BMJ Open

Developing a machine learning algorithm to predict probability of retear and functional outcomes in patients undergoing rotator cuff repair surgery: protocol for a retrospective, multicentre study

Laurens J H Allaart [1,2] Sanne van Spanning,[2,3] Laurent Lafosse,[1] Thibault Lafosse,[1] Alexandre Ladermann,[4,5] George S Athwal,[6] Laurent A M Hendrickx,[7,8] Job N Doornberg,[8,9] Michel P J van den Bekerom [2] Geert Alexander Buijze,[1,7] on behalf of the Machine Learning Consortium

For numbered affiliations see end of article.

Correspondence to
Laurens J H Allaart;
laurensallaart@gmail.com

ABSTRACT

Introduction The effectiveness of rotator cuff tear repair surgery is influenced by multiple patient-related, pathology-centred and technical factors, which is thought to contribute to the reported retear rates between 17% and 94%. Adequate patient selection is thought to be essential in reaching satisfactory results. However, no clear consensus has been reached on which factors are most predictive of successful surgery. A clinical decision tool that encompassed all aspects is still to be made. Artificial intelligence (AI) and machine learning algorithms use complex self-learning models that can be used to make patient-specific decision-making tools. The aim of this study is to develop and train an algorithm that can be used as an online available clinical prediction tool, to predict the risk of retear in patients undergoing rotator cuff repair.

Methods and analysis This is a retrospective, multicentre, cohort study using pooled individual patient data from multiple studies of patients who have undergone rotator cuff repair and were evaluated by advanced imaging for healing at a minimum of 6 months after surgery. This study consists of two parts. Part one: collecting all potential factors that might influence retear risks from retrospective multicentre data, aiming to include more than 1000 patients worldwide. Part two: combining all influencing factors into a model that can clinically be used as a prediction tool using machine learning.

Ethics and dissemination For safe multicentre data exchange and analysis, our Machine Learning Consortium adheres to the WHO regulation 'Policy on Use and Sharing of Data Collected by WHO in Member States Outside the Context of Public Health Emergencies'. The study results will be disseminated through publication in a peer-reviewed journal. Institutional Review Board approval does not apply to the current study protocol.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ This study aims to calculate a patient-specific retear-chance after rotator cuff repair surgery.
⇒ Creating an online-available tool that predicts retear chances can help both medical professionals and patients in clinical decision-making on rotator cuff repair surgery.
⇒ Included data will be gathered from previously published databases of all authors included in the Machine Learning Consortium, aiming to include data from over 1000 patients.
⇒ This study does have the limitation of being retrospective and therefore the study is dependent on the recordkeeping of each individual hospital.

INTRODUCTION

Despite technical advances of rotator cuff repair, the rate of unhealed or re-torn rotator cuff tears remains high, with percentages ranging between 10% and 94%.[1] A myriad of patient-related,[2] pathology-centred[3] and technical factors[4] influence this adverse outcome.

Patient selection is thought to be essential, however there is no consensus on which of the numerous potentially influential factors are most important for the prediction of satisfactory postoperative results.[5] Furthermore, the value of preoperative optimisation of potential patient-related influential factors including comorbidities, metabolic deficiencies and intoxications remains questionable. The increasing worldwide interest in these factors is confirmed by development of preoperative screening and optimisation programmes aiming for smoking cessation,

diabetes control, use of statins in hyperlipidaemia and vitamin D deficiency supplementing.[2][6] However, the majority of shoulder surgeons seems to limit decision-making to more basic, previously established predictive factors including age, functional demand and pathology-specific grading. Despite the many different classification systems that have been developed to facilitate decision-making, a patient-specific decision tool is still lacking.[7][8] This, in combination with the fact that existing research commonly evaluates a single treatment option between homogenic groups, makes it almost impossible for surgeons to preoperatively indicate a reliable chance of satisfactory results.

