THE HIP SOCIETY

2020 Frank Stinchfield Award: Identifying who will fail following irrigation and debridement for prosthetic joint infection

A MACHINE LEARNING-BASED VALIDATED TOOL

**N. Shohat,
K. Goswami,
T. L. Tan,
M. Yayac,
A. Soriano,
R. Sousa,
M. Wouthuyzen-Bakker,
J. Parvizi,
On behalf of the ESCMID Study Group of Implant Associated Infections (ESGIAI) and the Northern Infection Network of Joint Arthroplasty (NINJA)**

From The Rothman Institute, Philadelphia, Pennsylvania, USA

Aims

Failure of irrigation and debridement (I&D) for prosthetic joint infection (PJI) is influenced by numerous host, surgical, and pathogen-related factors. We aimed to develop and validate a practical, easy-to-use tool based on machine learning that may accurately predict outcome following I&D surgery taking into account the influence of numerous factors.

Methods

This was an international, multicentre retrospective study of 1,174 revision total hip (THA) and knee arthroplasties (TKA) undergoing I&D for PJI between January 2005 and December 2017. PJI was defined using the Musculoskeletal Infection Society (MSIS) criteria. A total of 52 variables including demographics, comorbidities, and clinical and laboratory findings were evaluated using random forest machine learning analysis. The algorithm was then verified through cross-validation.

Results

Of the 1,174 patients that were included in the study, 405 patients (34.5%) failed treatment. Using random forest analysis, an algorithm that provides the probability for failure for each specific patient was created. By order of importance, the ten most important variables associated with failure of I&D were serum CRP levels, positive blood cultures, indication for index arthroplasty other than osteoarthritis, not exchanging the modular components, use of immunosuppressive medication, late acute (haematogenous) infections, methicillin-resistant *Staphylococcus aureus* infection, overlying skin infection, polymicrobial infection, and older age. The algorithm had good discriminatory capability (area under the curve = 0.74). Cross-validation showed similar probabilities comparing predicted and observed failures indicating high accuracy of the model.

Conclusion

This is the first study in the orthopaedic literature to use machine learning as a tool for predicting outcomes following I&D surgery. The developed algorithm provides the medical profession with a tool that can be employed in clinical decision-making and improve patient care. Future studies should aid in further validating this tool on additional cohorts.

Cite this article: *Bone Joint J* 2020;102-B(7 Supple B):11–19.

Introduction

Among the options for treating prosthetic joint infection (PJI), irrigation and debridement (I&D) is associated with the least morbidity. Currently I&D is reserved for patients presenting with acute PJI, with variable definitions.[1-5] However, failure rates for this intervention are reported to vary between 30% and 80%.[6-14] Besides the effect on the patient and the healthcare burden, it has been suggested that an unsuccessful I&D may compromise the outcome of subsequent exchange arthroplasty.[15,16] Thus, it is essential to determine which group of patients are likely to have a high failure rate following I&D and reserve this procedure for the most appropriate candidates.

Historically, the time from index arthroplasty and acuteness of symptoms have been the predominant factors that have determined the utility of I&D. The time from index arthroplasty is believed to be important as a longer period with infection

Correspondence should be sent to J. Parvizi; email: javadparvizi@gmail.com

© 2020 The British Editorial Society of Bone & Joint Surgery

doi:10.1302/0301-620X.102B7.BJJ-2019-1628.R1 $2.00

Bone Joint J 2020;102-B(7 Supple B):11–19.

allows for biofilm to form and mature.[17] The latter notion has been the main impetus for choosing time intervals from index arthroplasty to define 'acute' PJI. However, recent data support the use of I&D as a reasonable option even more than four weeks from index arthroplasty, as long as it is performed within one week of symptoms, and modular components can be exchanged during the procedure.[18,19] During the recent International Consensus Meeting (ICM), most delegates believed that a binary division between acute and chronic PJI based on time from index arthroplasty was illogical.[20]

The ICM concluded that when deciding on the treatment plan for patients presenting with PJI in the early period following arthroplasty, surgeons should focus on factors that have been found to affect the outcome of I&D, which include host-related factors, the implant integrity, and the type of infecting organism to name a few.[20-22] Prior investigators have attempted to develop 'prognostic' classifications that could be used to determine the role of I&D. A study by Tornero et al[23] introduced the KLIC classification for early acute (post-surgical) infections and another study by Wouthuyzen-Bakker et al[24] proposed the CRIME-80 classification for late acute (haematogenous) infections. Both of these criteria have been widely applied.

Recent developments in machine learning have opened the door into more comprehensive, accurate, and user-friendly platforms that may help surgeons in decision-making. Using data from multiple centres across the USA and Europe, the aim of this study was to develop an algorithm suitable for use in everyday clinical practice that could predict the probability of failure/success in patients undergoing I&D, taking into account the influence of a large number of variables.

Methods

This was a retrospective, multicentre study including 27 centres throughout the USA and Europe (Supplementary Table i). All centres involved in this study have a dedicated team for treating PJI including infectious disease (ID) specialists and microbiologists. Following Institutional Review Board approval, we extracted data pertaining to total hip (THA) and knee arthroplasty (TKA) patients who developed early acute (post-surgical) and late acute (haematogenous) PJI between January 2005 and December 2017 and who were treated with I&D.

Early acute (post-surgical) PJIs were defined as infections occurring within three months of index arthroplasty and which were treated within the same time period. Late acute (also known as haematogenous) PJIs were considered infections occurring longer than three months from index arthroplasty, presenting as abrupt symptoms in patients who were previously asymptomatic, and lasting for less than three weeks prior to I&D. Only patients meeting the Musculoskeletal Infection Society (MSIS) criteria for infection were included.[25] Patients undergoing I&D who did not meet these aforementioned definitions were excluded from the study, as well as those with missing surgical or demographic data, and those with less than one-year follow-up. Overall, 1,174 patients underwent I&D due to early acute (n = 790) and late acute (n = 384) PJI, which included 565 hips and 609 knees.

