

BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

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:	<i>BMJ Open</i>
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

SCHOLARONE™
Manuscripts

1
2
3 1 **TITLE:** Developing a machine learning algorithm to predict the probability of aseptic loosening of
4 2 the glenoid component after anatomic total shoulder arthroplasty: a protocol for a retrospective,
5 3 multicentre study.
6
7
8
9 4

10 5 **MESH KEYWORDS:** “Arthroplasty, Replacement, Shoulder”, “Shoulder Prosthesis”, “Glenoid
11 6 Cavity”, “Prosthesis Failure”, “Reoperation”, “Artificial Intelligence”, “Machine Learning”
12
13
14 7

16 8 **ABSTRACT**

18 9 **Introduction**

20 10 Despite technological advancements in recent years, glenoid component loosening remains a common
21 11 complication after anatomic total shoulder arthroplasty (ATSA) and is one of the main causes for
22 12 revision surgery. Increasing emphasis is placed on the prevention of glenoid component failure.
23 13 Previous studies have successfully predicted range of motion, patient-reported outcomes, and short-
24 14 term complications after ATSA using machine learning methods, but an accurate predictive model for
25 15 (glenoid component) revision is currently lacking. This study aims to use a large international
26 16 database to accurately predict aseptic loosening of the glenoid component after ATSA using machine
27 17 learning algorithms.
28
29
30
31
32
33
34

35 18 **Methods and analysis**

36
37 19 For this multi-centre retrospective study, individual patient data will be compiled from previously
38 20 published studies reporting revision of ATSA. A systematic literature search will be performed in
39 21 Medline (PubMed) identifying all studies reporting outcomes of ATSA. Authors will be contacted and
40 22 invited to participate in the Machine Learning Consortium by sharing their anonymised databases. All
41 23 databases reporting revisions after ATSA will be included and individual patients with a follow-up
42 24 less than 2 years or a fracture as the indication for ATSA will be excluded. First, features (predictive
43 25 variables) will be identified using a random forest feature selection. The resulting features from the
44 26 compiled database will be used to train various machine learning algorithms (Stochastic Gradient
45 27 Boosting, Random Forest, Support Vector Machine, Neural Network and Elastic-Net Penalized
46 28 Logistic Regression). The developed and validated algorithms will be evaluated across discrimination
47 29 (c-statistic), calibration, the Brier score, and the decision curve analysis. The best-performing
48 30 algorithm will be used to create an open-access online prediction tool.
49
50
51
52
53
54
55
56

57 31 **Ethics and dissemination**

58
59
60

1
2
3 32 Data will be collected adhering to the World Health Organisation (WHO) regulation on data sharing.
4 33 The study will be published in a peer-reviewed journal. An Institutional Review Board (IRB) review
5 34 is not applicable.
6
7
8
9 35

10 36 **STRENGTHS AND LIMITATIONS OF THIS STUDY**

- 13 37 - A large international database will be collected, which increases accuracy, validity, and
14 38 external applicability.
- 16 39 - A clinical prediction model using machine learning algorithms will be developed to estimate
17 40 the probability of aseptic loosening of the glenoid after ATSA.
- 19 41 - An open-access prediction tool based on the best performing algorithm will be made available
20 42 online that can guide medical professionals in personalised treatment decision-making.
- 22 43 - The study is dependent on data provided by third parties, which is a potential source of bias.
- 24 44 - Input variables will be selected and categorised based on completeness and uniformity across
25 45 data sources, potentially decreasing the amount of detail in the study.

28 46 29 30 47 **INTRODUCTION**

31
32 48 Anatomic total shoulder arthroplasty (ATSA) is used for glenohumeral arthropathy causing pain
33 49 and/or a reduction in range of motion. Despite technological advancements in recent years, glenoid
34 50 component loosening remains a common complication after ATSA and is one of the main causes for
35 51 revision surgery. Glenoid loosening can be a trying complication to manage and the optimal course of
36 52 treatment remains unclear.[1]

37 53 Consequently, increasing emphasis is placed on the prevention of loosening. The predicted chance of
38 54 glenoid component failure plays an important role in clinical decision-making such as patient
39 55 selection for ATSA or which implants and techniques to use. Several previous studies assessing risk
40 56 factors of glenoid component loosening identified patient, treatment and prosthesis characteristics
41 57 related to glenoid component loosening. For example, male sex and a higher critical shoulder angle
42 58 have been associated with higher rates of loosening and revisions.[2,3] Glenoid retroversion did not
43 59 impact implant survivorship in one study of ATSA with minimal, noncorrective reaming.[4]
44 60 However, a larger degree of retroversion may have more impact. Several previous studies have also
45 61 identified aspects of the glenoid component design that correlated with the rate of loosening, such as
46 62 whether the poly-ethylene is cross-linked, whether the component is pegged or keeled, or the usage of
47 63 cement.[5–9] In spite of these studies identifying influential factors, accurate prediction of aseptic
48 64 loosening of the glenoid component remains a challenge with conventional methods.

1
2
3 65 In recent years, machine learning or artificial intelligence has been used with increasing precision to
4 66 predict outcomes after ATSA. A previous study using machine learning was able to accurately predict
5 67 range of motion and patient-reported outcomes after ATSA. The most influential factors they reported
6 68 were follow-up time, pre-operative range of motion and patient-reported outcome measures (PROMs),
7 69 patients' sex, and surgery on the dominant upper limb.[10] Another study was able to accurately
8 70 predict the improvement in ASES scores after shoulder arthroplasty using machine learning.[11] The
9 71 most relevant predictive factors were pre-operative ASES scores, pre-operative pain scores, Walch
10 72 classification, fatty infiltration in the supra- and infraspinatus, and age. A previous study using
11 73 artificial intelligence to predict patient satisfaction two years after shoulder arthroplasty found
12 74 baseline Single Assessment Numeric Evaluation score, exercise and activity, workers' compensation
13 75 status, diagnosis, symptom duration prior to surgery, body mass index, age, smoking status, anatomic
14 76 versus reverse TSA, and diabetes to be predicting factors.[12] Two studies report predictive models
15 77 on short-term complications after ATSA. One study using machine learning to predict complications
16 78 and 30-day unplanned readmissions found that a history of implant complication, severe chronic
17 79 kidney disease, teaching hospital status, coronary artery disease and male sex were the most important
18 80 features.[13] The machine learning model found teaching hospital status and male sex as a markedly
19 81 more important predictor compared to a logistic regression analysis of the same data. Another study
20 82 on short-term complications after total shoulder arthroplasty found percentage haematocrit, BMI, and
21 83 operative time were of highest importance in outcome prediction.[14] These studies demonstrate that
22 84 machine learning may provide accurate predictions for the outcomes after ATSA. Machine learning is
23 85 most effective with large amounts of data and is very dependent on the amount of detail. Furthermore,
24 86 the algorithm needs to be widely applicable; a varied and international database provides the highest
25 87 external validity.

26
27
28
29
30
31
32
33
34
35
36
37
38
39
40 88 To our knowledge, there are no studies predicting the long-term complications such as aseptic
41 89 loosening of the glenoid component using advanced machine learning techniques. Furthermore,
42 90 previous machine learning studies are limited in accuracy and validity due to the sample size and
43 91 homogeneity. Therefore, this study aims to develop a clinical prediction model for aseptic loosening
44 92 of the glenoid component using machine learning algorithms trained on a large international database
45 93 using clustered data. The large combined dataset is less prone to overfitting, and allows direct
46 94 validation of models across a range of populations and settings, thereby increasing
47 95 generalisability.[15] The predictive algorithm will be made available for clinical use through a
48 96 publicly available online prediction tool.

49
50
51
52
53
54
55 97

56 98 **METHODS AND ANALYSIS**

57 99 **Data collection**

1
2
3 100 For this multi-centre retrospective study individual patient data (IPD) will be collected from
4 101 previously published studies reporting failure and revision of ATSA. A systematic literature search
5 102 will be performed in Medline (PubMed) identifying all studies that report a cohort of ATSA including
6 103 revision as an outcome, the full search term is available in appendix 1. All original studies reporting
7 104 revision or failure rates after ATSA will be included. Reviews and letters to the editor will be
8 105 excluded, as well as studies published in languages other than English, Italian, Dutch, and French.
9 106 Authors will be requested to share the databases used for the identified studies. After sharing their
10 107 data, the authors will be included in in the Machine Learning Consortium. Inclusion criteria for
11 108 individual patients within the provided databases are a minimum age of 18 years and a minimum
12 109 follow-up of 2 years. Patients that underwent ATSA with a fracture as the indication or patients that
13 110 underwent concomitant procedures such as a cuff repair, tendon transfer or bone graft will be
14 111 excluded. The aim is to combine the IPD from previously published studies to create a large
15 112 international cohort which can be used to train a machine learning algorithm to predict aseptic
16 113 loosening of the glenoid component after anatomic total shoulder arthroplasty. Based on previous
17 114 studies, we estimate a glenoid revision rate of approximately 2%, resulting in a required cohort of at
18 115 least 5000 patients to achieve sufficient power for a model with up to 10 predictive variables.[6,16]

29 116 **Data curation and missing data**

30 117 Completeness across data sources will be assessed for each variable in the compiled multi-centre
31 118 database and variables with sufficient completeness (>70% complete) will be selected as input for the
32 119 machine learning algorithms. Variables with >30% missing data will be excluded. For the remaining
33 120 variables, missing data will be completed by imputation using multivariate imputation by chained
34 121 equations (MICE).[17] Uniformity in reporting will be assessed for each variable. If possible,
35 122 variables will be adjusted or categorised to ensure uniform reporting. In case uniformity of the
36 123 reported variable across data sources cannot be achieved without guaranteeing correctness, the
37 124 variable will be excluded. Eighty percent of the data will be randomly selected (stratified by outcome)
38 125 and used as the training dataset for developing the algorithms and the remaining 20% will be used as
39 126 the test dataset to assess the algorithm's performance. Data curation and imputation will be performed
40 127 using R (R foundation for statistical computing, Vienna, Austria).

49 128 **Variable Selection**

50 129 The primary outcome is a revision of the glenoid component for aseptic glenoid loosening. The input
51 130 variables for both methods are dependent on the uniformity and completeness of the gathered data but
52 131 will include demographics (eg. age, sex and ethnicity), patient-specific factors (eg, preoperative Body
53 132 Mass Index, comorbidity, smoking, dominance, previous surgery), disease-specific factors (eg,
54 133 affected side, indication, Walch classification, fatty-infiltration of cuff muscles) and surgical
55 134 characteristics (eg, corrective reaming, component design and type, component materials, cementing,

1
2
3 135 and sizes). Before training the machine learning models, relevant variables will be selected using
4 136 random forest algorithms with recursive selection.[18] At least 10 events for each predictor variable
5 137 will be included in the model, adhering to the rule of thumb in predictive models of binary
6 138 variables.[19]

10 139 **Development of prediction models**

11
12 140 The following machine learning algorithms were chosen for modelling based on prior research [20–
13 141 24]: Stochastic Gradient Boosting (SGM), Random Forest (RF), Support Vector Machine(SVM),
14 142 Neural Network (NN) and Elastic-Net Penalized Logistic Regression (PLR). The algorithms will be
15 143 trained on the training dataset with ten-fold cross-validation repeated 3 times. Cross-validation means
16 144 dividing data into a selected number of groups, also called folds. First, the data will be divided into 10
17 145 equally sized folds. Then, the algorithms will be trained on 9 of the 10 folds (90% of the training data)
18 146 and tested on the remaining fold (10% of the training data). Results will be averaged across all
19 147 repetitions of this sequence. Machine learning algorithms will be developed using Python (The
20 148 Python Software Foundation, Fredericksburg, United States of America). Hyper parameter tuning will
21 149 be performed as recommended in the Python libraries. The statistician who performs the machine
22 150 learning analysis will be blinded to the origin of the data, but the anonymised data source will be
23 151 available to be included as a potential confounding factor.

