BMJ Open

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

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2023-074700
Article Type:	Protocol
Date Submitted by the Author:	14-Apr-2023
Complete List of Authors:	Macken, Arno; Erasmus Medical Center, Department of Orthopaedics and Sports Medicine; Clinique Generale Annecy, Orthopaedic Surgery Macken, Loïc; Vrije Universiteit Amsterdam, Faculty of Science Oosterhoff, Jacobien; Delft University of Technology, Department of Engineering Systems and Services Boileau, Pascal; Centre Hospitalier Universitaire de Nice Athwal, George; Schulich School of Medicine and Dentistry Doornberg, Job; University Medical Centre Groningen, Orthopaedic Surgery Lafosse, Laurent; Clinique Générale Annecy, Orthopaedic Surgery Lafosse, Thibault; Clinique Generale Annecy, Orthopaedic surgery van den Bekerom, Michel; Vrije Universiteit Amsterdam, Department of Human Movement Sciences; OLVG, Department of Orthopaedic Surgery Buijze, Geert Alexander; Clinique Générale Annecy, Orthopaedic Surgery; Hôpital Lapeyronie, Department of Orthopedic Surgery
Keywords:	ORTHOPAEDIC & TRAUMA SURGERY, Shoulder < ORTHOPAEDIC & TRAUMA SURGERY, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS, Clinical Decision-Making, Risk Factors

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

MESH KEYWORDS: "Arthroplasty, Replacement, Shoulder", "Shoulder Prosthesis", "Glenoid Cavity", "Prosthesis Failure", "Reoperation", "Artificial Intelligence", "Machine Learning"

ABSTRACT

Introduction

Despite technological advancements in recent years, glenoid component loosening remains a common complication after anatomic total shoulder arthroplasty (ATSA) and is one of the main causes for 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 multi-centre 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 Penalized 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 World Health Organisation (WHO) regulation on data sharing. The study will be published in a peer-reviewed journal. An Institutional Review Board (IRB) review is not applicable.

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 ATSA.
- 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

Anatomic 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 for 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 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, noncorrective 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 poly-ethylene 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, pre-operative range of motion and patient-reported outcome measures (PROMs), patients' sex, and surgery on the dominant upper limb.[10] Another study was able to accurately predict the improvement in ASES scores after shoulder arthroplasty using machine learning.[11] The most relevant predictive factors were pre-operative ASES scores, pre-operative pain scores, Walch classification, fatty infiltration in the supra- and infraspinatus, and age. A previous study using artificial intelligence to predict patient satisfaction two 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, age, smoking status, anatomic versus reverse TSA, 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 to 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 multi-centre 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, the full search term is available in appendix 1. All original studies reporting revision or failure rates after 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 databases used for the identified studies. After sharing their data, the authors will be included in 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 that underwent ATSA with a fracture as the indication or patients that 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 anatomic total shoulder arthroplasty. Based on previous studies, we estimate a glenoid revision rate of approximately 2%, resulting in a required cohort of at least 5000 patients to achieve sufficient power for a model with up to 10 predictive variables.[6,16]

Data curation and missing data

Completeness across data sources will be assessed for each variable in the compiled multi-centre 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 (MICE).[17] 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. Eighty percent of the data will be randomly selected (stratified by outcome) and used as the training dataset for developing the algorithms and the remaining 20% will be used as the test dataset to assess the algorithm's performance. 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 Body Mass Index, 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.[18] 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.[19]

Development of prediction models

The following machine learning algorithms were chosen for modelling based on prior research [20–24]: Stochastic Gradient Boosting (SGM), Random Forest (RF), Support Vector Machine(SVM), Neural Network (NN) and Elastic-Net Penalized Logistic Regression (PLR). The algorithms will be trained on the training dataset with ten-fold cross-validation repeated 3 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, United States of America). Hyper parameter 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, calibration slope and intercept and the overall performance with the Brier score.[25]

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.[20,26] 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).[20,25,26] 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.[25]

In addition, the decision curve analysis will be performed and visualized 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.[20,25,27] The net benefit is a weighted average of true positives and false positives, formula = sensitivity x prevalence – (1-specificity) x (1 – prevalence) x (odds at the threshold probability). The decision curve of the model will be compared to decision curves of treating everyone as being at risk and treating no one as being at risk.[20]

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

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.

Patients and public involvement

Patients and the public were not involved in the making of this protocol.