Artificial intelligence and machine learning (ML) is believed to facilitate a more patient-specific approach and will allow us to move to the next level of evidence-based medicine: personalised patient-care. Clinical prediction tools, incorporating patient-specific factors to predict outcome probabilities will provide guidance to both clinicians and patients.[9][10] Within orthopaedic (oncology) surgery, prediction tools based on ML algorithms, have already been successfully implemented to predict patient-specific 5-year survival in patients with chondrosarcoma.[11] Furthermore, based on a series of 422 patients undergoing lumbar discectomy, Staartjes *et al* demonstrated deep learning algorithms to be superior to standard regression models in predicting patient-reported outcome measures (PROMs).[9]

Aim of this study
The aim of this study is to develop and train a ML algorithm in order to create a clinical prediction tool to be used in clinical practice by predicting retear-chance of the rotator cuff as well as chance of clinical improvement based on preoperative patient data. The prediction tool will be free and online available.

METHODS AND ANALYSIS
This is a retrospective, multicentre, cohort study.

The primary and secondary outcome measures will be implemented as features for the prediction algorithm.

Primary outcome measures
► Rotator cuff retear rates at minimum 6 months follow-up as measured on MRI, arthro-CT and/or ultrasound (yes vs no, defined by Sugaya grades 1–3 as no retear and grade 4–5 as retear[12]).
► Enduring satisfactory functional outcome defined as achievement (yes vs no) and maintenance (yes vs no) of the PROM-specific minimal clinical important difference (MCID)[13] in numeric rating scales of PROMs from baseline at 2–5 years follow-up after repair. PROMs include the Constant-Murley score, American Shoulder and Elbow Surgeons score (ASES), University of California at Los Angeles shoulder score (UCLA), Oxford Shoulder Score (OSS), Western Ontario Rotator Cuff index (WORC), Disabilities of Arm, Shoulder and Hand score (DASH).

Secondary outcome measures
► Adverse events graded as the possibility of none/minor versus moderate/severe complication as defined in accordance to Felsch *et al*.[14] Adverse events classify as moderate/severe from Felsch class III onwards, which means other surgical or radiological intervention was needed or unexpected hospital admission was necessary. Adverse events will be differentiated into three groups: infection, revision surgery or other.

Model development
The development of the prediction model will be performed based on the steps described by Steyerberg and Vergouwe[15]:
1. Data collection.
2. Data inspection.
3. Coding of predictors.
4. Model specification.
5. Model estimation and performance.
6. Model validation.
7. Model presentation.

Data collection
Step one will involve contacting authors from previously published studies in order to collect and combine their (raw) individual patient data into a central database. All randomised controlled trials comparing any surgical technique, add-on biological intervention or rehabilitation protocols concerning rotator cuff surgery will be included. In addition, cohorts evaluating risk factors of surgical techniques after rotator cuff repair will be included. This retrospective review will therefore incorporate patients with all types of tears and concomitant procedures (eg, biceps tenodesis or tenotomy and acromioclavicular resection). Exclusion criteria for all studies will be the lack of postoperative evaluation by ultrasound, contrast-enhanced CT or MRI at minimally 6 months after surgery or publication date from before 2005. Relevant studies will be identified using a systematic approach primarily searching the online PubMed database according to the search terms found in online supplemental file 1. As there is no golden standard for sample size or power calculations for prediction models, and we are fully dependent on contributed data, we aim to include at least 1000 patients worldwide.[15]

Problem definition and data inspection
All contributed data sets will be formatted into one central database. As data is commonly collected in .csv (Microsoft Excel) or .sav (SPSS) files, formatting will be performed with the dplyr package for R software. All raw data of the different variables will be separately reviewed for inaccuracies and other defects. This process will focus on uniformisation of possible inconsistencies in the collected data, for example, follow-up times into a standardised format as 'days after surgery'. Categorical data will be translated into English or corrected for typographs. Continuous variables will be screened for outliers by visualisation in

the ggplot package. Impossible values or uninterpretable syntax errors will be excluded from the central database.

