BMJ Open

# Development and training of a machine learning algorithm to identify patients at risk for recurrence following an arthroscopic Bankart repair (CLEARER): protocol for a retrospective, multicentre, cohort study

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

**To cite:** van Spanning SH, Verweij LPE, Allaart LJH, *et al*. Development and training of a machine learning algorithm to identify patients at risk for recurrence following an arthroscopic Bankart repair (CLEARER): protocol for a retrospective, multicentre, cohort study. *BMJ Open* 2022;**12**:e055346. doi:10.1136/bmjopen-2021-055346

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2021-055346).

Received 13 July 2021
Accepted 28 April 2022



For numbered affiliations see end of article.

**Correspondence to**
Ms Sanne H van Spanning; shvanspanning@gmail.com

## ABSTRACT

**Introduction** Shoulder instability is a common injury, with a reported incidence of 23.9 per 100 000 person-years. There is still an ongoing debate on the most effective treatment strategy. Non-operative treatment has recurrence rates of up to 60%, whereas operative treatments such as the Bankart repair and bone block procedures show lower recurrence rates (16% and 2%, respectively) but higher complication rates (<2% and up to 30%, respectively). Methods to determine risk of recurrence have been developed; however, patient-specific decision-making tools are still lacking. Artificial intelligence and machine learning algorithms use self-learning complex models that can be used to make patient-specific decision-making tools. The aim of the current study is to develop and train a machine learning algorithm to create a prediction model to be used in clinical practice—as an online prediction tool—to estimate recurrence rates following a Bankart repair.

**Methods and analysis** This is a multicentre retrospective cohort study. Patients with traumatic anterior shoulder dislocations that were treated with an arthroscopic Bankart repair without remplissage will be included. This study includes two parts. Part 1, collecting all potential factors influencing the recurrence rate following an arthroscopic Bankart repair in patients using multicentre data, aiming to include data from >1000 patients worldwide. Part 2, the multicentre data will be re-evaluated (and where applicable complemented) using machine learning algorithms to predict outcomes. Recurrence will be the primary outcome measure.

**Ethics and dissemination** For safe multicentre data exchange and analysis, our Machine Learning Consortium adhered 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. No Institutional Review Board is required for this study.

## STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ Data will be obtained from global databases of all authors included in the Machine Learning Consortium, aiming to include data from over 1000 patients.
⇒ Retrospective studies are less suitable to train machine learning algorithms than prospective studies due to missing data through incomplete record keeping and possible confounding factors.
⇒ Studies with different designs will be included. By combining data gathered by different studies to create one database, definitions may differ and, therefore, make it impossible to pool some of the data.
⇒ Due to the collection of individual patient data by previously published studies, variation in definitions may cause a significant source of bias.

## INTRODUCTION

Anterior shoulder dislocation is a common injury, with a reported incidence of 23.9 per 100 000 person-years.[1] Shoulder dislocations limit patients in their daily routine and participation in sports, cause irreversible damage to the shoulder joint and are associated with high costs.[2 3] There is an ongoing debate on the most effective treatment strategy to prevent recurrence. Non-operative treatment of first-time dislocations has recurrence rates of up to 60%, whereas operative treatment such as the arthroscopic labrum repair and bone block procedures have lower recurrence rates (16% and 2%, respectively).[4 5] However, the complication rates for bone block procedures compared with arthroscopic labrum repair (up to 30% and <2%, respectively) are higher and therefore preoperative

counselling with determination of the most suitable treatment is important in avoiding unnecessary risk of complications.[6 7] Methods to determine risk of recurrence have been developed, including the instability severity index score (ISIS), glenoid morphology (ie, concavity, version, inclination), an off-track Hill-Sachs lesion and translation of the humeral head.[8–12] However, a patient-specific decision-making tool is still lacking.

