Aims
No predictive model has been published to forecast operating time for total knee arthroplasty (TKA). The aims of this study were to design and validate a predictive model to estimate operating time for robotic-assisted TKA based on demographic data, and evaluate the added predictive power of CT scan-based predictors and their impact on the accuracy of the predictive model.

Methods
A retrospective study was conducted on 1,061 TKAs performed from January 2016 to December 2019 with an image-based robotic-assisted system. Demographic data included age, sex, height, and weight. The femoral and tibial mechanical axis and the osteophyte volume were calculated from CT scans. These inputs were used to develop a predictive model aimed to predict operating time based on demographic data only, and demographic and 3D patient anatomy data.

Results
The key factors for predicting operating time were the surgeon and patient weight, followed by 12 anatomical parameters derived from CT scans. The predictive model based only on demographic data showed that 90% of predictions were within 15 minutes of actual operating time, with 73% within ten minutes. The predictive model including demographic data and CT scans showed that 94% of predictions were within 15 minutes of actual operating time and 88% within ten minutes.

Conclusion
The primary factors for predicting robotic-assisted TKA operating time were surgeon, patient weight, and osteophyte volume. This study demonstrates that incorporating 3D patient-specific data can improve operating time predictions models, which may lead to improved operating room planning and efficiency.

Cite this article: Bone Jt Open 2022;3-5:383–389.

Keywords: Total knee arthroplasty, Operating time, Predictive model, CT scan, Image-based robotic-assisted surgery

Introduction
Total knee arthroplasty (TKA) operating time varies across different patients and is dependent on several parameters including previous surgery, range of motion (ROM), BMI, age, and bone quality. Operating time has previously been explored in TKA, with multiple studies indicating that prolonged duration can increase patient risk for various adverse outcomes, such as infection. Various analyses have focused on factors influencing operating time to better understand how to improve forecasts. A recent study has assessed the factors associated with variations in operating time. An accurate forecast for predicting operating time may allow for more efficient surgical scheduling, can ensure that appropriate operating room (OR) staff are available for potentially more difficult cases, and can avoid case cancellation. While a surgeon may accurately predict their mean operating time based on historic cases, single surgery prediction can be less accurate. Predictive
modelling is a discipline where algorithms are deployed to generate estimates for a specific target output. These models learn to identify the relationships between a set of features, also called predictors (patient age, BMI, level of fitness, etc.) and the selected target (e.g. occurrence of a myocardial infarction). Statistical models and machine-learning techniques identify complex target-predictor relationships in a specific dataset. Predictive models are usually deployed in contexts where the prediction of the output is difficult or time-demanding. When the likelihood of an adverse event in the future is high, a proactive intervention can be deployed to modify the course of action and avoid such an adverse event.

The increasing availability of large digital healthcare datasets has facilitated the application of predictive models in different healthcare settings such as cognitive impairment detection, prediction of heart failure patient readmission, or medical image segmentation. The combination of digital healthcare datasets and 3D images can generate new factors that may contribute to improved prediction of operating time for complex TKA. 3D imaging has shown improvements in surgical accuracy and preoperative planning, which may translate to more accurate operating time predictions.

Thus, the purposes of our study were to design and validate a predictive model to estimate operating time for TKA based on demographic data, and evaluate the added predictive power of CT scan-based predictors and their impact on the accuracy of the predictive model. Our two hypotheses were: demographic data can be used to generate operating time prediction in TKA; and adding CT-based 3D image patient-specific features will improve operating time prediction accuracy.

