2018-11

Development and external validation of a prognostic model for predicting poor outcome in people with acute ankle sprains: the SPRAINED study.

Keene, D

Health Technology Assessment
NIHR Journals Library

Development and external validation of a prognostic model for predicting poor outcome in people with acute ankle sprains: the SPRAINED study

Authors: David J. Keene[1], Michael M. Schlüssel[1], Jacqueline Thompson[1], Daryl A. Hagan[1], Mark A. Williams[2], Chris Byrne[3], Steve Goodacre[4], Matthew Cooke[5], Steve Gwilym[1,6], Phillip Hormbrey[6], Jennifer Bostock[7], Kirstie Haywood[5], David Wilson[6], Gary S. Collins[1] and Sarah E. Lamb[1,5] on behalf of the SPRAINED study group.

[1] Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom
[2] Department of Sport, Health Sciences and Social Work, Oxford Brookes University, Oxford, United Kingdom
[3] Faculty of Health and Human Sciences, University of Plymouth, Plymouth, United Kingdom
[4] School of Health and Related Research, University of Sheffield, Sheffield, United Kingdom
[5] Warwick Medical School, Unversity of Warwick, Coventry, United Kingdom
[6] Emergency Department, Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom
[7] Patient and Public Involvement representative

Corresponding author: Dr David Keene

Contact details:

Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Kadoorie Research Centre, Level 3, John Radcliffe Hospital, OX3 9DU

david.keene@ndorms.ox.ac.uk

+44 (0) 1865 223121

Competing interests:

Steve Goodacre is a member of the HTA Clinical Trials Board, HTA EESC Methods Group, HTA Funding Boards Policy Group (formerly CSG), HTA IP Methods Group, HTA Post board funding teleconference and HTA Prioritisation Group.

David Wilson declares personal fees from Oxford University.

Gary S Collins is a member of the HTA Commissioning Board.

Sarah E Lamb is Co-Director of Oxford Clinical Trials Unit and Professor of Rehabilitation at Warwick Clinical Trials Unit, both receiving funding from NIHR. Prof Lamb is also a member of the HTA Additional Capacity Funding Board, HTA End of Life Care and Add on Studies, HTA Prioritisation Group and the HTA Trauma Board.

Keywords: ankle, sprain, injury, prognosis, systematic review, nominal group technique, observational cohort, prognostic model

Abstract

Background

Ankle sprains are very common injuries. Although recovery can occur within weeks, around a third of patients have longer-term problems.

Objectives

To develop and externally validate a prognostic model for identifying people at increased risk of poor outcome after an acute ankle sprain.

Design

Development of a prognostic model in a clinical trial cohort dataset and external validation in a prospective cohort study.

Setting

Emergency departments in the UK.

Participants

Adults with an acute ankle sprain (within 7 days of injury).

Sample size

There were 584 clinical trial participants in the development dataset and 682 recruited for the external validation study.

Predictors

Candidate predictor variables were chosen based on availability in the clinical data set, clinical consensus, face-validity, a systematic review of the literature, data quality and plausibility of predictiveness of the outcomes.

Main outcome measures

Models were developed to predict two composite outcomes representing poor outcome. Outcome 1 was the presence of at least one of the following symptoms at 9 months after injury: persistent pain, functional difficulty or lack of confidence. Outcome 2 included the same symptoms as outcome 1 with the addition of recurrence of injury. Rates of poor outcome in the external dataset were lower than in the development dataset, 7 vs. 20% for outcome 1 and 16 vs. 24% for outcome 2, respectively.

Analysis

Multiple imputation was used to handle missing data. Logistic regression models, together with multivariable fractional polynomials, were used to select variables and identify

respective transformations of continuous predictors that best predicted the outcome based on a nominal alpha of 0.157, chosen to minimise over fitting. Predictive accuracy was evaluated by assessing model discrimination (c-statistic) and calibration (flexible calibration plot).

Results

Performance of the prognostic models in development dataset

Outcome 1 model combined c-statistic across the 50 imputed data sets was 0.74 (95%CI 0.70 to 0.79), with good model calibration across the imputed data sets. Outcome 2 model combined c-statistic across the 50 imputed data sets was 0.70 (95%CI 0.65 to 0.74), with good model calibration across the imputed data sets. Updating these models, which used baseline data collected at the emergency department, with an additional variable at 4 weeks after the injury (pain when bearing weight on the ankle) improved the discriminatory ability (c-statistic 0.77; 95%CI 0.73 to 0.82 for outcome 1 and 0.75; 95%CI 0.71 to 0.80 for outcome 2) and calibration of both models.

Performance of the models in the external dataset

The outcome 1 model combined c-statistic across the 50 imputed data sets was 0.72 (95%CI 0.66 to 0.79), with a calibration plot intercept of -0.71 (95%CI -0.98 to 0.44) and slope of 1.13 (95% CI 0.76 to 1.50) across the imputed data sets. The outcome 2 model combined c-statistic across the 50 imputed data sets was 0.63 (95%CI 0.58 to 0.69), with calibration plot intercept of -0.08 (95%CI -0.27 to 0.11) and slope of 1.03 (95%CI 0.65 to 1.42) across the imputed data sets. The updated models with the additional pain variable at 4 weeks had improved discriminatory ability over the baseline models but not better calibration.

Conclusions

The SPRAINED prognostic models performed reasonably, and showed benefit when compared to not using any model, so may assist clinical-decision making when managing and advising ankle sprain patients in the emergency department setting. The models use predictors that are simple to obtain.

Limitations

Data used to develop the prognostic models were from a randomised controlled trial so were

not originally intended to fulfil this aim. However, it was the best dataset available, with data on the symptoms and clinical events of interest.

Future work

Further model refinement, including re-calibration or identifying additional predictors may be required. Impact of implementing and using either model in clinical practice, in terms of acceptability and uptake by clinicians, and on patient outcomes should be investigated.

Study registration

Current Controlled Trials ISRCTN12726986

Funding details

NIHR Health Technology Assessment programme.

Word count: 500

Table of Contents

List of Tables

List of Figures

List of Boxes

List of Abbreviations

A&E	Accident & Emergency
ADL	Activities of Daily Living
AIC	Akaike Information Criterion
AUC	Area Under Curve
AUC ROC	Area Under Receiver Operating Characteristic Curve
BMI	Body Mass Index
CAST	Collaborative Ankle Support Trial
CRF	Clinical Dataset Form
ED	Emergency Department
EPV	Events-per-Variable
EQ-5D	EuroQuol Health related quality of life
FAOS	Foot and Ankle Outcome Score
FLP	Functional Limitation Profile
GCP	Good Clinical Practice
GP	General Practitioner
ICF	Informed Consent form
LOWESS	Locally weighted scatterplot smoothing
MFP	Multivariable Fractional Polynomial
MICE	Multiple Imputation by Chained Equations
MRC	Medical Research Council
MRI	Magnetic Resonance Imaging
NGT	Modified Nominal Group Technique
NHS	National Health Service
NIHR	National Institute for Health Research
OCTRU	Oxford Clinical Trials Research Unit
PIL	Participant Information Leaflet
PPI	Patient and Public Involvement
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROGRESS	Prognosis Research Strategy
QUIPS	Quality In Prognosis Studies
REC	Research Ethics Committee

ROM	Range of Motion
SMG	Study Management Group
SPRAINED	Synthesising a clinical Prognostic Rule for Ankle Injuries in the Emergency Department
SSC	Study Steering Committee
TRIPOD	Transparent Reporting of Multivariable Prediction Models for Individual Prognosis or Diagnosis
UK	United Kingdom
VAS	Visual Analogue Scale

Scientific Summary

Background

Ankle sprains are one of the most common musculoskeletal injuries. Although recovery can occur within weeks, up to a third of patients still have problems with their ankle at one year. In the acute phase there is no reliable way of establishing which patients are at risk of having a poor outcome.

Objectives

To develop prognostic models to be used in an acute setting to identify people at increased risk of poor outcome following an acute ankle sprain, and to evaluate the performance of these prognostic models in a prospective external validation study.

Methods

Research programme

A systematic review of prognostic factors for poor outcome after ankle sprain was conducted, followed by an expert consensus meeting, then development of prognostic models and external validation using data from a prospective observational cohort study.

Systematic review

The review was regsiterd on the PROSPERO database: CRD42014014471. Electronic databases were searched (AMED, EMBASE, Psych Info, CINAHL and SportDiscus, PubMed, Cochrane Register of Clinical Trials, Physiotherapy Evidence Database [PEDro]). Studies that had participants with acute ankle sprain, a longitudinal design and assessment of at lease one baseline prognostic factor were included. Eligibility assessments, and data extraction and risk of bias assessments (using the QUIPS tool) for included studies, were completed by two independent reviewers. A narrative synthesis was conducted.

Consensus meeting

A range of key stakeholders, including patient and public representatives, healthcare professionals, and clinical researchers representing a range of stakeholders involved in ankle sprain care and research in the UK NHS were invited to a one day consensus meeting.

A modified nominal group technique was used to facilitate the consensus process. The participants were divided into three groups (for which participants were pre-assigned to ensure a mixture of types of clinician, researchers and patient representatives) and were asked to rank mportant prognostic factors, some of which were nominated in the pre-meeting questionnaire. Discussions were immediately followed by a plenary session to feedback results of the group discussions to the entire group. A final session was a final voting process, in which each participant indicated whether each factor should be included in the prognostic model or not. The number of votes allowed was limited to 10 per individual. This was completed independently on paper questionnaires. Factors with 70% or more agreement across participants were considered as critically important to consider in the validation study

Development of the models

Data sources

Individual participant data from the existing Collaborative Ankle Support Trial (CAST [HTA ref. no. 01/14/10]) trial database were used to develop two prognostic models for poor outcome after ankle sprain. The CAST study was a pragmatic multicentre randomised controlled trial, with blinded assessment of the outcome, designed to estimate the clinical and cost-effectiveness of three different types of mechanical ankle support (Aircast brace, Bledsoe boot, or 10-day below-knee cast) for the initial management of severe ankle sprain (defined as an injury of grade 2 or 3, without fracture) compared with a double-layer tubular compression bandage.