Coding of predictors

For each primary outcome, a logistic regression will be performed including all available variables in the central database to identify the variables with the highest predictive values. The data points available include patient demographic (sex, age), patient-specific factors (body mass index, dominance, sport/activity level, workers compensation), pathology-specific factors (eg, tear size and location), surgical technique and add-on interventions. For a complete overview of all variables see online supplemental file 2. The variables with the highest predictive values will be used as the algorithms labels.

Missing data

As the main database will comprise data from multiple studies, we expect many cases of missing data. The approach to missing data will differ depending on the type of variable. Variables with less than 5% missing data will be replaced by imputation.[16] Missing data on any surgical technique or add-on intervention is expectable as interventions outside the scope of a study would not be mentioned (or briefly mentioned in the exclusions part). Therefore, this kind of missing data will be transformed to 'No'. Overall availability of variables will be presented according to current guidelines.[17] Any variances between hospitals will be reported.

Model specification

Algorithms to be trained

Based on previous studies,[18 19] the following algorithms are likely to result in accurate prediction models for our primary outcomes: (1) Bayes Point Machine, (2) Boosted Decision Tree (3) Penalised Logistical Regression, (4) Neural Network and (5) Support Vector Machine. In order to recognise patterns related to each outcome, the ML algorithms will have to be trained separately for each outcome.

Model estimation and performance

Assessing the performance of the algorithms

The performance of the ML-algorithms will be assessed and compared based on (1) model discrimination; (2) calibration and (3) overall model performance (Brier Score) according to Steyerberg's structured 'ABCD-methodology' for clinical prediction rules.[15 20]

The model's predicted probability will be plotted against the actual observed probability to calculate calibration of a model. Perfect models will have calibration intercepts of 0, and calibration slopes of 1.27. The overall performance of the model will be assessed with the Brier Score. A perfect Brier score, indicating total accuracy, is a score of 0. The lowest possible score is a Brier Score of 1.26. Accuracy, sensitivity, specificity and area under the receiver operated characteristic-curve will be measures for a model's ability to distinguish patients with the primary outcome from those without.

Model validation

Internal validation

Internal validation of our algorithms will be performed by 10-fold cross validation. This means that instead of dividing the main data set into one training set and one testing set, this process will be 10 times randomly repeated and the results will be averaged. This has as main advantage that all individual patient records are used as training and testing data simultaneously, which results in higher accuracy of predictions as well as lower chance of bias. The cross validation will be performed using the trainControl function from the Caret library for R.

External validation

Before incorporating the best performing algorithm, we aim to have the algorithm externally validated. The same performance metrics could be calculated as described above. However, this would involve collaboration with partners that have adequate data and are willing to share. As no agreements currently have been made, the external validation is outside the scope of this study.

Model presentation

The best performing algorithm will be deployed as an open-access probability calculator and used to design a clinical decision rule. To simulate the clinical scenario to which a decision rule would be most applicable, thresholds shall be selected based on patients with clinical symptoms of a retear or with an unsatisfactory functional outcome.

Patient and public involvement

None.

ETHICS AND DISSEMINATION

For safe multicentre data exchange and analysis, our Machine Learning Consortium adheres to the WHO regulation 'Policy on Use and Sharing of Data Collected by WHO in Member States Outside the Context of Public Health Emergencies'.[21] As Institutional Review Board (IRB) approval has been acquired for each of the included studies and data are anonymised as in conventional meta-analyses, additional IRB approval is not required for the current study protocol. The technical appendix, statistical code and final data set will be published with the study results.

Current status

The study has currently entered the data-collection phase, which is expected to last until March 2023. Re-evaluation of the data using ML algorithms to predict outcomes will start in April 2023, after which the algorithms can be externally validated. The expected time for study completion is by late 2023.