31 152 **Model Performance**

32 153 After training all models, the model performance will be analysed according to a proposed framework
33 154 by Steyerberg et al. including discrimination with the c-statistic, calibration slope and intercept and
34 155 the overall performance with the Brier score.[25]

35 156 The c-statistic (area under the curve of a receiver operating characteristic curve) is a score ranging
36 157 from 0.50 to 1.0 with 1.0 indicating the highest discrimination score and 0.50 indicating the lowest.
37 158 The higher the discrimination score, the better the model's ability to distinguish patients with and
38 159 without the outcome of interest.[20,26] A calibration plot plots the estimated versus the observed
39 160 probabilities for the primary outcome. A perfect calibration plot has an intercept of 0 (<0 reflects
40 161 overestimation and >0 reflects underestimation of the probability of the outcome) and a slope of 1
41 162 (model is performing similarly in training and test datasets).[20,25,26] The null-model Brier score,
42 163 which equals the probability of glenoid revision in the dataset, will be used to benchmark the
43 164 algorithm's Brier score. A Brier score lower than the null-model Brier score indicates superior
44 165 performance of the prediction model to this null benchmark. Perfect prediction would have a Brier
45 166 score of 0, whereas a Brier score of 1 would indicate the poorest possible prediction.[25]

46 167 In addition, the decision curve analysis will be performed and visualized to investigate the net benefit
47 168 (weighted average of true positives and false positives) of the conducted algorithms over the range of

1
2
3 169 risk thresholds for clinical decision-making.[20,25,27] The net benefit is a weighted average of true
4 170 positives and false positives, formula = sensitivity x prevalence – (1-specificity) x (1 – prevalence) x
5 171 (odds at the threshold probability). The decision curve of the model will be compared to decision
6 172 curves of treating everyone as being at risk and treating no one as being at risk.[20]
7
8
9

10 173 Due to the large heterogeneity of the compiled dataset from different international sources and the
11 174 internal validation of the prediction models, the generalisability of the model can be intrinsically
12 175 confirmed using the above-mentioned performance tests. Therefore, it is not strictly necessary to
13 176 externally validate the final algorithm.
14
15
16

17 177 **Open-access clinical prediction tool**

18
19 178 The best performing prediction algorithm will be used to create an open-access clinical prediction
20 179 tool, in the form of a publicly available web application accessible on desktops, tablets, and
21 180 smartphones.
22
23
24

25 181 **Patients and public involvement**

26
27 182 Patients and the public were not involved in the making of this protocol.
28

29 183 **Statistical analysis**

30
31 184 Categorical variables will be described as absolute numbers with frequencies, and continuous
32 185 variables as medians with interquartile ranges (IQR). The model performance metrics will be
33 186 calculated with 95% confidence interval (CI). Given the retrospective study design, post hoc power
34 187 analyses will be conducted to evaluate the sample size of the study with an alpha value of 0.05.
35
36
37

38 188 **Guidelines**

39
40
41 189 The study set-up will be performed following the Transparent Reporting of Multivariable Prediction
42 190 Models for Individual Prognosis or Diagnosis Guideline for Clustered data (TRIPOD-Cluster).[15]
43

44 191 **ETHICS AND DISSEMINATION**

45
46
47 192 For safe multicentre data exchange and analysis, our Machine Learning Consortium will adhere to the
48 193 World Health Organisation (WHO) regulation ‘Policy on Use and Sharing of Data Collected by WHO
49 194 in Member States Outside the Context of Public Health Emergencies’.[28] The study results will be
50 195 disseminated through publication in a peer-reviewed journal. An Institutional Review Board (IRB)
51 196 approval has been obtained for each of the included studies and the provided data are anonymised and
52 197 de-identified, no IRB review is required for this study. Patient consent for publication is not
53 198 applicable to this study.
54
55
56
57

58 199 **CURRENT STATUS**

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

202 **DISCUSSION**

203 For an informed decision when considering ATSA, it is important to be able to make an accurate
204 prediction of arthroplasty failure. Previous studies have demonstrated several factors that affect
205 complications and revision after ATSA, including male sex, comorbidities such as chronic kidney
206 disease or coronary artery disease, percentage haematocrit, a higher critical shoulder angle, teaching
207 hospital status, operative time, and the material and design of the prosthesis.[2,3,5–9,13,14]
208 Psychological studies have shown that in human judgement only a limited amount of variables can be
209 taken into account, and that prediction models are generally more accurate and less subject to
210 bias.[29] Machine learning algorithms have been shown to be an effective method in developing
211 patient-specific prediction tools, which may complement human judgement when counselling patients
212 in clinic.[30] Creating an online tool for aseptic loosening of the glenoid component after ATSA can
213 help guide surgeons in selecting patients who will most benefit from this treatment, and considering
214 alternatives in cases of high risk estimates.

215 The strength of this project is the large amount of data that will be gathered from authors participating
216 in the Machine Learning Consortium, aiming to include a minimum of 5000 patients in total. Using a
217 large, heterogenous international database for development of the algorithm and prediction tool will
218 result in high external validity and may improve applicability world-wide.[15] However, in gathering
219 data retrospectively from various sources, the study is subject to variances in the included variables.
220 Low completeness and large variability of reporting may introduce bias. However, only variables that
221 are consistently reported in multiple data sources will be included in the final analysis, variables will
222 be categorised to increase uniformity and missing data will be imputed where possible. The exclusion
223 and categorisation of variables will have to be balanced with the amount of detail in the final analysis.
224 Furthermore, the accuracy of data collection is dependent on third parties providing the data, the
225 method of data collection cannot be verified for all sources. However, the data source will be
226 considered as a confounder. Furthermore, the variety in data sources will increase the external
227 applicability of the algorithm.

228

229 **REFERENCES**

- 230 1 Sharareh B, Yao JJ, Matsen FA, *et al*. What is the Optimal Management of a Loose Glenoid
231 Component after Anatomic Total Shoulder Arthroplasty: A Systematic Review. *J shoulder Elb*
232 *Surg* 2022 2023;**32(3)**:653-661. doi:10.1016/J.JSE.2022.10.006
- 233 2 Tabeayo E, Chan PH, Prentice HA, *et al*. The association between critical shoulder angle and

- 1
2
3 234 revision following anatomic total shoulder arthroplasty: a matched case-control study. *J*
4 *shoulder Elb Surg* 2022;**31**:1796–802. doi:10.1016/J.JSE.2021.08.003
5 235
6
7 236 3 Aibinder W, Schoch B, Parsons M, *et al.* Risk factors for complications and revision surgery
8 237 after anatomic and reverse total shoulder arthroplasty. *J Shoulder Elb Surg* 2021;**30**:e689–701.
9 238 doi:10.1016/J.JSE.2021.04.029
10
11
12 239 4 Dekker TJ, Grantham WJ, Lacheta L, *et al.* Glenoid retroversion does not impact clinical
13 240 outcomes or implant survivorship after total shoulder arthroplasty with minimal, noncorrective
14 241 reaming. *JSES Int* 2022;**6**:596–603. doi:10.1016/J.JSEINT.2022.02.011
15
16
17 242 5 Page RS, Alder-Price AC, Rainbird S, *et al.* Reduced Revision Rates in Total Shoulder
18 243 Arthroplasty With Crosslinked Polyethylene: Results From the Australian Orthopaedic
19 244 Association National Joint Replacement Registry. *Clin Orthop Relat Res* 2022;**480**:1940–9.
20 245 doi:10.1097/CORR.0000000000002293
21
22
23 246 6 Dillon MT, Chan PH, Prentice HA, *et al.* The association between glenoid component design
24 247 and revision risk in anatomic total shoulder arthroplasty. *J shoulder Elb Surg* 2020;**29**:2089–
25 248 96. doi:10.1016/J.JSE.2020.02.024
26
27
28 249 7 Welsher A, Gohal C, Madden K, *et al.* A comparison of pegged vs. keeled glenoid components
29 250 regarding functional and radiographic outcomes in anatomic total shoulder arthroplasty: a
30 251 systematic review and meta-analysis. *JSES open access* 2019;**3**:136-144.e1.
31 252 doi:10.1016/J.JSES.2019.04.002
32
33
34 253 8 Page RS, Pai V, Eng K, *et al.* Cementless versus cemented glenoid components in
35 254 conventional total shoulder joint arthroplasty: analysis from the Australian Orthopaedic
36 255 Association National Joint Replacement Registry. *J shoulder Elb Surg* 2018;**27**:1859–65.
37 256 doi:10.1016/J.JSE.2018.03.017
38
39
40 257 9 Sharplin PK, Frampton CMA, Hirner M. Cemented vs. uncemented glenoid fixation in total
41 258 shoulder arthroplasty for osteoarthritis: a New Zealand Joint Registry study. *J shoulder Elb*
42 259 *Surg* 2020;**29**:2097–103. doi:10.1016/J.JSE.2020.03.008
43
44
45 260 10 Kumar V, Roche C, Overman S, *et al.* What Is the Accuracy of Three Different Machine
46 261 Learning Techniques to Predict Clinical Outcomes After Shoulder Arthroplasty? *Clin Orthop*
47 262 *Relat Res* 2020;**478**:2351–63. doi:10.1097/CORR.0000000000001263
48
49
50 263 11 McLendon PB, Christmas KN, Simon P, *et al.* Machine Learning Can Predict Level of
51 264 Improvement in Shoulder Arthroplasty. *JBJS Open Access* 2021;**6**.
52 265 doi:10.2106/JBJS.OA.20.00128
53
54
55
56
57
58
59
60

