university of groningen

University Medical Center Groningen

University of Groningen

Developing a machine learning algorithm to predict the probability of aseptic loosening of the glenoid component after anatomical total shoulder arthroplasty

Macken, Arno Alexander; Macken, Loïc C.; Oosterhoff, Jacobien H.F.; Boileau, Pascal; Athwal, George S.; Doornberg, Job N.; Lafosse, Laurent; Lafosse, Thibault; van den Bekerom, Michel P.J.; Buijze, Geert Alexander

Published in:
BMJ Open

DOI:
10.1136/bmjopen-2023-074700

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2023

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Macken, A. A., Macken, L. C., Oosterhoff, J. H. F., Boileau, P., Athwal, G. S., Doornberg, J. N., Lafosse, L., Lafosse, T., van den Bekerom, M. P. J., & Buijze, G. A. (2023). Developing a machine learning algorithm to predict the probability of aseptic loosening of the glenoid component after anatomical total shoulder arthroplasty: protocol for a retrospective, multicentre study. *BMJ Open*, *13*(10), Article e074700. https://doi.org/10.1136/bmjopen-2023-074700

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

BMJ Open

Developing a machine learning algorithm to predict the probability of aseptic loosening of the glenoid component after anatomical total shoulder arthroplasty: protocol for a retrospective, multicentre study

Arno Alexander Macken ,[1,2] Loïc C Macken,[3] Jacobien H F Oosterhoff,[4] Pascal Boileau,[5] George S Athwal,[6] Job N Doornberg,[7] Laurent Lafosse,[2] Thibault Lafosse,[2] Michel P J van den Bekerom,[8,9] Geert Alexander Buijze[2,10]

To cite: Macken AA, Macken LC, Oosterhoff JHF, *et al*. Developing a machine learning algorithm to predict the probability of aseptic loosening of the glenoid component after anatomical total shoulder arthroplasty: protocol for a retrospective, multicentre study. *BMJ Open* 2023;**13**:e074700. doi:10.1136/bmjopen-2023-074700

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2023-074700).

Received 14 April 2023
Accepted 25 September 2023

Check for updates

© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to
Arno Alexander Macken;
arnomacken@gmail.com

ABSTRACT

Introduction Despite technological advancements in recent years, glenoid component loosening remains a common complication after anatomical total shoulder arthroplasty (ATSA) and is one of the main causes of revision surgery. Increasing emphasis is placed on the prevention of glenoid component failure. Previous studies have successfully predicted range of motion, patient-reported outcomes and short-term complications after ATSA using machine learning methods, but an accurate predictive model for (glenoid component) revision is currently lacking. This study aims to use a large international database to accurately predict aseptic loosening of the glenoid component after ATSA using machine learning algorithms.

Methods and analysis For this multicentre, retrospective study, individual patient data will be compiled from previously published studies reporting revision of ATSA. A systematic literature search will be performed in Medline (PubMed) identifying all studies reporting outcomes of ATSA. Authors will be contacted and invited to participate in the Machine Learning Consortium by sharing their anonymised databases. All databases reporting revisions after ATSA will be included, and individual patients with a follow-up less than 2 years or a fracture as the indication for ATSA will be excluded. First, features (predictive variables) will be identified using a random forest feature selection. The resulting features from the compiled database will be used to train various machine learning algorithms (stochastic gradient boosting, random forest, support vector machine, neural network and elastic-net penalised logistic regression). The developed and validated algorithms will be evaluated across discrimination (c-statistic), calibration, the Brier score and the decision curve analysis. The best-performing algorithm will be used to create an open-access online prediction tool.

Ethics and dissemination Data will be collected adhering to the WHO regulation on data sharing. An Institutional Review Board review is not applicable. The study results will be published in a peer-reviewed journal.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ A large international database will be collected, which increases accuracy, validity and external applicability.
⇒ A clinical prediction model using machine learning algorithms will be developed to estimate the probability of aseptic loosening of the glenoid after anatomical total shoulder arthroplasty.
⇒ An open-access prediction tool based on the best-performing algorithm will be made available online that can guide medical professionals in personalised treatment decision-making.
⇒ The study is dependent on data provided by third parties, which is a potential source of bias.
⇒ Input variables will be selected and categorised based on completeness and uniformity across data sources, potentially decreasing the amount of detail in the study.