The self-learning complex models used by artificial intelligence (AI) and machine learning (ML) algorithms express high levels of intelligence without human error and are therefore highly suitable to be used for interpretation of images, pathology slides and patient-specific decision-making tools.[13–17] Hendrickx and colleagues recently developed a prediction model based on ML algorithms to estimate acute and late complications after intramedullary nailing of a tibial shaft fracture.[16] In other words, the authors were able to use the computationally intensive methods of ML, to go from the 'traditionally' reported overall complication rate of a cohort to calculate the probability of a specific patient complication rate. This study resulted in an online prediction tool.

### Aim and objectives

The aim of the current study is to develop and train a ML algorithm to create a prediction model to be used in clinical practice—as an online prediction tool—to estimate recurrence rates following a Bankart repair. No studies have yet been published applying ML algorithms to systematically reviewed/collected data in this field.

## METHODS AND ANALYSIS

### Study design

This multicentre retrospective cohort study includes two parts.

### Part 1 — Collecting data

Part 1 involves collecting individual patient data of published studies that evaluated potential factors predisposing recurrence following an arthroscopic Bankart repair without remplissage. The authors of these studies will be contacted by email and will be included in the Machine Learning Consortium when they provide the original patient data of their cohort. Through this process, we aim to combine the individual patient data from the published studies and create an international cohort of over 1000 patients. The current study will use the collected patient data to create a ML algorithm that can estimate the probability of recurrence for an individual patient. To make a reliable algorithm, it is estimated that the data should include at least 100 recurrences. With a recurrence rate of 12% following arthroscopic Bankart repairs, it was estimated that a minimum of 1000 patients would be sufficient.[18] To identify relevant studies, a systematic approach was used searching PubMed, Embase/Ovid, Cochrane Database of Systematic Reviews/Wiley, Cochrane Central Register of Controlled Trials/Wiley, CINAHL/Ebsco and Web of Science/Clarivate according to the search terms used in Verweij *et al* (see online supplemental appendix 1 for the search strategy) from inception up to July 2021.[19] The systematic review by Verweij *et al* is completed and submitted for publication separately. All studies reporting on risk factors for recurrence following Bankart repairs were included. Studies published in languages other than English, Dutch and French were excluded. The inclusion criteria are patients treated with arthroscopic Bankart repair without remplissage for traumatic anterior shoulder instability with a minimum of 2-year follow-up. Shoulder instability is defined as either a complete dislocation or subluxation.[20] Exclusion criteria include patients who underwent previous stabilisation procedures or other surgical procedures to the ipsilateral shoulder than arthroscopic Bankart repair and patients with posterior, multidirectional or voluntary habitual instability.

### Part 2 — Machine learning

Part 2, the multicentre data will be re-evaluated (and where applicable complemented) using ML algorithms to predict outcomes. The statistician who performs the ML analysis will be blinded to the origin of the data.

#### Training data and test data

Eighty percent (80%) of all (>1000) patients included in the Machine Learning Consortium Database will be randomly allocated to the training data set and 20% to the test data set.

#### Output variables

Each ML algorithm will be trained to recognise patterns related to recurrence rates.

#### Input variables

For the primary outcome, a Random-Forest algorithm will be used to identify the variables with the highest predictive value from all available data points in the Machine Learning Consortium Database. The data points available include demographics (age, sex and ethnicity aiming to include >1000 patients with balanced demographics), patient-specific factors (eg, preoperative Body Mass Index, comorbidity, dominance), disease-specific factors (eg, affected side, number of preoperative dislocations, associated lesions) and surgical characteristics (eg, time from injury to surgery, surgeon level) (see online supplemental appendix 2 for the complete list of factors that will be collected from the electronic medical records).

#### Algorithms to be trained

It is not possible to know what ML algorithm will be most suitable to calculate recurrence following an arthroscopic Bankart repair.[21] However, based on previous studies, the following algorithms will be tested as prediction models for recurrence rates: Decision Tree Models; Support Vector Machine; Neural Network; Bayes Point Machine; Logistic Regression.[16 22–27]

### Training and testing of the algorithms

For each ML algorithm, 10-fold cross-validation will be repeated three times on the training data set (80%), to train the algorithms in recognising patterns related to recurrence following an arthroscopic Bankart repair, and to subsequently assess their predictive performance based on the following performance characteristics: area under the receiver operating characteristic (ROC) curve, calibration (calibration slope, calibration intercept) and Brier score will be calculated.[28] The model's predicted probability is 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.[29] 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.[28] The remaining 20% of the data will be used as a test set to assess the performance of the best performing ML algorithms based on 'unseen' data. The technical appendix, statistical code and data set will be published.