- 1
2
3 266 12 Polce EM, Kunze KN, Fu MC, *et al.* Development of supervised machine learning algorithms
4 267 for prediction of satisfaction at 2 years following total shoulder arthroplasty. *J Shoulder Elb*
5 268 *Surg* 2021;**30**:e290–9. doi:10.1016/j.jse.2020.09.007
6
7
8 269 13 Devana SK, Shah AA, Lee C, *et al.* Development of a Machine Learning Algorithm for
9 270 Prediction of Complications and Unplanned Readmission Following Primary Anatomic Total
10 271 Shoulder Replacements. *J Shoulder Elb Arthroplast* 2022;**6**:247154922210754.
11 272 doi:10.1177/24715492221075444
12
13
14
15 273 14 Gowd AK, Agarwalla A, Amin NH, *et al.* Construct validation of machine learning in the
16 274 prediction of short-term postoperative complications following total shoulder arthroplasty. *J*
17 275 *Shoulder Elb Surg* 2019;**28**:e410–21. doi:10.1016/J.JSE.2019.05.017
18
19
20
21 276 15 Debray TPA, Collins GS, Riley RD, *et al.* Transparent reporting of multivariable prediction
22 277 models developed or validated using clustered data: TRIPOD-Cluster checklist. *BMJ*
23 278 2023;**380**:e071018. doi:10.1136/BMJ-2022-071018
24
25
26 279 16 Papadonikolakis A, Neradilek MB, Matsen FA. Failure of the glenoid component in anatomic
27 280 total shoulder arthroplasty :A systematic review of the english-language literature between
28 281 2006 and 2012. *J Bone Jt Surg* 2013;**95**:2205–12. doi:10.2106/JBJS.L.00552
29
30
31
32 282 17 van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations
33 283 in R. *J Stat Softw* 2011;**45**:1–67. doi:10.18637/JSS.V045.I03
34
35
36 284 18 DeBaun MR, Chavez G, Fithian A, *et al.* Artificial Neural Networks Predict 30-Day Mortality
37 285 After Hip Fracture: Insights From Machine Learning. *J Am Acad Orthop Surg* 2021;**29**:977–
38 286 83. doi:10.5435/JAAOS-D-20-00429
39
40
41 287 19 Riley RD, Ensor J, Snell KIE, *et al.* Calculating the sample size required for developing a
42 288 clinical prediction model. *BMJ* 2020;**368**. doi:10.1136/BMJ.M441
43
44
45 289 20 Oosterhoff JHF, Savelberg ABMC, Karhade AV, *et al.* Development and internal validation of
46 290 a clinical prediction model using machine learning algorithms for 90 day and 2 year mortality
47 291 in femoral neck fracture patients aged 65 years or above. *Eur J Trauma Emerg Surg*
48 292 2022;**48**:4669–82. doi:10.1007/s00068-022-01981-4
49
50
51
52 293 21 Oosterhoff JHF, Karhade A V., Oberai T, *et al.* Prediction of Postoperative Delirium in
53 294 Geriatric Hip Fracture Patients: A Clinical Prediction Model Using Machine Learning
54 295 Algorithms. *Geriatr Orthop Surg Rehabil* 2021;**12**. doi:10.1177/21514593211062277
55
56
57 296 22 Bongers MER, Thio QCBS, Karhade A V., *et al.* Does the SORG Algorithm Predict 5-year
58 297 Survival in Patients with Chondrosarcoma? An External Validation. *Clin Orthop Relat Res*
59
60

- 2019;**477**:2296–303. doi:10.1097/CORR.0000000000000748
- 299 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
- 302 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
- 305 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
- 308 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
- 311 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
- 314 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
- 319 29 Daniel Kahneman, Olivier Sibony, Cass R. Sunstein. Noise: A Flaw in Human Judgment *Little, Brown Spark; Hachette Book Group*, version 1, 2021
- 321 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

324

325 AUTHOR'S CONTRIBUTIONS

326 Authors AM, GB, and MP contributed to the conception, methods, planning, and writing and
327 reviewing of the protocol. LM and JO contributed to the statistical and machine learning methods of
328 the study. JO, PB, GA, JD, LL, and TL reviewed the draft of the protocol and provided supervision.
329 All authors revised the final version of the protocol and gave approval for publication.

1
2
3 **330 ACKNOWLEDGEMENTS**
4

5 331 The authors would like to thank Dr. Laurens Allaart and Dr. Sanne van Spanning for providing
6
7 332 examples and inspiration for this study based on their previous work.
8

9 **333 FUNDING STATEMENT**
10

11 334 This work was supported by Stichting Prof. Michaël-van Vloten fonds, VSBfonds grant number
12
13 335 VSB.22/00098, Prins Bernard Cultuurfonds grant number 40042437/NOD, and Stichting Erasmus
14
15 336 Trustfonds grant number 97030.2021.101.577/193/RB.
16

17 **337 COMPETING INTERESTS STATEMENT**
18

19 338 LL received consultancy fees from Depuy Stryker and royalties from Depuy-Synthes. TL received
20
21 339 consultancy fees from Depuy Mitek and Stryker. GB received consultancy fees from Depuy-Synthes
22
23 340 and Stryker. The remaining authors certify that he or she has received no funding or has commercial
24
25 341 associations that might pose a conflict of interest in connection with the submitted article.
26

27 342
28

29 343
30

31 **344 APPENDIX 1**
32

33 345 Search strategy:
34

35 346 ("Arthroplasty, Replacement, Shoulder"[Mesh] OR "Total shoulder arthroplasty"[tiab] OR "Anatomic
36
37 347 shoulder arthroplasty"[tiab] OR "TSA"[tiab] OR "ATSA"[tiab] OR "Total shoulder prosthes*"[tiab]
38
39 348 OR "Anatomic shoulder prosthes*"[tiab]) AND ("Reoperation"[Mesh] OR "reoperation"[tiab] OR
40
41 349 "re-operation"[tiab] OR "revision"[tiab] OR "survival"[tiab] OR "implant-survival"[tiab] OR
42
43 350 "Prosthesis Failure"[Mesh] OR "failure"[tiab] OR "loosening"[tiab] OR "aseptic"[tiab] OR
44
45 351 "Postoperative Complications"[Mesh] OR "complication"[tiab])
46

47 352 Filters: Clinical study, Clinical trial, Controlled clinical trial, Comparative study, Dataset
48
49
50
51
52
53
54
55
56
57
58
59
60



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 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:	<i>BMJ Open</i>
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

SCHOLARONE™
Manuscripts

1
2
3 1 **TITLE:** Developing a machine learning algorithm to predict the probability of aseptic loosening of
4 2 the glenoid component after anatomic total shoulder arthroplasty: a protocol for a retrospective,
5 3 multicentre study.

6
7
8 4 **AUTHORS:** Arno A. Macken^{1,2}, Loïc C. Macken³, Jacobien H.F. Oosterhoff⁴, Pascal Boileau⁵,
9 5 George S. Athwal⁶, Job N. Doornberg⁷, Laurent Lafosse¹, Thibault Lafosse¹, Michel P.J. van den
10 6 Bekerom^{8,9}, Geert Alexander Buijze^{1,10,11}, on behalf the Machine Learning Consortium.

- 11 7 1. Alps Surgery Institute, Clinique Générale d'Annecy, France
- 12 8 2. Department of Orthopaedic Surgery and Sports medicine, Erasmus Medical Centre,
13 9 Rotterdam, the Netherlands
- 14 10 3. Faculty of Science, Vrije Universiteit Amsterdam, the Netherlands
- 15 11 4. Department of Engineering Systems and Services, Faculty of Technology Policy and
16 12 Management, Delft University of Technology, the Netherlands
- 17 13 5. Institut de Chirurgie Réparatrice, Locomoteur & Sports, Centre Hospitalier Universitaire de
18 14 Nice, France
- 19 15 6. Roth McFarlane Hand and Upper Limb Center, Schulich School of Medicine and Dentistry,
20 16 London, Ontario, Canada
- 21 17 7. Departments of Orthopaedic and Trauma Surgery, University Medical Center Groningen, the
22 18 Netherlands
- 23 19 8. Department of Human Movement Sciences, Faculty of Behavioural and Movement Sciences,
24 20 Vrije Universiteit Amsterdam, Amsterdam Movement Sciences, Amsterdam, the Netherlands
- 25 21 9. Shoulder and Elbow Unit, Joint Research, Department of Orthopaedic Surgery, OLVG,
26 22 Amsterdam, The Netherlands
- 27 23 10. Department of Orthopedic Surgery, Amsterdam Movement Sciences, Amsterdam UMC,
28 24 location AMC, University of Amsterdam, Amsterdam, The Netherlands
- 29 25 11. Department of Orthopedic Surgery, Montpellier University Medical Center, Lapeyronie
30 26 Hospital, University of Montpellier, Montpellier, France

31 27 Corresponding author: Arno A. Macken

32 28 4 Chem. de la Tour la Reine, 74000 Annecy, France

33 29 +33 4 50 33 04 69

34 30 arnomacken@gmail.com

35
36
37 32 **MESH KEYWORDS:** “Arthroplasty, Replacement, Shoulder”, “Shoulder Prosthesis”, “Glenoid
38 33 Cavity”, “Prosthesis Failure”, “Reoperation”, “Artificial Intelligence”, “Machine Learning”

39
40
41
42
43
44
45
46
47
48
49
50
51 34
52
53 35 **WORD COUNT:** 2972
54
55 36
56
57 37
58
59
60 38

39 **ABSTRACT**

40 **Introduction**

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

49 **Methods and analysis**

50 For this multi-centre retrospective study, individual patient data will be compiled from previously
51 published studies reporting revision of ATSA. A systematic literature search will be performed in
52 Medline (PubMed) identifying all studies reporting outcomes of ATSA. Authors will be contacted and
53 invited to participate in the Machine Learning Consortium by sharing their anonymised databases. All
54 databases reporting revisions after ATSA will be included and individual patients with a follow-up
55 less than 2 years or a fracture as the indication for ATSA will be excluded. First, features (predictive
56 variables) will be identified using a random forest feature selection. The resulting features from the
57 compiled database will be used to train various machine learning algorithms (Stochastic Gradient
58 Boosting, Random Forest, Support Vector Machine, Neural Network and Elastic-Net Penalized
59 Logistic Regression). The developed and validated algorithms will be evaluated across discrimination
60 (c-statistic), calibration, the Brier score, and the decision curve analysis. The best-performing
61 algorithm will be used to create an open-access online prediction tool.

62 **Ethics and dissemination**

63 Data will be collected adhering to the World Health Organisation (WHO) regulation on data sharing.
64 The study will be published in a peer-reviewed journal. An Institutional Review Board (IRB) review
65 is not applicable.

66

67 **STRENGTHS AND LIMITATIONS OF THIS STUDY**

- 68 - A large international database will be collected, which increases accuracy, validity, and
69 external applicability.
- 70 - A clinical prediction model using machine learning algorithms will be developed to estimate
71 the probability of aseptic loosening of the glenoid after ATSA.

- 1
2
3 72 - An open-access prediction tool based on the best performing algorithm will be made available
4 73 online that can guide medical professionals in personalised treatment decision-making.
5
6 74 - The study is dependent on data provided by third parties, which is a potential source of bias.
7
8 75 - Input variables will be selected and categorised based on completeness and uniformity across
9 76 data sources, potentially decreasing the amount of detail in the study.
10
11
12 77

14 78 **INTRODUCTION**

16 79 Anatomic total shoulder arthroplasty (ATSA) is used for glenohumeral arthropathy causing pain
17 80 and/or a reduction in range of motion. Despite technological advancements in recent years, glenoid
18 81 component loosening remains a common complication after ATSA and is one of the main causes for
19 82 revision surgery. Glenoid loosening can be a trying complication to manage and the optimal course of
20 83 treatment remains unclear.[1]

24 84 Consequently, increasing emphasis is placed on the prevention of loosening. The predicted chance of
25 85 glenoid component failure plays an important role in clinical decision-making such as patient
26 86 selection for ATSA or which implants and techniques to use. Several previous studies assessing risk
27 87 factors of glenoid component loosening identified patient, treatment and prosthesis characteristics
28 88 related to glenoid component loosening. For example, male sex and a higher critical shoulder angle
29 89 have been associated with higher rates of loosening and revisions.[2,3] Glenoid retroversion did not
30 90 impact implant survivorship in one study of ATSA with minimal, noncorrective reaming.[4]
31 91 However, a larger degree of retroversion may have more impact. Several previous studies have also
32 92 identified aspects of the glenoid component design that correlated with the rate of loosening, such as
33 93 whether the poly-ethylene is cross-linked, whether the component is pegged or keeled, or the usage of
34 94 cement.[5–9] In spite of these studies identifying influential factors, accurate prediction of aseptic
35 95 loosening of the glenoid component remains a challenge with conventional methods.