The trial population comprised 584 individuals aged 16 years or older, attending emergency departments in the UK with an ankle sprain and an inability to fully bear weight on the injured ankle at the time of presentation to the emergency department (ED) and their review clinic appointment (the trial's baseline assessment). People were excluded if they presented with an ankle fracture (apart from a flake fractures of less than 2mm), any other recent fracture, any contraindication to any of the four arms of the trial, poor skin viability

preventing splinting or casting, or if their injury occurred more than 7 days before the first presentation at the recruiting ED. Participants were followed up at one, three and nine months after randomisation.

Candidate predictors

Twenty-three candidate predictor variables collected during the enrollment and baseline assessments of CAST were examined: age, sex, pain, previous injury, ankle stability tests, weight bearing ability, and severity of presenting clinical signs and symptoms. These candidate predictor variables were chosen based on clinical consensus, face-validity, systematic review of the literature, data quality and whether they were plausibly predictive of the outcomes.

Outcomes

The first prognostic model was developed to predict a composite outcome representing the presence of at least one of the following symptoms at 9 months after injury: persistent pain, functional difficulty or lack of confidence (outcome 1).

The second model was developed to predict a composite outcome representing the presence of at least one of the following symptoms or clinical events at 9 months after injury: persistent pain, functional difficulty, lack of confidence or recurrence of injury (outcome 2).

Sample size

Based on the CAST data set, between 20% (116/584) and 24% (140/584) of people attending an ED for an acute ankle sprain experienced a poor outcome at nine months. As this was the first study aiming to produce prognostic models to predict poor outcome after ankle sprain, we relaxed the recommendation of five events-per-variable (EPV) for the number of variables in a logistic regression models. We included 23 candidate predictors (with a total of 35 degrees of freedom) in the full model, which meant an EPV ratio of approximately 3 (116/35) and 4 (140/35) for Outcomes 1 and 2, respectively.

Analysis

Multiple imputation was used to handle missing data with 50 imputed datasets created. Based on a logistic regression model, multivariable fractional polynomials were used to select variables and identify respective transformations of continuous variables that best predicted the outcome. Inclusion of predictors in the final models were based on a nominal alpha of 0.157 (equivalent to the Akaike Information Cirterion) to reduce the risk of over fitting. Shrinkage of the regression coefficients and intercepts was performed based on heuristic shrinkage factors to correct for optimism. Predictive accuracy of the models was evaluated by assessing model discrimination (quantified by the c-statistic) and model calibration (flexible calibration plot).

External validation of the model

A prospective cohort study recruited people with acute ankle sprain attending one of ten NHS emergency departments across England over a period of nine months (from July 2015 to March 2016). There was no planned treatment allocation as in a randomised controlled trial and EDs provided usual care in accordance with local protocols. Data collection took place at the time of participant's presentation to any of the study recruiting sites (baseline) and subsequently at 4 weeks, 4 and 9 months after the initial injury. People with an acute ankle sprain (<7 days old) of any severity, aged 16 years or over and were invited to take part of the study. People with an ankle fracture (except a flake fracture <2mm) or other recent (<3 months) lower limb fracture were excluded. During this part of the study, a pilot of dynamic consent was also included in the later stages of recruitment. This gave participants an opportunity to use a website to interact with study information and update their preferences.

Results

Systematic review

Searches identified 4173 reports, with eight reports identified from additional sources. Thirty-six reports were assessed in full-text screening and nine studies were included in the review. One study was judged to be at low risk of bias, five at moderate risk, and three studies at high risk. [1-3] Incomplete and/or inadequate reporting standards was a common issue, for example it was difficult to identify whether prognostic factors were eliminated due to low statistical reasons or poor clinical utility. No studies reported on performance of the prognostic models using methods to assess internal or external validation. Across the included studies, a wide range of prognostic factors were investigated. The prognostic factors that were analysed

varied considerably between studies, with no common framing for prognostic factors being investigated across the studies. The identified studies and risk of bias assements were summarised to participants of the consensus meeting.

Consensus meeting

Thirty particioants attended the meeting. The final consenses voting identifies eight baseline factors that were deemed critical for the identification of people likely to have a poor recovery. These factors spanned pre-injury, sociodemographic, psychosocial and clinical assessment factors, encompassing a holistic biopsychosocial model of recovery. These factors were included in the data collection at baseline for the prospective observational study.

Performance of the prognostic models in development dataset

The first model predicted the presence of either persistent pain, functional difficulty or lack of confidence at nine months, included age, body mass index, pain when resting, pain when bearing weight, days from injury to assessment, whether the injury is a recurrent sprain and the ability to bear any weight on the injured ankle (outcome 1). The apparent performance on a complete-case analysis of the CAST data set showed a c-statistic of 0.82 (95% CI 0.75 to 0.89). The combined c-statistic across the 50 imputed data sets was 0.74 (95% CI 0.70 to 0.79), with good model calibration across the imputed data sets.

The second model predicted the presence of either persistent pain, functional difficulty, lack of confidence or recurrence of injury at nine months, included pain when resting, pain when bearing weight, days from injury to assessment, ability to bear any weight on the injured ankle, and whether the injury is a recurrent sprain (outcome 2). The apparent performance on a complete-case analysis of the CAST data set showed a c-statistic of 0.73 (95% CI: 0.66 to 0.81). The combined c-statistic across the 50 imputed data sets was 0.70 (95% CI 0.65 to 0.74), with good model calibration across the imputed data sets.

Updating these models, which used baseline data collected at the emergency department, with an additional variable at 4 weeks after the injury (pain when bearing weight on the ankle)

improved the the predictions of the models when compared using decision curve analysis plots.

Performance of the models in the external data set

Discrimination of the model for outcome 1 was similar to that observed in the development dataset (combined c-statistic across the 50 imputed data sets = 0.72; 95% CI: 0.66 to 0.79), but calibration was poor (combined calibration plot intercept = -0.71; 95% CI: -0.98 to 0.44 and slope = 1.13; 95% CI: 0.76 to 1.50). The model for outcome 2, combined c-statistic across the 50 imputed data sets was 0.63 (95% CI: 0.58 to 0.69); calibration plot intercept was -0.08 (95% CI: -0.27 to 0.11) and slope 1.03 (95% CI: 0.65 to 1.42). Discrimination of the updated model for outcome 1 was better (combined c-statistic = 0.78; 95% CI: 0.72 to 0.84), but calibration did not improve substantially (combined calibration plot intercept = -0.51; 95% CI: -0.78 to 0.24 and slope = 1.17; 95% CI: 0.86 to 1.48). The combined c-statistic for the updated model for outcome 2 was 0.64 (95% CI: 0.59 to 0.69); calibration plot intercept was 0.19 (95% CI: -0.01 to -0.38) and slope 0.68 (95% CI: 0.46 to 0.91). Finally, model performance was not better for the subgroup of participants with more severe injuries (ankle sprains of grades 2 and 3). All models were recalibrated (had their regression coefficients and intercepts re-estimated) using the external validation dataset.

A sub-study to pilot dynamic consent recruited 22 participants in the later phase of the prospective cohort study. Eight participants accessed their dynamic consent online webpage, none changed their consent decisions during the study.

Conclusions

Both models and respective updates provided good predictions of poor outcome for people with acute ankle sprain on the population used in their derivation. There was a slight decrease in model discrimination for both models when evaluated in prospectively collected external validation cohort. The models predicting presence of either persistent pain, functional difficulty, lack of confidence or recurrence of injury showed good calibration, whilst there was miscalibration of the model predicting persistent pain, functional difficulty or lack of

confidence. Recalibration of the models may be required to improve the accuracy of the predicted risks in other populations (within and outside the UK).

Implications for healthcare

The SPRAINED study prognostic models performed reasonably and showed benefit when compared to not using any model (consider all patients high risk of poor outcome), so may assist clinical-decision making when assessing and advising people with ankle sprains in the emergency department setting and when deciding on on-going management. The models benefit from using predictors that are simple to obtain during routine clinical assessment.

Recommendations for research

Further research to evaluate the performance of the models in other settings. Further refinement of the models, including external validation of the re-calibrated models or identifying additional predictors may be required. The impact of implementing and using either model in clinical practice, in terms of acceptability and uptake by emergency department staff and their impact on patient outcomes should be investigated.

Study registration

Current Controlled Trials ISRCTN12726986

Word count: 1295

Plain English Summary

Sprains of the ankle joint ligaments are very common injuries. Most people recover within a few weeks but up to 1 out of 3 have a poor outcome. A poor outcome includes problems such as ongoing pain, difficulties moving about, lack of confidence and further sprains. It is challenging to work out which people will recover or not as when people come into emergency departments for assessment the ankle is often too sore to examine thoroughly.

We developed a tool to help predict which people are at greater risk of a poor outcome. A tool like this would be useful as it would have the potential to assist clinical decision making and could help identify the people with an acute ankle sprain who could benefit from rehabilitation and monitoring.

The tool includes participant and injury characteristics such as age and the severity of pain reported. The tool had good accuracy within a group of participants that had been involved in a clinical trial. To see how the tool performed in another group of participants, we recruited 682 participants from ten emergency departments in the UK. We collected information on the participant and injury characteristics when the participant attended the emergency department and again 9 months afterwards. The research indicated the tool had a moderate ability to predict what would happen in the future. There were limitations to the accuracy of the predictions of the tool. However, our analyses suggest that using the tool was better than the scenario of not using a tool to identify people at risk of a poor outcome after ankle sprain.

To make use of the tool in clinical settings it would benefit from being set up on an web-based application or a similar mobile platform to enable clinicans to enter information about a patient and obtain a calculated a risk score. The prediction tool could also be improved by further research to see how well it performs in routine clinical care and in other settings.