A total of 52 variables that included patient characteristics, comorbidities, clinical presentation, organism profile, and surgical treatment were collected (Table I). A manual chart review was performed on all patients meeting the above inclusion criteria. Time from index surgery to I&D was documented in patients with early acute (post-surgical) infections and time from development of symptoms to I&D was documented for late acute (haematogenous) infections only. Patient characteristics (age, sex), body mass index (BMI), smoking, alcohol, index surgery (primary/revision), and comorbidities (ischaemic heart disease, heart failure, hypertension, chronic obstructive pulmonary disease (COPD), chronic renal failure, liver cirrhosis, active malignancy, rheumatoid arthritis (RA), use of oral anticoagulants, and immunosuppression medication) were evaluated. The indication for index arthroplasty was dichotomized to osteoarthritis and causes other than osteoarthritis (including RA, fracture, and osteonecrosis of femoral head). The use of cement during index surgery was also documented. Clinical findings (persistent wound drainage, skin necrosis, skin infection, and a sinus track) and laboratory results (serum CRP, serum white blood cell count (WBC), as well as positive blood cultures) were also recorded. The details of I&D surgery, as much as possible, were also extracted to include variables such as exchange of modular components. The type of infecting organism was also recorded and evaluated. Fever was defined as any value above 38.0°C prior to surgery. Four variables (dementia, HIV, use of a plastic surgical flap, and endocarditis) were removed during the analysis due to their rarity. An additional variable (pacemaker) was removed due to too many missing data points.

Failure was defined if any of the following conditions were met: subsequent prosthesis removal during the follow-up period after I&D; use of suppressive antibiotic therapy due to persistent clinical or laboratory signs of infection; and reinfection of the index joint with the same initial organism(s) or different organism(s). Treatment success was defined as the ability to retain the initial arthroplasty hardware, with no clinical or laboratory signs and symptoms of infection, and without the need for suppressive antibiotic treatment at a minimum one-year follow-up.

Statistical analysis. Analyses were performed with an objective to create an algorithm that predicts the probability for failure of I&D. First, the most appropriate model for analysis was examined. Four models were created: random forest analysis (trying several tuning parameters), logistic regression with all variables, stepwise (forward) logistic regression, stepwise (forward) logistic regression including all interactions with the variable: timing of infection (early acute (post-surgical) and late acute (haematogenous)). Area under the curve (AUC), sensitivity, and specificity were obtained for each model. Random forest showed the highest AUC (0.74) and was therefore the model of choice for continued analysis.

The random forest is a type of machine learning classification algorithm consisting of many decision trees. It uses bagging and feature randomness when building each individual tree to try to create an uncorrelated forest of trees whose prediction by committee is more accurate than that of any individual tree. By resampling of both observation and covariates for constructing each tree, it is robust for outliers. In addition, the tuning parameters were chosen based on k-fold cross-validation, which makes it even more robust for outliers and reduces the possibility of

Table I. Patient demographics, characteristics, clinical presentation, and organism profile stratified based on treatment outcome (success vs failure).

Variable	Failure (n = 405)	Success (n = 769)	p-value
Timing			
Acute (postoperative), n (%) (n = 790)	276 (34.9)	514 (65.1)	0.695*
Mean time from index surgery to I&D, days (SD)	13.7 (12.9)	14.4 (15.1)	0.358†
Acute haematogenous, n (%) (n = 384)	129 (33.6)	255 (66.4)	0.691*
Mean time from symptoms to I&D, days (SD)	2.7 (7.9)	2.0 (5.2)	0.101†
Demographics and comorbidities			
Mean age, yrs (SD)	70.5 (12.5)	69.9 (11.6)	0.432†
Sex (male), n (%)	204 (50.4)	334 (43.4)	0.027*
Mean body mass index (BMI), kg/m^2 (SD)	30.5 (6.4)	30.9 (6.6)	0.267†
Smoking, n (%)	93 (23.0)	201 (26.1)	0.257*
Alcohol, n (%)	148 (36.5)	274 (35.6)	0.798*
Joint (knee), n (%)	197 (48.6)	412 (53.6)	0.111*
Hypertension, n (%)	246 (60.7)	475 (61.8)	0.753*
Ischaemic heart disease, n (%)	70 (17.3)	85 (11.1)	0.004*
Heart failure, n (%)	50 (12.3)	84 (10.9)	0.499*
Oral anticoagulants, n (%)	81 (20.0)	119 (15.5)	0.060*
Diabetes mellitus, n (%)	95 (23.5)	146 (19.0)	0.080*
Chronic obstructive pulmonary disease, n (%)	72 (17.8)	104 (13.5)	0.048*
Chronic renal failure, n (%)	36 (8.9)	53 (6.9)	0.246*
Liver cirrhosis, n (%)	20 (4.9)	21 (2.7)	0.065*
Active malignancy, n (%)	43 (10.6)	96 (12.5)	0.393*
Rheumatoid arthritis (RA), n (%)	39 (9.6)	46 (6.0)	0.025*
Immunosuppression medications, n (%)	63 (15.6)	75 (9.8)	0.004*
History of infected prosthesis, n (%)			
Index surgery was a revision	96 (23.7)	155 (20.2)	0.182*
Index surgery used cemented prosthesis	296 (73.1)	535 (69.6)	0.224*
Indication for primary arthroplasty			< 0.001 *
Osteoarthritis	307 (31.0)	684 (69.0)	
Other (RA, fracture, or malignancy)	98 (53.6)	85 (46.4)	
Clinical findings, n (%)			
Wound leakage	220 (54.3)	398 (51.8)	0.424*
Skin necrosis	92 (22.7)	165 (21.5)	0.656*
Skin infection	140 (34.6)	180 (23.4)	< 0.001*
Fistula	90 (22.2)	212 (27.6)	0.049*
Fever (> 38°C)	100 (24.7)	161 (20.9)	0.161*
Laboratory findings			
Mean serum CRP, mg/dl (SD)	14.9 (11.3)	11.5 (11.5)	< 0.001 †
Mean serum WBC, ×10^9/l (SD)	13.0 (30.3)	10.5 (8.3)	0.109†
Positive blood cultures, n (%)	121 (29.9)	144 (18.7)	< 0.001 *
Operative factors, n (%)			
Exchange of mobile component	174 (43.0)	410 (53.3)	< 0.001*
Organism profile, n (%)			
Gram-positive			
MSSA	80 (19.8)	135 (17.6)	0.382*
MRSA	128 (31.6)	170 (22.1)	< 0.001*
Staphylococcus epidermidis	91 (22.5)	194 (25.2)	0.316*
Streptococcus spp	66 (16.3)	128 (16.6)	0.934*
Enterococcus spp	45 (11.1)	83 (10.8)	0.922*
Gram-negative			
Escherichia coli	32 (7.9)	49 (6.4)	0.334*
Enterobacter spp	14 (3.5)	26 (3.4)	1.000*
Pseudomonas spp	13 (3.2)	35 (4.6)	0.352*
Proteus spp	18 (4.4)	23 (3.0)	0.241*
Candida spp	5 (1.2)	5 (0.7)	0.327*
Polymicrobial	128 (31.6)	201 (26.1)	0.046*

I&D, irrigation and debridement; MSSA, methicillin-sensitive *Staphylococcus aureus*; MRSA, methicillin-resistant *Staphylococcus aureus*; WBC, white blood cell count.
*Chi-squared test.
†Independent-samples *t*-test.

overfitting. Cross-validation is a powerful preventive tool to avoid overfitting as it uses initial training data to generate multiple mini train-test splits and then uses these splits to tune the model.

