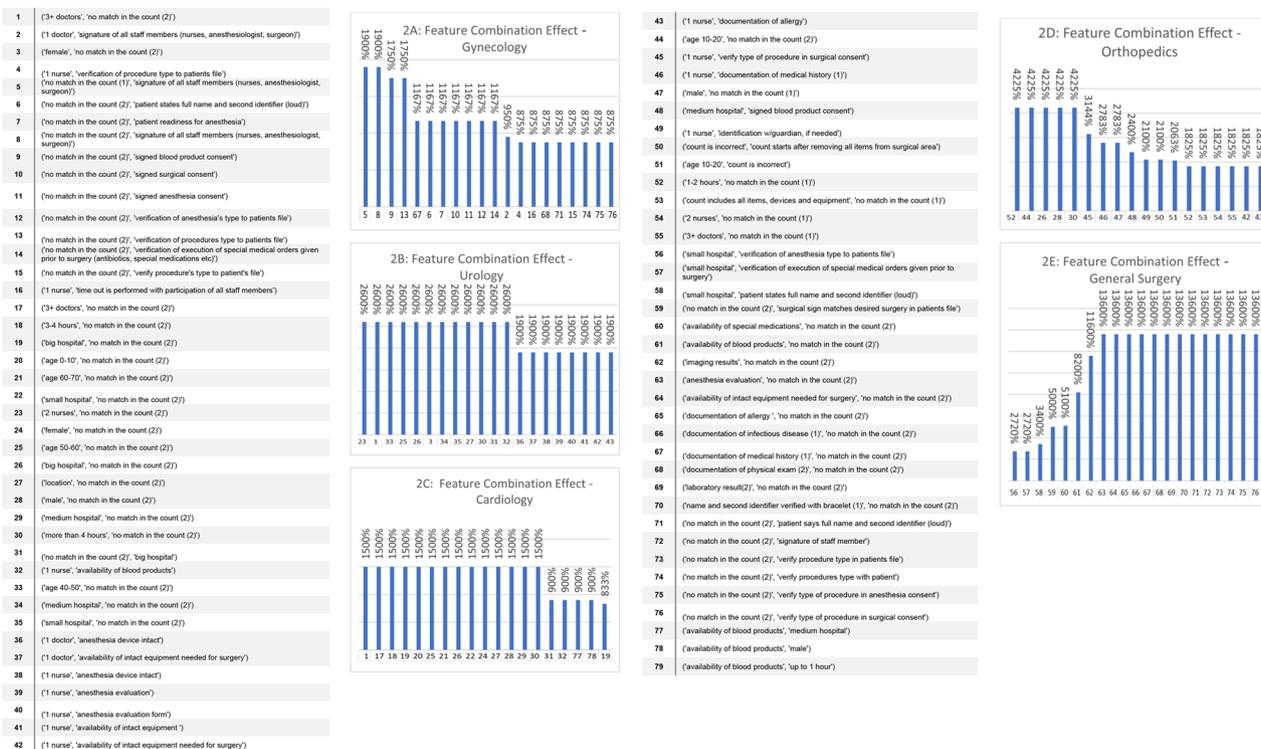


Fig. 2 Effect of two features' combination on prediction by surgical departments

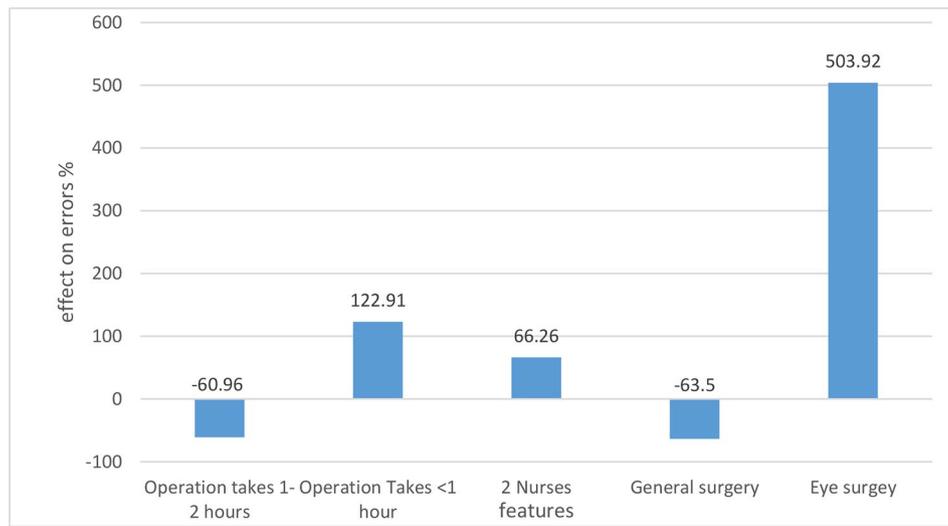


Fig. 3 Features affecting the wrong surgery site (type A)