Word count: 237

1. Introduction

1.1 Background

1.1.1 Incidence and costs

Ankle sprains are one of the most common musculoskeletal injuries. Between 3 to 5% of people who attend an emergency department (ED) in the UK do so as a result of sustaining a sprained ankle.[1] The vast majority of sprains are of the lateral (outside) ligaments, and vary from minor stretching (Grade I) to a complete tear (Grade III).[2] Recent systematic reviews conclude that approximately 30% of people still have problems one year after an ankle sprain, depending on the outcome measured and perhaps more importantly the sampling frame.[3, 4] Many studies are restrictive in their sampling frame, either to elite athletes or excluding younger and older people. Studies also have variable inception and follow up points which further complicates interpretation. In a large multi-centre randomised clinical trial conducted in EDs in the UK there was an estimate of 30% for poor outcome at 9 months.[5] Other studies agree that recovery plateaus around 9 months, and that residual disability after this point is likely to be persistent.[6] One potential consequence of ankle sprain, chronic ankle instability (CAI), is implicated in the development of ankle osteoarthritis, even without an acute osteochondral lesion.[7]

1.1.2 Usual clinical pathway

Assessment of the injury in the acute phase is challenging as the ankle is often so swollen and painful that it cannot easily be examined. Most people are advised to rest, elevate the ankle, apply ice and compression, and are often issued with crutches if bearing weight is difficult. The Ottawa guidance,[8] can be used to reduce the requirement for imaging without missing significant fractures. Where clinicians are concerned about the degree of injury, most health care providers operate a system of review within one week in a trauma clinic or equivalent injury service. This time frame allows some resolution of swelling, and greater certainty in ascertainment of injury severity and presence of other significant mechanical derangement.[9] Treatment options at this stage include further watchful waiting, diagnostics, intensive physiotherapy, and immobilisation. Surgery may be considered at this stage, although most

centres would initiate a test of conservative management first. We have previously published a survey of practice, which remains a reasonable reflection of current management in the UK.[1]

1.2 Value of a prognostic model

In this report we utilise the terms to describe different types of prognostic research recommended in the Prognosis Research Strategy (PROGRESS) framework.[10-12] A prognostic factor is '…any measure that, among people with a given health condition (that is, a start point), is associated with a subsequent clinical outcome (an endpoint)'.[12] A prognostic model is '…a formal combination of multiple predictors from which risks of a specific endpoint can be calculated for individual patients'.[10]

A prognostic model is indicated to identify people likely to experience poor outcome after ankle sprain. There are several ways in which better prognostic information could yield benefit to the NHS and to patients. Firstly, the ability to decide whether an early review is merited, and avoid unnecessary appointments. Secondly, the ability to target treatments and diagnostics more effectively and earlier in the recovery pathway. Finally, it could offer reassurance that people with ankle sprains who are not followed up are likely to be on a positive recovery trajectory. The high volume of people who have an ankle sprain is a key issue for management; cost savings will accrue if treatments are more efficiently targeted. Any prognostic model needs to be simple to complete in the ED, ideally administered in a single assessment.

1.3 Requirements of a prognostic model

To be considered useful, a prognostic model should be clinically meaningful, accurate (well calibrated with good discrimination) and generalizable (have been evaluated on a separate data set, referred to as external validation). Many prognostic models are developed using datasets that are too small, are not sufficiently generalizable, have questionable methodological quality (in particular no or limited evaluation of predictive accuracy) and use inadequate statistical methods.[10-12] Other issues in developing a prognostic model are variable selection, handling of missing data, timing and method (self-report versus clinical examination).

1.4 Existing prognostic models

Hiller et al. [13] authored a systematic review of factors associated with the risk of sustaining an ankle sprain but there are few studies evaluating risk of poor recovery after the injury. Other than recurrent sprain, few studies of post-injury recovery have considered wider pre-dispositional factors. In 2008, Van Rijn et al [3] published a systematic review of the clinical course and prognostic factors for recovery following ankle sprain. They found just one eligible study [14] which concluded that high levels of sports activity was a prognostic factor for residual symptoms (n=150).

To the best of our knowledge there are no externally validated prognostic models for acute ankle sprain (see our Systematic Review in Chapter 3 for details). Studies that have developed a prognostic model are of limited generalisability due to highly selective patient populations (e.g. exclusion of some the more severe types of injury, exclusion of older people and/or sole inclusion of athletic/military populations). We identified only one study judged as being of high methodological quality, but a limited number of candidate prognostic factors were assessed. Therefore, development of a new prognostic model, using robust methods, considering a range of plausible prognostic factors, and conducting external validation, is indicated.

Polzer et al. [4] developed a prognostic algorithm and treatment pathway, but substantial sections were based on expert judgements. A robustly developed and validated prognostic model could help better target treatment and improve outcomes for people who have an ankle sprain.[10] There are treatment options available for people who have poor prognosis. The most solidly evidenced based is physiotherapy [15]. Other options include surgical reconstruction of ligaments.[16]

1.6 Aim of the SPRAINED study

The aim of the SPRAINED study was to develop and validate a prognostic model for use in EDs for people with acute ankle sprain in order to identify those in whom recovery may be substantially prolonged or incomplete.

2. Overview of Methods

The development of a prognostic model for ankle sprains required a research programme that was conducted in two stages and used a variety of research methods. In order to facilitate an understanding of the development and validation of the prognostic model, the methods used across the research programme are outlined within this chapter. A full description of the methods for the different stages of the research are contained in their respective chapters.

2.1 Summary of Study Design

The SPRAINED study had two stages, summarised in Figure *1*.

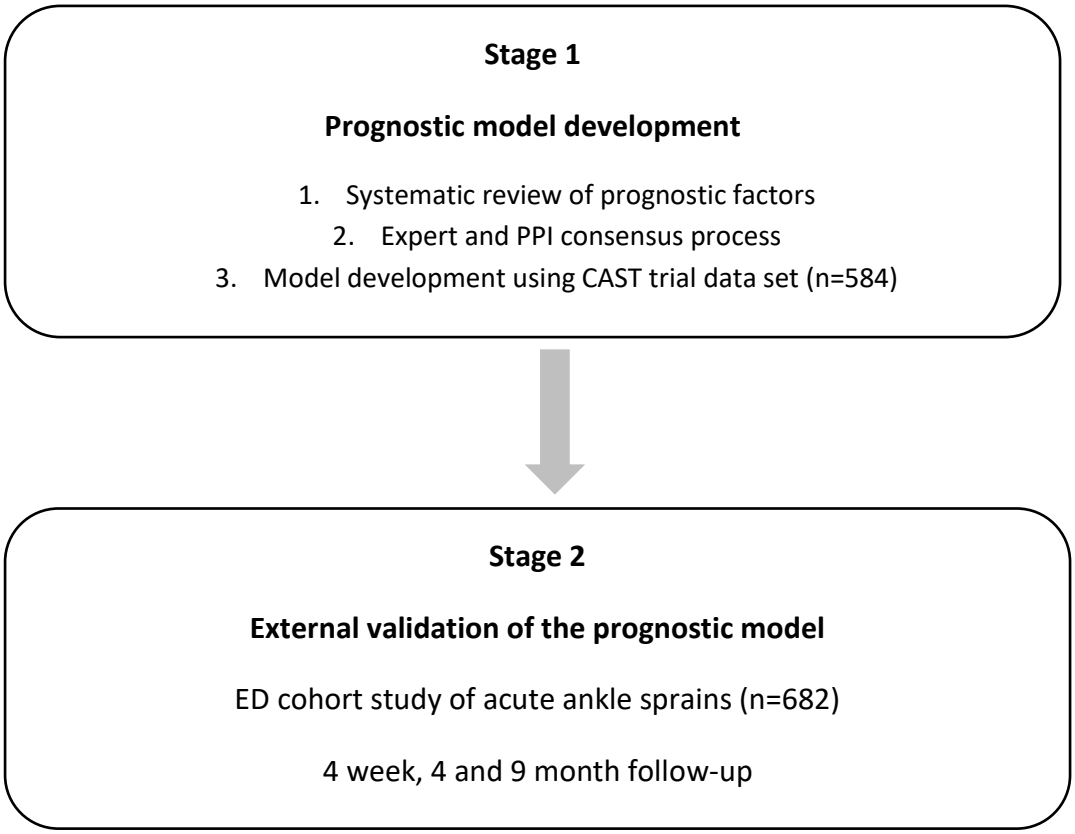


Figure 1: Stages of the SPRAINED study

2.2 Systematic review of the literature

A systematic review was conducted to identify prognostic factors of poor outcome following acute ankle sprain to identify variables which could be considered from the array available in the dataset described below (see section 2.4) and in the external validation study (see section 2.5).

2.3 Expert consensus process

A Modified Nominal Group Technique was used to gain consensus and information on preferences. Briefing papers were prepared and circulated to clinicians, patient and public representatives and clinical researchers, which contained lay summaries of the preliminary modelling elements completed (see below) and prognostic factors identified in the systematic review. The consensus element was achieved by a face-to-face meeting at which small groups were facilitated to answer a pre-specified set of questions. Two steps were used in this process, the first one for identification of issues and general discussion, and the second for resolution and consensus.

2.4 Developing a multivariable prognostic model from the Collaborative Ankle Support Trial (CAST) dataset.

CAST was the largest registered randomised clinical trial of interventions for moderate to severe ankle sprains to date worldwide (n=584).[17] Data was collected on a large number of candidate prognostic factors, including those identified as potentially important by clinical guidelines and consensus, and in previous multivariable analyses. The central research team had access to data at ED presentation, 2 to 3 days later, then at 1, 3 and 9 months after randomisation. Candidate prognostic factors were identified and included in multivariable models.

2.5 External validation of prognostic model in a prospective observational cohort study

We conducted a prospective observation study of 682 participants across 10 EDs across England between 20th July 2015 and 17th March 2016. In this final part of the research, the prognostic model developed in the earlier work was externally validated and recalibrated. A baseline proforma was used to obtain participant and clinical data on the candidate predictor variables, completed by the ED clinician at initial attendance. Follow up data were collected from participants at 4 and 9 months via telephone, postal or online questionnaires, and

included capture of persistent symptoms, the validated Foot and Ankle Outcome Score [18], health service resource use and health related quality of life (EuroQol EQ-5D-3L)[19]. An overview of this part of the study is contained in Figure 2. Data collected at baseline and 4 weeks after the injury were minimal, including mainly information on the predictors selected to compose the prognostic models developed for the two outcomes of interest. Data was also collected on a few baseline candidate predictors not present in the CAST dataset to determine whether the prognostic validity of the models could be improved by the addition of this extra information.