Statistical analysis

Categorical variables will be described as absolute numbers with frequencies, and continuous variables as medians with interquartile ranges (IQR). The model performance metrics will be calculated with 95% confidence interval (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 (TRIPOD-Cluster).[15]

ETHICS AND DISSEMINATION

For safe multicentre data exchange and analysis, our Machine Learning Consortium will adhere to the World Health Organisation (WHO) regulation 'Policy on Use and Sharing of Data Collected by WHO in Member States Outside the Context of Public Health Emergencies'.[28] The study results will be disseminated through publication in a peer-reviewed journal. 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 IRB review is required for this study. Patient consent for publication is not applicable to this study.

CURRENT STATUS

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

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.[29] 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.[30] 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 5000 patients in total. Using a large, heterogenous international database for development of the algorithm and prediction tool will result in high external validity and may improve applicability world-wide.[15] However, 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.

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 Elb Surg* 2022 2023;**32(3)**:653-661. doi:10.1016/J.JSE.2022.10.006

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 Elb Surg* 2022;**31**:1796–802. doi:10.1016/J.JSE.2021.08.003

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 Elb Surg* 2021;**30**:e689–701. doi:10.1016/J.JSE.2021.04.029

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. doi:10.1016/J.JSEINT.2022.02.011

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. doi:10.1097/CORR.0000000000002293

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 Elb Surg* 2020;**29**:2089–96. doi:10.1016/J.JSE.2020.02.024

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.e1. doi:10.1016/J.JSES.2019.04.002

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 Elb Surg* 2018;**27**:1859–65. doi:10.1016/J.JSE.2018.03.017

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 Elb Surg* 2020;**29**:2097–103. doi:10.1016/J.JSE.2020.03.008

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. doi:10.1097/CORR.0000000000001263

11 McLendon PB, Christmas KN, Simon P, *et al.* Machine Learning Can Predict Level of Improvement in Shoulder Arthroplasty. *JBJS Open Access* 2021;**6**. doi:10.2106/JBJS.OA.20.00128

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 Elb Surg* 2021;**30**:e290–9. doi:10.1016/j.jse.2020.09.007

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**:247154922210754. doi:10.1177/24715492221075444

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 Elb Surg* 2019;**28**:e410–21. doi:10.1016/J.JSE.2019.05.017

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. doi:10.1136/BMJ-2022-071018

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 Jt Surg* 2013;**95**:2205–12. doi:10.2106/JBJS.L.00552

17 van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *J Stat Softw* 2011;**45**:1–67. doi:10.18637/JSS.V045.I03

18 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. doi:10.5435/JAAOS-D-20-00429

19 Riley RD, Ensor J, Snell KIE, *et al.* Calculating the sample size required for developing a clinical prediction model. *BMJ* 2020;**368**. doi:10.1136/BMJ.M441

20 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. doi:10.1007/s00068-022-01981-4

21 Oosterhoff JHF, Karhade A V., 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**. doi:10.1177/21514593211062277

22 Bongers MER, Thio QCBS, Karhade A V., *et al.* Does the SORG Algorithm Predict 5-year Survival in Patients with Chondrosarcoma? An External Validation. *Clin Orthop Relat Res*

2019;**477**:2296–303. doi:10.1097/CORR.0000000000000748

23 Karhade A V., Ogink PT, Thio QCBS, *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. doi:10.1016/J.SPINEE.2019.06.002

24 Karhade A V., Ogink PT, Thio QCBS, *et al.* Discharge Disposition After Anterior Cervical Discectomy and Fusion. *World Neurosurg* 2019;**132**:e14–20. doi:10.1016/J.WNEU.2019.09.026

25 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. doi:10.1097/EDE.0b013e3181c30fb2

26 Oosterhoff JHF, Gravesteijn BY, Karhade A V., *et al.* Feasibility of Machine Learning and Logistic Regression Algorithms to Predict Outcome in Orthopaedic Trauma Surgery. *J Bone Jt Surg - Am Vol* 2022;**104**:544–51. doi:10.2106/JBJS.21.00341

27 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. doi:10.1093/EURHEARTJ/EHU207

28 World Health Organization. Who policy statement on Data Sharing [Internet]. 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; [cited 2023Apr13]. Available from: https://www.who.int/docs/default-source/publishing-policies/who-policy-statement-on-data-sharing.pdf?sfvrsn=71a41493_0

29 Daniel Kahneman, Olivier Sibony, Cass R. Sunstein. Noise: A Flaw in Human Judgment *Little, Brown Spark; Hachette Book Group*, version **1**, 2021

30 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. doi:10.2106/JBJS.20.00903

AUTHOR'S CONTRIBUTIONS

Authors AM, GB, and MP contributed to the conception, methods, planning, and writing and reviewing of the protocol. LM and JO contributed to the statistical and machine learning methods of the study. JO, PB, GA, JD, 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.