DISCUSSION

Due to the wide variety of pathological factors at the origin of rotator cuff tears and the numerous surgical approaches

to repair, optimal decision-making remains challenging. Smaller case series often provide heterogeneous data on this topic, however the largest and most recent meta-analysis to date including 2611 patients with a mean follow-up of 25 months has somewhat demystified the matter. Patients with a full-thickness rotator cuff retear exhibited significantly lower functional outcome scores and strength compared with patients with an intact or partially torn rotator cuff.[22] This is corroborated by the findings of rotator cuff repair with more than 10 years follow-up, showing clinical superiority of structural tendon integrity in partial cuff tears.[23–25] Progressive osteoarthritic changes are significantly more common in patients with repair failures.[24] The most recent randomised controlled trial comparing surgical repair to conservative treatment for degenerative rotator cuff tears showed that only operated patients without retear had an improvement exceeding the MCID in functional outcome at 1-year follow-up.[26] Findings from the latest meta-analysis on this comparative topic conclude that as the success rate of conservative treatment may be high, judicious selection of patients who are most likely to benefit from surgery is key.[27] It is extremely difficult to combine all these factors into a clinical decision related to one specific patient. Creating a free online available clinical prediction tool that takes all these factors into account will assist physicians in selecting which patients with rotator cuff tears will benefit from a repair. In addition, the aimed size (more than 1000 patients) of the database that will be used to design and train the prediction tool might provide new insights on which biological or biomechanical factors influence outcomes after rotator cuff repair the most. Awareness of these factors would be the essential first step to incorporating them in future treatment strategies and eventually improving outcomes. The main limitation of this study is that it is a retrospective, multicentre study. This means this study is dependent on the quality of recordkeeping in the different participating hospitals. This may lead to variance in recorded variables and therefore missing data.

Author affiliations
[1]Orthopaedic Surgery, Clinique Générale Annecy, Annecy, Auvergne-Rhône-Alpes, France
[2]Department of Human Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
[3]Shoulder and Elbow Unit, Joint Research, Department of Orthopaedic Surgery, OLVG, Amsterdam, The Netherlands
[4]Division of Orthopaedics and Trauma Surgery, La Tour Hopital Prive SA, Meyrin, Switzerland
[5]Faculty of Medicine, University of Geneva, Geneve, Switzerland
[6]Roth McFarlane Hand and Upper Limb Center, Schulich School of Medicine and Dentistry, London, Ontario, Canada
[7]Department of Orthopedic Surgery, University of Amsterdam, Amsterdam, The Netherlands
[8]Orthopaedic and Trauma Surgery, Flinders University, Adelaide, South Australia, Australia
[9]Orthopaedic Surgery, University Medical Centre Groningen, Groningen, The Netherlands

ORCID iDs
Laurens J H Allaart http://orcid.org/0000-0002-7211-183X
Michel P J van den Bekerom http://orcid.org/0000-0002-5146-6148

REFERENCES

1. Zhao J, Luo M, Pan J, *et al*. Risk factors affecting rotator cuff retear after arthroscopic repair: a meta-analysis and systematic review. *J Shoulder Elbow Surg* 2021;30:2660–70.
2. Zumstein MA, Lädermann A, Raniga S, *et al*. The biology of rotator cuff healing. *Orthop Traumatol Surg Res* 2017;103:S1–10.
3. Kunze KN, Rossi LA, Beletsky A, *et al*. Does the use of knotted versus knotless transosseous equivalent rotator cuff repair technique influence the incidence of retears? A systematic review. *Arthroscopy* 2020;36:1738–46.
4. Rossi LA, Chahla J, Verma NN, *et al*. Rotator cuff retears. *JBJS Rev* 2020;8:e0039.
5. Griffiths S, Yohannes AM. Surgical referral criteria for degenerative rotator cuff tears: a Delphi questionnaire study. *Musculoskeletal Care* 2014;12:82–91.
6. Yang Y, Qu J. The effects of hyperlipidemia on rotator cuff diseases: a systematic review. *J Orthop Surg Res* 2018;13:204.
7. Lädermann A, Burkhart SS, Hoffmeyer P, *et al*. Classification of full-thickness rotator cuff lesions: a review. *EFORT Open Rev* 2016;1:420–30.
8. Lee CS, Davis SM, Doremus B, *et al*. Interobserver agreement in the classification of partial-thickness rotator cuff tears

using the snyder classification system. *Orthop J Sports Med* 2016;4:2325967116667058.