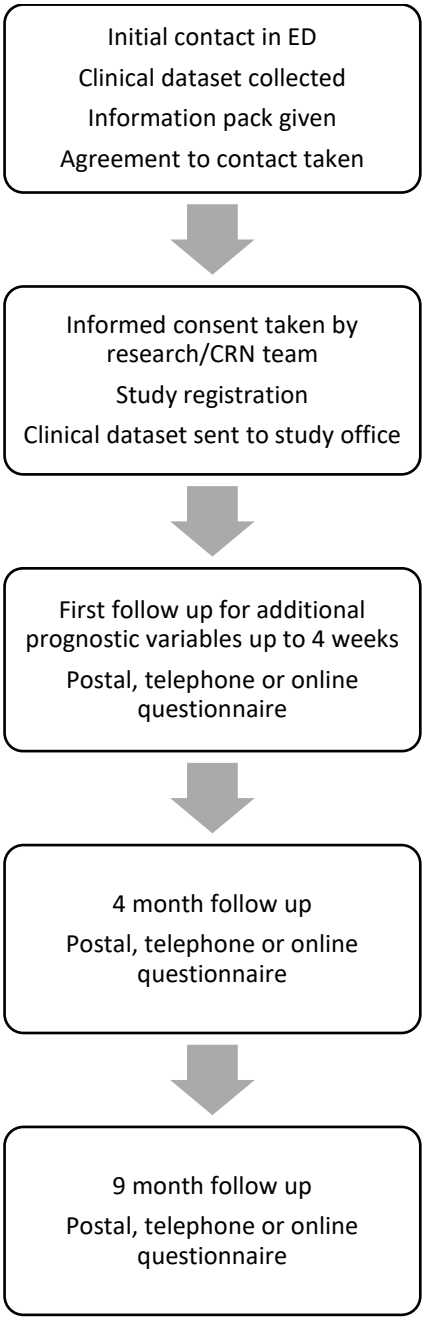


Figure 2: Flowchart of the SPRAINED cohort study

2.6 Pilot of sub-study of dynamic consent

Towards the end of recruitment for the external validation study, participants were offered the opportunity to join a dynamic consent pilot study. This gave participants an opportunity to

use a website to interact with study information and update their preferences. Details of this sub-study can be found in Appendix D.

2.7 Patient and public involvement (PPI)

The SPRAINED study recruited four PPI representatives from a process of open advert on the People in Research website, South Central Research Design Service e-bulletin, and the John Radcliffe Hospital Emergency Department in Oxford. Our appointed PPI representatives had experienced an ankle sprain and accessed NHS ED services. One representative agreed to be the PPI lead representative and is a co-applicant.

In order to develop and refine our application, we held a programme development meeting with our PPI representatives. Our representatives reviewed and contributed to ideas and provided feedback on our programmes of work including who the team should consist of; the experience of service use from the PPI perspective; the relevance of our proposed outcomes; acceptability of the research methods and the role of PPI input in developing and guiding the full application and research programme. We sought input on what were important outcomes and these influenced the make-up of our composite outcome measure.

PPI representatives were involved in piloting the pre-consensus meeting questionnaire and participated in the consensus meeting. We also had input from the lead PPI representative on interpretation of the results and in planning dissemination during a study management group meeting and they were involved in reviewing the report.

2.8 Ethical approval and monitoring

Ethics approval for the SPRAINED study was given by the National Research Ethics Committee (REC) (London - Chelsea), REC number 15/LO/0538, on 10th April 2015. This trial was conducted in accordance with the ethical principles that have their origin in the Declaration of Helsinki, and that are consistent with Good Clinical Practice (GCP) and the applicable requirements as stated in the Research Governance Framework for Health and Social Care. The Sponsor for the study (University of Oxford) reviewed study documents prior to ethics submission.

The Oxford Clinical Trials Research Unit's (OCTRU) assisted collaborating sites to obtain the necessary approvals to allow the study to take place within their NHS Trusts. The study was monitored and audited in accordance with the current approved protocol, GCP, relevant

regulations and standard operating procedures. A monitoring plan was developed according to OCTRU's standard operating procedures.

Study Steering Committee

The Study Steering Committee (SSC) provided overall supervision of the study on the behalf of the funder and was chaired by an Independent Member. The SSC abided by the OCTRU Standard Operating Procedure (accredited by the UKCRC Clinical Trials Unit Registration process) and SSC Charter. The SSC monitored study progress and advised on scientific credibility.

Study Management Group

The Study Management Group (SMG) was made up of SPRAINED study investigators, and staff working on the project within OCTRU and the Critical Care, Trauma and Rehabilitation Trials Group. This group oversaw the day-to-day running of the trial and met regularly.

2.9 Reporting

The Chief Investigator submitted progress reports throughout the study period to the REC Committee, host organisation and Sponsor.

The description of the development and external validation of the two models followed the Transparent Reporting of Multivariable Prediction Models for Individual Prognosis or Diagnosis (TRIPOD) statement.[20]

2.10 Summary of changes to the study protocol and analysis plan

The changes to the study protocol are summarised in Table 1. The planned analysis was refined during the programme of research in line with methodological developments and in response to the findings between the development and external validation stages of the study, these included:

- The primary outcome to represent ‘poor outcome’ after ankle sprain was clarified and pre-specified in the analysis plan. This was a result of the development of the research, considering the current literature, and expert and patient involvement input. The final definitions were two different combinations of clinical features reported at nine months after injury:
    - Outcome 1: was the presence of at least one of the following symptoms at 9 months after injury: persistent pain, functional difficulty or lack of confidence.
    - Outcome 2: included the same symptoms as outcome 1 with the addition of recurrence of injury.
- Net reclassification improvement (NRI) and integrated discrimination improvement (IDI) was not done, instead a decision curve analysis was undertaked (see Chapter 5)
- Decision curve analysis was not used to investigate the incremental value of a multivariable model with additional predictors not present in the development phase, as these predictors never reached that stage (see Chapter 6, section 6.3.4 Model recalibration)
- More than 15 candidate predictors were chosen for inclusion in the multivariable logistic regression models (see Chapter 5, sections 5.2.7 Sample size considerations and 5.2.8 Data modelling)
- Selection of the predictors in the final multivariable model was made by selecting those predictors with $p<0.157$ (equivalent to Akaike Information Criterion) instead of backwards elimination with $p<0.2$ as stopping rule, to minimise overfitting (see Chapter 5, section 5.2.8 Data modelling).
- Internal validation using bootstrap was not done (not being possible without suppressing one or more of the strategies used to prevent overfitting). Instead, the estimation of heuristic shrinkage factors for each developed model was performed and these were used to correct intercepts and beta coefficients for optimism (see Chapter 5, sections 5.2.10 Assessment of model performance and 5.2.11 Shrinkage)
- Model presentation was not simplified to a scoring system. The final models developed were fairly simple, with only a few predictors commonly screened in

clinical routine so instead, the equations with respective regression coefficients and intercepts were presented.

Table 1: Changes to the protocol during the study by version number.

Amendment No.	Protocol Version No.	Date issued	Details of Changes made
1	2.0	11 November 2015	Added information on Dynamic Consent bolt-on study.
2	3.0	03 March 2016	Clarification that follow-up time points are from study registration.
3	4.0	28 July 2016	Addition of electronic/online methods of data collection taking place for all follow up time points.

3. Systematic review

3.1 Introduction

A systematic review of prognostic factors for poor outcome following acute ankle sprain was conducted with the aim of identifying candidate variables which could be considered in the SPRAINED study. In this chapter the methods, results and key findings of the systematic review that contributed to the prognostic model development are detailed.

3.2 Methods

The review protocol was registered on PROSPERO: https://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42014014471.

3.2.1 Search strategy

Searches of the following electronic databases were conducted from inception to September 2016: AMED, EMBASE, Psych Info (via Ovid), CINAHL and SportDiscus (via EBSCOHost), PubMed, Cochrane Register of Clinical Trials. Relevant National Institutes of Health Medical Subject Headings (MeSH) terms were used where appropriate in these databases. Search strings containing terms for the health condition or body region were used in: Physiotherapy Evidence Database (PEDro), International Foot and Ankle Biomechanics, International Ankle Symposium, and Open Grey. No language restrictions were applied and the reference lists of included studies were screened for potentially relevant studies. The search strategy is available in Appendix A.

3.2.2 Eligibility criteria

Studies were eligible for inclusion if they had:

- A sample, or a separately analysed sub-group, with a clinical diagnosis of acute (less than or equal to 7 days from injury to assessment) lateral ankle ligament sprain
- A longitudinal design, with at least one follow-up time point
- Statistical assessment of at least one baseline prognostic factor on recovery outcomes.

Excluded were studies that included participants with ankle fracture (excluding flake fracture of less than 2mm) and other recent (less than 3 months since injury) lower limb injuries.

3.2.3 Data extraction

Titles and abstract were screened by two reviewers (JT, CB, MW). The Rayyan systematic review web application was used to manage screening.[21] Full-text articles for potentially eligible records were independently reviewed by two reviewers (JT, CB, MW). Data extraction and risk of bias assessments were completed independently by two reviewers (JT, CB). Discrepancies between reviewers were resolved by discussion, or in consultation with a third reviewer (MMS, DJK).

3.2.4 Risk of bias assessment

Study quality was assessed using the Quality In Prognosis Studies (QUIPS) tool,[22] which considers the six following domains of validity and risk of bias in prognostic factor studies:

1. study participation
2. study attrition
3. prognostic factor measurement
4. confounding measurement and account
5. outcome measurement
6. analysis and reporting

3.2.5 Data synthesis and reporting

A narrative synthesis was conducted, meta-analysis being considered inappropriate due to heterogeneity in the prognostic factors, outcome measures, follow-up durations, and limited number of studies. Follow-up time points from injury were grouped as short-term (≤8 weeks), medium-term (≤4 months), and long-term (>4 months).

3.3 Results

Searches identified 4173 reports, with eight reports identified from additional sources. Figure 3 shows the PRISMA flow diagram. Thirty-six reports were assessed in full-text screening. Of these, 27 were excluded, the remaining nine studies were included in the review.[23-31]

3.3.1 Study characteristics

Table 2 illustrates the characteristics of the included studies. Six studies were prospective cohorts, three were retrospective analyses of randomised controlled trials.[26, 28, 30] Three studies were based in the Netherlands, three in the USA, and one in other countries, England,

Northern Ireland, and Germany. Median sample size was 33 (range 20-553), and follow-up data ranged from one day to 12 months after injury. Three studies recruited high school or university athletes, the remainder were based in primary or secondary care.