### Methods

#### Data sources and population

A retrospective study was conducted on 1,061 robotic-assisted primary TKAs from January 2016 to December 2019. Data were obtained from two arthroplasty fellowship-trained orthopaedic surgeons at two different hospital centres within the USA (South County Health, Rhode Island, and The Orthopaedic Center, Oklahoma). The Stryker Performance Solutions (SPS) Hospital Reported Outcomes database (Stryker, USA) was used, which consists of de-identified real-world data from electronic health records including patient demographic information and patient-specific 3D data from CT scans. The robotic-assisted MAKO (Stryker) system is an image-based system, using a preoperative CT scan to create a 3D model of the patient’s bony anatomy used for the surgical planning. This robotic platform provides real-time 3D feedback to the surgeon on the implant position and limb alignment, as well as virtual flexion and extension gaps. This robotic-assisted system is semiautonomous, with a robotic arm extended by a saw, controlled and manipulated by the surgeon.

Demographic data included age, sex, height, weight, and laterality. Preoperative CT scans were obtained in this same patient cohort, and femoral and tibial axis angles (medial proximal tibial angles (MPTA) and lateral distal femoral angles (L DFA)) and osteophyte volume were automatically extracted from the 3D images as described below. Operating time was defined as incision to closure, which were obtained from each hospital’s electronic health record system. The time prediction models were trained using two datasets: demographic data only, and demographic and 3D patient anatomy (e.g. osteophyte volume) data. Table I provides an overview of the patient population. An exhaustive description of each predictor is available in Supplementary Table I.

#### Osteophyte volume detection

An osteophyte detection algorithm by Imorphics (Stryker, UK) segmented osteophyte volumes. The algorithm, based on training data provided by expert segmenters, segmented the outer surface of the cortical bone and an osteophyte-free surface. Subtracting the osteophyte-free surface from the outer surface generated the osteophyte surface (Figure 1). Related techniques, using atlases of healthy bones instead of an explicit osteophyte-free surface, have been validated in previous studies.

Osteophyte volumes were automatically generated per case for the femur and tibia and split into four regions: anterior medial, anterior lateral, posterior medial, and posterior lateral. The partition used orthogonal planes...
derived from the epicondylar axis and a femur shaft centre point for the femur, along with an axis connecting the deepest point per tibial plateau (resection landmarks) and a tibia shaft centre point for the tibia (Figure 2).

**Model development.** Figure 3 outlines the architecture used to develop each model. Prior to modelling, the input data were analyzed using cross-correlation tabulation, and any inputs with >0.8 correlation were removed to ensure the model was parsimonious. For example, BMI correlated 0.84 with weight (Figure 4), and as BMI is a function of height and weight, removing this feature would simplify the dataset.

The initial dataset underwent an 85:15 split between train and test. As this dataset was skewed towards one surgeon (885 to 176), the split was stratified to ensure a representative proportion of cases from both surgeons. Three models were trained on these datasets: linear regression, random forest regression, and CatBoost regression. We use the training dataset to train and tune the model’s hyperparameters. The training dataset underwent a five-fold cross-validation wherein the training dataset was split into five subsets (one subset for validation and four for training the algorithm). This was repeated five times until each subset has been used once for validation (Figure 3). Each five-fold cross-validation was performed with a different set of hyperparameters, and the optimum combination of parameters was then retained for the next phase. After hyperparameter tuning per model and dataset, we trained the final model on the original test split with the defined hyperparameter values. To compare the $R^2$ values of the model, we used the bootstrap resampling technique. The training dataset population underwent resampling with arthroplasties and was repeated 100 times. This technique allowed us to estimate summary statistics, such as mean and standard deviation (SD); therefore, these confidence intervals were used to compare against the other models. Feature importance values were normalized so that the sum of importance values equaled 100. The variable importance evaluates the difference in model performance between the original model, and the model performance with that feature excluded.

**Results**

**Predictive factors of operating time prediction.** The overall mean operating time was 70.9 minutes (SD 18.9). The
most important features in our operating time prediction model were the surgeon and patient weight, followed by 12 anatomical parameters derived from 3D CT scan, particularly osteophyte volume (Table II). The predictive model based only on demographic data showed that 90% of predictions were within 15 minutes of actual operating time, with 73% within ten minutes and 40% within five minutes. Model 2, which included demographic data and CT scan data (including osteophyte volume), showed that 94% of predictions were within 15 minutes of actual operating time, 88% within ten minutes, and 78% within five minutes. Predicted versus actual operating time performance of this model is presented in Figure 5.