Data were analyzed based on the premise that clinicians will have all datasets available when using the algorithm in practice. Missing data were filled by five imputations using the multiple imputations by chained questions (MICE) procedure and each imputed dataset was then fitted.[26] The fitting was done through k-fold cross-validation, which is a statistical method used to estimate the performance of machine learning models on future data without an external validation dataset. For the training model, the data were randomly divided into eight folds. Seven of these folds were used as the learning dataset to construct the training model, while the one remaining fold was used as the test dataset to determine model performance. This process was repeated seven times, each time with a new fold acting as the test dataset. The general AUC of the model is the mean AUC over the eight folds. The relative importance of each variable was examined. Timing of infection (early acute (post-surgical) vs late acute (haematogenous)) and type of joint (hips vs knees) were analyzed in the general model as well as in separate subgroup analysis. Statistical significance was set at a p-value < 0.05.

Results

Of the 1,174 PJIs that were included in the study, 405 patients (34.5%) failed to have their infection eradiated following I&D. There were significant differences in patient demographics, characteristics, clinical presentation, and organism profile between patients who failed compared to those who did not fail treatment (Table I). Risk factors for failure in the initial univariate analysis were male sex (p = 0.027, chi-squared test), ischaemic heart disease (p = 0.004, chi-squared test), COPD (p = 0.048, chi-squared test), RA (p = 0.025, chi-squared test), use of immunosuppressive medication (p = 0.004, chi-squared test), indication other than osteoarthritis for index surgery (p < 0.001, chi-squared test), overlying skin infection (p < 0.001, chi-squared test), presence of a sinus tract (p = 0.049, chi-squared test), higher serum CRP levels (p < 0.001, independent-samples

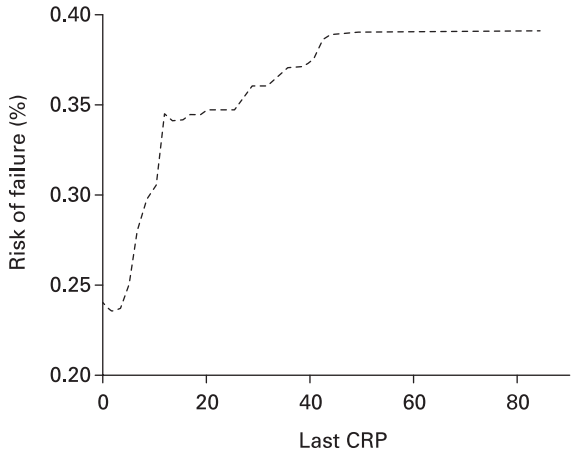

Fig. 1

Partial dependence plot of CRP levels on risk for failure.

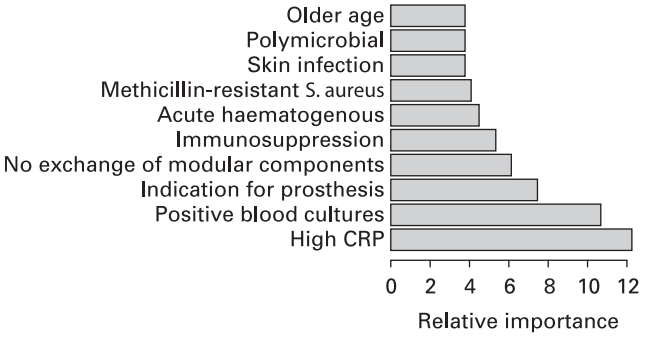

Fig. 2

Random forest analysis showing the ten most important factors associated with failure by order of importance. *S. aureus, Staphylococcus aureus*.

t-test), positive blood cultures ($p < 0.001$, chi-squared test), not exchanging the modular component ($p < 0.001$, chi-squared test), methicillin-resistant *Staphylococcus aureus* (MRSA) ($p < 0.001$, chi-squared test), and polymicrobial ($p = 0.046$, chi-squared test) infections.

Random forest analysis for the entire cohort showed that the ten most important factors associated with failure were the following (by order of importance): higher CRP levels (Figure 1), positive blood cultures, indication for index surgery other than osteoarthritis, not exchanging the modular component, use of immunosuppressive medication, late acute (haematogenous) infections, MRSA, overlying skin infection, polymicrobial infection, and older age (Figure 2). While the order of importance remained relatively similar in patients with early acute (post-surgical) infections, patients with late acute (haematogenous) infections exhibited several differences compared to the early acute (post-surgical) cohort, as well as the overall cohort. In patients with late acute (haematogenous) infections, days of symptoms prior to I&D, use of immunosuppressive medications, and *Staphylococcus epidermidis* infections were the three most important factors predicting failure (Figure 3).

Stratifying patients by type of arthroplasty (TKA and THA) and running a random forest on each group separately showed similar results to the entire cohort in regard to two of the three most important variables: high CRP levels and positive blood cultures. However, indication other than osteoarthritis for index surgery, use of cement in index surgery, and persistent wound drainage were more important indicators for failure in the hip cohort while not exchanging the modular component and time from index surgery to I&D were more important in the knee cohort. (Figure 4).

The algorithm created by random forest was tested using cross-validation. Comparison of predicted (expected) and observed (empirical) failure showed similar probabilities, indicating high accuracy of the model (Table II). Patients were grouped into categories according to predicted probability of failure and the proportion of actual failure in each category was examined; high agreement was seen (Figure 5).