### External validation of the best-performing algorithm

Before incorporation into an online open access decision-making tool, the best-performing algorithm will be externally validated in a prospective database. The same performance metrics will be calculated as described above.

### Open-access clinical prediction tool

An open-access clinical prediction tool will be developed using the best performing algorithm.

### Patients and public involvement

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

## Current status

Currently, the study is at the finishing stage of collection data from global databases. Re-evaluation of the data using ML algorithms to predict outcomes will start in March 2022. The expected time of completion is by the end of 2022.

## ETHICS AND DISSEMINATION

For safe multicentre data exchange and analysis, our Machine Learning Consortium adhered to the WHO regulation 'Policy on Use and Sharing of Data Collected by WHO in Member States Outside the Context of Public Health Emergencies'.[30] The study results will be disseminated through publication in a peer-reviewed journal. No Institutional Review Board is required for this study.

## DISCUSSION

Operative treatment significantly reduces the risk of recurrent shoulder instability compared with non-operative treatment.[31] Patients with first-time dislocations who receive operative treatment are most often treated with labrum repair.[31] Risk factors associated with failure of an arthroscopic Bankart repair include young age (≤30 years), participation in competitive sports, multiple preoperative dislocations, >6 months surgical delay from first-time dislocation to surgery, ISIS >3 and associated lesions (Hill-Sachs, glenoid bone loss and anterior labrum periosteal sleeve avulsion (ALPSA) lesions.[19] It is impossible to take all these risk factors into account and make an objective decision on what treatment is most suitable. Several prediction tools have been developed to help counselling patients; however, these tools only provide an indicative overall score and are not patient specific.[8–12] AI and ML algorithms have shown potential to make a patient-specific decision tool.[16] Creating an online prediction tool for recurrence following an arthroscopic Bankart repair can help guide surgeons in selecting patients who benefit from this procedure. Patients with first-time anterior shoulder dislocations receive proper evidence-based information only in 29% of the cases.[32] An online prediction tool might elevate these numbers and makes it possible for shared decision-making based on objective measures.

The strength of this study is the great amount of data that will be gathered. Data will be obtained from global databases of all authors included in the Machine Learning Consortium, aiming to include data of >1000 patients. This study does have the limitation of being retrospective and therefore the study is dependent on the recordkeeping of each individual hospital. This may lead to a variance in listed variables per database, resulting in missing data. In addition, blinding of participants and personnel may have been addressed differently in every institute. Moreover, only risk factors that were identified in the literature were included.

**Author affiliations**
[1]Orthopaedic Surgery, OLVG, Amsterdam, Noord-Holland, The Netherlands
[2]Hand, Upper Limb, Peripheral Nerve, Brachial Plexus and Microsurgery Unit, Alps Surgery Institute, Annecy, France
[3]Department of Human Movement Sciences, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
[4]Orthopedic Surgery, Amsterdam Movement Sciences, Amsterdam UMC Locatie AMC, Amsterdam, North Holland, The Netherlands
[5]Academic Center for Evidence-based Sports Medicine (ACES), Amsterdam UMC Locatie AMC, Amsterdam, North Holland, The Netherlands
[6]Amsterdam Collaboration for Health and Safety in Sports (ACHSS), International Olympic Committee (IOC) Research Centre, Amsterdam UMC, Amsterdam, Netherlands
[7]Department of Orthopaedic & Trauma Surgery, Flinders Medical Centre, Flinders University, Adelaide, South Australia, Australia
[8]Roth McFarlane Hand and Upper Limb Center, Schulich School of Medicine and Dentistry, London, Ontario, Canada
[9]Department of Orthopaedic Surgery, Montpellier University Medical Center, Montpellier, Languedoc-Roussillon, France