INTRODUCTION

Anatomical total shoulder arthroplasty (ATSA) is used for glenohumeral arthropathy causing pain and/or a reduction in range of motion. Despite technological advancements in recent years, glenoid component loosening remains a common complication after ATSA and is one of the main causes of revision surgery. Glenoid loosening can be a trying complication to manage, and the optimal course of treatment remains unclear.[1]

Consequently, increasing emphasis is placed on the prevention of loosening. The predicted chance of glenoid component failure plays an important role in clinical decision-making such as patient selection for ATSA or which implants and techniques to use. Several previous studies assessing

BMJ Open: first published as 10.1136/bmjopen-2023-074700 on 18 October 2023. Downloaded from http://bmjopen.bmj.com/ on November 29, 2023 at University of Groningen. Protected by copyright.

risk factors of glenoid component loosening identified patient, treatment and prosthesis characteristics related to glenoid component loosening. For example, male sex and a higher critical shoulder angle have been associated with higher rates of loosening and revisions.[2 3] Glenoid retroversion did not impact implant survivorship in one study of ATSA with minimal, non-corrective reaming.[4] However, a larger degree of retroversion may have more impact. Several previous studies have also identified aspects of the glenoid component design that correlated with the rate of loosening, such as whether the polyethylene is cross-linked, whether the component is pegged or keeled, or the usage of cement.[5–9] In spite of these studies identifying influential factors, accurate prediction of aseptic loosening of the glenoid component remains a challenge with conventional methods.

In recent years, machine learning or artificial intelligence has been used with increasing precision to predict outcomes after ATSA. A previous study using machine learning was able to accurately predict range of motion and patient-reported outcomes after ATSA. The most influential factors they reported were follow-up time, preoperative range of motion and patient-reported outcome measures, patients' sex and surgery on the dominant upper limb.[10] Another study was able to accurately predict the improvement in American Shoulder and Elbow Surgeons (ASES) scores after shoulder arthroplasty using machine learning.[11] The most relevant predictive factors were preoperative ASES scores, preoperative pain scores, Walch classification, fatty infiltration in the supraspinatus and infraspinatus, and age. A previous study using artificial intelligence to predict patient satisfaction 2 years after shoulder arthroplasty found baseline Single Assessment Numeric Evaluation score, exercise and activity, workers' compensation status, diagnosis, symptom duration prior to surgery, body mass index (BMI), age, smoking status, anatomical versus reverse total shoulder arthroplasty and diabetes to be predicting factors.[12] Two studies report predictive models on short-term complications after ATSA. One study using machine learning to predict complications and 30-day unplanned readmissions found that a history of implant complication, severe chronic kidney disease, teaching hospital status, coronary artery disease and male sex were the most important features.[13] The machine learning model found teaching hospital status and male sex as a markedly more important predictor compared with a logistic regression analysis of the same data. Another study on short-term complications after total shoulder arthroplasty found percentage haematocrit, BMI and operative time were of highest importance in outcome prediction.[14] These studies demonstrate that machine learning may provide accurate predictions for the outcomes after ATSA. Machine learning is most effective with large amounts of data and is very dependent on the amount of detail. Furthermore, the algorithm needs to be widely applicable; a varied and international database provides the highest external validity.

To our knowledge, there are no studies predicting the long-term complications such as aseptic loosening of the glenoid component using advanced machine learning techniques. Furthermore, previous machine learning studies are limited in accuracy and validity due to the sample size and homogeneity. Therefore, this study aims to develop a clinical prediction model for aseptic loosening of the glenoid component using machine learning algorithms trained on a large international database using clustered data. The large combined dataset is less prone to overfitting, and allows direct validation of models across a range of populations and settings, thereby increasing generalisability.[15] The predictive algorithm will be made available for clinical use through a publicly available online prediction tool.