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.

FUNDING STATEMENT

This work was supported by Stichting Prof. 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 STATEMENT

LL received consultancy fees from Depuy Stryker and royalties from Depuy-Synthes. TL received consultancy fees from Depuy Mitek and Stryker. GB received consultancy fees from Depuy-Synthes and Stryker. The remaining authors certify that he or she has received no funding or has commercial associations that might pose a conflict of interest in connection with the submitted article.

APPENDIX 1

Search strategy:

("Arthroplasty, Replacement, Shoulder"[Mesh] OR “Total shoulder arthroplasty”[tiab] OR “Anatomic shoulder arthroplasty”[tiab] OR “TSA”[tiab] OR “ATSA”[tiab] OR “Total shoulder prosthes*”[tiab] OR “Anatomic shoulder prosthes*”[tiab]) AND ("Reoperation"[Mesh] OR “reoperation”[tiab] OR “re-operation”[tiab] OR “revision”[tiab] OR “survival”[tiab] OR “implant-survival”[tiab] OR "Prosthesis Failure"[Mesh] OR “failure”[tiab] OR “loosening”[tiab] OR “aseptic”[tiab] OR "Postoperative Complications"[Mesh] OR “complication”[tiab])

Filters: Clinical study, Clinical trial, Controlled clinical trial, Comparative study, Dataset

TRIPOD Checklist: Prediction Model Development and Validation

Section/Topic	Item		Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	**X**
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	**X**
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	**X**
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	**X**
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	**X**
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	**X**
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	**X**
	5b	D;V	Describe eligibility criteria for participants.	**X**
	5c	D;V	Give details of treatments received, if relevant.	**X**
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	**X**
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	**X**
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	**X**
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	**X**
Sample size	8	D;V	Explain how the study size was arrived at.	**X**
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	**X**
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	**X**
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	**X**
	10c	V	For validation, describe how the predictions were calculated.	**X**
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	**X**
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	**X**
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	**X**
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	**X**
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	**N/A**
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	**N/A**
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	**N/A**
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	**N/A**
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	**N/A**
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	**N/A**
	15b	D	Explain how to the use the prediction model.	**N/A**
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	**N/A**
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	**N/A**
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	**X**
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	**X**
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	**N/A**
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	**X**
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	**X**
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	**X**

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.

BMJ Open

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

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2023-074700.R1
Article Type:	Protocol
Date Submitted by the Author:	02-Sep-2023
Complete List of Authors:	Macken, Arno; Erasmus Medical Center, Department of Orthopaedics and Sports Medicine; Clinique Generale Annecy, Orthopaedic Surgery Macken, Loïc; Vrije Universiteit Amsterdam, Faculty of Science Oosterhoff, Jacobien; Delft University of Technology, Department of Engineering Systems and Services Boileau, Pascal; Centre Hospitalier Universitaire de Nice Athwal, George; Schulich School of Medicine and Dentistry Doornberg, Job; University Medical Centre Groningen, Orthopaedic Surgery Lafosse, Laurent; Clinique Générale Annecy, Orthopaedic Surgery Lafosse, Thibault; Clinique Generale Annecy, Orthopaedic surgery van den Bekerom, Michel; Vrije Universiteit Amsterdam, Department of Human Movement Sciences; OLVG, Department of Orthopaedic Surgery Buijze, Geert Alexander; Clinique Générale Annecy, Orthopaedic Surgery; Hôpital Lapeyronie, Department of Orthopedic Surgery
Primary Subject Heading:	Surgery
Secondary Subject Heading:	Health informatics, Research methods
Keywords:	ORTHOPAEDIC & TRAUMA SURGERY, Shoulder < ORTHOPAEDIC & TRAUMA SURGERY, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS, Clinical Decision-Making, Risk Factors

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

AUTHORS: Arno A. 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[1], Thibault Lafosse[1], Michel P.J. van den Bekerom[8,9], Geert Alexander Buijze[1,10,11], on behalf the Machine Learning Consortium.