9 Staartjes VE, de Wispelaere MP, Vandertop WP, *et al*. Deep learning-based preoperative predictive analytics for patient-reported outcomes following lumbar discectomy: feasibility of center-specific modeling. *Spine J* 2019;19:853–61.

10 Choi E, Sanyal N, Ding VY, *et al*. Development and validation of a risk prediction model for second primary lung cancer. *J Natl Cancer Inst* 2022;114:87–96.

11 Thio QCBS, Karhade AV, Ogink PT, *et al*. Can machine-learning techniques be used for 5-year survival prediction of patients with chondrosarcoma? *Clin Orthop Relat Res* 2018;476:2040–8.

12 Sugaya H, Maeda K, Matsuki K, *et al*. Functional and structural outcome after arthroscopic full-thickness rotator cuff repair: single-row versus dual-row fixation. *Arthroscopy* 2005;21:1307–16.

13 Dabija DI, Jain NB. Minimal clinically important difference of shoulder outcome measures and diagnoses: a systematic review. *Am J Phys Med Rehabil* 2019;98:671–6.

14 Felsch Q, Mai V, Durchholz H, *et al*. Complications within 6 months after arthroscopic rotator cuff repair: registry-based evaluation according to a core event set and severity grading. *Arthroscopy* 2021;37:50–8.

15 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.

16 Donders ART, van der Heijden GJMG, Stijnen T, *et al*. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol* 2006;59:1087–91.

17 Nijman S, Leeuwenberg AM, Beekers I, *et al*. Missing data is poorly handled and reported in prediction model studies using machine learning: a literature review. *J Clin Epidemiol* 2022;142:218–29.

18 Machine Learning Consortium, on behalf of the SPRINT and FLOW Investigators. A machine learning algorithm to identify patients with tibial shaft fractures at risk for infection after operative treatment. *J Bone Joint Surg Am* 2021;103:532–40.

19 Wolpert DH. The lack of a priori distinctions between learning algorithms. *Neural Computation* 1996;8:1341–90.

20 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.

21 Data policy. 2022 Available: https://www.who.int/about/policies/publishing/data-policy

22 Yang J, Robbins M, Reilly J, *et al*. The clinical effect of a rotator cuff retear: a meta-analysis of arthroscopic single-row and double-row repairs. *Am J Sports Med* 2017;45:733–41.

23 Heuberer PR, Smolen D, Pauzenberger L, *et al*. Longitudinal long-term magnetic resonance imaging and clinical follow-up after single-row arthroscopic rotator cuff repair: clinical superiority of structural tendon integrity. *Am J Sports Med* 2017;45:1283–8.

24 Plachel F, Siegert P, Rüttershoff K, *et al*. Long-term results of arthroscopic rotator cuff repair: a follow-up study comparing single-row versus double-row fixation techniques. *Am J Sports Med* 2020;48:1568–74.

25 Carbonel I, Martinez AA, Calvo A, *et al*. Single-row versus double-row arthroscopic repair in the treatment of rotator cuff tears: a prospective randomized clinical study. *Int Orthop* 2012;36:1877–83.

26 Lambers Heerspink FO, van Raay J, Koorevaar RCT, *et al*. Comparing surgical repair with conservative treatment for degenerative rotator cuff tears: a randomized controlled trial. *J Shoulder Elbow Surg* 2015;24:1274–81.

27 Schemitsch C, Chahal J, Vicente M, *et al*. Surgical repair versus conservative treatment and subacromial decompression for the treatment of rotator cuff tears: a meta-analysis of randomized trials. *Bone Jt J* 2019;101-B:1100–6.