3.3.2 Risk of bias assessment

Figure 4 shows the outcome of the risk of bias assessments. One study was judged to be at low risk of bias,[26] five at moderate risk of bias,[23, 25, 28, 30, 31] and three studies at high risk of bias.[24, 27, 29] Incomplete and/or inadequate reporting standards was a common issue, for example it was difficult to identify whether prognostic factors were eliminated due to low statistical reasons or poor clinical utility. No studies reported on performance of the prognostic models using methods to assess internal or external validation.

3.3.3 Prognostic factors identified

Prognostic factors included in the final models for each included study are shown in Table 3 (short-term), Table 4 (medium-term) and Table 5 (long-term).

Prognostic factors for short-term recovery (≤8 weeks)

Five studies investigated prognostic factors for short-term recovery (Table 3).[23-25, 29, 30] de Bie et al.[23] reported that having a baseline Ankle Function Score (AFS) ≤ 35 was a prognostic factor for non-recovery at two weeks. A combination of an AFS ≤ 35, higher severity grading by a doctor, and a higher palpation / ligament stress test score were included in the final model for the four week time point. van der Wees et al. [29] reported that a baseline AFS ≤ 40 was a prognostic factor for non-recovery at two weeks. Wilson & Gansneder [24] reported that greater range of motion loss and a greater extent of swelling were prognostic factors for a longer duration of disability. They also reported that greater functional limitations measured on an objective six-item weight-bearing activity score and a self-reported current athletic ability rating as prognostic factors.[24] Combining these ankle impairment and functional limitation prognostic factors produced an additive effect and explained 59% of the variance in disability duration.[24] Cross et al.[25] reported the baseline prognostic factors of lower self-reported physical function, self-reported global function, and objectively measured ambulation status as being associated with a greater number of days to return to sport.

O'Connor et al.[30] reported that baseline prognostic factors of greater age, more severe injury grade, and poorer weight bearing status were associated with lower subjective ankle function at four weeks.

Prognostic factors for medium-term recovery (≤4 months)

O'Connor et al.[30] reported greater age, poorer weight bearing status, and non-inversion injury mechanism were prognostic factors for poorer subjective function at 4 months follow-up (Table 4). The authors also identified medial joint line pain on palpation and pain on WB during ankle dorsiflexion at 4 weeks as prognostic factors for poorer subjective function at 4 months.[30]

Prognostic factors for long-term recovery (>4 months)

Three studies reported prognostic factors for long-term recovery (Table 5). [26-28] Akacha et al.[26] demonstrated that higher age and female sex were prognostic factors for slower and incomplete recovery.[26] Langner et al.[27] reported that more severe grading of injury, greater number of injured ligaments, and presence of a bone bruise determined (all determined with MRI) were associated with greater time to return to sports activities. van Middelkoop et al.[28] reported that none of the candidate prognostic factors measured at baseline were associated with outcome at 12 months follow-up.

Table 6 is on overview of all prognostic factors investigated, at what time point they were assessed, and indicates if the methods used within the study did or did not find evidence of an association between the variable and the outcome.

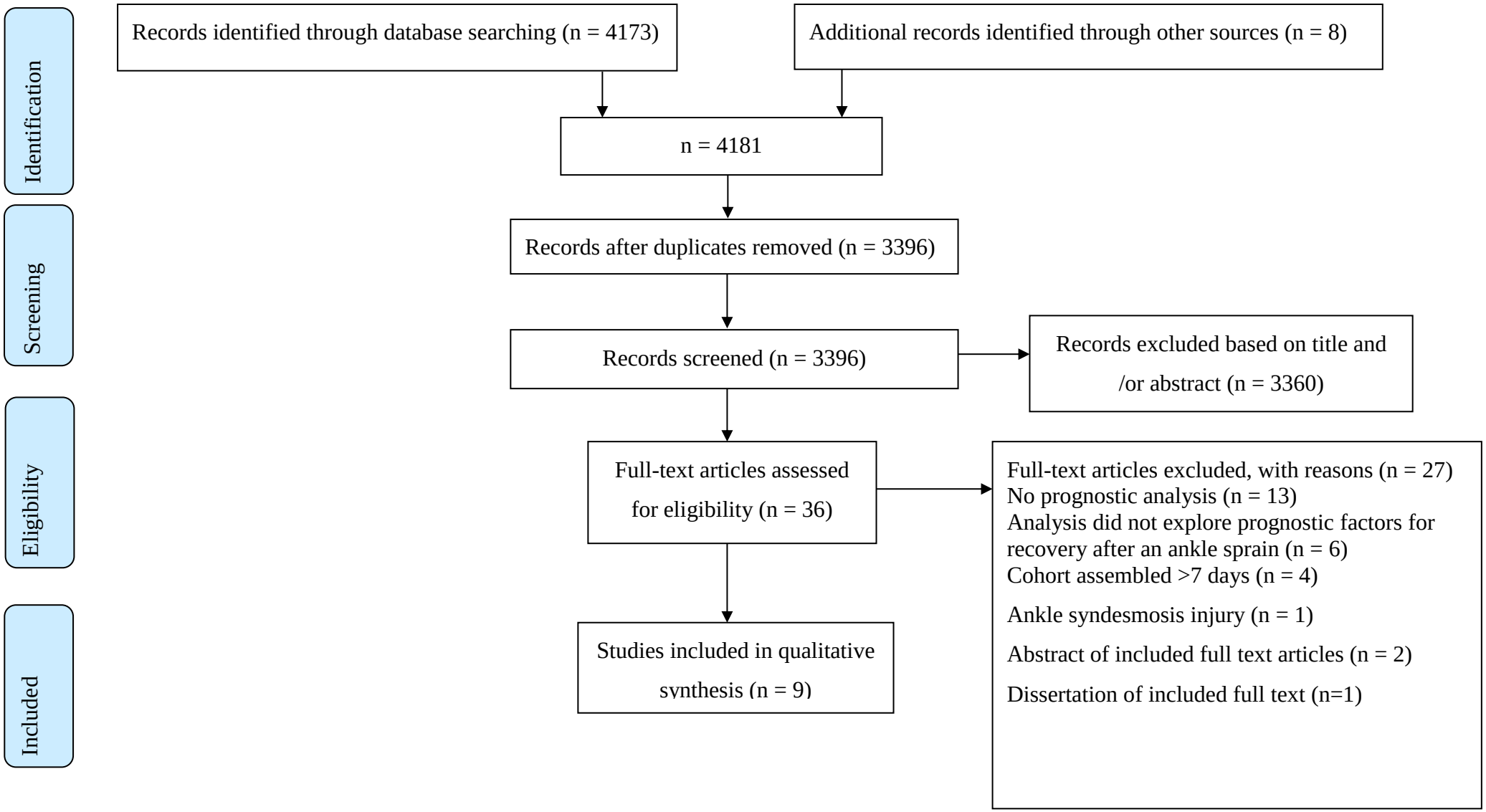


Figure 3: PRISMA flow diagram.

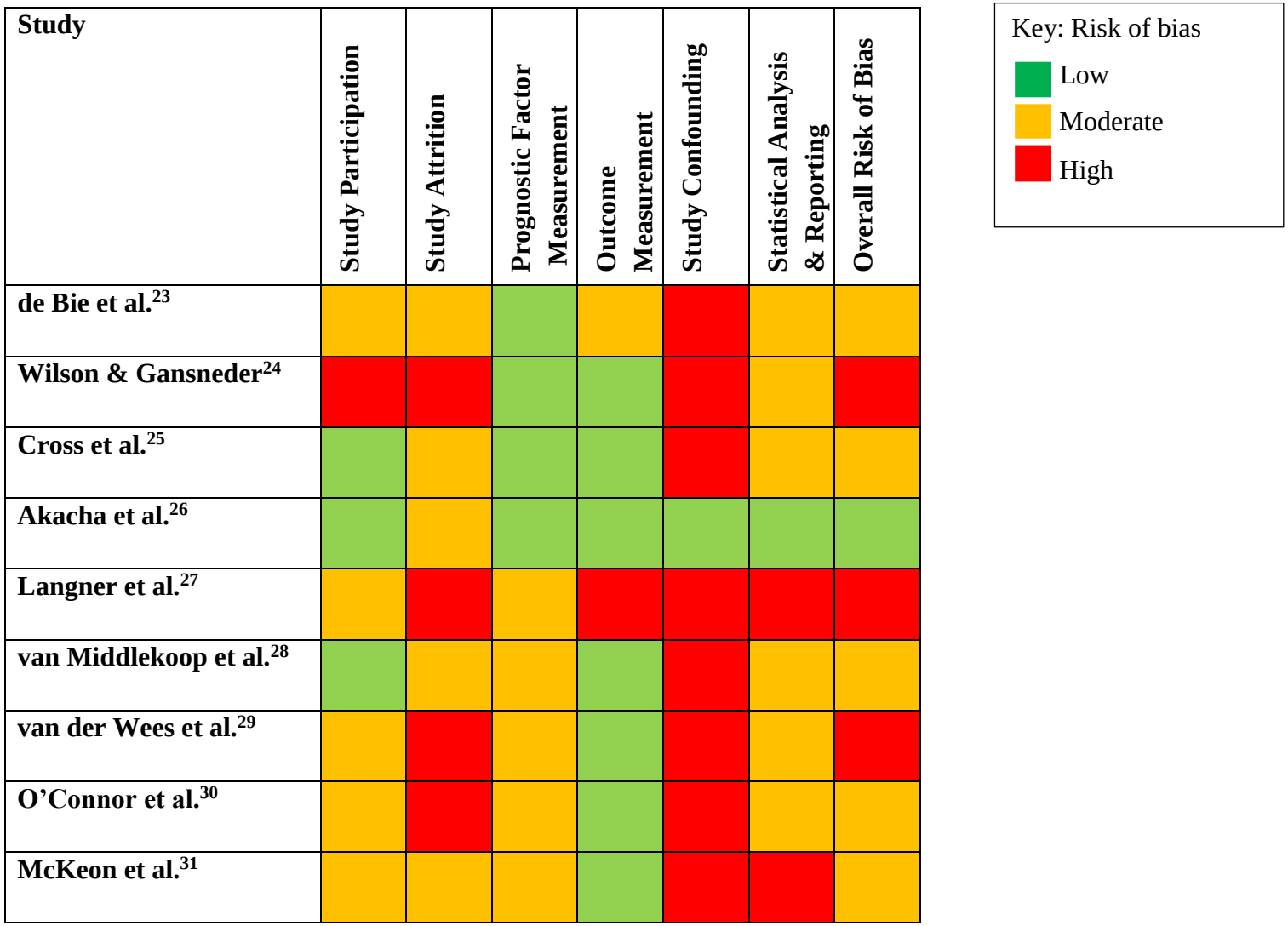


Figure 4: **Risk of bias assessment of the nine included studies according to the Quality in Prognosis Studies (QUIPS) tool.**