1. Alps Surgery Institute, Clinique Générale d'Annecy, France
2. Department of Orthopaedic Surgery and Sports medicine, Erasmus Medical Centre, Rotterdam, the Netherlands
3. Faculty of Science, Vrije Universiteit Amsterdam, the Netherlands
4. Department of Engineering Systems and Services, Faculty of Technology Policy and Management, Delft University of Technology, the Netherlands
5. Institut de Chirurgie Réparatrice, Locomoteur & Sports, Centre Hospitalier Universitaire de Nice, France
6. Roth McFarlane Hand and Upper Limb Center, Schulich School of Medicine and Dentistry, London, Ontario, Canada
7. Departments of Orthopaedic and Trauma Surgery, University Medical Center Groningen, the Netherlands
8. Department of Human Movement Sciences, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam Movement Sciences, Amsterdam, the Netherlands
9. Shoulder and Elbow Unit, Joint Research, Department of Orthopaedic Surgery, OLVG, Amsterdam, The Netherlands
10. Department of Orthopedic Surgery, Amsterdam Movement Sciences, Amsterdam UMC, location AMC, University of Amsterdam, Amsterdam, The Netherlands
11. Department of Orthopedic Surgery, Montpellier University Medical Center, Lapeyronie Hospital, University of Montpellier, Montpellier, France

Corresponding author: Arno A. Macken

4 Chem. de la Tour la Reine, 74000 Annecy, France

+33 4 50 33 04 69

arnomacken@gmail.com

MESH KEYWORDS: "Arthroplasty, Replacement, Shoulder", "Shoulder Prosthesis", "Glenoid Cavity", "Prosthesis Failure", "Reoperation", "Artificial Intelligence", "Machine Learning"

WORD COUNT: 2972

ABSTRACT

Introduction

Despite technological advancements in recent years, glenoid component loosening remains a common complication after anatomic total shoulder arthroplasty (ATSA) and is one of the main causes for 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 multi-centre 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 Penalized 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 World Health Organisation (WHO) regulation on data sharing. The study will be published in a peer-reviewed journal. An Institutional Review Board (IRB) review is not applicable.

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 ATSA.

- 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

Anatomic 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 for 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 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, noncorrective 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 poly-ethylene 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, pre-operative range of motion and patient-reported outcome measures (PROMs), patients' sex, and surgery on the dominant upper limb.[10] Another study was able to accurately predict the improvement in ASES scores after shoulder arthroplasty using machine learning.[11] The most relevant predictive factors were pre-operative ASES scores, pre-operative pain scores, Walch classification, fatty infiltration in the supra- and infraspinatus, and age. A previous study using artificial intelligence to predict patient satisfaction two 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, age, smoking status, anatomic

versus reverse TSA, 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 to 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 multi-centre 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 full search term is available in 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 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 that underwent ATSA with a fracture as the indication or patients that 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 anatomic total shoulder arthroplasty. 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 5 predictive variables.

Data curation and missing data

Completeness across data sources will be assessed for each variable in the compiled multi-centre 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 (MICE).[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 data set 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 ML 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 Body Mass Index, 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 ML 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 ML 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 (SGM), Random Forest (RF), Support Vector Machine (SVM), Neural Network (NN) and Elastic-Net Penalized Logistic Regression (PLR). The algorithms will be trained on the training dataset with ten-fold cross-validation repeated 3 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, United States of America). Hyper parameter 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 visualized 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 x prevalence – (1-specificity) x (1 – prevalence) x (odds at the threshold probability). The decision curve of the model will be compared to 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 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.

Patients and public involvement

Patients and the public were not involved in the making of this protocol.

Statistical analysis

Categorical variables will be described as absolute numbers with frequencies, and continuous variables as medians with interquartile ranges (IQR). The model performance metrics will be calculated with 95% confidence interval (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 (TRIPOD-Cluster).[15]

ETHICS AND DISSEMINATION

For safe multicentre data exchange and analysis, our Machine Learning Consortium will adhere to the World Health Organisation (WHO) regulation 'Policy on Use and Sharing of Data Collected by WHO in Member States Outside the Context of Public Health Emergencies'.[30] The study results will be disseminated through publication in a peer-reviewed journal. 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 is 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.

CURRENT STATUS

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 of 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, heterogenous international database for development of the algorithm and prediction tool will result in high external validity and may improve applicability world-wide.[15] However, most ML techniques require a larger sample size to achieve an accurate prediction compared to traditional regression models. The minimum events per variable is not clearly defined and differs 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, ML 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.