**Collaborators** T Flinkillä, S Nakagawa, M Loppini, B R Waterman, B Owens, M Gotoh, H Nakamura, L Rossi, I Pasqualini, M Scheibel, M Minkus, J S Shaha, M A Ruiz Ibán, R T Li, A Lin, Y V Kleinlugtenbelt, J M Woodmass, P MacDonald, J Phadnis, A Stone, C Hatrick, T P van Iersel

**Contributors** SHvS, GAB, MPJvdB, LPEV and LJHA contributed to the conception, overall design and planning of the study. LAMH and JND contributed to the conception and design of the methods section, primarily focussing on the machine learning section and data analysis. GSA, TL and LL contributed to the design of

the methods section and primarily focussed on how the data should be collected and interpreted. SHvS, GAB, MPJvdB and LPEV contributed to writing the protocol. All authors revised this version of the protocol and gave final approval for it to be published. All authors ensure that questions related to the accuracy or integrity of any part of this protocol are appropriately investigated and resolved.

**Funding** The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

**Competing interests** GSA reports as 'financial activities outside the submitted work' to be a consultant for ConMed Linvatec. LL is a consultant for Depuy Stryker, received royalties from Depuy. TL is consultant for Depuy Mitek and Stryker. GAB received consultancy fees from Depuy-Synthes and Research Funds from SECEC, Vivalto Santé. The remaining authors certify that neither he or she has funding or commercial associations that might pose a conflict of interest in connection with the submitted article.