36 96 In recent years, machine learning or artificial intelligence has been used with increasing precision to
37 97 predict outcomes after ATSA. A previous study using machine learning was able to accurately predict
38 98 range of motion and patient-reported outcomes after ATSA. The most influential factors they reported
39 99 were follow-up time, pre-operative range of motion and patient-reported outcome measures (PROMs),
40 100 patients' sex, and surgery on the dominant upper limb.[10] Another study was able to accurately
41 101 predict the improvement in ASES scores after shoulder arthroplasty using machine learning.[11] The
42 102 most relevant predictive factors were pre-operative ASES scores, pre-operative pain scores, Walch
43 103 classification, fatty infiltration in the supra- and infraspinatus, and age. A previous study using
44 104 artificial intelligence to predict patient satisfaction two years after shoulder arthroplasty found
45 105 baseline Single Assessment Numeric Evaluation score, exercise and activity, workers' compensation
46 106 status, diagnosis, symptom duration prior to surgery, body mass index, age, smoking status, anatomic

1
2
3 107 versus reverse TSA, and diabetes to be predicting factors.[12] Two studies report predictive models
4 108 on short-term complications after ATSA. One study using machine learning to predict complications
5 109 and 30-day unplanned readmissions found that a history of implant complication, severe chronic
6 110 kidney disease, teaching hospital status, coronary artery disease and male sex were the most important
7 111 features.[13] The machine learning model found teaching hospital status and male sex as a markedly
8 112 more important predictor compared to a logistic regression analysis of the same data. Another study
9 113 on short-term complications after total shoulder arthroplasty found percentage haematocrit, BMI, and
10 114 operative time were of highest importance in outcome prediction.[14] These studies demonstrate that
11 115 machine learning may provide accurate predictions for the outcomes after ATSA. Machine learning is
12 116 most effective with large amounts of data and is very dependent on the amount of detail. Furthermore,
13 117 the algorithm needs to be widely applicable; a varied and international database provides the highest
14 118 external validity.

15
16 119 To our knowledge, there are no studies predicting the long-term complications such as aseptic
17 120 loosening of the glenoid component using advanced machine learning techniques. Furthermore,
18 121 previous machine learning studies are limited in accuracy and validity due to the sample size and
19 122 homogeneity. Therefore, this study aims to develop a clinical prediction model for aseptic loosening
20 123 of the glenoid component using machine learning algorithms trained on a large international database
21 124 using clustered data. The large combined dataset is less prone to overfitting, and allows direct
22 125 validation of models across a range of populations and settings, thereby increasing
23 126 generalisability.[15] The predictive algorithm will be made available for clinical use through a
24 127 publicly available online prediction tool.

25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

129 **METHODS AND ANALYSIS**

130 **Data collection**

131 For this multi-centre retrospective study individual patient data (IPD) will be collected from
132 previously published studies reporting failure and revision of ATSA. A systematic literature search
133 will be performed in Medline (PubMed) identifying all studies that report a cohort of ATSA including
134 revision as an outcome, published between January 2000 and June 2023. The full search term is
135 available in appendix 1. All original studies reporting revision or failure rates after primary ATSA
136 will be included. Reviews and letters to the editor will be excluded, as well as studies published in
137 languages other than English, Italian, Dutch, and French. Authors will be requested to share the
138 anonymised databases used for the identified studies. Only de-identified databases used for previous
139 studies are included, authors are not required to gather additional data or access patient files. After
140 sharing their data, the authors will be included in in the Machine Learning Consortium. Inclusion
141 criteria for individual patients within the provided databases are a minimum age of 18 years and a

1
2
3 142 minimum follow-up of 2 years. Patients that underwent ATSA with a fracture as the indication or
4 143 patients that underwent concomitant procedures such as a cuff repair, tendon transfer or bone graft
5 144 will be excluded. The aim is to combine the IPD from previously published studies to create a large
6 145 international cohort which can be used to train a machine learning algorithm to predict aseptic
7 146 loosening of the glenoid component after anatomic total shoulder arthroplasty. Based on previous
8 147 studies, we estimate a glenoid revision rate of approximately 2%. [6,16] The minimum number of
9 148 events per variable to achieve sufficient accuracy differs per model and is not clearly defined for each
10 149 technique. [17,18] We aim to include at least 30 events per variable, resulting in a sample size of 7500
11 150 patients for a model with up to 5 predictive variables.

151 **Data curation and missing data**

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

163 **Variable Selection**

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

174 **Development of prediction models**

1
2
3 175 Different ML models result in varying performance metrics based on the type of input data
4 (continuous, categorical, dichotomous). Due to the variation in type of input variables in the dataset,
5 176
6 177 several different ML techniques will be used and compared based on model performance. The
7
8 178 following machine learning algorithms were chosen for modelling based on prior research [23–27]:
9
10 179 Stochastic Gradient Boosting (SGM), Random Forest (RF), Support Vector Machine (SVM), Neural
11
12 180 Network (NN) and Elastic-Net Penalized Logistic Regression (PLR). The algorithms will be trained
13
14 181 on the training dataset with ten-fold cross-validation repeated 3 times. Cross-validation means
15
16 182 dividing data into a selected number of groups, also called folds. First, the data will be divided into 10
17
18 183 equally sized folds. Then, the algorithms will be trained on 9 of the 10 folds (90% of the training data)
19
20 184 and tested on the remaining fold (10% of the training data). Results will be averaged across all
21
22 185 repetitions of this sequence. Machine learning algorithms will be developed using Python (The
23
24 186 Python Software Foundation, Fredericksburg, United States of America). Hyper parameter tuning will
25
26 187 be performed as recommended in the Python libraries. The statistician who performs the machine
27
28 188 learning analysis will be blinded to the origin of the data, but the anonymised data source will be
29
30 189 available to be included as a potential confounding factor.

190 **Model Performance**

31
32 191 After training all models, the model performance will be analysed according to a proposed framework
33
34 192 by Steyerberg et al. including discrimination with the c-statistic, positive predictive value (PPV), true
35
36 193 positive rate (TPR), precision-recall curve, calibration slope and intercept and the overall performance
37
38 194 with the Brier score.[28]

39
40 195 The c-statistic (area under the curve of a receiver operating characteristic curve) is a score ranging
41
42 196 from 0.50 to 1.0 with 1.0 indicating the highest discrimination score and 0.50 indicating the lowest.
43
44 197 The higher the discrimination score, the better the model's ability to distinguish patients with and
45
46 198 without the outcome of interest.[20,23] The PPV is the proportion of true positive outcomes over the
47
48 199 number of predicted positive outcomes. The TPR is the proportion of true positive outcomes over the
49
50 200 number of observed positive outcomes. The precision recall curve is a plot of the PPV versus the
51
52 201 TPR. A calibration plot plots the estimated versus the observed probabilities for the primary outcome.
53
54 202 A perfect calibration plot has an intercept of 0 (<0 reflects overestimation and >0 reflects
55
56 203 underestimation of the probability of the outcome) and a slope of 1 (model is performing similarly in
57
58 204 training and test datasets).[20,23,28] The null-model Brier score, which equals the probability of
59
60 205 glenoid revision in the dataset, will be used to benchmark the algorithm's Brier score. A Brier score
206
207 206 lower than the null-model Brier score indicates superior performance of the prediction model to this
208
209 207 null benchmark. Perfect prediction would have a Brier score of 0, whereas a Brier score of 1 would
210
211 208 indicate the poorest possible prediction.[28]

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

215 Due to the large heterogeneity of the compiled dataset from different international sources and the
216 internal validation of the prediction models, the generalisability of the model can be intrinsically
217 confirmed using the above-mentioned performance tests. Therefore, it is not strictly necessary to
218 externally validate the final algorithm. However, this study's primary aim is model development.
219 External validation in a specific setting is advised before applying the algorithm to clinical practice.

220 **Open-access clinical prediction tool**

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

224 **Patients and public involvement**

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

226 **Statistical analysis**

227 Categorical variables will be described as absolute numbers with frequencies, and continuous
228 variables as medians with interquartile ranges (IQR). The model performance metrics will be
229 calculated with 95% confidence interval (CI). Given the retrospective study design, post hoc power
230 analyses will be conducted to evaluate the sample size of the study with an alpha value of 0.05.

231 **Guidelines**

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

234 **ETHICS AND DISSEMINATION**

235 For safe multicentre data exchange and analysis, our Machine Learning Consortium will adhere to the
236 World Health Organisation (WHO) regulation 'Policy on Use and Sharing of Data Collected by WHO
237 in Member States Outside the Context of Public Health Emergencies'.[30] The study results will be
238 disseminated through publication in a peer-reviewed journal. An Institutional Review Board (IRB)
239 approval has been obtained for each of the included studies and the provided data are anonymised and
240 de-identified, no additional prospective data is collected and contributing authors are not required to

1
2
3 241 access any patient files, no IRB review is required for this study. Patient consent for publication is not
4
5 242 applicable to this study.

6 7 243 **CURRENT STATUS**

8
9 244 Data collection for this project is currently ongoing. The analysis will start in December 2023. The
10
11 245 expected time of completion for the project is July of 2024.

12 13 246 **DISCUSSION**

14
15 247 For an informed decision when considering ATSA, it is important to be able to make an accurate
16
17 248 prediction of arthroplasty failure. Previous studies have demonstrated several factors that affect
18
19 249 complications and revision after ATSA, including male sex, comorbidities such as chronic kidney
20
21 250 disease or coronary artery disease, percentage haematocrit, a higher critical shoulder angle, teaching
22
23 251 hospital status, operative time, and the material and design of the prosthesis.[2,3,5–9,13,14]
24
25 252 Psychological studies have shown that in human judgement only a limited amount of variables can be
26
27 253 taken into account, and that prediction models are generally more accurate and less subject to
28
29 254 bias.[31] Machine learning algorithms have been shown to be an effective method in developing
30
31 255 patient-specific prediction tools, which may complement human judgement when counselling patients
32
33 256 in clinic.[32] Creating an online tool for aseptic loosening of the glenoid component after ATSA can
34
35 257 help guide surgeons in selecting patients who will most benefit from this treatment, and considering
36
37 258 alternatives in cases of high risk estimates.

38
39 259 The strength of this project is the large amount of data that will be gathered from authors participating
40
41 260 in the Machine Learning Consortium, aiming to include a minimum of 7500 patients in total. Using a
42
43 261 large, heterogenous international database for development of the algorithm and prediction tool will
44
45 262 result in high external validity and may improve applicability world-wide.[15] However, most ML
46
47 263 techniques require a larger sample size to achieve an accurate prediction compared to traditional
48
49 264 regression models. The minimum events per variable is not clearly defined and differs per technique.
50
51 265 Furthermore, in gathering data retrospectively from various sources, the study is subject to variances
52
53 266 in the included variables. Low completeness and large variability of reporting may introduce bias.
54
55 267 However, only variables that are consistently reported in multiple data sources will be included in the
56
57 268 final analysis, variables will be categorised to increase uniformity and missing data will be imputed
58
59 269 where possible. The exclusion and categorisation of variables will have to be balanced with the
60
270 amount of detail in the final analysis. Furthermore, the accuracy of data collection is dependent on
271 third parties providing the data, the method of data collection cannot be verified for all sources.
272 However, the data source will be considered as a confounder. Furthermore, the variety in data sources
273 will increase the external applicability of the algorithm. Last, ML prediction models for a
274 dichotomous outcome are limited to risk classification, the individual risk must be interpreted in the
275 clinical context when used for medical decision making.