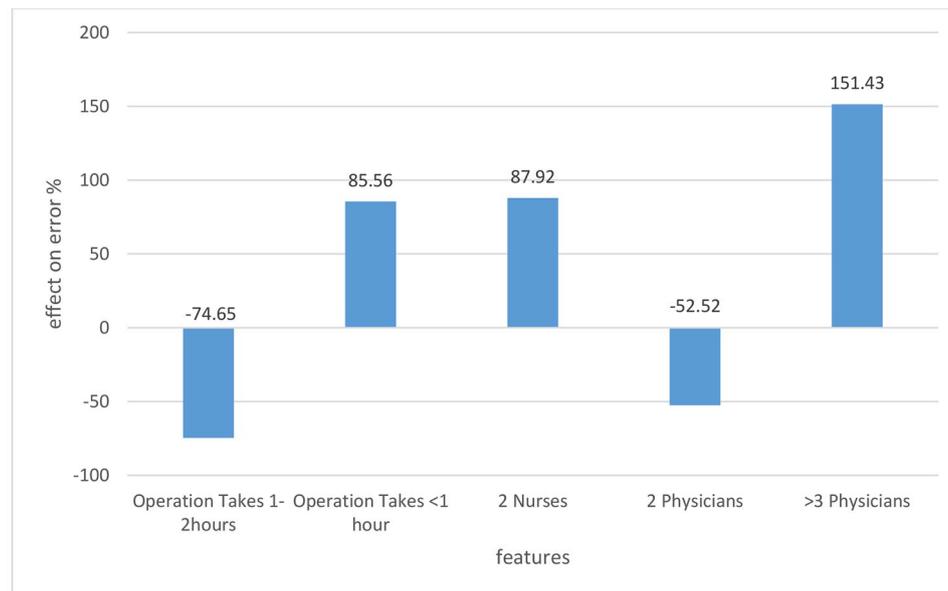


Fig. 4 Features affecting retained foreign items during surgery (type B)

by 60% and 74%, respectively. The surgical department that was most affected regarding the occurrence of type A Never Events was Ophthalmology, with a prevalence of 504%, while General Surgery was associated with a decrease of 63% in type A (Fig. 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% (Fig. 4).

Discussion

Surgical errors are a serious public health problem and uncovering their causes is challenging [29]. In this study, we sought to uncover factors that contribute to Never Events by using machine learning methods to identify heretofore unknown contributors, as machine learning can be used to automate searches for patterns not seen when using traditional methods [18, 30].

The checklists used in the OR, mainly the surgical safety checklist and surgical count, aim to implement strict work processes in order to prevent errors. Despite their widespread use, the incidence of Never Events has not significantly decreased [31, 32], probably because their occurrence is related to human error and not to the

system errors that have been identified as contributing to Never Events. Such system errors are dependent on staff behavior [31, 33]. For example, in our study, discrepancy in the surgical count was found to be a contributing factor to Never Events, while fascia closure after a correct surgical count or staff's agreement to time out were protective factors for prevention of Never Events. Another study supported the impact of the human factor in performing safety standards and occurrence of Never Events and classified them into four categories: preconditions for action, unsafe actions, oversight and supervisory factors, and organization influences [6]. Additional studies have described the lack of safety standards implementation in the OR as arising from a lack of communication and note the lack of empirical evidence relating to barriers to their implementation [29, 34]. Our findings revealed the contribution of discrepancies in the surgical count to occurrence of Never Events. 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 also explain our finding of a higher probability of type B errors in General Surgery and Urology, which involve these 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 h increased the chances for a Never Event, which can be explained by partial compliance with the safety standards. In shorter surgeries, staff may rush and skip some phases of the checklists [38] and the complex surgical devices used during the surgery 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 could be due to the difficulty of performing a time out because the surgeon's hands are sterilized and they cannot review charts, or perhaps because doing so is not made a priority [40]. The decreased risk 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 Never Events is a lack of communication among members of the surgical team [33], which may explain our finding that the number of staff participating in the surgery had a proportional increasing/decreasing effect on Never Event occurrence – and outcome likely affected by lack of communication.

We recognize that the current study is limited by the quantity, quality, and diversity of the data used. Our

samples come from two distinct sources: prospective observations and retrospective investigations of Never Events, the latter consisting of a small number of Never Events compared to the relatively high number of observations analyzed. We believe that these limitations are inherent in the problem studied, as performing prospective analyses of Never Events is virtually impossible due to their infrequency, and the number of Never Events is nominally small. To mitigate some of these concerns, we have used grounded statistical techniques enabling us to train adequate models and estimate feature importance. Nevertheless, given the above, the impact of features 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 Never Events as they are accumulated. In other work, we will explore the use of transferable learning about Never Events 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

In this study, we used machine learning methods to reveal unknown contributing factors to occurrence or prevention of Never Events based on surgery's characteristics, including type, length, and staff presence. We also quantified the contribution of the use of safety standards to occurrence of Never Events.