Figure 6 illustrates the proportion (percentage) of patients in each probability group. In 556 patients (47.3%) the algorithm predicted a risk lower than 30% for failure (relatively low risk) and indeed only 121 patients in this category failed (21.8%). The predicted risk for failure was 30% to 50% (intermediate risk) in 463 patients (39.4%) and of those, 174 (37.6%) failed. A total of 155 patients (13.2%) had a high probability (above 50%) for failure and failure indeed occurred in 97 of these patients (62.5%). Examples of different clinical scenarios and the algorithm predictions are demonstrated in Table III.

Discussion

I&D is an appealing surgical procedure for management of acute PJI as it carries a low morbidity. However, the outcome of this surgical procedure is unpredictable and the reported failure rate of the procedure varies greatly.[27-34] Numerous studies have sought to identify specific risk factors that are associated with failure, and criteria for failure have also been proposed by prior studies. Prior criteria, namely KLIC and CRIME-80, were introduced that evaluated the importance of preoperative factors for failure of I&D.[23,24] The latter classification systems have been applied orthopaedically for the last few years. The present multicentre, international study was designed and conducted to take advantage of recent developments in a machine learning algorithm that is patient-specific and can more accurately predict probability for failure/success of treatment. This study represents a major advancement in decision-making prior to surgery and offers a for future machine learning studies in the field of orthopaedics as another step towards personalized medicine.

The major difference between machine learning and simple statistics is their purpose. Machine learning models are designed to make the most accurate predictions possible, whereas statistical modelling is more about finding relationships between variables and the significance of those relationships. For many cases, especially in research, the point of our model has been to characterize the relationship between the data and our outcome variable, but not to make predictions about future data. The purpose of machine learning is to obtainin a model that can make repeatable predictions. To investigate the best modality for predicting outcome we compared four types of analyses. Random forest analysis showed the highest AUC and thus was

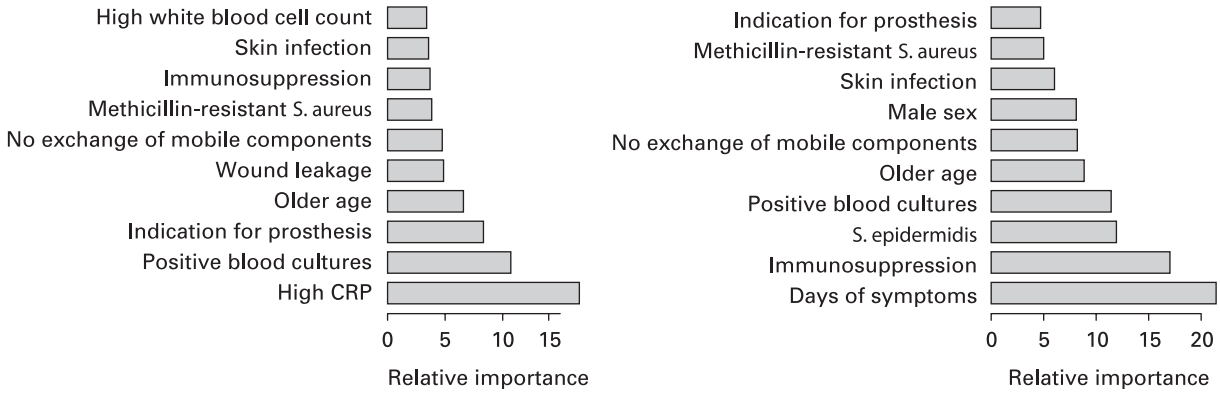

Fig. 3

Random forest analysis stratified based on acute (postoperative; left) and acute haematogenous (right) infections. *S. aureus, Staphylococcus aureus; S. epidermidis, Staphylococcus epidermidis*.

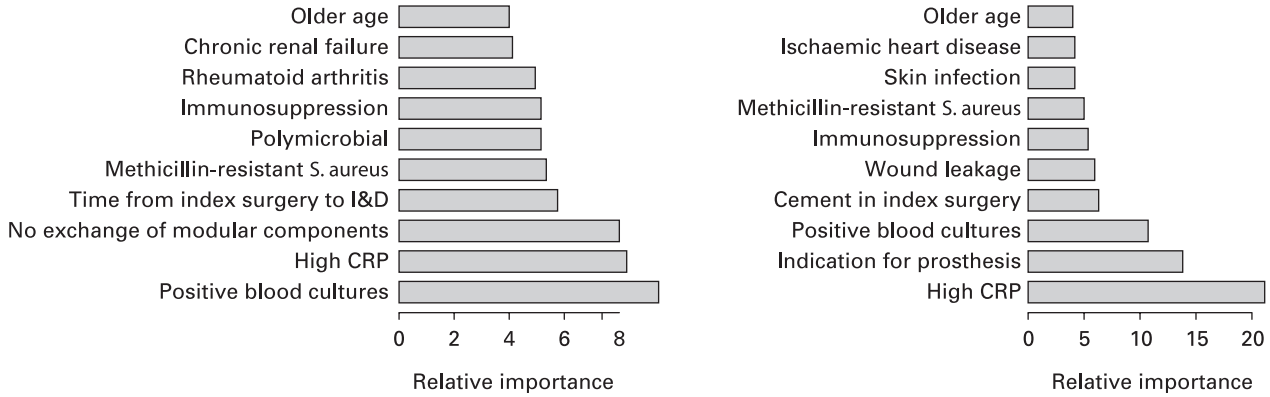

Fig. 4

Random forest analysis stratified based on knee (left) and hip (right) infections. I&D, irrigation and debridement. *S. aureus, Staphylococcus aureus*.

Table II. Cross-validation of the random forest algorithm comparing predicted and observed probability of failure. The number of patients and the failure rate in each probability group is presented.