**Model results.** The $R^2$ values were significantly improved across all three machine-learning models (Linear Regression, Random Forest Regression, and CatBoost Regression) once 3D patient-specific information was implemented into the dataset ($p < 0.001$ across all three models). Between all three models, the CatBoost regression model achieved superior performance with an $R^2$ value of 0.764 in dataset 2 and 0.722 in dataset 1 (Table III). The random forest model achieved an $R^2$ value of 0.732 in dataset 2 and 0.705 in dataset 1; the linear regression model achieved an $R^2$ value of 0.708 in dataset 2 and 0.686 in dataset 1 (Table III).

We derived the optimum model (CatBoost Regression on Dataset 2) using the entire training dataset (original 85% split without bootstrapping) with the optimal tuned hyperparameters from the ‘training, tuning and evaluation segment’ of the process (Figure 3).
Discussion

Inefficient surgical scheduling is linked with worse clinical outcomes for patients, as well as staff dissatisfaction and avoidable costs for the hospitals. As with many other surgical interventions, TKA procedures are often scrutinized in terms of efficiency with a specific focus on OR use. The main finding of this study is the importance of 3D patient-specific data, particularly osteophyte volume, to improve operating time prediction in robotic-assisted TKA. Surgeon, patient weight, and osteophyte volume were the most significant predictive features in our operating time robotic-assisted TKA prediction model.

In the context of real-life deployment, our model achieved a high level of accuracy in predicting operating time. Indeed, when investigating the applications of machine-learning in surgical case duration, Tuwatatanurak et al. defined the threshold of 15 minutes as clinically significant error. However, they defined case duration as the time from entry into the OR to exit from the OR rather than incision to closure. In total, 94% of the predictions generated by the highest performing model in our study were within 15 minutes of the actual operating time. Prior knowledge of operating time for each TKA procedure has the potential to improve OR schedule management. Surgeon can optimize the surgical team composition to deal with complex procedures and maximize OR time efficiency. Longer cases can be identified in advance and scheduling changed accordingly to avoid unexpected delays on the day of surgery and last case of the day cancellations. Inaccurate estimations of operating time can lead to inefficient surgical scheduling, which results in both under- and overuse of OR time.

In our study, the key factor in predicting TKA operating time was the surgeon. In literature, only one study described a method to predict TKA operating time, which was composed of operating time estimates self-reported by the surgeon compared to their historical averages. They found that when self-reporting, surgeons overestimated their own operating time by an average of 18 minutes per case. However, when incorporating a model based on historical averages for each surgeon, these predictions showed a significant reduction in overestimations with mean estimation bias reduced to -0.1 to 0.3 minutes (p < 0.001), regardless of the number of

---

Table II. Feature importance from the CatBoost model with 3D patient anatomy data.

| Feature                | Variable importance, % |
|------------------------|------------------------|
| Surgeon ID             | 77.7                   |
| Weight                 | 4.5                    |
| Posterolateral femur   | 1.8                    |
| Anterolateral tibia    | 1.9                    |
| Anteromedial femur     | 1.9                    |
| Anteromedial tibia     | 1.1                    |
| MPTA                   | 1.1                    |
| Sex                    | 1.0                    |
| Posteromedial tibia    | 0.8                    |
| LDFA                   | 0.8                    |
| Age                    | 0.8                    |
| JLCA                   | 1.2                    |
| Height                 | 0.5                    |
| Side                   | 0.1                    |

sdHKa, diseased hip-knee-ankle angle; hKa, hip-knee-ankle angle; JLCA, joint line convergence angle; JLO, joint line obliquity; LDFA, lateral distal femoral axis; MPTA, medial proximal tibial axis.