METHODS AND ANALYSIS

Data collection

For this multicentre, retrospective study, individual patient data (IPD) will be collected from previously published studies reporting failure and revision of ATSA. A systematic literature search will be performed in Medline (PubMed) identifying all studies that report a cohort of ATSA including revision as an outcome, published between January 2000 and June 2023. The limit was set at January 2000 to increase the likelihood of the dataset that was used for the study still being available. The minimum required data retention period varies between countries but is generally 20 years or less. The full search strategy is available in online supplemental appendix 1. All original studies reporting revision or failure rates after primary ATSA will be included. Reviews and letters to the editor will be excluded, as well as studies published in languages other than English, Italian, Dutch and French. Authors will be requested to share the anonymised databases used for the identified studies. Only de-identified databases used for previous studies are included; authors are not required to gather additional data or access patient files. After sharing their data, the authors will be included in the Machine Learning Consortium. Inclusion criteria for individual patients within the provided databases are a minimum age of 18 years and a minimum follow-up of 2 years. Patients who underwent ATSA with a fracture as the indication or patients who underwent concomitant procedures such as a cuff repair, tendon transfer or bone graft will be excluded. The aim is to combine the IPD from previously published studies to create a large international cohort which can be used to train a machine learning algorithm to predict aseptic loosening of the glenoid component after ATSA. Based on previous studies, we estimate a glenoid revision rate of approximately 2%.[6 16] The minimum number of events per variable to achieve sufficient accuracy differs per model and is not clearly defined for each technique.[17 18] We aim to include at least 30 events per variable, resulting in a sample size of 7500 patients for a model with up to five predictive variables.

BMJ Open: first published as 10.1136/bmjopen-2023-074700 on 18 October 2023. Downloaded from http://bmjopen.bmj.com/ on November 29, 2023 at University of Groningen. Protected by copyright.

Data curation and missing data

Completeness across data sources will be assessed for each variable in the compiled multicentre database, and variables with sufficient completeness (>70% complete) will be selected as input for the machine learning algorithms. Variables with >30% missing data will be excluded. For the remaining variables, missing data will be completed by imputation using multivariate imputation by chained equations.[19] Uniformity in reporting will be assessed for each variable. If possible, variables will be adjusted or categorised to ensure uniform reporting. In case uniformity of the reported variable across data sources cannot be achieved without guaranteeing correctness, the variable will be excluded. Each dataset will be split into training (80%) and test (20%) subsets, stratified by outcome. Fivefold cross-validation of the training set will be used to develop the machine learning models.[20] Data curation and imputation will be performed using R (R foundation for statistical computing, Vienna, Austria).

Variable selection

The primary outcome is a revision of the glenoid component for aseptic glenoid loosening. The input variables for both methods are dependent on the uniformity and completeness of the gathered data but will include demographics (eg, age, sex and ethnicity), patient-specific factors (eg, preoperative BMI, comorbidity, smoking, dominance, previous surgery), disease-specific factors (eg, affected side, indication, Walch classification, fatty infiltration of cuff muscles) and surgical characteristics (eg, corrective reaming, component design and type, component materials, cementing and sizes). Before training the machine learning models, relevant variables will be selected using random forest algorithms with recursive selection.[21] At least 10 events for each predictor variable will be included in the model, adhering to the rule of thumb in predictive models of binary variables.[22]

Development of prediction models

Different machine learning models result in varying performance metrics based on the type of input data (continuous, categorical, dichotomous). Due to the variation in type of input variables in the dataset, several different machine learning techniques will be used and compared based on model performance. The following machine learning algorithms were chosen for modelling based on prior research[23–27]: stochastic gradient boosting, random forest, support vector machine, neural network and elastic-net penalised logistic regression. The algorithms will be trained on the training dataset with 10-fold cross-validation repeated three times. Cross-validation means dividing data into a selected number of groups, also called folds. First, the data will be divided into 10 equally sized folds. Then, the algorithms will be trained on 9 of the 10 folds (90% of the training data) and tested on the remaining fold (10% of the training data). Results will be averaged across all repetitions of this sequence. Machine learning algorithms will be developed using Python (The Python Software Foundation, Fredericksburg, USA). Hyperparameter tuning will be performed as recommended in the Python libraries. The statistician who performs the machine learning analysis will be blinded to the origin of the data, but the anonymised data source will be available to be included as a potential confounding factor.