Predicted probability for failure (grouped)	Number	Failure, n (%)	Mean predicted failure (SD)	Mean observed failure
0.0 to 0.1	63	7 (11.1)	0.08 (0.01)	0.11
0.1 to 0.2	211	41 (19.4)	0.15 (0.03)	0.19
0.2 to 0.3	282	73 (25.9)	0.25 (0.03)	0.26
0.3 to 0.4	277	101 (36.5)	0.35 (0.03)	0.36
0.4 to 0.5	186	86 (46.2)	0.45 (0.03)	0.46
0.5 to 0.6	100	58 (58.0)	0.54 (0.03)	0.58
0.6 to 0.7	44	30 (68.2)	0.64 (0.03)	0.68
0.7 to 0.8	10	8 (80.0)	0.74 (0.03)	0.80
0.8 to 0.9	1	1 (100)	1.00 (N/A)	1.00

N/A, not applicable.

chosen for creating the algorithm. The major disadvantage of these kinds of models is that they are not easily interpretable (i.e. black box analysis) and it remains unclear how the algorithm predicts outcome. Therefore, there is always a need to validate the stability of machine learning models to see whether the captured patterns from the data are indeed correct. We used k-fold cross-validation for this purpose. In this method only a subset of the data is used for training the random forest model, and the remaining subset data are then used to validate the model. The big advantage that comes with k-fold cross-validation is that it is much less prone to selection bias since training and testing is performed on several different parts. Performing eight-fold cross-validation allowed us to be even more certain of the robustness of our model since training and testing were performed on many different sub-datasets.

Machine learning has been introduced into many areas within the healthcare system with the potential to revolutionize the medical landscape.[35,36] The main difference between common approaches to data analysis and machine learning is that in the latter, a model learns from observations instead of being programmed with predefined rules. By using decision trees and algorithms for learning from observations, a model is then created that will generalize the information so that an assumption can be achieved correctly with input that have not been seen before. In cases such as predicting outcome of treatment, where accuracy is of extreme importance, the ability of machine learning to find arrangements through millions of features and examples is what distinguishes it from common analysis. In this

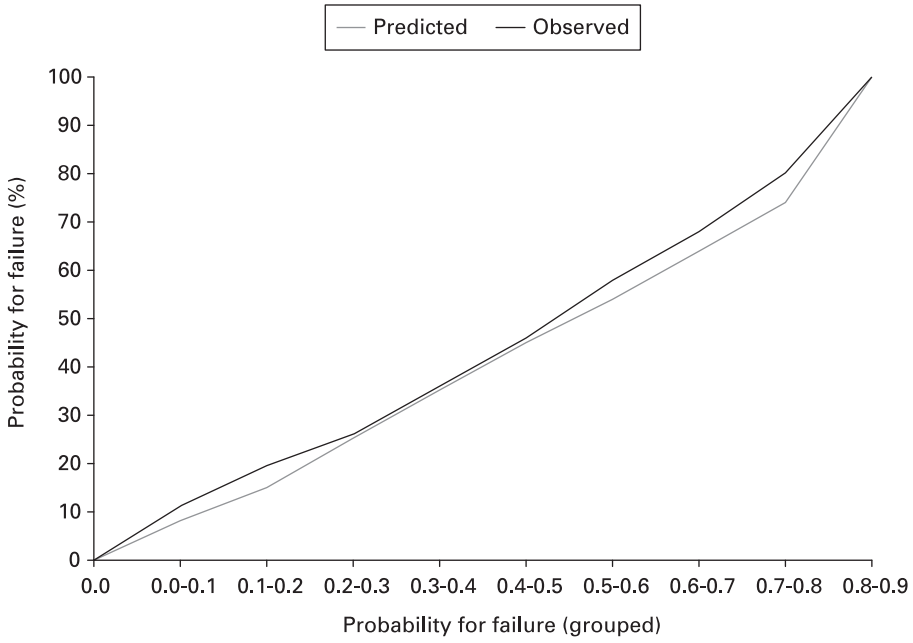

Fig. 5

Predicted probability of failure versus observed probability for failure.

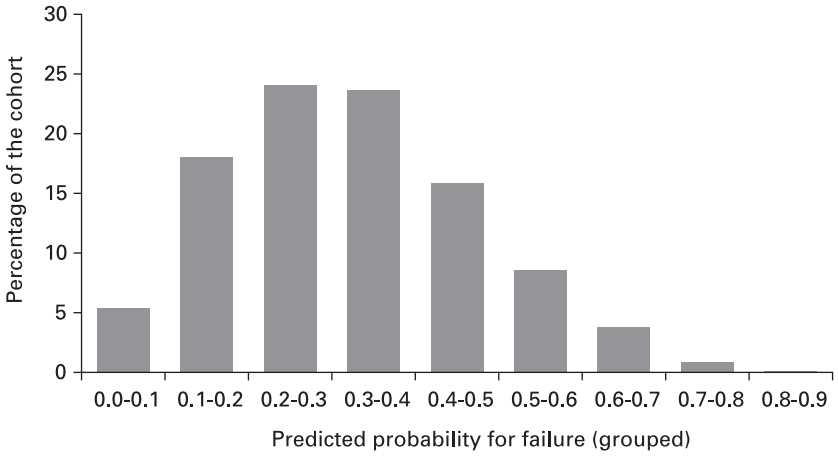

Fig. 6

Proportion (percentage) of patients in each probability group.

study, machine learning through random forest analysis created an algorithm to predict success/failure of I&D based on a wide range of comorbidities, clinical scenarios, and presentations, as well as physical findings and laboratory results that may appear in thousands of different combinations. This algorithm can be used in daily practice by easily entering a computer-based software or telephone application (Figure 7).

Past studies have acknowledged several risk factors associated with failure of I&D, including increasing age, RA, use of immunosuppressive drugs, elevated CRP, bacteraemia, as well as infections by MRSA and polymicrobial infections.[37-40] While the nature of random forest ('black box') analysis, such as the one used in the present study, does not allow us to fully understand the decision trees generated to reach the final algorithm, looking at the random forest relative importance does provide insight into its decision-making. All ten of the most important variables that were pointed out by the random forest have been previously reported as variables that are associated with failure, providing reassurance that the algorithm is not only accurate, but consistent with accepted risk factors. Since previous studies noted differences in treatment failure between early acute (post-surgical) and late acute (haematogenous) PJI,[40,41] we performed a subgroup analysis investigating these cohorts separately. Indeed, the groups differed in the order of importance given to each variable and the algorithm takes this into account as well.

Table III. Example patient characteristics, expected probability for failure, and actual outcome.