276

277 REFERENCES

- 278 1 Sharareh B, Yao JJ, Matsen FA, *et al.* What is the Optimal Management of a Loose Glenoid
279 Component after Anatomic Total Shoulder Arthroplasty: A Systematic Review. *J shoulder Elb*
280 *Surg* Published Online First: November 2022. doi:10.1016/J.JSE.2022.10.006
- 281 2 Tabeayo E, Chan PH, Prentice HA, *et al.* The association between critical shoulder angle and
282 revision following anatomic total shoulder arthroplasty: a matched case-control study. *J*
283 *shoulder Elb Surg* 2022;**31**:1796–802. doi:10.1016/J.JSE.2021.08.003
- 284 3 Aibinder W, Schoch B, Parsons M, *et al.* Risk factors for complications and revision surgery
285 after anatomic and reverse total shoulder arthroplasty. *J Shoulder Elb Surg* 2021;**30**:e689–701.
286 doi:10.1016/J.JSE.2021.04.029
- 287 4 Dekker TJ, Grantham WJ, Lacheta L, *et al.* Glenoid retroversion does not impact clinical
288 outcomes or implant survivorship after total shoulder arthroplasty with minimal, noncorrective
289 reaming. *JSES Int* 2022;**6**:596–603. doi:10.1016/J.JSEINT.2022.02.011
- 290 5 Page RS, Alder-Price AC, Rainbird S, *et al.* Reduced Revision Rates in Total Shoulder
291 Arthroplasty With Crosslinked Polyethylene: Results From the Australian Orthopaedic
292 Association National Joint Replacement Registry. *Clin Orthop Relat Res* 2022;**480**:1940–9.
293 doi:10.1097/CORR.0000000000002293
- 294 6 Dillon MT, Chan PH, Prentice HA, *et al.* The association between glenoid component design
295 and revision risk in anatomic total shoulder arthroplasty. *J shoulder Elb Surg* 2020;**29**:2089–
296 96. doi:10.1016/J.JSE.2020.02.024
- 297 7 Welsher A, Gohal C, Madden K, *et al.* A comparison of pegged vs. keeled glenoid components
298 regarding functional and radiographic outcomes in anatomic total shoulder arthroplasty: a
299 systematic review and meta-analysis. *JSES open access* 2019;**3**:136-144.e1.
300 doi:10.1016/J.JSES.2019.04.002
- 301 8 Page RS, Pai V, Eng K, *et al.* Cementless versus cemented glenoid components in
302 conventional total shoulder joint arthroplasty: analysis from the Australian Orthopaedic
303 Association National Joint Replacement Registry. *J shoulder Elb Surg* 2018;**27**:1859–65.
304 doi:10.1016/J.JSE.2018.03.017
- 305 9 Sharplin PK, Frampton CMA, Hirner M. Cemented vs. uncemented glenoid fixation in total
306 shoulder arthroplasty for osteoarthritis: a New Zealand Joint Registry study. *J shoulder Elb*
307 *Surg* 2020;**29**:2097–103. doi:10.1016/J.JSE.2020.03.008

- 1
2
3 308 10 Kumar V, Roche C, Overman S, *et al.* What Is the Accuracy of Three Different Machine
4 Learning Techniques to Predict Clinical Outcomes After Shoulder Arthroplasty? *Clin Orthop*
5 309 *Relat Res* 2020;**478**:2351–63. doi:10.1097/CORR.0000000000001263
6 310
7
8 311 11 McLendon PB, Christmas KN, Simon P, *et al.* Machine Learning Can Predict Level of
9 312 Improvement in Shoulder Arthroplasty. *JBJS Open Access* 2021;**6**.
10 313 doi:10.2106/JBJS.OA.20.00128
11
12
13
14 314 12 Polce EM, Kunze KN, Fu MC, *et al.* Development of supervised machine learning algorithms
15 315 for prediction of satisfaction at 2 years following total shoulder arthroplasty. *J Shoulder Elb*
16 316 *Surg* 2021;**30**:e290–9. doi:10.1016/j.jse.2020.09.007
17
18
19 317 13 Devana SK, Shah AA, Lee C, *et al.* Development of a Machine Learning Algorithm for
20 318 Prediction of Complications and Unplanned Readmission Following Primary Anatomic Total
21 319 Shoulder Replacements. <https://doi.org/10.1177/24715492221075444>
22 320 2022;**6**:247154922210754. doi:10.1177/24715492221075444
23
24
25
26 321 14 Gowd AK, Agarwalla A, Amin NH, *et al.* Construct validation of machine learning in the
27 322 prediction of short-term postoperative complications following total shoulder arthroplasty. *J*
28 323 *Shoulder Elb Surg* 2019;**28**:e410–21. doi:10.1016/J.JSE.2019.05.017
29
30
31
32 324 15 Debray TPA, Collins GS, Riley RD, *et al.* Transparent reporting of multivariable prediction
33 325 models developed or validated using clustered data: TRIPOD-Cluster checklist. *BMJ*
34 326 2023;**380**:e071018. doi:10.1136/BMJ-2022-071018
35
36
37 327 16 Papadonikolakis A, Neradilek MB, Matsen FA. Failure of the glenoid component in anatomic
38 328 total shoulder arthroplasty :A systematic review of the english-language literature between
39 329 2006 and 2012. *J Bone Jt Surg* 2013;**95**:2205–12. doi:10.2106/JBJS.L.00552
40
41
42
43 330 17 Van Der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry:
44 331 A simulation study for predicting dichotomous endpoints. *BMC Med Res Methodol* 2014;**14**:1–
45 332 13. doi:10.1186/1471-2288-14-137/FIGURES/13
46
47
48 333 18 Steyerberg EW. Clinical Prediction Models; A Practical Approach to Development,
49 334 Validation, and Updating. Springer 2019
50
51
52 335 19 van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations
53 336 in R. *J Stat Softw* 2011;**45**:1–67. doi:10.18637/JSS.V045.I03
54
55
56 337 20 Oosterhoff JHF, Gravesteijn BY, Karhade A V., *et al.* Feasibility of Machine Learning and
57 338 Logistic Regression Algorithms to Predict Outcome in Orthopaedic Trauma Surgery. *J Bone Jt*
58 339 *Surg - Am Vol* 2022;**104**:544–51. doi:10.2106/JBJS.21.00341
59
60

- 1
2
3 340 21 DeBaun MR, Chavez G, Fithian A, *et al.* Artificial Neural Networks Predict 30-Day Mortality
4 341 After Hip Fracture: Insights From Machine Learning. *J Am Acad Orthop Surg* 2021;**29**:977–
5 342 83. doi:10.5435/JAAOS-D-20-00429
6
7
8 343 22 Riley RD, Ensor J, Snell KIE, *et al.* Calculating the sample size required for developing a
9 344 clinical prediction model. *BMJ* 2020;**368**. doi:10.1136/BMJ.M441
10
11
12 345 23 Oosterhoff JHF, Savelberg ABMC, Karhade AV, *et al.* Development and internal validation of
13 346 a clinical prediction model using machine learning algorithms for 90 day and 2 year mortality
14 347 in femoral neck fracture patients aged 65 years or above. *Eur J Trauma Emerg Surg*
15 348 2022;**48**:4669–82. doi:10.1007/s00068-022-01981-4
16
17
18 349 24 Oosterhoff JHF, Karhade A V., Oberai T, *et al.* Prediction of Postoperative Delirium in
19 350 Geriatric Hip Fracture Patients: A Clinical Prediction Model Using Machine Learning
20 351 Algorithms. *Geriatr Orthop Surg Rehabil* 2021;**12**. doi:10.1177/21514593211062277
21
22
23 352 25 Bongers MER, Thio QCBS, Karhade A V., *et al.* Does the SORG Algorithm Predict 5-year
24 353 Survival in Patients with Chondrosarcoma? An External Validation. *Clin Orthop Relat Res*
25 354 2019;**477**:2296–303. doi:10.1097/CORR.0000000000000748
26
27
28 355 26 Karhade A V., Ogink PT, Thio QCBS, *et al.* Development of machine learning algorithms for
29 356 prediction of prolonged opioid prescription after surgery for lumbar disc herniation. *Spine J*
30 357 2019;**19**:1764–71. doi:10.1016/J.SPINEE.2019.06.002
31
32
33 358 27 Karhade A V., Ogink PT, Thio QCBS, *et al.* Discharge Disposition After Anterior Cervical
34 359 Discectomy and Fusion. *World Neurosurg* 2019;**132**:e14–20.
35 360 doi:10.1016/J.WNEU.2019.09.026
36
37
38 361 28 Steyerberg EW, Vickers AJ, Cook NR, *et al.* Assessing the performance of prediction models:
39 362 A framework for traditional and novel measures. *Epidemiology* 2010;**21**:128–38.
40 363 doi:10.1097/EDE.0b013e3181c30fb2
41
42
43 364 29 Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for
44 365 development and an ABCD for validation. *Eur Heart J* 2014;**35**:1925–31.
45 366 doi:10.1093/EURHEARTJ/EHU207
46
47
48 367 30 World Health Organization. Who policy statement on Data Sharing [Internet]. Policy on use
49 368 and sharing of data collected in Member States by the World Health Organization (WHO)
50 369 outside the context of public health emergencies. WHO; [cited 2023Apr13]. Available from:
51 370 [https://www.who.int/docs/default-source/publishing-policies/who-policy-statement-on-data-](https://www.who.int/docs/default-source/publishing-policies/who-policy-statement-on-data-sharing.pdf?sfvrsn=71a41493_0)
52 371 [sharing.pdf?sfvrsn=71a41493_0](https://www.who.int/docs/default-source/publishing-policies/who-policy-statement-on-data-sharing.pdf?sfvrsn=71a41493_0)
53
54
55
56
57
58
59
60

1
2
3 372 31 Daniel Kahneman, Olivier Sibony, Cass R. Sunstein. *Noise: A Flaw in Human Judgment*
4 373 Little, Brown Spark; Hachette Book Group, version 1, 2021

5
6
7 374 32 A Machine Learning Algorithm to Identify Patients with Tibial Shaft Fractures at Risk for
8 375 Infection After Operative Treatment. *J Bone Joint Surg Am* 2021;**103**:532–40.
9 376 doi:10.2106/JBJS.20.00903
10
11

12 377

14 378 **AUTHOR'S CONTRIBUTIONS**

15
16
17 379 Authors AM, GB, and MP contributed to the conception, methods, planning, and writing and
18 380 reviewing of the protocol. LM and JO contributed to the statistical and machine learning methods of
19 381 the study. JO, PB, GA, JD, LL, and TL reviewed the draft of the protocol and provided supervision.
20 382 All authors revised the final version of the protocol and gave approval for publication.
21
22

23 383 **ACKNOWLEDGEMENTS**

24
25
26 384 The authors would like to thank Dr. Laurens Allaart and Dr. Sanne van Spanning for providing
27 385 examples and inspiration for this study based on their previous work.
28
29

30 386 **FUNDING STATEMENT**

31
32 387 This work was supported by Stichting Prof. Michaël-van Vloten fonds, VSBfonds grant number
33 388 VSB.22/00098, Prins Bernard Cultuurfonds grant number 40042437/NOD, and Stichting Erasmus
34 389 Trustfonds grant number 97030.2021.101.577/193/RB.
35
36

37 390 **COMPETING INTERESTS STATEMENT**

38
39
40 391 LL received consultancy fees from Depuy Stryker and royalties from Depuy-Synthes. TL received
41 392 consultancy fees from Depuy Mitek and Stryker. GB received consultancy fees from Depuy-Synthes
42 393 and Stryker. The remaining authors certify that he or she has received no funding or has commercial
43 394 associations that might pose a conflict of interest in connection with the submitted article.
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 **1 APPENDIX 1**
4