Model performance

After training all models, the model performance will be analysed according to a proposed framework by Steyerberg *et al* including discrimination with the c-statistic, positive predictive value (PPV), true positive rate (TPR), precision-recall curve, calibration slope and intercept and the overall performance with the Brier score.[28]

The c-statistic (area under the curve of a receiver operating characteristic curve) is a score ranging from 0.50 to 1.0 with 1.0 indicating the highest discrimination score and 0.50 indicating the lowest. The higher the discrimination score, the better the model's ability to distinguish patients with and without the outcome of interest.[20 23] The PPV is the proportion of true positive outcomes over the number of predicted positive outcomes. The TPR is the proportion of true positive outcomes over the number of observed positive outcomes. The precision recall curve is a plot of the PPV versus the TPR. A calibration plot plots the estimated versus the observed probabilities for the primary outcome. A perfect calibration plot has an intercept of 0 (<0 reflects overestimation and >0 reflects underestimation of the probability of the outcome) and a slope of 1 (model is performing similarly in training and test datasets).[20 23 28] The null-model Brier score, which equals the probability of glenoid revision in the dataset, will be used to benchmark the algorithm's Brier score. A Brier score lower than the null-model Brier score indicates superior performance of the prediction model to this null benchmark. Perfect prediction would have a Brier score of 0, whereas a Brier score of 1 would indicate the poorest possible prediction.[28]

In addition, the decision curve analysis will be performed and visualised to investigate the net benefit (weighted average of true positives and false positives) of the conducted algorithms over the range of risk thresholds for clinical decision-making.[23 28 29] The net benefit is a weighted average of true positives and false positives, formula=sensitivity×prevalence−(1−specificity)×(1−prevalence)×(odds at the threshold probability). The decision curve of the model will be compared with decision curves of treating everyone as being at risk and treating no one as being at risk.[23]

Due to the large heterogeneity of the compiled dataset from different international sources and the internal

BMJ Open: first published as 10.1136/bmjopen-2023-074700 on 18 October 2023. Downloaded from http://bmjopen.bmj.com/ on November 29, 2023 at University of Groningen. Protected by copyright.

validation of the prediction models, the generalisability of the model can be intrinsically confirmed using the above-mentioned performance tests. Therefore, it is not strictly necessary to externally validate the final algorithm. However, this study's primary aim is model development. External validation in a specific setting is advised before applying the algorithm to clinical practice.

Open-access clinical prediction tool

The best-performing prediction algorithm will be used to create an open-access clinical prediction tool, in the form of a publicly available web application accessible on desktops, tablets and smartphones.

Statistical analysis

Categorical variables will be described as absolute numbers with frequencies, and continuous variables as medians with IQRs. The model performance metrics will be calculated with 95% CI. Given the retrospective study design, post hoc power analyses will be conducted to evaluate the sample size of the study with an alpha value of 0.05.

Guidelines

The study set-up will be performed following the Transparent Reporting of Multivariable Prediction Models for Individual Prognosis or Diagnosis Guideline for Clustered data.[15]

Patient and public involvement

None.

Ethics and dissemination

For safe multicentre data exchange and analysis, our Machine Learning Consortium will adhere to the WHO regulation 'Policy on Use and Sharing of Data Collected by WHO in Member States Outside the Context of Public Health Emergencies'.[30] An Institutional Review Board (IRB) approval has been obtained for each of the included studies and the provided data are anonymised and de-identified; no additional prospective data are collected and contributing authors are not required to access any patient files. No IRB review is required for this study. Patient consent for publication is not applicable to this study.