Timing	Age, yrs	BMI, kg/m²	Comorbidity	History	Clinical finding	Laboratory findings	Mobile exchange	Organism	Probability for failure	Outcome
Acute (28 days)	75.1	29.7	Alcohol, hypertension, diabetes	TKA, OA, + cement	None	CRP* = 0.79, WBC† = 8.7	+	*Staphylococcus epidermidis*	0.03	Success
Acute (21 days)	69.9	33.8	Diabetes	TKA, OA, - cement	None	CRP = 3.7, WBC = 4.3	+	*Staphylococcus epidermidis*	0.16	Success
Acute haematogenous (one day)	84.0	26.9	Hypertension, immunosuppression	TKA, OA, + cement	Necrosis, fistula	CRP = 10.9, WBC = 9.2, + fever	+	Gram (-)	0.28	Success
Acute (eight days)	92.1	22.8	Alcohol, RA	TKA, index revision, + cement	Necrosis, fistula	CRP = 20.8, WBC =16.8, + fever + blood cx	+	*Streptococcus* spp.	0.37	Success
Acute haematogenous (30 days)	76.2	36.7	Smoking, alcohol,	TKA, index revision, + cement	Necrosis, fistula	CRP = 31.9, WBC = 8.7, + fever + blood cx	+	None	0.44	Failed
Acute (22 days)	73.3	26.2	COPD, CRF, immunosuppression	THA, index revision, + cement	Wound leakage, skin infection	CRP = 41.7, WBC = 7.8, + fever	+	MRSA	0.56	Failed
Acute (15 days)	77.7	34.7	Hypertension	THA, OA, + cement	Skin infection	CRP = 36.4, WBC = 6.1, + fever + blood cx	+	Polymicrobial (MRSA, gram (-))	0.63	Failed
Acute (16 days)	85.1	24.4	Hypertension	THA, not OA, + cement	Wound leakage, skin infection	CRP = 22.6, WBC = 12.4, + fever, + blood cx	-	MRSA	0.76	Failed
Acute haematogenous (six days)	80.1	24.2	IHD, diabetes, COPD	THA, not OA, + cement	Wound leakage, skin infection	CRP = 47.6, WBC = 22.9, + fever, + blood cx	-	MRSA	0.86	Failed

*Units mg/dl.
†Units $\times 10^9$/l.
CRF, chronic renal failure; COPD, chronic obstructive pulmonary disease; CX, cultures; IHD, ischemic heart disease; MRSA, methicillin-resistant *Staphylococcus aureus*; OA, osteoarthritis; RA, rheumatoid arthritis; THA, total hip arthroplasty; TKA, total knee arthroplasty; WBC, white blood cell count

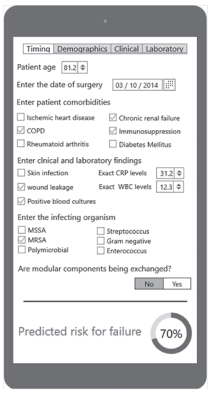

Fig. 7

Mockup example of a computer-based software or phone application that can be used to predict failure of treatment.

The algorithm was tested using cross-validation and was found to be accurate as predictions and true observations overlapped in the vast majority of cases. In 63 patients with less than 10% probability for failure, indeed failure rate stood at 11.1% (7/63), in patients with probability of failure of 10% to 20%, probability of failure was 19.4% (41/211), when probability was 20% to 30%, 25.9% (73/282) failed, and so on. Notably, 55 patients (4.7%) had a probability of failure above 60% and of those 70.9% (39/55) failed. There were no patients with greater than 90% probability of failure. Currently surgeons take into account and counsel patients for whom the risk of failure lies roughly between 30% and 80% without a practical tool to distinguish one from the other. Knowing that a certain patient has less than 10% chance of failure or greater than 50% risk for failure should help clinicians greatly in treatment planning and counselling. Perhaps in the future with additional variables and larger cohorts, we will be able to limit the proportion of outlying patients, which will further help surgeons in decision-making. In the meantime, the provided algorithm provides a clinical tool never seen before.

There were several limitations to our study. First, the retrospective design and multicentre nature of the investigation over a long period of time is likely to have introduced biases associated with differences in management and treatment protocols. These differences could also explain the relatively high failure rate seen in this study. We could not rely on data regarding antibiotic treatment, especially use of rifampicin combinations and presence of fluoroquinolone in the antibiotic protocol, which

are known to positively impact treatment outcomes.[42,43] We could also not account for the quality of I&D performed nor specify whether a specialized surgeon was present during the procedure, which may also have contributed to success/failure. That being said, all centres involved in this study have a dedicated team for treating PJI including ID-specialists and microbiologists. Secondly, as mentioned earlier due to the endless observations and decision trees, 'black box' analysis did not allow us to comprehend fully the nature of the algorithm. However, the relative importance and the cross-validation reassured us that the developed algorithm was indeed accurate and well balanced. Thirdly, organism and blood cultures were important factors affecting outcome, however many times these variables are not known to the clinician prior to surgery. In these cases, the predicted outcome may be less informative. Finally, there are many possible variables that were not assessed in the present study and their inclusion may have improved the predictive capabilities. Future studies to incorporate these factors into the algorithm may further refine this tool.

To conclude, we successfully created and validated an easy-to-use, practical, and accurate tool for predicting outcome following I&D. To our knowledge this is the first study to use machine learning in the orthopaedic literature. We believe this tool can be used in clinical practice to improve decision-making and patient counselling. The model needs to be validated in an external cohort of patients to confirm its accuracy.

Take home message

- Machine learning is a valid tool for predicting outcomes following I&D surgery.
- This study provides a validated easy to use tool for clinical decision making.

Supplementary material

Patient distribution and failure rate among the 27 centres that took part in the study.