---

Table III. Comparing R² values across three different model types on the two datasets (with and without 3D patient anatomy data).

| Dataset     | Linear regression (SD) | Random Forest (SD) | CatBoost (SD) |
|-------------|------------------------|--------------------|---------------|
| Dataset 1*  | 0.686 (0.027)          | 0.705 (0.026)      | 0.722 (0.030) |
| Dataset 2†  | 0.708 (0.028)          | 0.732 (0.024)      | 0.764 (0.026) |

*Demographic data only,
†Including 3D patient anatomy data.
SD, standard deviation.
cases performed. Nevertheless, this model was less reliable for complex cases because only one feature (own historical operating times) was assessed in the prediction, contrary to our study. Other studies found that surgeon experience has an impact on the operating time, without exact prediction of this difference.23

Cases that are considered complex TKAs are usually due to high BMI or the presence of complex deformities. Several studies have demonstrated a relationship between operating time and risk of perioperative complications due to high BMI.24,25 Gadinsky et al26 reported a mean operating time of 74 minutes (SD 14) for normal BMI versus 90 minutes (SD 18) for a BMI > 40 kg/m² in a case-control study on 454 TKAs.24 However, there are no studies that have used BMI to predict operating time in TKA. The combination of BMI/weight and surgeon experience in operating on patients with a wide range of BMIs must be incorporated in models to increase model generalizability to other geographies.

Complex deformities are highlighted within preoperative imaging. 3D imaging, such as CT scans, brings relevant data for TKA planning, such as bone deformity, osteophyte volume, and bone quality.17 These data have shown a demonstrable benefit for TKA with patient-specific instrumentation (PSI).18 With the development of image-based robotic-assisted TKA, the benefits of preoperative 3D imaging have been also realized. Several studies suggest that image-based robotic-assisted TKA improves surgical accuracy.19,26 This study continues to show additional advantages of 3D patient-specific data with a specific focus on operating time prediction. Two types of data were derived from 3D patient scans in our operating time prediction model: patient morphology (knee joint angles) and osteophyte volume. The former can be measured on preoperative radiographs. The patient morphology was defined in our study by the angles of frontal deformity, as measured usually on standard radiographs. Therefore, these measurements on CT scans were likely not very helpful in improving the operating time. If more complex measurements would be considered (such as the femoral rotation, in the axial plane) or in a population with complex deformities in two or three plans, the CT scan data could probably be more beneficial for the surgical planning or the optimization of operating time. The osteophytes require 3D imaging to be quantified. The presence of osteophytes necessitates extra operating time and can lead to complications during knee exposure. Decreased ROM caused by anterior or posterior osteophytes can be a secondary factor in longer operating time. Posterior osteophyte removal and its impact on the extension gap may lead to additional bony cuts or soft-tissue releases to achieve a balanced knee after TKA. Some studies have assessed the impact of osteophyte removal on ROM and knee exposure.27,28 No study has assessed the impact of osteophytes on operating time. Our predictive model was based on accurate quantification of osteophyte volume and their localization, which is only attainable if derived from 3D images. These data represent a proxy for exposure difficulties, potentially limited ROM, and time for osteophyte removal.21

Our findings should be considered in light of the key limitations of this study. The dataset is composed of a limited number of cases which were performed by only two orthopaedic surgeons. A dataset containing more TKA cases, and encompassing several surgeons and centres, is warranted to extend its use to other geographies and healthcare systems. Furthermore, this was not a comparative study using demographic data and preoperative radiographs to assess the benefit of 3D CT over conventional 2D radiographs. Finally, several parameters were not considered in this model. For example, the model did not integrate bone quality or bone loss, which can increase the difficulty of performing TKA, although these characteristics are likely to be more relevant for predicting the risk of early complications after TKA rather than operating time.29

With the rise of new technologies, particularly in robotic-assisted surgery, it becomes crucial to use preoperative and intraoperative data to improve surgeon practice. Overall, this study has shown that 3D patient-specific data can improve machine-learning models to predict operating time of robotic-assisted TKA. The most influential factors for this TKA operating time prediction model were surgeon, patient weight, and osteophyte volume. A preoperative CT scan has the potential to enable more accurate operating time estimation, which in turn may improve surgical scheduling.