Table 2: **Key characteristics of included studies.**

Study	Design	Setting	Sample size	Sample characteristics	Time from injury to assessment	Injury severity	Follow-up
de Bie et al.[23]	Prospective cohort	Netherlands 1 x Hospital FAD	N = 35 at baseline N = 33 at 2 weeks N = 31 at 4 weeks	General population N = 22 M, 13 F 28 SD 10 (13-59) y	NR	NR	2 weeks 4 weeks
Wilson & Gansneder[24]	Prospective cohort	USA 1 x University	N = 24 at baseline N = 21 at follow-up	Athletes N = 13 M, 8 F 20 SD 2 y	67.8 SD 15.2 hours	Grade I, II	11.9 SD 6.6 days
Cross et al.[25]	Prospective cohort	USA 1 x University	N = 20 at baseline N = 20 at follow-up	Athletes N = 7 M, 13 F 19 SD 1 (18-21) y	≤ 24 hours	NR	14.7 SD 8.8 (3–40) days
Akacha et al.[26]	Retrospective cohort	England 8 x Hospital ED	N = 584 at baseline N = 553 at 4 weeks, 12 weeks, & 9 months	General population N = 321 M, 232 F 30 SD 11 (16-72) y	≤ 7 days	Severe (NWB at 3 days)	4 weeks 12 weeks 9 months
Langner et al.[27]	Prospective cohort	Germany 1 x Hospital ED	N = 38 at baseline N = 26 at 6 months N = NR at 12 months	General population N = 18 M, 20 F 38 SD 13 (20-75) y	< 24 hours	ATFL Grade I (27%), II (27%), III (46%)	6 months 12 months
van Middelkoop et al.[28]	Retrospective cohort	Netherlands 32 x GP 1 x Hospital ED	N = 102 at baseline N = 95 at 3 months N = 80 at 12 months	General population N = 59 M, 43 F 37 SD 12 (18-60) y	≤ 7 days	Mild (42%), moderate or severe (44%), unknown (14%)	3 months 12 months
van der Wees et al.[29]	Prospective cohort	Netherlands 20 x Primary care Physiotherapists	N = 107 at baseline N = 33 at 2 weeks	General population N = 65 M, 42 F 32 SD 14 y	8.7 SD 8.9 days ≤ 5 days N = 53 > 5 days N = 54	Light (50%), severe (50%)	2 weeks

Study	Design	Setting	Sample size	Sample characteristics	Time from injury to assessment	Injury severity	Follow-up
O'Connor et al.[30]	Retrospective cohort	Northern Ireland 1 x Hospital ED 1 x University sports injury clinic	N = 101 at baseline N = NR at 4 weeks N = 85 at 4 months	General population, athletes. N = 69 M, 31 F 27 SD 10 (16-58) y	< 7 days 40 SD 36 hours	Grade I (26%), II (63%), II+ (11%)	4 weeks 4 months
McKeon et al.[31]	Prospective cohort	USA 7 x High schools	N = 204 sprains at baseline N = 198 sprains in analysis	High school athletes	≤ 24 hours	Time to return to play. same day (23.7%), next day (21.2%), 3 days (29.3%), 7 days (11.6%), 10 days (8.6%), >22 days (5.6%)	Time to return to play. same day, next day, 3 days, 7 days, 10 days, 21 days, >22 days.

Abbreviations. FAD, first aid department; N, number; M, males; F, females; y, years; NR, not reported; NWB, non-weight bearing status; ED, emergency department; ATFL, anterior talofibular ligament; GP, general practice primary care; SD, standard deviation..

Table 3: **Prognostic factors for short-term (≤ 8 weeks) outcome in acute lateral ankle sprain.**

Study	Primary outcome measure	Variables in final model	Analysis	Prognostic factors in final models associated with short-term outcome
de Bie et al.[23]	Healed or not healed at 2 & 4 weeks. Healed = AFS >75 (0-100) & palpation / ligament stress test score <2 (0-12).	AFS (0-100) ≤35, doctor severity grading (0-10), palpation / ligament stress test score (0-12).	Multivariable logistic regression	2 weeks - baseline AFS ≤ 35 predicted recovery status. Sensitivity = 97%, specificity = 100%. 4 weeks - combined baseline AFS ≤ 35, severity grading, & palpation / ligament stress test score predicted recovery status. Sensitivity = 81%, specificity = 80%.
Wilson & Gansneder[24]	Number of days to return to full sports practice or competition (11.9 SD 6.6 days).	Joint swelling (ml), sagittal plane ROM loss (°), objective WB activity score (0-6), self-reported athletic ability score (VAS, 0-100).	Hierarchical regression	Combined swelling (β = -0.02) & ROM loss (β = -0.08). R^2 = 0.34, p = 0.023. Combined WB activity score (β = -0.55) & self-reported ability score (β = -0.39). R^2 = 0.33, p = 0.004. Combined swelling, ROM loss, WB activity score, & self-reported athletic ability score. R^2 = 0.59; p = 0.001.
Cross et al.[25]	Number of days to return to sport (14.7 SD 8.8 days).	SF36PF (0-100), Self-reported global function (0-100%), objective ambulation status (1-7).	Univariate regression, stepwise multivariable regression.	SF36PF - R^2 = 0.28, p = 0.016. Self-reported global function. R^2 = 0.22, p = .036. Objective ambulation status. R^2 = 0.22, p = 0.019. Combined SF36PF, self-reported global function, & objective ambulation status - R^2 = 0.34, p < .01.

Study	Primary outcome measure	Variables in final model	Analysis	Prognostic factors in final models associated with short-term outcome
van der Wees et al.[29]	Global perceived effect ≥2 (1 = recovered, 2-7 = not recovered) at 2 weeks.	AFS (0-100) ≤40.	Sensitivity & specificity	2 weeks - baseline AFS ≤ 40 predicted recovery status. Sensitivity = 76%, specificity = 63%.
O'Connor et al.[30]	Karlsson function score (0-100) at 4 weeks.	Age (years), injury grade (1, 2, 2+), WB status (FWB, FWB with pain, PWB, NWB).	Univariate regression, step-wise multivariable regression.	4 weeks - combined age (β = -0.32, p = .001), injury grade (β = -0.23, p = 0.003), & WB status (β = -0.34, p = 0.038). R^2 = 0.34, p < 0.01.

Abbreviations. AFS, ankle function score; °, degrees; VAS, visual analogue scale; β, standardised beta; ROM, range of motion; WB, weight-bearing; SF36PF, short form-36 physical function scale; FWB, full weight-bearing status; PWB, partial weight-bearing status; NWB, non-weight-bearing status; SD, standard deviation.

Table 4: **Prognostic factors for medium-term (≤ 4 months) outcome.**

Study	Primary outcome measure	Variables in final model	Analysis	Prognostic factors in final models associated with medium-term outcome
O'Connor et al.[30]	Karlsson ankle function score (0-100) at 4 months.	Baseline. Age (years); WB status (FWB, FWB with pain, PWB, NWB); injury mechanism (inversion / other). 4 weeks - pain on WB ankle DF; medial joint line pain (yes/no).	Univariate regression, step-wise multivariable regression.	4 months - baseline combined age (β = -0.26, p = 0.01), WB status (β = -0.23, p = 0.25), & injury mechanism (β = -0.25, p = 0.17), adjusted R^2 = 0.34, p < 0.01. 4 months. 4 week combined pain on WB ankle DF (β = 0.60, p < 0.001), medial joint line pain (β = 0.24, p = 0.07), adjusted R^2 = 0.49, p < 0.01.

Abbreviations. WB, weight-bearing; FWB, full weight-bearing status; PWB, partial weight-bearing status; NWB, non-weight-bearing status; DF, dorsiflexion, β, standardised beta; SD, standard deviation.

Table 5: **Prognostic factors for long-term (> 4 months) outcome.**

Study	Primary outcome measure	Variables in final model	Analysis	Prognostic factors in final models associated with long-term outcome
Akacha et al.[26]	FAOS-S (0-100, 0 = extreme symptoms, 100 = no symptoms).	Age, sex.	Non-linear mixed model	Greater age and female sex associated with slower and incomplete recovery. Greater age (β = -0.01, 95% CI -0.12 to -0.004) Female (β = -0.06, 95% CI -0.01 to -0.002)
Langner et al.[27]	Time to return to sports activities.	MRI grading of ligamentous injury (1-3, 1 = stretching, 2 = partial tear, 3 = complete tear); number of injured ligaments; presence of bone bruise.	Multivariable regression	MRI grading of ligamentous injury, R^2 = 0.45, p < 0.01. Number of injured ligaments, R^2 = 0.35, p < 0.01. Bone bruise, R^2 = 0.32, p < 0.01.
Van Middelkoop et al.[28]	Self-reported recovery (NRS, 0-10. 0 = not recovered; 10 = completely recovered) at 12 months.	Re-sprain within 3 months; pain at rest at 3 months (NRS, 0-10).	Multivariable regression	12 months. Re-sprain within 3 months (β = -1.64, 95% CI -3.11 to -0.16); pain at rest at 3 months (β = -0.69, 95% CI -1.08 to -0.29).