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 Elb Surg* Published Online First: November 2022. doi:10.1016/J.JSE.2022.10.006

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 Elb Surg* 2022;**31**:1796–802. doi:10.1016/J.JSE.2021.08.003

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 Elb Surg* 2021;**30**:e689–701. doi:10.1016/J.JSE.2021.04.029

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. doi:10.1016/J.JSEINT.2022.02.011

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. doi:10.1097/CORR.0000000000002293

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 Elb Surg* 2020;**29**:2089–96. doi:10.1016/J.JSE.2020.02.024

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.e1. doi:10.1016/J.JSES.2019.04.002

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 Elb Surg* 2018;**27**:1859–65. doi:10.1016/J.JSE.2018.03.017

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 Elb Surg* 2020;**29**:2097–103. doi:10.1016/J.JSE.2020.03.008

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. doi:10.1097/CORR.0000000000001263

11 McLendon PB, Christmas KN, Simon P, *et al.* Machine Learning Can Predict Level of Improvement in Shoulder Arthroplasty. *JBJS Open Access* 2021;**6**. doi:10.2106/JBJS.OA.20.00128

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 Elb Surg* 2021;**30**:e290–9. doi:10.1016/j.jse.2020.09.007

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. *https://doi.org/101177/24715492221075444* 2022;**6**:247154922210754. doi:10.1177/24715492221075444

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 Elb Surg* 2019;**28**:e410–21. doi:10.1016/J.JSE.2019.05.017

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. doi:10.1136/BMJ-2022-071018

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 Jt Surg* 2013;**95**:2205–12. doi:10.2106/JBJS.L.00552

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**:1–13. doi:10.1186/1471-2288-14-137/FIGURES/13

18 Steyerberg EW. Clinical Prediction Models; A Practical Approach to Development, Validation, and Updating. Springer 2019

19 van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *J Stat Softw* 2011;**45**:1–67. doi:10.18637/JSS.V045.I03

20 Oosterhoff JHF, Gravesteijn BY, Karhade A V., *et al.* Feasibility of Machine Learning and Logistic Regression Algorithms to Predict Outcome in Orthopaedic Trauma Surgery. *J Bone Jt Surg - Am Vol* 2022;**104**:544–51. doi:10.2106/JBJS.21.00341

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. doi:10.5435/JAAOS-D-20-00429

22 Riley RD, Ensor J, Snell KIE, *et al.* Calculating the sample size required for developing a clinical prediction model. *BMJ* 2020;**368**. doi:10.1136/BMJ.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. doi:10.1007/s00068-022-01981-4

24 Oosterhoff JHF, Karhade A V., 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**. doi:10.1177/21514593211062277

25 Bongers MER, Thio QCBS, Karhade A V., *et al.* Does the SORG Algorithm Predict 5-year Survival in Patients with Chondrosarcoma? An External Validation. *Clin Orthop Relat Res* 2019;**477**:2296–303. doi:10.1097/CORR.0000000000000748

26 Karhade A V., Ogink PT, Thio QCBS, *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. doi:10.1016/J.SPINEE.2019.06.002

27 Karhade A V., Ogink PT, Thio QCBS, *et al.* Discharge Disposition After Anterior Cervical Discectomy and Fusion. *World Neurosurg* 2019;**132**:e14–20. doi:10.1016/J.WNEU.2019.09.026

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. doi:10.1097/EDE.0b013e3181c30fb2

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. doi:10.1093/EURHEARTJ/EHU207

30 World Health Organization. Who policy statement on Data Sharing [Internet]. 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; [cited 2023Apr13]. Available from: https://www.who.int/docs/default-source/publishing-policies/who-policy-statement-on-data-sharing.pdf?sfvrsn=71a41493_0

31 Daniel Kahneman, Olivier Sibony, Cass R. Sunstein. Noise: A Flaw in Human Judgment Little, Brown Spark; Hachette Book Group, version 1, 2021

32 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. doi:10.2106/JBJS.20.00903

AUTHOR'S CONTRIBUTIONS

Authors AM, GB, and MP contributed to the conception, methods, planning, and writing and reviewing of the protocol. LM and JO contributed to the statistical and machine learning methods of the study. JO, PB, GA, JD, 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.

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.

FUNDING STATEMENT

This work was supported by Stichting Prof. 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 STATEMENT

LL received consultancy fees from Depuy Stryker and royalties from Depuy-Synthes. TL received consultancy fees from Depuy Mitek and Stryker. GB received consultancy fees from Depuy-Synthes and Stryker. The remaining authors certify that he or she has received no funding or has commercial associations that might pose a conflict of interest in connection with the submitted article.