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

**Patient consent for publication** Not applicable.

**Provenance and peer review** Not commissioned; externally peer reviewed.



**ORCID iD**

Michel P J van den Bekerom http://orcid.org/0000-0002-5146-6148

## REFERENCES

1 Zacchilli MA, Owens BD. Epidemiology of shoulder dislocations presenting to emergency departments in the United States. *J Bone Joint Surg Am* 2010;92:542–9.
2 Verweij LPE, Pruijssen EC, Kerkhoffs GMMJ, *et al*. Treatment type may influence degree of post-dislocation shoulder osteoarthritis: a systematic review and meta-analysis. *Knee Surg Sports Traumatol Arthrosc* 2021;29:2312–24.
3 Zaremski JL, Galloza J, Sepulveda F, *et al*. Recurrence and return to play after shoulder instability events in young and adolescent athletes: a systematic review and meta-analysis. *Br J Sports Med* 2017;51:177–84.
4 Hovelius L, Rahme H. Primary anterior dislocation of the shoulder: long-term prognosis at the age of 40 years or younger. *Knee Surg Sports Traumatol Arthrosc* 2016;24:330–42.
5 Hurley ET, Lim Fat D, Farrington SK, *et al*. Open versus arthroscopic Latarjet procedure for anterior shoulder instability: a systematic review and meta-analysis. *Am J Sports Med* 2019;47:1248–53.
6 Williams HLM, Evans JP, Furness ND, *et al*. It's not all about Redislocation: a systematic review of complications after anterior shoulder stabilization surgery. *Am J Sports Med* 2019;47:3277–83.
7 Griesser MJ, Harris JD, McCoy BW, *et al*. Complications and re-operations after Bristow-Latarjet shoulder stabilization: a systematic review. *J Shoulder Elbow Surg* 2013;22:286–92.
8 Balg F, Boileau P. The instability severity index score. A simple pre-operative score to select patients for arthroscopic or open shoulder stabilisation. *J Bone Joint Surg Br* 2007;89:1470–7.
9 Di Giacomo G, Itoi E, Burkhart SS. Evolving concept of bipolar bone loss and the Hill-Sachs lesion: from "engaging/non-engaging" lesion to "on-track/off-track" lesion. *Arthroscopy* 2014;30:90–8.
10 Matsumura N, Oki S, Fukasawa N, *et al*. Glenohumeral translation during active external rotation with the shoulder abducted in cases with glenohumeral instability: a 4-dimensional computed tomography analysis. *J Shoulder Elbow Surg* 2019;28:1903–10.
11 Moroder P, Damm P, Wierer G, *et al*. Challenging the current concept of critical Glenoid bone loss in shoulder instability: does the size measurement really tell it all? *Am J Sports Med* 2019;47:688–94.
12 Yamamoto N, Itoi E, Abe H, *et al*. Contact between the glenoid and the humeral head in abduction, external rotation, and horizontal extension: a new concept of glenoid track. *J Shoulder Elbow Surg* 2007;16:649–56.
13 Chilamkurthy S, Ghosh R, Tanamala S, *et al*. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *Lancet* 2018;392:2388–96.
14 Ehteshami Bejnordi B, Veta M, Johannes van Diest P, *et al*. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 2017;318:2199–210.
15 Gulshan V, Peng L, Coram M, *et al*. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus Photographs. *JAMA* 2016;316:2402–10.
16 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.
17 Ting DSW, Cheung CY-L, Lim G, *et al*. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA* 2017;318:2211–23.
18 Yapp LZ, Nicholson JA, Robinson CM. Primary arthroscopic stabilization for a first-time anterior dislocation of the shoulder: long-term follow-up of a randomized, double-blinded trial. *J Bone Joint Surg Am* 2020;102:460–7.
19 Verweij LPE, van Spanning SH, Grillo A, *et al*. Age, participation in competitive sports, bony lesions, ALPSA lesions, >1 preoperative dislocations, surgical delay and ISIS score >3 are risk factors for recurrence following arthroscopic Bankart repair: a systematic review and meta-analysis of 4584 shoulders. *Knee Surg Sports Traumatol Arthrosc* 2021;29:4004–14.
20 Alkaduhimi H, Connelly JW, van Deurzen DFP, *et al*. High variability of the definition of recurrent Glenohumeral instability: an analysis of the current literature by a systematic review. *Arthrosc Sports Med Rehabil* 2021;3:e951–66.
21 Wolpert DH. The lack of a priori distinctions between learning algorithms. *Neural Comput* 1996;8:1341–90.
22 Karhade AV, Thio QCBS, Ogink PT, *et al*. Development of machine learning algorithms for prediction of 30-day mortality after surgery for spinal metastasis. *Neurosurgery* 2019;85:E83–91.
23 Fernandez-Delgado EC M, Barro S, Amorim D. Do we need hundreds of classifiers to solve real world classification problems? *J Mach Learn Res* 2014;15:3133–81.
24 Maroco J, Silva D, Rodrigues A, *et al*. Data mining methods in the prediction of dementia: a real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests. *BMC Res Notes* 2011;4:299.
25 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.
26 Wainer J. Comparison of 14 different families of classification algorithms on 115 binary datasets. *arXiv:1606.00930v1 [cs.LG]* 2016.
27 Oosterhoff JHF, Gravesteijn BY, Karhade AV. Feasibility of machine learning and logistic regression algorithms to predict outcome in orthopaedic trauma surgery. *J Bone Joint Surg Am* 2021.
28 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.
29 Stevens RJ, Poppe KK. Validation of clinical prediction models: what does the "calibration slope" really measure? *J Clin Epidemiol* 2020;118:93–9.
30 World Health Organization. WHO data policy, 2019. Available: https://www.who.int/publishing/datapolicy/en/ [Accessed 7 Sep 2019].
31 van Spanning SH, Verweij LPE, Priester-Vink S, *et al*. Operative versus Nonoperative treatment following first-time anterior shoulder dislocation: a systematic review and meta-analysis. *JBJS Rev* 2021;9. doi:10.2106/JBJS.RVW.20.00232. [Epub ahead of print: 23 09 2021].
32 Hutyra CA, Streufert B, Politzer CS, *et al*. Assessing the effectiveness of evidence-based medicine in practice: a case study of first-time anterior shoulder dislocations. *J Bone Joint Surg Am* 2019;101:e6.