**Take home message**

- 3D patient-specific data can improve operating time predictions models, which may lead to improved operating room planning and efficiency.
- The primary features for predicting robotic-assisted total knee arthroplasty operating time were surgeon, patient weight, and osteophyte volume.

**Supplementary material**

Table reporting an exhaustive description of each predictor of this predictive model.

**References**

1. George J, Mahmood B, Sultan AA, et al. How fast should a total knee arthroplasty be performed? An analysis of 140,199 surgeries. J Arthroplasty. 2018;33(8):2616–2622.
2. Anis HK, Dodhi N, Klika AK, et al. Is operative time a predictor for post-operative infection in primary total knee arthroplasty? J Arthroplasty. 2019;34(7S):S331–S338.
3. Naranje S, Lendway L, Mehle S, Giss DJ. Does operative time affect infection rate in primary total knee arthroplasty? Clin Orthop Relat Res. 2015;473(1):64–69.
4. Obata R, Bito H, Ohmura M, et al. The effects of prolonged low-flow sevoflurane anesthesia on renal and hepatic function. Anesth Analg. 2000;91(5):1262–1268.
5. Hernandez AJ, Almeida A de, Fávaro E, Sguizzato GT. The influence of tourniquet use and operative time on the incidence of deep vein thrombosis in total knee arthroplasty. Clinics (Sao Paulo). 2012;67(9):1053–1057.
6. Vorhies JS, Wang Y, Herndon JH, Maloney WJ, Huddleston JI. Decreased length of stay after TKA is not associated with increased readmission rates in a national Medicare sample. Clin Orthop Relat Res. 2012;470(1):166–171.
7. Goudie EB, Robinson C, Walmsley P, Brenkel I. Changing trends in total knee replacement. Eur J Orthop Surg Traumatol. 2017;27(4):539–544.

8. Theelen L, Bischoff C, Grimm B, Heyligers IC. Current practice of orthopaedic surgical skills training raises performance of supervised residents in total knee arthroplasty to levels equal to those of orthopaedic surgeons. Perspect Med Educ. 2018;7(2):126–132.

9. Haughom BD, Scharier WW, Hellman MD, Yi PH, Levine BR. Resident involvement does not influence complication after total hip arthroplasty: an analysis of 13,109 cases. J Arthroplasty. 2014;29(10):1919–1924.

10. Anis HK, Sodzi N, Coste M, et al. A comparison of peri-operative outcomes between elective and non-elective total hip arthroplasties. Ann Transl Med. 2019;7(4):78.

11. Acuña AJ, Samuel LT, Karnuta JM, Sultan AA, Swiergosz AM, Kamath AF. What factors influence operative time in total knee arthroplasty? A 10-year analysis in a national sample. J Arthroplasty. 2020;35(3):821–827.

12. Wu A, Huang CC, Weaver MJ, Urman RD. Use of historical surgical times to predict duration of primary total knee arthroplasty. J Arthroplasty. 2016;31(12):2768–2772.

13. Bini SA, Shah RF, Bendich I, Patterson JT, Hwang KM, Zaid MB. Machine learning algorithms can use wearable sensor data to accurately predict six-week patient-reported outcome scores following joint replacement in a prospective trial. J Arthroplasty. 2019;34(10):2242–2247.

14. Suk H, Lee SW, Shen D. Alzheimer’s Disease Neuroimaging Initiative. Latent feature representation with stacked auto-encoder for AD/MCI diagnosis. Brain Struct Funct. 2020;225(2):641–659.

15. Amarnasingham R, Moore BJ, Tabak YP, et al. An automated model to identify heart failure patients at risk for 30-day readmission or death using electronic medical record data. Med Care. 2010;48(11):881–888.