5 2 Search strategy:
6

7 ("Arthroplasty, Replacement, Shoulder"[Mesh] OR "Total shoulder arthroplasty"[tiab] OR "Anatomic
8 shoulder arthroplasty"[tiab] OR "TSA"[tiab] OR "ATSA"[tiab] OR "Total shoulder prosthes*"[tiab]
9 OR "Anatomic shoulder prosthes*"[tiab]) AND ("Reoperation"[Mesh] OR "reoperation"[tiab] OR
10 "re-operation"[tiab] OR "revision"[tiab] OR "survival"[tiab] OR "implant-survival"[tiab] OR
11 "Prosthesis Failure"[Mesh] OR "failure"[tiab] OR "loosening"[tiab] OR "aseptic"[tiab] OR
12 "Postoperative Complications"[Mesh] OR "complication"[tiab])
13
14
15
16
17

18 9 Filters: Clinical study, Clinical trial, Controlled clinical trial, Comparative study, Dataset
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



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 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:	<i>BMJ Open</i>
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

SCHOLARONE™
Manuscripts

1
2
3 **1 Developing a machine learning algorithm to predict the probability of aseptic loosening of the**
4 **glenoid component after anatomic total shoulder arthroplasty: protocol for a retrospective,**
5 **2**
6 **3 multicentre study**
7

8 Arno A. Macken^{1,2}, Loïc C. Macken³, Jacobien H.F. Oosterhoff⁴, Pascal Boileau⁵, George S. Athwal⁶,
9 Job N. Doornberg⁷, Laurent Lafosse¹, Thibault Lafosse¹, Michel P.J. van den Bekerom^{8,9}, Geert
10 Alexander Buijze^{1,10,11}, on behalf the Machine Learning Consortium
11
12
13

- 14 7 1. Alps Surgery Institute, Clinique Générale d'Annecy, France
15 8 2. Department of Orthopaedic Surgery and Sports medicine, Erasmus Medical Centre,
16 9 Rotterdam, Netherlands
17 10 3. Faculty of Science, Vrije Universiteit Amsterdam, Netherlands
18 11 4. Department of Engineering Systems and Services, Faculty of Technology Policy and
19 12 Management, Delft University of Technology, Netherlands
20 13 5. Institut de Chirurgie Réparatrice, Locomoteur & Sports, Centre Hospitalier Universitaire de
21 14 Nice, France
22 15 6. Roth McFarlane Hand and Upper Limb Center, Schulich School of Medicine and Dentistry,
23 16 London, Ontario, Canada
24 17 7. Departments of Orthopaedic and Trauma Surgery, University Medical Center Groningen,
25 18 Netherlands
26 19 8. Department of Human Movement Sciences, Faculty of Behavioural and Movement Sciences,
27 20 Vrije Universiteit Amsterdam, Amsterdam Movement Sciences, Amsterdam, Netherlands
28 21 9. Shoulder and Elbow Unit, Joint Research, Department of Orthopaedic Surgery, OLVG,
29 22 Amsterdam, Netherlands
30 23 10. Department of Orthopedic Surgery, Amsterdam Movement Sciences, Amsterdam UMC,
31 24 location AMC, University of Amsterdam, Amsterdam, Netherlands
32 25 11. Department of Orthopedic Surgery, Montpellier University Medical Center, Lapeyronie
33 26 Hospital, University of Montpellier, Montpellier, France

35
36 27 Correspondence to:

37 28 Arno A. Macken

38 29 4 Chem. de la Tour la Reine, 74000 Annecy, France

39 30 +33 4 50 33 04 69

40 31 arnomacken@gmail.com
41 32

42
43 33 **KEYWORDS:** “Arthroplasty, Replacement, Shoulder”, “Shoulder Prosthesis”, “Glenoid Cavity”,
44 34 “Prosthesis Failure”, “Reoperation”, “Artificial Intelligence”, “Machine Learning”
45 35
46 36

47 37
48 38
49 36 **WORD COUNT:** 3028
50 37
51 38

52 39
53 40
54 38 **ABSTRACT**
55 39

56 39 **Introduction**
57 40

58 40 Despite technological advancements in recent years, glenoid component loosening remains a common
59 41 complication after anatomic total shoulder arthroplasty (ATSA) and is one of the main causes for
60 41

1
2
3 42 revision surgery. Increasing emphasis is placed on the prevention of glenoid component failure.
4
5 43 Previous studies have successfully predicted range of motion, patient-reported outcomes, and short-
6
7 44 term complications after ATSA using machine learning methods, but an accurate predictive model for
8
9 45 (glenoid component) revision is currently lacking. This study aims to use a large international
10
11 46 database to accurately predict aseptic loosening of the glenoid component after ATSA using machine
12
13 47 learning algorithms.

13 48 **Methods and analysis**

15 49 For this multi-centre retrospective study, individual patient data will be compiled from previously
16
17 50 published studies reporting revision of ATSA. A systematic literature search will be performed in
18
19 51 Medline (PubMed) identifying all studies reporting outcomes of ATSA. Authors will be contacted and
20
21 52 invited to participate in the Machine Learning Consortium by sharing their anonymised databases. All
22
23 53 databases reporting revisions after ATSA will be included and individual patients with a follow-up
24
25 54 less than 2 years or a fracture as the indication for ATSA will be excluded. First, features (predictive
26
27 55 variables) will be identified using a random forest feature selection. The resulting features from the
28
29 56 compiled database will be used to train various machine learning algorithms (Stochastic Gradient
30
31 57 Boosting, Random Forest, Support Vector Machine, Neural Network and Elastic-Net Penalized
32
33 58 Logistic Regression). The developed and validated algorithms will be evaluated across discrimination
34
35 59 (c-statistic), calibration, the Brier score, and the decision curve analysis. The best-performing
36
37 60 algorithm will be used to create an open-access online prediction tool.

35 61 **Ethics and dissemination**

37 62 Data will be collected adhering to the World Health Organisation (WHO) regulation on data sharing.
38
39 63 An Institutional Review Board (IRB) review is not applicable. The study results will be published in a
40
41 64 peer-reviewed journal.

45 66 **STRENGTHS AND LIMITATIONS OF THIS STUDY**

- 47 67 - A large international database will be collected, which increases accuracy, validity, and
48
49 68 external applicability.
50
51 69 - A clinical prediction model using machine learning algorithms will be developed to estimate
52
53 70 the probability of aseptic loosening of the glenoid after anatomic total shoulder arthroplasty.
54
55 71 - An open-access prediction tool based on the best performing algorithm will be made available
56
57 72 online that can guide medical professionals in personalised treatment decision-making.
58
59 73 - The study is dependent on data provided by third parties, which is a potential source of bias.
60
61 74 - Input variables will be selected and categorised based on completeness and uniformity across
62
63 75 data sources, potentially decreasing the amount of detail in the study.

76

77 INTRODUCTION

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

83 Consequently, increasing emphasis is placed on the prevention of loosening. The predicted chance of
84 glenoid component failure plays an important role in clinical decision-making such as patient
85 selection for ATSA or which implants and techniques to use. Several previous studies assessing risk
86 factors of glenoid component loosening identified patient, treatment and prosthesis characteristics
87 related to glenoid component loosening. For example, male sex and a higher critical shoulder angle
88 have been associated with higher rates of loosening and revisions.[2,3] Glenoid retroversion did not
89 impact implant survivorship in one study of ATSA with minimal, noncorrective reaming.[4]
90 However, a larger degree of retroversion may have more impact. Several previous studies have also
91 identified aspects of the glenoid component design that correlated with the rate of loosening, such as
92 whether the poly-ethylene is cross-linked, whether the component is pegged or keeled, or the usage of
93 cement.[5–9] In spite of these studies identifying influential factors, accurate prediction of aseptic
94 loosening of the glenoid component remains a challenge with conventional methods.

95 In recent years, machine learning or artificial intelligence has been used with increasing precision to
96 predict outcomes after ATSA. A previous study using machine learning was able to accurately predict
97 range of motion and patient-reported outcomes after ATSA. The most influential factors they reported
98 were follow-up time, pre-operative range of motion and patient-reported outcome measures (PROMs),
99 patients' sex, and surgery on the dominant upper limb.[10] Another study was able to accurately
100 predict the improvement in ASES scores after shoulder arthroplasty using machine learning.[11] The
101 most relevant predictive factors were pre-operative ASES scores, pre-operative pain scores, Walch
102 classification, fatty infiltration in the supra- and infraspinatus, and age. A previous study using
103 artificial intelligence to predict patient satisfaction two years after shoulder arthroplasty found
104 baseline Single Assessment Numeric Evaluation score, exercise and activity, workers' compensation
105 status, diagnosis, symptom duration prior to surgery, body mass index, age, smoking status, anatomic
106 versus reverse TSA, and diabetes to be predicting factors.[12] Two studies report predictive models
107 on short-term complications after ATSA. One study using machine learning to predict complications
108 and 30-day unplanned readmissions found that a history of implant complication, severe chronic
109 kidney disease, teaching hospital status, coronary artery disease and male sex were the most important
110 features.[13] The machine learning model found teaching hospital status and male sex as a markedly

1
2
3 111 more important predictor compared to a logistic regression analysis of the same data. Another study
4 112 on short-term complications after total shoulder arthroplasty found percentage haematocrit, BMI, and
5 113 operative time were of highest importance in outcome prediction.[14] These studies demonstrate that
6 114 machine learning may provide accurate predictions for the outcomes after ATSA. Machine learning is
7 115 most effective with large amounts of data and is very dependent on the amount of detail. Furthermore,
8 116 the algorithm needs to be widely applicable; a varied and international database provides the highest
9 117 external validity.

10
11 118 To our knowledge, there are no studies predicting the long-term complications such as aseptic
12 119 loosening of the glenoid component using advanced machine learning techniques. Furthermore,
13 120 previous machine learning studies are limited in accuracy and validity due to the sample size and
14 121 homogeneity. Therefore, this study aims to develop a clinical prediction model for aseptic loosening
15 122 of the glenoid component using machine learning algorithms trained on a large international database
16 123 using clustered data. The large combined dataset is less prone to overfitting, and allows direct
17 124 validation of models across a range of populations and settings, thereby increasing
18 125 generalisability.[15] The predictive algorithm will be made available for clinical use through a
19 126 publicly available online prediction tool.

20
21
22
23
24
25
26
27
28
29
30 127

31 32 128 **METHODS AND ANALYSIS**

33 34 129 **Data collection**

35
36 130 For this multi-centre retrospective study individual patient data (IPD) will be collected from
37 131 previously published studies reporting failure and revision of ATSA. A systematic literature search
38 132 will be performed in Medline (PubMed) identifying all studies that report a cohort of ATSA including
39 133 revision as an outcome, published between January 2000 and June 2023. The limit was set at January
40 134 2000 to increase the likelihood of the dataset that was used for the study still being available. The
41 135 minimum required data retention period varies between countries but is generally 20 years or less.
42 136 The full search strategy is available in appendix 1. All original studies reporting revision or failure
43 137 rates after primary ATSA will be included. Reviews and letters to the editor will be excluded, as well
44 138 as studies published in languages other than English, Italian, Dutch, and French. Authors will be
45 139 requested to share the anonymised databases used for the identified studies. Only de-identified
46 140 databases used for previous studies are included, authors are not required to gather additional data or
47 141 access patient files. After sharing their data, the authors will be included in in the Machine Learning
48 142 Consortium. Inclusion criteria for individual patients within the provided databases are a minimum
49 143 age of 18 years and a minimum follow-up of 2 years. Patients that underwent ATSA with a fracture as
50 144 the indication or patients that underwent concomitant procedures such as a cuff repair, tendon transfer
51 145 or bone graft will be excluded. The aim is to combine the IPD from previously published studies to

1
2
3 146 create a large international cohort which can be used to train a machine learning algorithm to predict
4 147 aseptic loosening of the glenoid component after anatomic total shoulder arthroplasty. Based on
5 148 previous studies, we estimate a glenoid revision rate of approximately 2%.[6,16] The minimum
6 149 number of events per variable to achieve sufficient accuracy differs per model and is not clearly
7 150 defined for each technique.[17,18] We aim to include at least 30 events per variable, resulting in a
8 151 sample size of 7500 patients for a model with up to 5 predictive variables.