References

1. **Van Kleunen JP**, **Knox D**, **Garino JP**, **Lee GC**. Irrigation and débridement and prosthesis retention for treating acute periprosthetic infections. *Clin Orthop Relat Res*. 2010;468(8):2024–2028.
2. **Kuiper JWP**, **Willink RT**, **Moojen DJF**, **van den Bekerom MP**, **Colen S**. Treatment of acute periprosthetic infections with prosthesis retention: review of current concepts. *World J Orthop*. 2014;5(5):667–676.
3. **Osmon DR**, **Berbari EF**, **Berendt AR**, **et al**. Infectious Diseases Society of America. Executive summary: diagnosis and management of prosthetic joint infection: clinical practice guidelines by the Infectious Diseases Society of America. *Clin Infect Dis*. 2013;56(1):1–10.
4. **Sousa R**, **Abreu MA**. Treatment of Prosthetic Joint Infection with Debridement, Antibiotics and Irrigation with Implant Retention - a Narrative Review. *J Bone Jt Infect*. 2018;3(3):108–117.
5. **Sendi P**, **Lötscher PO**, **Kessler B**, **et al**. Debridement and implant retention in the management of hip periprosthetic joint infection: outcomes following guided and rapid treatment at a single centre. *Bone Joint J*. 2017;99-B(3):330–336.
6. **Konigsberg BS**, **Della Valle CJ**, **Ting NT**, **Qiu F**, **Sporer SM**. Acute hematogenous infection following total hip and knee arthroplasty. *J Arthroplasty*. 2014;29(3):469–472.
7. **Marculescu CE**, **Berbari EF**, **Hanssen AD**, **et al**. Outcome of prosthetic joint infections treated with debridement and retention of components. *Clin Infect Dis*. 2006;42(4):471–478.
8. **Cobo J**, **Miguel LGS**, **Euba G**, **et al**. Early prosthetic joint infection: outcomes with debridement and implant retention followed by antibiotic therapy. *Clinical Microbiology and Infection*. 2011;17(11):1632–1637.
9. **Westberg M**, **Grøgaard B**, **Snorrason F**. Early prosthetic joint infections treated with debridement and implant retention: 38 primary hip arthroplasties prospectively recorded and followed for median 4 years. *Acta Orthop*. 2012;83(3):227–232.
10. **Koyonos L**, **Zmistowski B**, **Della Valle CJ**, **Parvizi J**. Infection control rate of irrigation and débridement for periprosthetic joint infection. *Clin Orthop Relat Res*. 2011;469(11):3043–3048.
11. **Byren I**, **Bejon P**, **Atkins BL**, **et al**. One hundred and twelve infected arthroplasties treated with 'DAIR' (debridement, antibiotics and implant retention): antibiotic duration and outcome. *J Antimicrob Chemother*. 2009;63(6):1264–1271.
12. **Azzam KA**, **Seeley M**, **Ghanem E**, **et al**. Irrigation and debridement in the management of prosthetic joint infection: traditional indications revisited. *J Arthroplasty*. 2010;25(7):1022–1027.
13. **Odum SM**, **Fehring TK**, **Lombardi AV**, **et al**. Periprosthetic infection Consortium. irrigation and debridement for periprosthetic infections: does the organism matter? *J Arthroplasty*. 2011;26(6 Suppl):114–118.
14. **Sukeik M**, **Patel S**, **Haddad FS**. Aggressive early débridement for treatment of acutely infected cemented total hip arthroplasty. *Clin Orthop Relat Res*. 2012;470(11):3164–3170.
15. **Sherrell JC**, **Fehring TK**, **Odum S**, **et al**. The Chitranjan Ranawat Award: fate of two-stage reimplantation after failed irrigation and débridement for periprosthetic knee infection. *Clin Orthop Relat Res*. 2011;469(1):18–25.
16. **Gardner J**, **Gioe TJ**, **Tatman P**. Can this prosthesis be saved?: implant salvage attempts in infected primary TKA. *Clin Orthop Relat Res*. 2011;469(4):970–976.
17. **Barberán J**, **Aguilar L**, **Carroquino G**, **et al**. Conservative treatment of staphylococcal prosthetic joint infections in elderly patients. *Am J Med*. 2006;119(11):e7–993.e10.
18. **Löwik CAM**, **Parvizi J**, **Jutte PC**, **et al**. Northern Infection Network Joint Arthroplasty (NINJA) and ESCMID study group for implant-associated infections (ESGIAI). Debridement, antibiotics and implant retention is a viable treatment option for early periprosthetic joint infection presenting more than four weeks after index arthroplasty. *Clin Infect Dis*. 2019. (Epub ahead of print) PMID: 31504331.
19. **Grammatopoulos G**, **Kendrick B**, **McNally M**, **et al**. Outcome Following Debridement, Antibiotics, and Implant Retention in Hip Periprosthetic Joint Infection-An 18-Year Experience. *J Arthroplasty*. 2017;32(7):2248–2255.
20. **Masters EA**, **Trombetta RP**, **de Mesy Bentley KL**, **et al**. Evolving concepts in bone infection: redefining "biofilm", "acute vs. chronic osteomyelitis", "the immune proteome" and "local antibiotic therapy". *Bone Res*. 2019;7(1):20.
21. **Saeed K**, **McLaren AC**, **Schwarz EM**, **et al**. 2018 international consensus meeting on musculoskeletal infection: summary from the biofilm Workgroup and consensus on biofilm related musculoskeletal infections. *J Orthop Res*. 2019;37(5):1007–1017.
22. **Schwarz EM**, **Parvizi J**, **Gehrke T**, **et al**. 2018 international consensus meeting on musculoskeletal infection: research priorities from the general assembly questions. *J Orthop Res*. 2019;37(5):997–1006.
23. **Tornero E**, **Morata L**, **Martínez-Pastor JC**, **et al**. KLIC-score for predicting early failure in prosthetic joint infections treated with debridement, implant retention and antibiotics. *Clin Microbiol Infect*. 2015;21(8):e9–786.e17.
24. **Wouthuyzen-Bakker M**, **Sebillotte M**, **Lomas J**, **et al**. ESCMID Study Group for Implant-Associated Infections (ESGIAI). Clinical outcome and risk factors for failure in late acute prosthetic joint infections treated with debridement and implant retention. *J Infect*. 2019;78(1):40–47.
25. **Parvizi J**, **Zmistowski B**, **Berbari EF**, **et al**. New definition for periprosthetic joint infection: from the Workgroup of the musculoskeletal infection Society. *Clin Orthop Relat Res*. 2011;469(11):2992–2994.
26. **Buuren Svan**, **Groothuis-Oudshoorn K**. mice : Multivariate Imputation by Chained Equations in *R. J Stat Softw*. 2011;45(3):1–67.
27. **Romanò CL**, **Manzi G**, **Logoluso N**, **Romanò D**. Value of debridement and irrigation for the treatment of peri-prosthetic infections. A systematic review. *Hip Int*. 2012;22 Suppl 8(8_suppl):19–24.
28. **Bradbury T**, **Fehring TK**, **Taunton M**, **et al**. The fate of acute methicillin-resistant Staphylococcus aureus periprosthetic knee infections treated by open debridement and retention of components. *J Arthroplasty*. 2009;24(6 Suppl):101–104.
29. **Hartman MB**, **Fehring TK**, **Jordan L**, **Norton HJ**. Periprosthetic knee sepsis. The role of irrigation and debridement. *Clin Orthop Relat Res*. 1991;273(273):113–118.
30. **Brandt CM**, **Sistrunk WW**, **Duffy MC**, **et al**. Staphylococcus aureus prosthetic joint infection treated with debridement and prosthesis retention. *Clin Infect Dis*. 1997;24(5):914–919.
31. **Burger RR**, **Basch T**, **Hopson CN**. Implant salvage in infected total knee arthroplasty. *Clin Orthop Relat Res*. 1991;(273):105–112.