16. Zhang W, Li R, Deng H, et al. Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. Neuroimage. 2015;108:214–224.

17. Sarielí E, Kajetanek C, Catonné Y. Comparison of custom cutting guides based on three-dimensional computerized CT-scan planning and a conventional ancillary system based on two-dimensional planning in total knee arthroplasty: a randomized controlled trial. Int Orthop. 2019;43(11):2242–2258.

18. Renson L, Poilvaech P, Van den Wynaert H. Improved alignment and operating room efficiency with patient-specific instrumentation for TKA. Knee. 2014;21(8):1216–1220.

19. Batailler C, Fernandez A, Swan J, et al. MAKO CT-based robotic arm-assisted system is a reliable procedure for total knee arthroplasty: a systematic review. Knee Surg Sports Traumatol Arthrosc. 2021;29(11):2529–2538.

20. Liu Z, Huang X, Hu L, et al. Effects of hinge-region natural polymorphisms on human immunodeficiency virus-type 1 protease structure, dynamics, and drug pressure evolution. J Biol Chem. 2016;291(43):22741–22756.

21. Morton AM, Akhbari B, Moore DC, Crisco JJ. Osteophyte volume calculation using dissimilarity-excluding Procrustes registration of archived bone models from healthy volunteers. J Orthop Res. 2020;38(6):1307–1315.

22. Tuwatananurak JP, Zadeh S, Xu X, et al. Machine learning can improve estimation of surgical case duration: a pilot study. J Med Vet. 2019;43(3):44.

23. Khamjua HS, Solano MA, Sterling RS, Oni JK, Chaudhry YP, Jones LC. Surgeon mean operative times in total knee arthroplasty in a variety of settings in a health system. J Arthroplasty. 2019;34(11):2569–2572.

24. Gadinsky NE, Manuel JB, Lyman S, Westrich GH. Increased operating room time in patients with obesity during primary total knee arthroplasty: conflicts for scheduling. J Arthroplasty. 2012;27(8):1171–1176.

25. Hanly RJ, Marvi SK, Whitehouse SL, Crawford RW. Morbid obesity in total knee arthroplasty: joint-specific variance in outcomes for operative time, length of stay, and readmission. J Arthroplasty. 2017;32(9):2712–2716.

26. Batailler C, Bordes M, Lording T, et al. Improved sizing with image-based robotic-assisted system compared to image-free and conventional techniques in medial unicompartmental knee arthroplasty. Bone Joint J. 2021;103-B(4):610–618.

27. Su EP. Fixed flexion deformity and total knee arthroplasty. J Bone Joint Surg Br. 2012;94-B(11 Suppl A):112–115.

28. Lein MA, Klassen A, Osima T, et al. Large osteophyte removal from the posterior femoral condyle significantly improves extension at the time of surgery in a total knee arthroplasty. J Orthop. 2020;19:76–83.

29. Huang CC, Jiang CC, Hsieh CH, Tsai CJ, Chiang H. Local bone quality affects the outcome of prosthetic total knee arthroplasty. J Orthop Res. 2016;34(2):240–248.

Author information:
A. Motesharei, PhD, Engineer, Stryker, Newbury, UK.
C. Batailler, MD, PhD, Orthopaedic Surgeon
S. Lustig, MD, Orthopaedic Surgeon
D. De Massari, PhD, Engineer, Stryker, Amsterdam, The Netherlands.
G. Vincent, PhD, Engineer, Imorphics Ltd, Manchester, UK.
A. F. Chen, MD, MBA, Orthopaedic Surgeon, Department of Orthopaedic Surgery, Brigham & Women’s Hospital, Boston, Massachusetts, USA.

Open access funding was provided by Lyon 1 University, Lyon, France.

Open access funding was provided by Lyon 1 University, Lyon, France.

Open access funding was provided by Lyon 1 University, Lyon, France.

Open access funding was provided by Lyon 1 University, Lyon, France.

© 2022 Author(s) et al. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial No Derivatives (CC BY-NC-ND 4.0) licence, which permits the copying and redistribution of the work only, and provided the original author and source are credited. See https://creativecommons.org/licenses/by-nc-nd/4.0/