13 152 **Data curation and missing data**

15 153 Completeness across data sources will be assessed for each variable in the compiled multi-centre
16 154 database and variables with sufficient completeness (>70% complete) will be selected as input for the
17 155 machine learning algorithms. Variables with >30% missing data will be excluded. For the remaining
18 156 variables, missing data will be completed by imputation using multivariate imputation by chained
19 157 equations (MICE).[19] Uniformity in reporting will be assessed for each variable. If possible,
20 158 variables will be adjusted or categorised to ensure uniform reporting. In case uniformity of the
21 159 reported variable across data sources cannot be achieved without guaranteeing correctness, the
22 160 variable will be excluded. Each data set will be split into training (80%) and test (20%) subsets,
23 161 stratified by outcome. Fivefold cross-validation of the training set will be used to develop the ML
24 162 models.[20] Data curation and imputation will be performed using R (R foundation for statistical
25 163 computing, Vienna, Austria).

33 164 **Variable selection**

35 165 The primary outcome is a revision of the glenoid component for aseptic glenoid loosening. The input
36 166 variables for both methods are dependent on the uniformity and completeness of the gathered data but
37 167 will include demographics (eg. age, sex and ethnicity), patient-specific factors (eg. preoperative Body
38 168 Mass Index, comorbidity, smoking, dominance, previous surgery), disease-specific factors (eg,
39 169 affected side, indication, Walch classification, fatty-infiltration of cuff muscles) and surgical
40 170 characteristics (eg. corrective reaming, component design and type, component materials, cementing,
41 171 and sizes). Before training the machine learning models, relevant variables will be selected using
42 172 random forest algorithms with recursive selection.[21] At least 10 events for each predictor variable
43 173 will be included in the model, adhering to the rule of thumb in predictive models of binary
44 174 variables.[22]

52 175 **Development of prediction models**

54 176 Different ML models result in varying performance metrics based on the type of input data
55 177 (continuous, categorical, dichotomous). Due to the variation in type of input variables in the dataset,
56 178 several different ML techniques will be used and compared based on model performance. The
57 179 following machine learning algorithms were chosen for modelling based on prior research [23–27]:

1
2
3 180 Stochastic Gradient Boosting (SGM), Random Forest (RF), Support Vector Machine (SVM), Neural
4
5 181 Network (NN) and Elastic-Net Penalized Logistic Regression (PLR). The algorithms will be trained
6
7 182 on the training dataset with ten-fold cross-validation repeated 3 times. Cross-validation means
8
9 183 dividing data into a selected number of groups, also called folds. First, the data will be divided into 10
10
11 184 equally sized folds. Then, the algorithms will be trained on 9 of the 10 folds (90% of the training data)
12
13 185 and tested on the remaining fold (10% of the training data). Results will be averaged across all
14
15 186 repetitions of this sequence. Machine learning algorithms will be developed using Python (The
16
17 187 Python Software Foundation, Fredericksburg, United States of America). Hyper parameter tuning will
18
19 188 be performed as recommended in the Python libraries. The statistician who performs the machine
20
21 189 learning analysis will be blinded to the origin of the data, but the anonymised data source will be
22
23 190 available to be included as a potential confounding factor.

21 191 **Model performance**

23 192 After training all models, the model performance will be analysed according to a proposed framework
24
25 193 by Steyerberg et al. including discrimination with the c-statistic, positive predictive value (PPV), true
26
27 194 positive rate (TPR), precision-recall curve, calibration slope and intercept and the overall performance
28
29 195 with the Brier score.[28]

30 196 The c-statistic (area under the curve of a receiver operating characteristic curve) is a score ranging
31
32 197 from 0.50 to 1.0 with 1.0 indicating the highest discrimination score and 0.50 indicating the lowest.
33
34 198 The higher the discrimination score, the better the model's ability to distinguish patients with and
35
36 199 without the outcome of interest.[20,23] The PPV is the proportion of true positive outcomes over the
37
38 200 number of predicted positive outcomes. The TPR is the proportion of true positive outcomes over the
39
40 201 number of observed positive outcomes. The precision recall curve is a plot of the PPV versus the
41
42 202 TPR. A calibration plot plots the estimated versus the observed probabilities for the primary outcome.
43
44 203 A perfect calibration plot has an intercept of 0 (<0 reflects overestimation and >0 reflects
45
46 204 underestimation of the probability of the outcome) and a slope of 1 (model is performing similarly in
47
48 205 training and test datasets).[20,23,28] The null-model Brier score, which equals the probability of
49
50 206 glenoid revision in the dataset, will be used to benchmark the algorithm's Brier score. A Brier score
51
52 207 lower than the null-model Brier score indicates superior performance of the prediction model to this
53
54 208 null benchmark. Perfect prediction would have a Brier score of 0, whereas a Brier score of 1 would
55
56 209 indicate the poorest possible prediction.[28]

53 210 In addition, the decision curve analysis will be performed and visualized to investigate the net benefit
54
55 211 (weighted average of true positives and false positives) of the conducted algorithms over the range of
56
57 212 risk thresholds for clinical decision-making.[23,28,29] The net benefit is a weighted average of true
58
59 213 positives and false positives, formula = sensitivity x prevalence – (1-specificity) x (1 – prevalence) x
60

1
2
3 214 (odds at the threshold probability). The decision curve of the model will be compared to decision
4
5 215 curves of treating everyone as being at risk and treating no one as being at risk.[23]
6

7 216 Due to the large heterogeneity of the compiled dataset from different international sources and the
8
9 217 internal validation of the prediction models, the generalisability of the model can be intrinsically
10
11 218 confirmed using the above-mentioned performance tests. Therefore, it is not strictly necessary to
12
13 219 externally validate the final algorithm. However, this study's primary aim is model development.
14
15 220 External validation in a specific setting is advised before applying the algorithm to clinical practice.

16 221 **Open-access clinical prediction tool**

17
18 222 The best performing prediction algorithm will be used to create an open-access clinical prediction
19
20 223 tool, in the form of a publicly available web application accessible on desktops, tablets, and
21
22 224 smartphones.

23 225 **Statistical analysis**

24
25 226 Categorical variables will be described as absolute numbers with frequencies, and continuous
26
27 227 variables as medians with interquartile ranges (IQR). The model performance metrics will be
28
29 228 calculated with 95% confidence interval (CI). Given the retrospective study design, post hoc power
30
31 229 analyses will be conducted to evaluate the sample size of the study with an alpha value of 0.05.

32 230 **Guidelines**

33
34
35 231 The study set-up will be performed following the Transparent Reporting of Multivariable Prediction
36
37 232 Models for Individual Prognosis or Diagnosis Guideline for Clustered data (TRIPOD-Cluster).[15]

38 233 **Patient and public involvement**

39
40
41 234 None.

42 43 235 **ETHICS AND DISSEMINATION**

44
45 236 For safe multicentre data exchange and analysis, our Machine Learning Consortium will adhere to the
46
47 237 World Health Organisation (WHO) regulation 'Policy on Use and Sharing of Data Collected by WHO
48
49 238 in Member States Outside the Context of Public Health Emergencies'.[30] An Institutional Review
50
51 239 Board (IRB) approval has been obtained for each of the included studies and the provided data are
52
53 240 anonymised and de-identified, no additional prospective data is collected and contributing authors are
54
55 241 not required to access any patient files, no IRB review is required for this study. Patient consent for
56
57 242 publication is not applicable to this study.

58
59 243 The study results will be disseminated through publication in a peer-reviewed journal. To facilitate
60
244 reproduction of the results and external validation of the algorithm, the (anonymous) code of the
245 developed predictive algorithms will be made available upon request with the authors.

1
2
3 246 Data collection for this project is currently ongoing. The analysis will start in December 2023. The
4
5 247 expected time of completion for the project is July 2024.

6 7 248 **DISCUSSION**

8
9 249 For an informed decision when considering ATSA, it is important to be able to make an accurate
10
11 250 prediction of arthroplasty failure. Previous studies have demonstrated several factors that affect
12
13 251 complications and revision after ATSA, including male sex, comorbidities such as chronic kidney
14
15 252 disease or coronary artery disease, percentage haematocrit, a higher critical shoulder angle, teaching
16
17 253 hospital status, operative time, and the material and design of the prosthesis.[2,3,5–9,13,14]
18
19 254 Psychological studies have shown that in human judgement only a limited amount of variables can be
20
21 255 taken into account, and that prediction models are generally more accurate and less subject to
22
23 256 bias.[31] Machine learning algorithms have been shown to be an effective method in developing
24
25 257 patient-specific prediction tools, which may complement human judgement when counselling patients
26
27 258 in clinic.[32] Creating an online tool for aseptic loosening of the glenoid component after ATSA can
28
29 259 help guide surgeons in selecting patients who will most benefit from this treatment, and considering
30
31 260 alternatives in cases of high risk estimates.

32
33 261 The strength of this project is the large amount of data that will be gathered from authors participating
34
35 262 in the Machine Learning Consortium, aiming to include a minimum of 7500 patients in total. Using a
36
37 263 large, heterogenous international database for development of the algorithm and prediction tool will
38
39 264 result in high external validity and may improve applicability world-wide.[15] However, most ML
40
41 265 techniques require a larger sample size to achieve an accurate prediction compared to traditional
42
43 266 regression models. The minimum events per variable is not clearly defined and differs per technique.
44
45 267 Furthermore, in gathering data retrospectively from various sources, the study is subject to variances
46
47 268 in the included variables. Low completeness and large variability of reporting may introduce bias.
48
49 269 However, only variables that are consistently reported in multiple data sources will be included in the
50
51 270 final analysis, variables will be categorised to increase uniformity and missing data will be imputed
52
53 271 where possible. The exclusion and categorisation of variables will have to be balanced with the
54
55 272 amount of detail in the final analysis. Furthermore, the accuracy of data collection is dependent on
56
57 273 third parties providing the data, the method of data collection cannot be verified for all sources.
58
59 274 However, the data source will be considered as a confounder. Furthermore, the variety in data sources
60
275 will increase the external applicability of the algorithm. Last, ML prediction models for a
276 dichotomous outcome are limited to risk classification, the individual risk must be interpreted in the
277 clinical context when used for medical decision making.

278

279

280 **CONTRIBUTORS**

281 Authors AM, GB, and MP contributed to the conception, methods, planning, and writing and
282 reviewing of the protocol. LM and JO contributed to the statistical and machine learning methods of
283 the study. JO, PB, GA, JD, LL, and TL reviewed the draft of the protocol and provided supervision.
284 All authors revised the final version of the protocol and gave approval for publication.

285 **ACKNOWLEDGEMENTS**

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

288 **FUNDING**

289 This work was supported by Stichting Prof. Michaël-van Vloten fonds, VSBfonds grant number
290 VSB.22/00098, Prins Bernard Cultuurfonds grant number 40042437/NOD, and Stichting Erasmus
291 Trustfonds grant number 97030.2021.101.577/193/RB.