Our results suggest that the existing, “one size fits all” safety approach should be adjusted to accommodate the surgery's characteristics. Specifically, each Operating Room should perform a risk assessment relative to the occurrence of Never Events during a specific surgery and make tailored adjustments in the safety standards or work environment to prevent them.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13037-023-00356-x>.

Supplementary Material 1

Supplementary Material 2

Acknowledgements

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

Author Contribution

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, and R.M. made a major contribution to the writing of the manuscript. All authors read and approved the final manuscript.

Declarations

Ethics approval and consent to participate

The research was approved by the "Helsinki" ethics committee of the Israel 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

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. This study was funded by grant #MOHIG 14-2019 from the Medical Research Fund for Health Services–Jerusalem. 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, and R.M. made a major contribution to the writing of the manuscript. All authors read and approved the final manuscript.

Received: 10 November 2022 / Accepted: 9 March 2023

Published online: 31 March 2023

References

- 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.
- 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.
- 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)
- 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)
- El Bardissi AW, Sundt TM. Human factors and operating room safety. *Surg Clin North Am*. 2012;92(1):21–35. <https://doi.org/10.1016/j.suc.2011.11.007>.
- 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. <https://doi.org/10.1016/j.surg.2015.03.053>.
- 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.
- 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.
- Görs C, Unbeck M, Nilsson U. Interprofessional team assessments of the patient safety climate in swedish operating rooms: a cross-sectional survey. *BMJ Open*. 2017;7:e015607. <https://doi.org/10.1136/bmjopen-2016-015607>. Ehrenberg A.
- Paige JT, Garbee DD, Bonanno LS, Kerdolff KE. Qualitative analysis of effective teamwork in the operating room (OR). *J Surg Ed*. 2021;78(3):967–79.
- Surgical Safety Checklist. The World Health Organization., January, 2009. (https://apps.who.int/iris/bitstream/handle/10665/44186/9789241598590_eng_Checklist.pdf?sessionid=1908B5C90ED0DC4F1362F25B6DE63AEA?sequence)
- 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.
- Urbach DR, Govindarajan A, Saskin R, Wilton AS, Baxter NN. Introduction of surgical safety checklists in Ontario, Canada. *N Engl J Med*. 2014;370(1):1029–38.
- Moppett IK, Moppett SH. Surgical caseload and the risk of surgical never events in England. *Anaesthesia*. 2016;71(1):17–30. <https://doi.org/10.1111/anae.13290>.
- OECD. Foreign body left in during procedure, 2017 (or nearest year). Quality and outcomes of care. OECD Publishing:Paris; 2019.
- 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.
- Doupe P, Faghmous J, Basu S. Machine learning for health services researchers. *Value Health*. 2019;22(7):808–15.
- 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.
- Shalev-Shwartz S, Ben-David S. Understanding machine learning: from theory to algorithms. Cambridge, England: Cambridge University Press; 2014.
- Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Mach Learn*. 2006;63(1):3–42.
- Gong J, Simon GE, Liu S. Machine learning discovery of longitudinal patterns of depression and suicidal ideation. *PLoS ONE*. 2019;14(9):e0222665.
- 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.
- Goldberg JL, Feldman DL. Implementing AORN recommended practices for prevention of retained surgical items. *AORN*. 2012;95(2):205–19.
- 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.
- Elkan C. The foundations of cost-sensitive learning. In: Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence. 2001; 17(1): 973–8.
- He H, Ma Y, editors. Imbalanced learning: foundations, algorithms, and applications. New Jersey: John Wiley & Sons; 2013.
- Hosmer DW, Lemeshow S. Applied logistic regression. 2nd ed. New York, NY: John Wiley & Sons; 2000. pp. 160–4.
- Fernández A, García S, Galar M, Prati RC, Krawczyk B, Herrera F. Learning from imbalanced data sets. Volume 10. Berlin: Springer; 2018.
- Nembrini S, König IR, Wright MN. The revival of the Gini importance? *Bioinformatics*. 2018;34(21):3711–8.
- Rodziewicz TL, Houseman B, Hipskind JE. Medical error prevention. Treasure Island, FL: Stat Pearls Publishing; 2020.
- Moshtaghi O, Haidar YM, Sahyouni R, et al. Wrong-site surgery in California, 2007–2014. *Otolaryngol Head Neck Surg*. 2017;157(1):48–52.
- 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.
- Koleva SI. A literature review exploring common factors contributing to never events in surgery. *J Perioper Pract*. 2020;30(9):256–64. <https://doi.org/10.1177/1750458919886182>.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- Elsley 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.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.