APPENDIX 1

Search strategy:

("Arthroplasty, Replacement, Shoulder"[Mesh] OR “Total shoulder arthroplasty”[tiab] OR “Anatomic shoulder arthroplasty”[tiab] OR “TSA”[tiab] OR “ATSA”[tiab] OR “Total shoulder prosthes*”[tiab] OR “Anatomic shoulder prosthes*”[tiab]) AND ("Reoperation"[Mesh] OR “reoperation”[tiab] OR “re-operation”[tiab] OR “revision”[tiab] OR “survival”[tiab] OR “implant-survival”[tiab] OR "Prosthesis Failure"[Mesh] OR “failure”[tiab] OR “loosening”[tiab] OR “aseptic”[tiab] OR "Postoperative Complications"[Mesh] OR “complication”[tiab])

Filters: Clinical study, Clinical trial, Controlled clinical trial, Comparative study, Dataset

TRIPOD Checklist: Prediction Model Development and Validation

Section/Topic	Item		Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	**X**
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	**X**
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	**X**
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	**X**
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	**X**
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	**X**
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	**X**
	5b	D;V	Describe eligibility criteria for participants.	**X**
	5c	D;V	Give details of treatments received, if relevant.	**X**
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	**X**
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	**X**
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	**X**
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	**X**
Sample size	8	D;V	Explain how the study size was arrived at.	**X**
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	**X**
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	**X**
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	**X**
	10c	V	For validation, describe how the predictions were calculated.	**X**
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	**X**
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	**X**
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	**X**
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	**X**
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	**N/A**
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	**N/A**
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	**N/A**
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	**N/A**
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	**N/A**
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	**N/A**
	15b	D	Explain how to the use the prediction model.	**N/A**
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	**N/A**
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	**N/A**
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	**X**
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	**X**
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	**N/A**
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	**X**
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	**X**
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	**X**

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.

BMJ Open

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

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2023-074700.R2
Article Type:	Protocol
Date Submitted by the Author:	23-Sep-2023
Complete List of Authors:	Macken, Arno; Erasmus Medical Center, Department of Orthopaedics and Sports Medicine; Clinique Generale Annecy, Orthopaedic Surgery Macken, Loïc; Vrije Universiteit Amsterdam, Faculty of Science Oosterhoff, Jacobien; Delft University of Technology, Department of Engineering Systems and Services Boileau, Pascal; Centre Hospitalier Universitaire de Nice Athwal, George; Schulich School of Medicine and Dentistry Doornberg, Job; University Medical Centre Groningen, Orthopaedic Surgery Lafosse, Laurent; Clinique Générale Annecy, Orthopaedic Surgery Lafosse, Thibault; Clinique Generale Annecy, Orthopaedic surgery van den Bekerom, Michel; Vrije Universiteit Amsterdam, Department of Human Movement Sciences; OLVG, Department of Orthopaedic Surgery Buijze, Geert Alexander; Clinique Générale Annecy, Orthopaedic Surgery; Hôpital Lapeyronie, Department of Orthopedic Surgery
Primary Subject Heading:	Surgery
Secondary Subject Heading:	Health informatics, Research methods
Keywords:	ORTHOPAEDIC & TRAUMA SURGERY, Shoulder < ORTHOPAEDIC & TRAUMA SURGERY, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS, Clinical Decision-Making, Risk Factors

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

Arno A. 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[1], Thibault Lafosse[1], Michel P.J. van den Bekerom[8,9], Geert Alexander Buijze[1,10,11], on behalf the Machine Learning Consortium

1. Alps Surgery Institute, Clinique Générale d'Annecy, France
2. Department of Orthopaedic Surgery and Sports medicine, Erasmus Medical Centre, Rotterdam, Netherlands
3. Faculty of Science, Vrije Universiteit Amsterdam, Netherlands
4. Department of Engineering Systems and Services, Faculty of Technology Policy and Management, Delft University of Technology, Netherlands
5. Institut de Chirurgie Réparatrice, Locomoteur & Sports, Centre Hospitalier Universitaire de Nice, France
6. Roth McFarlane Hand and Upper Limb Center, Schulich School of Medicine and Dentistry, London, Ontario, Canada
7. Departments of Orthopaedic and Trauma Surgery, University Medical Center Groningen, Netherlands
8. Department of Human Movement Sciences, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam Movement Sciences, Amsterdam, Netherlands
9. Shoulder and Elbow Unit, Joint Research, Department of Orthopaedic Surgery, OLVG, Amsterdam, Netherlands
10. Department of Orthopedic Surgery, Amsterdam Movement Sciences, Amsterdam UMC, location AMC, University of Amsterdam, Amsterdam, Netherlands
11. Department of Orthopedic Surgery, Montpellier University Medical Center, Lapeyronie Hospital, University of Montpellier, Montpellier, France