292 **COMPETING INTERESTS**

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

298 **REFERENCES**

- 299 1 Sharareh B, Yao JJ, Matsen FA, *et al.* What is the Optimal Management of a Loose Glenoid
300 Component after Anatomic Total Shoulder Arthroplasty: A Systematic Review. *J shoulder Elb*
301 *Surg* Published Online First: November 2022. doi:10.1016/J.JSE.2022.10.006
- 302 2 Tabeayo E, Chan PH, Prentice HA, *et al.* The association between critical shoulder angle and
303 revision following anatomic total shoulder arthroplasty: a matched case-control study. *J*
304 *shoulder Elb Surg* 2022;**31**:1796–802. doi:10.1016/J.JSE.2021.08.003
- 305 3 Aibinder W, Schoch B, Parsons M, *et al.* Risk factors for complications and revision surgery
306 after anatomic and reverse total shoulder arthroplasty. *J Shoulder Elb Surg* 2021;**30**:e689–701.
307 doi:10.1016/J.JSE.2021.04.029
- 308 4 Dekker TJ, Grantham WJ, Lacheta L, *et al.* Glenoid retroversion does not impact clinical
309 outcomes or implant survivorship after total shoulder arthroplasty with minimal, noncorrective
310 reaming. *JSES Int* 2022;**6**:596–603. doi:10.1016/J.JSEINT.2022.02.011

- 1
2
3 311 5 Page RS, Alder-Price AC, Rainbird S, *et al.* Reduced Revision Rates in Total Shoulder
4 312 Arthroplasty With Crosslinked Polyethylene: Results From the Australian Orthopaedic
5 313 Association National Joint Replacement Registry. *Clin Orthop Relat Res* 2022;**480**:1940–9.
6 314 doi:10.1097/CORR.0000000000002293
- 7
8
9
10 315 6 Dillon MT, Chan PH, Prentice HA, *et al.* The association between glenoid component design
11 316 and revision risk in anatomic total shoulder arthroplasty. *J Shoulder Elb Surg* 2020;**29**:2089–
12 317 96. doi:10.1016/J.JSE.2020.02.024
- 13
14
15 318 7 Welsher A, Gohal C, Madden K, *et al.* A comparison of pegged vs. keeled glenoid components
16 319 regarding functional and radiographic outcomes in anatomic total shoulder arthroplasty: a
17 320 systematic review and meta-analysis. *JSES open access* 2019;**3**:136-144.e1.
18 321 doi:10.1016/J.JSES.2019.04.002
- 19
20
21
22 322 8 Page RS, Pai V, Eng K, *et al.* Cementless versus cemented glenoid components in
23 323 conventional total shoulder joint arthroplasty: analysis from the Australian Orthopaedic
24 324 Association National Joint Replacement Registry. *J Shoulder Elb Surg* 2018;**27**:1859–65.
25 325 doi:10.1016/J.JSE.2018.03.017
- 26
27
28
29 326 9 Sharplin PK, Frampton CMA, Hirner M. Cemented vs. uncemented glenoid fixation in total
30 327 shoulder arthroplasty for osteoarthritis: a New Zealand Joint Registry study. *J Shoulder Elb*
31 328 *Surg* 2020;**29**:2097–103. doi:10.1016/J.JSE.2020.03.008
- 32
33
34
35 329 10 Kumar V, Roche C, Overman S, *et al.* What Is the Accuracy of Three Different Machine
36 330 Learning Techniques to Predict Clinical Outcomes After Shoulder Arthroplasty? *Clin Orthop*
37 331 *Relat Res* 2020;**478**:2351–63. doi:10.1097/CORR.0000000000001263
- 38
39
40 332 11 McLendon PB, Christmas KN, Simon P, *et al.* Machine Learning Can Predict Level of
41 333 Improvement in Shoulder Arthroplasty. *JBJS Open Access* 2021;**6**.
42 334 doi:10.2106/JBJS.OA.20.00128
- 43
44
45
46 335 12 Polce EM, Kunze KN, Fu MC, *et al.* Development of supervised machine learning algorithms
47 336 for prediction of satisfaction at 2 years following total shoulder arthroplasty. *J Shoulder Elb*
48 337 *Surg* 2021;**30**:e290–9. doi:10.1016/j.jse.2020.09.007
- 49
50
51 338 13 Devana SK, Shah AA, Lee C, *et al.* Development of a Machine Learning Algorithm for
52 339 Prediction of Complications and Unplanned Readmission Following Primary Anatomic Total
53 340 Shoulder Replacements. <https://doi.org/10.1177/24715492221075444>
54 341 2022;**6**:247154922210754. doi:10.1177/24715492221075444
- 55
56
57
58 342 14 Gowd AK, Agarwalla A, Amin NH, *et al.* Construct validation of machine learning in the
59 343 prediction of short-term postoperative complications following total shoulder arthroplasty. *J*

- 1
2
3 344 *Shoulder Elb Surg* 2019;**28**:e410–21. doi:10.1016/J.JSE.2019.05.017
4
5 345 15 Debray TPA, Collins GS, Riley RD, *et al.* Transparent reporting of multivariable prediction
6 346 models developed or validated using clustered data: TRIPOD-Cluster checklist. *BMJ*
7 347 2023;**380**:e071018. doi:10.1136/BMJ-2022-071018
8
9
10 348 16 Papadonikolakis A, Neradilek MB, Matsen FA. Failure of the glenoid component in anatomic
11 349 total shoulder arthroplasty :A systematic review of the english-language literature between
12 350 2006 and 2012. *J Bone Jt Surg* 2013;**95**:2205–12. doi:10.2106/JBJS.L.00552
13
14
15
16 351 17 Van Der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry:
17 352 A simulation study for predicting dichotomous endpoints. *BMC Med Res Methodol* 2014;**14**:1–
18 353 13. doi:10.1186/1471-2288-14-137/FIGURES/13
19
20
21 354 18 Steyerberg EW. Clinical Prediction Models; A Practical Approach to Development,
22 355 Validation, and Updating. Springer 2019
23
24
25 356 19 van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations
26 357 in R. *J Stat Softw* 2011;**45**:1–67. doi:10.18637/JSS.V045.I03
27
28
29 358 20 Oosterhoff JHF, Gravesteijn BY, Karhade A V., *et al.* Feasibility of Machine Learning and
30 359 Logistic Regression Algorithms to Predict Outcome in Orthopaedic Trauma Surgery. *J Bone Jt*
31 360 *Surg - Am Vol* 2022;**104**:544–51. doi:10.2106/JBJS.21.00341
32
33
34 361 21 DeBaun MR, Chavez G, Fithian A, *et al.* Artificial Neural Networks Predict 30-Day Mortality
35 362 After Hip Fracture: Insights From Machine Learning. *J Am Acad Orthop Surg* 2021;**29**:977–
36 363 83. doi:10.5435/JAAOS-D-20-00429
37
38
39 364 22 Riley RD, Ensor J, Snell KIE, *et al.* Calculating the sample size required for developing a
40 365 clinical prediction model. *BMJ* 2020;**368**. doi:10.1136/BMJ.M441
41
42
43 366 23 Oosterhoff JHF, Savelberg ABMC, Karhade AV, *et al.* Development and internal validation of
44 367 a clinical prediction model using machine learning algorithms for 90 day and 2 year mortality
45 368 in femoral neck fracture patients aged 65 years or above. *Eur J Trauma Emerg Surg*
46 369 2022;**48**:4669–82. doi:10.1007/s00068-022-01981-4
47
48
49
50 370 24 Oosterhoff JHF, Karhade A V., Oberai T, *et al.* Prediction of Postoperative Delirium in
51 371 Geriatric Hip Fracture Patients: A Clinical Prediction Model Using Machine Learning
52 372 Algorithms. *Geriatr Orthop Surg Rehabil* 2021;**12**. doi:10.1177/21514593211062277
53
54
55
56 373 25 Bongers MER, Thio QCBS, Karhade A V., *et al.* Does the SORG Algorithm Predict 5-year
57 374 Survival in Patients with Chondrosarcoma? An External Validation. *Clin Orthop Relat Res*
58 375 2019;**477**:2296–303. doi:10.1097/CORR.0000000000000748
59
60

- 1
2
3 376 26 Karhade A V., Ogink PT, Thio QCBS, *et al.* Development of machine learning algorithms for
4 377 prediction of prolonged opioid prescription after surgery for lumbar disc herniation. *Spine J*
5 378 2019;**19**:1764–71. doi:10.1016/J.SPINEE.2019.06.002
6
7
8 379 27 Karhade A V., Ogink PT, Thio QCBS, *et al.* Discharge Disposition After Anterior Cervical
9 380 Discectomy and Fusion. *World Neurosurg* 2019;**132**:e14–20.
10 381 doi:10.1016/J.WNEU.2019.09.026
11
12
13 382 28 Steyerberg EW, Vickers AJ, Cook NR, *et al.* Assessing the performance of prediction models:
14 383 A framework for traditional and novel measures. *Epidemiology* 2010;**21**:128–38.
15 384 doi:10.1097/EDE.0b013e3181c30fb2
16
17
18 385 29 Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for
19 386 development and an ABCD for validation. *Eur Heart J* 2014;**35**:1925–31.
20 387 doi:10.1093/EURHEARTJ/EHU207
21
22
23 388 30 World Health Organization. Who policy statement on Data Sharing [Internet]. Policy on use
24 389 and sharing of data collected in Member States by the World Health Organization (WHO)
25 390 outside the context of public health emergencies. WHO; [cited 2023Apr13]. Available from:
26 391 [https://www.who.int/docs/default-source/publishing-policies/who-policy-statement-on-data-](https://www.who.int/docs/default-source/publishing-policies/who-policy-statement-on-data-sharing.pdf?sfvrsn=71a41493_0)
27 392 [sharing.pdf?sfvrsn=71a41493_0](https://www.who.int/docs/default-source/publishing-policies/who-policy-statement-on-data-sharing.pdf?sfvrsn=71a41493_0)
28
29
30 393 31 Daniel Kahneman, Olivier Sibony, Cass R. Sunstein. *Noise: A Flaw in Human Judgment*
31 394 Little, Brown Spark; Hachette Book Group, version 1, 2021
32
33 395 32 A Machine Learning Algorithm to Identify Patients with Tibial Shaft Fractures at Risk for
34 396 Infection After Operative Treatment. *J Bone Joint Surg Am* 2021;**103**:532–40.
35 397 doi:10.2106/JBJS.20.00903
36
37
38
39
40
41
42 398
43
44
45 399
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 **1 APPENDIX 1**
4

5 2 Search strategy:
6

7 ("Arthroplasty, Replacement, Shoulder"[Mesh] OR "Total shoulder arthroplasty"[tiab] OR "Anatomic
8 shoulder arthroplasty"[tiab] OR "TSA"[tiab] OR "ATSA"[tiab] OR "Total shoulder prosthes*"[tiab]
9 OR "Anatomic shoulder prosthes*"[tiab]) AND ("Reoperation"[Mesh] OR "reoperation"[tiab] OR
10 "re-operation"[tiab] OR "revision"[tiab] OR "survival"[tiab] OR "implant-survival"[tiab] OR
11 "Prosthesis Failure"[Mesh] OR "failure"[tiab] OR "loosening"[tiab] OR "aseptic"[tiab] OR
12 "Postoperative Complications"[Mesh] OR "complication"[tiab])
13
14
15
16
17

18 9 Filters: Clinical study, Clinical trial, Controlled clinical trial, Comparative study, Dataset
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



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 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.