The study results will be disseminated through publication in a peer-reviewed journal. To facilitate reproduction of the results and external validation of the algorithm, the (anonymous) code of the developed predictive algorithms will be made available upon request with the authors.

Data collection for this project is currently ongoing. The analysis will start in December 2023. The expected time of completion for the project is July 2024.

DISCUSSION

For an informed decision when considering ATSA, it is important to be able to make an accurate prediction of arthroplasty failure. Previous studies have demonstrated several factors that affect complications and revision after ATSA, including male sex, comorbidities such as chronic kidney disease or coronary artery disease, percentage haematocrit, a higher critical shoulder angle, teaching hospital status, operative time, and the material and design of the prosthesis.[2 3 5–9 13 14] Psychological studies have shown that in human judgement, only a limited amount of variables can be taken into account, and that prediction models are generally more accurate and less subject to bias.[31] Machine learning algorithms have been shown to be an effective method in developing patient-specific prediction tools, which may complement human judgement when counselling patients in clinic.[32] Creating an online tool for aseptic loosening of the glenoid component after ATSA can help guide surgeons in selecting patients who will most benefit from this treatment and considering alternatives in cases of high-risk estimates.

The strength of this project is the large amount of data that will be gathered from authors participating in the Machine Learning Consortium, aiming to include a minimum of 7500 patients in total. Using a large, heterogeneous international database for development of the algorithm and prediction tool will result in high external validity and may improve applicability worldwide.[15] However, most machine learning techniques require a larger sample size to achieve an accurate prediction compared with traditional regression models. The minimum events per variable are not clearly defined and differ per technique. Furthermore, in gathering data retrospectively from various sources, the study is subject to variances in the included variables. Low completeness and large variability of reporting may introduce bias. However, only variables that are consistently reported in multiple data sources will be included in the final analysis; variables will be categorised to increase uniformity and missing data will be imputed where possible. The exclusion and categorisation of variables will have to be balanced with the amount of detail in the final analysis. Furthermore, the accuracy of data collection is dependent on third parties providing the data; the method of data collection cannot be verified for all sources. However, the data source will be considered as a confounder. Furthermore, the variety in data sources will increase the external applicability of the algorithm. Last, as machine learning prediction models for a dichotomous outcome are limited to risk classification, the individual risk must be interpreted in the clinical context when used for medical decision-making.

Author affiliations
[1]Department of Orthopaedics and Sports Medicine, Erasmus Medical Center, Rotterdam, The Netherlands
[2]Alps Surgery Institute, Clinique Generale Annecy, Annecy, France

BMJ Open: first published as 10.1136/bmjopen-2023-074700 on 18 October 2023. Downloaded from http://bmjopen.bmj.com/ on November 29, 2023 at University of Groningen. Protected by copyright.

[3]Faculty of Science, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
[4]Department of Engineering Systems and Services, Delft University of Technology, Delft, The Netherlands
[5]Institut de Chirurgie Réparatrice, Locomoteur & Sport, Centre Hospitalier Universitaire de Nice, Nice, France
[6]Roth McFarlane Hand and Upper Limb Center, Schulich School of Medicine and Dentistry, London, Ontario, Canada
[7]Orthopaedic Surgery, University Medical Centre Groningen, Groningen, The Netherlands
[8]Department of Human Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
[9]Department of Orthopaedic Surgery, OLVG, Amsterdam, The Netherlands
[10]Department of Orthopedic Surgery, Hôpital Lapeyronie, Montpellier, France

Acknowledgements The authors would like to thank Dr Laurens Allaart and Dr Sanne van Spanning for providing examples and inspiration for this study based on their previous work.

Contributors AAM, GAB and MPJvdB contributed to the conception, methods, planning, and writing and reviewing of the protocol. LCM and JHFO contributed to the statistical and machine learning methods of the study. JHFO, PB, GSA, JND, LL and TL reviewed the draft of the protocol and provided supervision. All authors revised the final version of the protocol and gave approval for publication.