Correspondence to:
Arno A. Macken
4 Chem. de la Tour la Reine, 74000 Annecy, France
+33 4 50 33 04 69
arnomacken@gmail.com

KEYWORDS: "Arthroplasty, Replacement, Shoulder", "Shoulder Prosthesis", "Glenoid Cavity", "Prosthesis Failure", "Reoperation", "Artificial Intelligence", "Machine Learning"

WORD COUNT: 3028

ABSTRACT

Introduction

Despite technological advancements in recent years, glenoid component loosening remains a common complication after anatomic total shoulder arthroplasty (ATSA) and is one of the main causes for

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 multi-centre 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 Penalized 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 World Health Organisation (WHO) regulation on data sharing. An Institutional Review Board (IRB) 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 anatomic 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

Anatomic 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 for 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 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, noncorrective 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 poly-ethylene 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, pre-operative range of motion and patient-reported outcome measures (PROMs), patients' sex, and surgery on the dominant upper limb.[10] Another study was able to accurately predict the improvement in ASES scores after shoulder arthroplasty using machine learning.[11] The most relevant predictive factors were pre-operative ASES scores, pre-operative pain scores, Walch classification, fatty infiltration in the supra- and infraspinatus, and age. A previous study using artificial intelligence to predict patient satisfaction two 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, age, smoking status, anatomic versus reverse TSA, 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 to 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 multi-centre 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 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 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 that underwent ATSA with a fracture as the indication or patients that 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 anatomic total shoulder arthroplasty. 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 5 predictive variables.

Data curation and missing data

Completeness across data sources will be assessed for each variable in the compiled multi-centre 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 (MICE).[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 data set 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 ML 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 Body Mass Index, 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 ML 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 ML 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 (SGM), Random Forest (RF), Support Vector Machine (SVM), Neural Network (NN) and Elastic-Net Penalized Logistic Regression (PLR). The algorithms will be trained on the training dataset with ten-fold cross-validation repeated 3 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, United States of America). Hyper parameter 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 visualized 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 x prevalence – (1-specificity) x (1 – prevalence) x

(odds at the threshold probability). The decision curve of the model will be compared to 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 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 interquartile ranges (IQR). The model performance metrics will be calculated with 95% confidence interval (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 (TRIPOD-Cluster).[15]

Patient and public involvement

None.

ETHICS AND DISSEMINATION

For safe multicentre data exchange and analysis, our Machine Learning Consortium will adhere to the World Health Organisation (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 is 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, heterogenous international database for development of the algorithm and prediction tool will result in high external validity and may improve applicability world-wide.[15] However, most ML techniques require a larger sample size to achieve an accurate prediction compared to traditional regression models. The minimum events per variable is not clearly defined and differs 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, ML 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.

CONTRIBUTORS

Authors AM, GB, and MP contributed to the conception, methods, planning, and writing and reviewing of the protocol. LM and JO contributed to the statistical and machine learning methods of the study. JO, PB, GA, JD, 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.

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.

FUNDING

This work was supported by Stichting Prof. 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. GB 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.

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 Elb Surg* Published Online First: November 2022. doi:10.1016/J.JSE.2022.10.006

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 Elb Surg* 2022;**31**:1796–802. doi:10.1016/J.JSE.2021.08.003

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 Elb Surg* 2021;**30**:e689–701. doi:10.1016/J.JSE.2021.04.029

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. doi:10.1016/J.JSEINT.2022.02.011

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. doi:10.1097/CORR.0000000000002293

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 Elb Surg* 2020;**29**:2089–96. doi:10.1016/J.JSE.2020.02.024

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.e1. doi:10.1016/J.JSES.2019.04.002

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 Elb Surg* 2018;**27**:1859–65. doi:10.1016/J.JSE.2018.03.017

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 Elb Surg* 2020;**29**:2097–103. doi:10.1016/J.JSE.2020.03.008

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. doi:10.1097/CORR.0000000000001263

11 McLendon PB, Christmas KN, Simon P, *et al.* Machine Learning Can Predict Level of Improvement in Shoulder Arthroplasty. *JBJS Open Access* 2021;**6**. doi:10.2106/JBJS.OA.20.00128

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 Elb Surg* 2021;**30**:e290–9. doi:10.1016/j.jse.2020.09.007

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. *https://doi.org/101177/24715492221075444* 2022;**6**:247154922210754. doi:10.1177/24715492221075444