Abbreviations. FAOS-S, foot and ankle outcome score symptoms subscale; β, standardised beta; 95% CI, 95% confidence interval; MRI, magnetic resonance imaging; NRS, numerical rating scale.

Table 6: Summary of all formally investigated prognostic factors across the included studies.

Prognostic factor assessed	de Bie et al.[23]	Wilson & Gansneder[24]	Cross et al. [25]	Akacha et al. [26]	Langner et al. [27]	van Middelkoop et al. [28]	van der Wees et al. [29]	O'Connor et al.[30]	McKeon et al.[31]
Age				LT ✓		LT ✗		ST ✓ MT ✓	
Ankle function score	ST ✓					LT ✗		✗	
Active range of motion for injured leg			ST ✗						
Active range of motion uninjured leg			ST ✗						
Body mass index						LT ✗		ST ✗	
Clinical severity grading	ST ✓							ST ✓	
Dorsiflexion muscle strength for injured leg			ST ✗						
Dorsiflexion muscle strength for uninjured leg			ST ✗						
Gait pattern						LT ✗			
Sex				LT ✓		LT ✗		✗	✗
Global function question			ST ✓						
Global perceived effect (GPE)							ST ✓		
Injury grade						LT ✗		ST ✓	

Prognostic factor assessed	de Bie et al.[23]	Wilson & Gansneder[24]	Cross et al. [25]	Akacha et al. [26]	Langner et al. [27]	van Middelkoop et al. [28]	van der Wees et al. [29]	O'Connor et al.[30]	McKeon et al.[31]
Instability						LT ✗			
Mechanism of injury								MT ✓	
Medial joint line pain on palpation								MT ✓	
Molander ankle score							ST ✓		
MRI grading of bone bruise					LT✓				
MRI grading of number of injured ligaments					LT✓				
MRI severity grading of ligamentous injury					LT✓				
Pain at rest						LT ✗			
Pain at rest at 3 months						LT✓			
Pain on weight bearing ankle dorsi-flexion								MT ✓	
Pain while running						LT ✗			
Pain while walking						LT ✗			
Palpation score	ST ✓								
Patient-specific complaints							ST ✓		
Plantar flexion muscle strength for involved leg			ST ✗						

Prognostic factor assessed	de Bie et al.[23]	Wilson & Gansneder[24]	Cross et al. [25]	Akacha et al. [26]	Langner et al. [27]	van Middelkoop et al. [28]	van der Wees et al. [29]	O'Connor et al.[30]	McKeon et al.[31]
Plantar flexion muscle strength for uninvolved leg			ST ✗						
Previous ankle sprain history								✗	ST ✗
Reduced range of motion		ST ✓							
Referrals									ST ✗
Re-sprain within 3 months						LT ✓			
Return to full sports activities					LT ✗				
Return to work on full duties					LT ✗				
Self-reported global function							ST ✓		
Self-reported athletic ability		ST ✓							
Self-reported physical limitations			ST ✓						
Setting						LT ✗			
Short Form-36 Physical Function Scale			ST ✗						
Side hop test								✗	
Sport load						LT ✗			ST ✗
Subjective recovery						LT ✗			
Swelling		ST ✓				LT ✗			

Prognostic factor assessed	de Bie et al.[23]	Wilson & Gansneder[24]	Cross et al. [25]	Akacha et al. [26]	Langner et al. [27]	van Middelkoop et al. [28]	van der Wees et al. [29]	O'Connor et al.[30]	McKeon et al.[31]
Treatment / randomisation group				LT ✗		LT ✗			
Visual analogue pain scale			ST ✗						
Activity score		ST ✓							
Weight bearing status	ST ✓		ST ✓					ST ✓ MT ✓	
Work load						LT ✗			

Key: ST, short-term, medium-term; MT, medium term; LT, long-term, red, study judged as being high risk of bias; amber, study judged as being medium risk of bias; green, study judged as being low risk of bias.

✓ Prognostic factor assessed, included in final models and evidence of statistical association with outcome

✗ Prognostic factor assessed and but not found to be statistically associated with outcome due to a range of given reason, primarily dropped in univariable analysis prior to multivariable modelling, or dropped during multivariable modelling analysis.

3.4 Discussion

Across the included studies, a wide range of prognostic factors were investigated. The prognostic factors that were analysed varied considerably between studies, with no common framing for prognostic factors being investigated across the studies. Due to the methodological issues identified in the majority of included studies it is important that the evidence of statistical associations between the candidate prognostic factors and outcomes reported should be interpreted with caution.

Age was identified as an independent prognostic factor in one study with low risk of bias [26], and another study [30] with moderate risk of bias. Higher baseline age was associated with poor recovery at short,[30] medium [30] and long term follow-up.[26] Injury severity was reported as a prognostic factor in two studies by clinical symptoms [23, 30], but in another study [27] MRI was used to grade severity. Clinical assessments may be subjective to some extent, but sensitive investigations such as MRI are not readily available in acute settings. Furthermore, the insufficient evidence for diagnostic imaging findings as prognostic factors highlights that structural pathology may not be indicative of clinical severity. A lack of association between structural changes in the ankle and persistent ankle impairments has been reported.[32]

Measures obtained somewhat later after injury appeared to have better prognostic value than in the early acute stage, indicating that the timing of measurement taken can influence the value of prognostic factors. The challenge of using measures taken at some delay after injury arguably have less clinical utility when developing a prognostic model to identify people at risk of poor outcome as this could delay decisions about monitoring and early intervention.

Limitations of ankle sprain prognostic factor studies

The majority of the studies included had short-term follow-up durations, when symptoms are still prominent and resolving, hence recovery at this stage can be quite variable. Methodological shortcomings were evident across the studies, none reported assessments of interval validity or attempted external validation of their models. Adjustments for confounding factors such as time since injury were not employed. Regression analyses were often not reported in sufficient detail to identify whether prognostic factors were eliminated due to small sample size or poor clinical utility. Two studies [23, 29] dichotomised a continuous outcome measure, with the cut-points used not being well justified or pre-specified.

The study judged as being high quality tended to report conservative estimates of associations between predictors and outcome.[26] However, a limited range of prognostic factors were investigated.

Although a wide range of prognostic factors have been investigated, the limitations of previous studies highlight the need for large scale studies that employ robust prognostic research methods [10] and adhere to recognised reporting guidelines.[20] The systematic review did provide some evidence to inform the decision making processes within the consensus exercise.

4. Consensus meeting

4.1 Introduction

In this chapter, we report the findings of a UK-based consensus meeting that sought to help determine which prognostic factors that should be considered as candidates in the SPRAINED prognostic model. There is no universally accepted method on how best to develop a prognostic model.[33] Current recommendations for this include using variables that have already demonstrated prognostic value (see Chapter 3) and including other clinically plausible variables.[34] Therefore our aim was to use a triangulation of methods to ensure a comprehensive representation of prognostic factors was considered for inclusion in the SPRAINED prognostic model. Firstly, we used the results from preliminary analyses of data from a previous large-scale clinical trial involving people with acute lateral ankle sprains attending EDs[5] to explore what prognostic factors could be important for predicting recovery at 9 months after injury (see Chapter 5 for details). Secondly, we used the results of a systematic literature review of studies investigating prognostic factors for recovery (see Chapter 3) to elucidate which prognostic factors had been previously identified and the level of evidence for these factors. Thirdly, we used a consensus meeting to triangulate these with clinical and patient/public opinion.

In order to optimise the development of the SPRAINED prognostic model, we aimed to obtain interpretations of these sources of evidence from a range of key stakeholders and achieve consensus on which baseline and delayed prognostic factors should be included in the prognostic model which was to be evaluated in the external validation study (Chapter 6).

4.2 Methods

A variety of methodologies for achieving consensus exist (e.g. Delphi methods, Discrete Choice Experiments, and face-to-face methods) but there is no agreed optimum approach on how to synthesize judgements when a state of uncertainty exists.[35] We chose to use a modified nominal group technique because it provides a structured scientific process, which incorporates private views of individual participants as well as facilitated discussion leading to an aggregated group judgement. The modified nominal group technique (NGT) was

originally reported by Delbecq and colleagues [36] and has since been refined and utilised in a range of musculoskeletal research settings, most notably the Outcome Measures in Rheumatology (OMERACT) initiative.[37] In modified NGT, individual participants express views via a questionnaire before a face-to-face meeting where findings are fed back, structured discussion is facilitated, and then a final vote is taken of individual views.[38]

4.2.1 Participants

We aimed to recruit a range of key stakeholders, including patient and public representatives, healthcare professionals, and clinical researchers in order to represent a range of parties involved in ankle sprain care and research in the UK NHS. We invited 30 individuals to participate, including a variety of healthcare professionals from across the UK who worked in ambulance services, general practice, radiology, emergency and trauma surgery departments, as well as clinical researchers. We also aimed to recruit patient and public representatives from the south central area of the UK who had experience of an ankle sprain or were able to represent an individual or group that had. We placed adverts in local Oxford supermarkets, the John Radcliffe Hospital Emergency and Trauma Outpatient Departments, the website https://www.peopleinresearch.org/ and the NIHR Research Design Service South Central's mailing list.

4.2.2 Facilitators

The SPRAINED study team facilitators were guided by a lead facilitator (KH) with experience in conducting modified NGT processes in musculoskeletal research [39]. Additional facilitators were provided with a standardised brief to follow during the meeting and supervised by the lead facilitator.

4.2.3 Consensus process

We conducted the consensus process in three main stages:

Preparation and supply of information

Participants were provided with an electronic information pack 10 days before a face-to-face meeting. This pack consisted of a summary of the SPRAINED study to date, findings from the systematic review of prognostic factors for acute ankle sprains (Chapter 3), preliminary findings from statistical modelling of the CAST trial dataset (Chapter 5) and a pre-meeting questionnaire (described in the next section).

Completion of pre-meeting questionnaire

The pre-meeting questionnaire was developed with two key sections (see Appendix B). Firstly, a section to gain the participant's opinions on which were important prognostic factors for recovery following acute ankle sprain. Data from the systematic review and statistical modelling was utilised to generate a list of 14 pre-defined factors. Participants were also given the facility to nominate unlisted factors. Response options were provided in the form of the 9-point GRADE scale (1 to 3 = not important; 4 to 6 important, but not critical; 7 to 9 critical) [40] with importance defined as 'how important do you think [prognostic variable] is a factor in recovering from an ankle sprain?' There was also a response box 'don't know' provided as an option.