Funding This work was supported by Stichting Professor Michaël-van Vloten fonds, (VSBfonds grant number VSB.22/00098), Prins Bernard Cultuurfonds (grant number 40042437/NOD) and Stichting Erasmus Trustfonds (grant number 97030.2021.101.577/193/RB).

Competing interests LL received consultancy fees from Depuy Stryker and royalties from Depuy-Synthes. TL received consultancy fees from Depuy Mitek and Stryker. GAB received consultancy fees from Depuy-Synthes and Stryker. The remaining authors certify that they have received no funding and have no commercial associations that might pose a conflict of interest in connection with the submitted article.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iD

Arno Alexander Macken http://orcid.org/0000-0002-7513-7437

REFERENCES

1 Sharareh B, Yao JJ, Matsen FA, *et al*. What is the optimal management of a loose glenoid component after anatomic total shoulder arthroplasty: a systematic review. *J Shoulder Elbow Surg* 2023;32:653–61.
2 Tabeayo E, Chan PH, Prentice HA, *et al*. The association between critical shoulder angle and revision following anatomic total shoulder arthroplasty: a matched case-control study. *J Shoulder Elbow Surg* 2022;31:1796–802.
3 Aibinder W, Schoch B, Parsons M, *et al*. Risk factors for complications and revision surgery after anatomic and reverse total shoulder arthroplasty. *J Shoulder Elbow Surg* 2021;30:e689–701.
4 Dekker TJ, Grantham WJ, Lacheta L, *et al*. Glenoid retroversion does not impact clinical outcomes or implant survivorship after total shoulder arthroplasty with minimal, noncorrective reaming. *JSES Int* 2022;6:596–603.
5 Page RS, Alder-Price AC, Rainbird S, *et al*. Reduced revision rates in total shoulder arthroplasty with crosslinked polyethylene: results from the Australian Orthopaedic Association national joint replacement registry. *Clin Orthop Relat Res* 2022;480:1940–9.
6 Dillon MT, Chan PH, Prentice HA, *et al*. The association between glenoid component design and revision risk in anatomic total shoulder arthroplasty. *J Shoulder Elbow Surg* 2020;29:2089–96.
7 Welsher A, Gohal C, Madden K, *et al*. A comparison of pegged vs. keeled glenoid components regarding functional and radiographic outcomes in anatomic total shoulder arthroplasty: a systematic review and meta-analysis. *JSES Open Access* 2019;3:136–144.
8 Page RS, Pai V, Eng K, *et al*. Cementless versus cemented glenoid components in conventional total shoulder joint arthroplasty: analysis from the Australian Orthopaedic Association national joint replacement registry. *J Shoulder Elbow Surg* 2018;27:1859–65.
9 Sharplin PK, Frampton CMA, Hirner M. Cemented vs. uncemented glenoid fixation in total shoulder arthroplasty for osteoarthritis: a New Zealand joint registry study. *J Shoulder Elbow Surg* 2020;29:2097–103.
10 Kumar V, Roche C, Overman S, *et al*. What is the accuracy of three different machine learning techniques to predict clinical outcomes after shoulder arthroplasty? *Clin Orthop Relat Res* 2020;478:2351–63.
11 McLendon PB, Christmas KN, Simon P, *et al*. Machine learning can predict level of improvement in shoulder arthroplasty. *JB JS Open Access* 2021;6:e20.00128.
12 Polce EM, Kunze KN, Fu MC, *et al*. Development of supervised machine learning algorithms for prediction of satisfaction at 2 years following total shoulder arthroplasty. *J Shoulder Elbow Surg* 2021;30:e290–9.
13 Devana SK, Shah AA, Lee C, *et al*. Development of a machine learning algorithm for prediction of complications and unplanned readmission following primary anatomic total shoulder replacements. *J Shoulder Elb Arthroplast* 2022;6:24715492221075444.
14 Gowd AK, Agarwalla A, Amin NH, *et al*. Construct validation of machine learning in the prediction of short-term postoperative complications following total shoulder arthroplasty. *J Shoulder Elbow Surg* 2019;28:e410–21.
15 Debray TPA, Collins GS, Riley RD, *et al*. Transparent reporting of multivariable prediction models developed or validated using clustered data: TRIPOD-cluster checklist. *BMJ* 2023;380:e071018.
16 Papadonikolakis A, Neradilek MB, Matsen FA. Failure of the glenoid component in anatomic total shoulder arthroplasty: a systematic review of the English-language literature between 2006 and 2012. *J Bone Joint Surg Am* 2013;95:2205–12.
17 van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. *BMC Med Res Methodol* 2014;14:137.
18 Steyerberg EW. *Clinical Prediction Models; A Practical Approach to Development, Validation, and Updating*. Cham: Springer, 2019.
19 van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. *J Stat Softw* 2011;45:1–67.
20 Oosterhoff JHF, Gravesteijn BY, Karhade AV, *et al*. Feasibility of machine learning and logistic regression algorithms to predict outcome in orthopaedic trauma surgery. *J Bone Joint Surg Am* 2022;104:544–51.
21 DeBaun MR, Chavez G, Fithian A, *et al*. Artificial neural networks predict 30-day mortality after hip fracture: insights from machine learning. *J Am Acad Orthop Surg* 2021;29:977–83.
22 Riley RD, Ensor J, Snell KIE, *et al*. Calculating the sample size required for developing a clinical prediction model. *BMJ* 2020;368:m441.
23 Oosterhoff JHF, Savelberg ABMC, Karhade AV, *et al*. Development and internal validation of a clinical prediction model using machine learning algorithms for 90 day and 2 year mortality in femoral neck fracture patients aged 65 years or above. *Eur J Trauma Emerg Surg* 2022;48:4669–82.
24 Oosterhoff JHF, Karhade AV, Oberai T, *et al*. Prediction of postoperative delirium in geriatric hip fracture patients: a clinical prediction model using machine learning algorithms. *Geriatr Orthop Surg Rehabil* 2021;12:21514593211062277.
25 Bongers MER, Thio QCBS, Karhade AV, *et al*. Does the SORG algorithm predict 5-year survival in patients with