32. **Schoifet SD**, **Morrey BF**. Treatment of infection after total knee arthroplasty by débridement with retention of the components. *J Bone Joint Surg Am*. 1990;72-A(9):1383–1390.
33. **Tattevin P**, **Crémieux AC**, **Pottier P**, **Huten D**, **Carbon C**. Prosthetic joint infection: when can prosthesis salvage be considered? *Clin Infect Dis*. 1999;29(2):292–295.
34. **Teeny SM**, **Dorr L**, **Murata G**, **Conaty P**. Treatment of infected total knee arthroplasty. irrigation and debridement versus two-stage reimplantation. *J Arthroplasty*. 1990;5(1):35–39.
35. **Rajkomar A**, **Dean J**, **Kohane I**. Machine learning in medicine. *N Engl J Med*. 2019;380(14):1347–1358.
36. **No authors listed**. Ascent of machine learning in medicine. *Nat Mater*. 2019;18(5):407.
37. **Choong PFM**, **Dowsey MM**, **Carr D**, **Daffy J**, **Stanley P**. Risk factors associated with acute hip prosthetic joint infections and outcome of treatment with a rifampinbased regimen. *Acta Orthop*. 2007;78(6):755–765.
38. **Bongartz T**, **Halligan CS**, **Osmon DR**, **et al**. Incidence and risk factors of prosthetic joint infection after total hip or knee replacement in patients with rheumatoid arthritis. *Arthritis Rheum*. 2008;59(12):1713–1720.
39. **Jämsen E**, **Huhtala H**, **Puolakka T**, **Moilanen T**. Risk factors for infection after knee arthroplasty. A register-based analysis of 43,149 cases. *J Bone Joint Surg Am*. 2009;91-A(1):38–47.
40. **Lora-Tamayo J**, **Murillo O**, **Iribarren JA**, **et al**. A large multicenter study of methicillin-susceptible and methicillin-resistant Staphylococcus aureus prosthetic joint infections managed with implant retention. *Clin Infect Dis*. 2013;56(2):182–194.
41. **Tande AJ**, **Palraj BR**, **Osmon DR**, **et al**. Clinical presentation, risk factors, and outcomes of hematogenous prosthetic joint infection in patients with Staphylococcus aureus bacteremia. *Am J Med*. 2016;129(2):e11–221.e20.
42. **Tornero E**, **Morata L**, **Martínez-Pastor JC**, **et al**. Importance of selection and duration of antibiotic regimen in prosthetic joint infections treated with debridement and implant retention. *J Antimicrob Chemother*. 2016;71(5):1395–1401.
43. **Rodríguez-Pardo D**, **Pigrau C**, **Lora-Tamayo J**, **et al**. REIPI group for the study of prosthetic infection. gram-negative prosthetic joint infection: outcome of a debridement, antibiotics and implant retention approach. A large multicentre study. *Clin Microbiol Infect*. 2014;20(11):O911–O919.

Author information:
N. Shohat, MD, Orthopedic Surgeon, Rothman Orthopaedic Institute at Thomas Jefferson University, Philadelphia, Pennsylvania, USA; Sackler Faculty of Medicine, Tel Aviv University, Ramat Aviv, Israel.

K. Goswami, MD, Orthopedic Resident

T. L. Tan, MD, Orthopedic Surgeon
M. Yayac, MD, Research Fellow
J. Parvizi, MD, FRCS, Orthopedic Surgeon
Rothman Orthopaedic Institute at Thomas Jefferson University, Philadelphia, Pennsylvania, USA.

A. Soriano, MD, Infectious Disease Specialist, Service of Infectious Diseases, Hospital Clínic, University of Barcelona, Barcelona, Spain.

R. Sousa, MD, Orthopaedic Surgeon, Centro Hospitalar Universitário do Porto, Porto, Portugal.

M. Wouthuyzen-Bakker, MD, Infectious Disease Specialist, Department of Medical Microbiology and Infection Prevention, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.

Author contributions:
N. Shohat: Gathered and analyzed the data, Wrote the manuscript.
K. Goswami: Gathered the data, Edited the manuscript.
T. L. Tan: Gathered the data, Edited the manuscript.
M. Yayac: Gathered and analyzed the data, Wrote the manuscript.
A. Soriano: Conceived the study, Gathered the data, Edited the manuscript.
R. Sousa: Conceived the study, Gathered the data, Edited the manuscript.
M. Wouthuyzen-Bakker: Conceived the study, Gathered the data, Edited the manuscript.

Funding statement:
No benefits in any form have been received or will be received from a commercial party related directly or indirectly to the subject of this article.

Acknowledgements:
Collaborators: Marine Sebillotte, Jose Lomas, Adrian Taylor, Eva Benavent Palomares, Oscar Murillo, Javier Cobo Reinoso, Rosa Escudero Sánchez, Marta Fernandez-Sampedro, Eric Senneville, Kaisa Huotari, José Maria Barbero Allende, Joaquín Garcia-Cañete, Jaime Lora-Tamayo, Matteo Carlo Ferrari, Danguole Vaznaisiene, Erlangga Yusuf, Craig Aboltins, Mauro José Salles, Natividad Benito, Andrea Vila, Maria Dolores Del Toro, Tobias Siegfried Kramer, Sabine Petersdorf, Vicens Diaz-Brito, Zeliha Kocak Tufan, Marisa Sanchez, Cédric Arvieux, Pierre Tattevin, Jaime Esteban, Matthew Scarborough, Paul Jutte, Joris Ploegmakers, Claudia Löwik, Bas Knobben, and Wierd Zijlstra.

Ethical review statement:
Informed consent was retrieved when required by the ethics committee of each participating centre.

This article was primary edited by G. Scott.

This paper was presented at The Hip Society 2019 Members Meeting in Kohler, Wisconsin, USA.