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 Elb Surg 2019;**28**:e410–21. doi:10.1016/J.JSE.2019.05.017

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. doi:10.1136/BMJ-2022-071018

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 Jt Surg* 2013;**95**:2205–12. doi:10.2106/JBJS.L.00552

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**:1–13. doi:10.1186/1471-2288-14-137/FIGURES/13

18 Steyerberg EW. Clinical Prediction Models; A Practical Approach to Development, Validation, and Updating. Springer 2019

19 van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *J Stat Softw* 2011;**45**:1–67. doi:10.18637/JSS.V045.I03

20 Oosterhoff JHF, Gravesteijn BY, Karhade A V., *et al.* Feasibility of Machine Learning and Logistic Regression Algorithms to Predict Outcome in Orthopaedic Trauma Surgery. *J Bone Jt Surg - Am Vol* 2022;**104**:544–51. doi:10.2106/JBJS.21.00341

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. doi:10.5435/JAAOS-D-20-00429

22 Riley RD, Ensor J, Snell KIE, *et al.* Calculating the sample size required for developing a clinical prediction model. *BMJ* 2020;**368**. doi:10.1136/BMJ.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. doi:10.1007/s00068-022-01981-4

24 Oosterhoff JHF, Karhade A V., 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**. doi:10.1177/21514593211062277

25 Bongers MER, Thio QCBS, Karhade A V., *et al.* Does the SORG Algorithm Predict 5-year Survival in Patients with Chondrosarcoma? An External Validation. *Clin Orthop Relat Res* 2019;**477**:2296–303. doi:10.1097/CORR.0000000000000748

26 Karhade A V., Ogink PT, Thio QCBS, *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. doi:10.1016/J.SPINEE.2019.06.002

27 Karhade A V., Ogink PT, Thio QCBS, *et al.* Discharge Disposition After Anterior Cervical Discectomy and Fusion. *World Neurosurg* 2019;**132**:e14–20. doi:10.1016/J.WNEU.2019.09.026

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. doi:10.1097/EDE.0b013e3181c30fb2

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. doi:10.1093/EURHEARTJ/EHU207

30 World Health Organization. Who policy statement on Data Sharing [Internet]. 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; [cited 2023Apr13]. Available from: https://www.who.int/docs/default-source/publishing-policies/who-policy-statement-on-data-sharing.pdf?sfvrsn=71a41493_0

31 Daniel Kahneman, Olivier Sibony, Cass R. Sunstein. Noise: A Flaw in Human Judgment Little, Brown Spark; Hachette Book Group, version 1, 2021

32 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. doi:10.2106/JBJS.20.00903

APPENDIX 1

Search strategy:

("Arthroplasty, Replacement, Shoulder"[Mesh] OR “Total shoulder arthroplasty”[tiab] OR “Anatomic shoulder arthroplasty”[tiab] OR “TSA”[tiab] OR “ATSA”[tiab] OR “Total shoulder prosthes*”[tiab] OR “Anatomic shoulder prosthes*”[tiab]) AND ("Reoperation"[Mesh] OR “reoperation”[tiab] OR “re-operation”[tiab] OR “revision”[tiab] OR “survival”[tiab] OR “implant-survival”[tiab] OR "Prosthesis Failure"[Mesh] OR “failure”[tiab] OR “loosening”[tiab] OR “aseptic”[tiab] OR "Postoperative Complications"[Mesh] OR “complication”[tiab])

Filters: Clinical study, Clinical trial, Controlled clinical trial, Comparative study, Dataset

TRIPOD Checklist: Prediction Model Development and Validation

Section/Topic	Item		Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	**X**
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	**X**
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	**X**
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	**X**
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	**X**
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	**X**
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	**X**
	5b	D;V	Describe eligibility criteria for participants.	**X**
	5c	D;V	Give details of treatments received, if relevant.	**X**
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	**X**
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	**X**
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	**X**
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	**X**
Sample size	8	D;V	Explain how the study size was arrived at.	**X**
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	**X**
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	**X**
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	**X**
	10c	V	For validation, describe how the predictions were calculated.	**X**
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	**X**
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	**X**
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	**X**
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	**X**
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	**N/A**
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	**N/A**
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	**N/A**
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	**N/A**
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	**N/A**
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	**N/A**
	15b	D	Explain how to the use the prediction model.	**N/A**
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	**N/A**
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	**N/A**
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	**X**
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	**X**
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	**N/A**
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	**X**
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	**X**
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	**X**

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.