BMJ Open: first published as 10.1136/bmjopen-2023-074700 on 18 October 2023. Downloaded from http://bmjopen.bmj.com/ on November 29, 2023 at University of Groningen. Protected by copyright.

chondrosarcoma? An external validation. *Clin Orthop Relat Res* 2019;477:2296–303.
26 Karhade AV, Ogink PT, Thio Q, *et al*. Development of machine learning algorithms for prediction of prolonged opioid prescription after surgery for lumbar disc herniation. *Spine J* 2019;19:1764–71.
27 Karhade AV, Ogink PT, Thio Q, *et al*. Discharge disposition after anterior cervical discectomy and fusion. *World Neurosurg* 2019;132:e14–20.
28 Steyerberg EW, Vickers AJ, Cook NR, *et al*. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;21:128–38.
29 Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J* 2014;35:1925–31.
30 World Health Organization. Who policy statement on data sharing. policy on use and sharing of data collected in member States by the world health Organization (WHO) outside the context of public health emergencies. WHO, Available: https://www.who.int/docs/default-source/publishing-policies/who-policy-statement-on-data-sharing.pdf?sfvrsn=71a41493_0 [Accessed 13 Apr 2023].
31 Kahneman D, Sibony O, Sunstein CR. *Noise: A Flaw in Human Judgment Little, Brown Spark, version 1*. Hachette Book Group, 2021.
32 Machine Learning Consortium, on behalf of the SPRINT and FLOW Investigators. A machine learning algorithm to identify patients with Tibial shaft fractures at risk for infection after operative treatment. *J Bone Joint Surg Am* 2021;103:532–40.

BMJ Open: first published as 10.1136/bmjopen-2023-074700 on 18 October 2023. Downloaded from http://bmjopen.bmj.com/ on November 29, 2023 at University of Groningen. Protected by copyright.