A second section was developed to enquire about when and how additional delayed information should be taken. This was informed by studies included in the systematic review (Chapter 3) that demonstrated information collected after baseline improved prognostic model accuracy. The questions were in the form 'If we were to collect further information like this, how many weeks after the initial visit do you think we should collect this information?' (response options ranged from '1 week' to '6 weeks') and 'How should we collect this information?' (Response options were Hospital visit; Postal Questionnaire; Online Questionnaire; Telephone Questionnaire). The pre-meeting questionnaire was piloted with two potential participants (one patient representative and one clinical researcher) who provided comment on structure, content and clarity.

The consensus process participants were asked to complete and return this in electronic form prior to the meeting. Data was analysed prior to the meeting in order to summarise the distribution of ratings for each prognostic factor, including the group median and interquartile range. The importance of a factor was deemed to be 'critical' if the group median score ranged between 7 and 9.[41]

Consensus meeting

This was a one-day meeting, held in Oxford, UK. The meeting had three sections.[42] At the start of the meeting a detailed explanation of the systematic review and preliminary statistical modelling was provided, followed by a summary of responses to the pre-meeting questionnaire (participants were also provided with copies of their own individual responses).

The second section consisted of two rounds of structured facilitator-led discussions that aimed to identify the most important prognostic factors measured initially, and which delayed

prognostic factors should be collected and how. The participants were divided into three groups (for which participants were pre-assigned to ensure a mixture of types of clinician, researchers and patient representatives) and were asked to rank a maximum of 10 important prognostic factors (from the 14 factors identified from the pre-meeting questionnaire) and five important additional prognostic factors (from the 20 nominated in the pre-meeting questionnaire). Ten points and five points respectively were awarded to the most important item, and one to the least important factor. Each round of group discussions were immediately followed by a plenary session to feedback results of the group discussions to the entire group.

Lastly, a session was convened where a final voting process was undertaken in which each participant indicated whether each factor should be included in the prognostic model or not. The number of votes allowed was limited to 10 per individual. This was completed independently on paper questionnaires and then collated. Factors with 70% or more agreement across participants were considered as critically important to consider in the validation study.[43].

4.3 Results

4.3.1 Participants

Of the 30 individuals invited, 25 clinicians and clinical researchers agreed to participate: paramedics (n=6), physiotherapists (n=6), emergency department nurses (n=4), emergency department consultants (n=5), radiology consultant (n=1), trauma and orthopaedic consultant (n=1) and clinical researchers (n=2). Three patient and public representatives responded to the advertisements, but only one was able to attend the consensus meeting. Seventeen individuals returned the pre-meeting electronic questionnaire, 18 attended the meeting and participated in the first two rounds of group discussions. Two participants were unable to complete the final round of individual voting. Hence, only 16 participants voted for the factors which had been prioritised throughout the day.

4.3.2 Pre-meeting questionnaire results

The results of the electronic pre-meeting questionnaire are shown in Table 7. Three baseline factors were rated as critically important (scoring between 7 and 9) and the remainder as important, but not critical (scoring between 4 and 6). Twenty additional factors were nominated by respondents, all of which were deemed critically important. There was a varied

response to when and how delayed prognostic factors should be collected. The most frequent preferences were 4 weeks' post-injury and by telephone method.

Table 7. Findings from the pre-meeting questionnaire including ratings of importance for baseline prognostic factors and additional nominated factors

Qu.	**Prognostic factor****	**Median (IQR)**	**Min, Max**
1	Time between injury and presenting to the Emergency Department	5 (4, 6)	1, 7
2	Pain severity	5 (4, 6)	2, 7
3	Pain on weight bearing	7 (4, 7)*	2, 8
4	Ability to fully weight bear in Emergency Department	6 (5, 7)	2, 9
5	Amount of ankle movement (dorsiflexion)	4.5 (3, 6)	2, 7
6	Amount of ankle movement (plantarflexion)	5 (3, 6)	2, 8
7	Abnormal imaging findings	6 (5, 8)	3, 9
8	Age	6 (5, 8)	2, 9
9	Body Mass Index	7 (5, 7)*	2, 8
10	Working status	5 (3.75, 6.25)	2, 9
11	Level of education	4 (3, 5)	1, 7
12	Mechanism of injury	6 (3.75, 7)	2, 8
13	Repeatedly sprained ankle previously	7 (5, 8)*	5, 9
14	Reporting of catching or locking of the ankle	5.5 (5, 6.25)	3, 7

*Rating critically important

**Other factors nominated in pre-meeting questionnaire: history of chronic pain (n=2), co-morbidities (including Osteoporosis)(n=3), sporting participation, swelling (n=2), anterior talofibular ligament vs posterior talofibular ligament injury, weight bearing status immediately post-injury, occult fracture, other soft tissue damage, syndesmotic sprain, anxiety (n=3), perception of injury severity, recovery expectations (n=2), desire to get better, self-efficacy, beliefs, coping styles, ability to exercise despite pain, requiring regular analgesia, physiotherapy/rehabilitation referral (n=3), adherence to advice.

4.3.3 Consensus meeting results

Eighteen participants, divided into three groups, participated in the two rounds of facilitated discussions and prioritisation exercises. Some groups were unable to agree on or did not use the maximum number of ranks. Priority rankings of the prognostic factors rated by the three groups of key stakeholders are shown in Table 8. Highest priority prognostic factors included repeatedly spraining ankle previously, older age and mechanism of injury. Only six of the 20 additional factors nominated in the pre-meeting questionnaire were deemed as priority for including in the prognostic model (occult fracture/diagnostic imaging result, history of chronic pain/problems, desire to get better, psychosocial factors about recovery, ability to weight bear immediately post-injury and self-efficacy). Following the facilitated discussions, sixteen participants completed the final vote as to which factors should be included in the prognostic model or not. Eight of the 14 originally proposed prognostic factors achieved agreement to be included in the prognostic model; pain intensity, pain intensity on weight bearing, ability to weight bear in the emergency department, age, body mass index, working status, mechanism of injury, and repeatedly sprained ankle previously. Only one additional factor nominated from the pre-meeting questionnaire achieved agreement for inclusion in the prognostic model - psychosocial recovery factors (Table 8). No delayed factors reached agreement for inclusion in the prognostic model.

Table 8: Results of final voting for prognostic factors. Dichotomous responses (yes/no).

	Results of meeting Section 2			**Results of Meeting Section 3**
Prognostic factor	**Priority rank Group 1 (10 highest, 1 lowest)**	**Priority rank Group 2 (10 highest, 1 lowest)**	**Priority rank Group 3 (10 highest, 1 lowest)**	**No. of votes for inclusion in prognostic model, n (%)**
Age	1	10	8	16 (100)*
Body Mass Index	7	-	7	16 (100)*

Repeatedly sprained ankle previously	8	9	10	16 (100)*
Ability to fully weight bear in Emergency Department	-	8	6	16 (100)*
Mechanism of injury	6	-	9	14 (88)*
Pain on weight bearing	10	-	4	14 (88)*
Working status	5	6	-	14 (88)*
Pain severity	-	-	3	13 (81)*
Time between injury and presenting to the Emergency Department	-	7	2	7 (44)
Amount of ankle movement (dorsiflexion)	-	-	5	7 (44)
Abnormal imaging findings	9	-	-	7 (44)
Amount of ankle movement (plantarflexion)	-	-	-	4 (25)
Level of education	-	-	-	2 (13)
Reporting of catching or locking of the ankle	-	-	-	0 (0)
Additional prognostic factors nominated in pre-meeting questionnaire				
	Priority rank	**Priority rank**	**Priority rank**	

	Group 1 (5 highest, 1 lowest)	**Group 2 (5 highest, 1 lowest)**	**Group 3 (5 highest, 1 lowest)**	
Psychosocial factors about recovery	2	5	-	12 (75)*
Occult fracture/diagnostic imaging result	5	-	-	
History of chronic pain/problems	4	3	-	
Desire to get better	3	-	-	
Ability to weight bear immediately post-injury	1	-	-	
Self-efficacy	-	4	-	

4.4 Discussion

This chapter described the consensus-based approach employed in the development of the SPRAINED study prognostic model. We identified eight baseline factors that were deemed critical for the identification of people likely to have a poor recovery. These factors span pre-injury, sociodemographic, psychosocial and clinical assessment factors, encompassing a holistic biopsychosocial model of recovery.[44]

Only one prognostic variable not included in the CAST dataset (Chapter 5) was deemed important enough to be added to the prognostic variables collected in the external validation study (Chapter 6) to enable a later investigation into this prognostic factor. It was agreed that participants should be asked how long they expected it to take to recover from their ankle sprain, which aimed to capture the person's psychological state and perceptions in the acute phase. No additional delayed factors were rated as critical to be included in the model.

The results of our meeting were strengthened by the use of a diverse group of clinical and research practitioners as well as a patient and public representative. We also had the opportunity to test the structure and content of the questions we presented to the group for voting. The limitations of our approach include the lower than anticipated number of patient participants with direct experience of short- or long-term limitations due to an ankle sprain. This may have provided a broader perspective relevant to this patient population. A limitation of the NGT is the short time constraints, limiting the re-iterations of the discussion process and time that participants have to reflect and achieve consensus. The pragmatic approach used may have influenced the length of the group discussions and consequently, the final results.

The findings of this consensus meeting were used in combination with the findings of the systematic review (Chapter 3) and the statistical analysis development (Chapter 5) to inform what additional factors could be included in the model assessed during the external validation study (Chapter 6). The main impact of the meeting was to strengthen the evidence regarding prognostic factors already considered candidates for the model, and importantly leading to the addition of a question to consider psychosocial status around expectation of recovery, as a reflection of wider beliefs and anxieties about the injury and recovery.

The size of the CAST data set was known ahead of all the modelling processes, and using simple rules, allowed us to pre-specify the number of variables that could plausibly be considered as candidates in the internal validation. The consensus exercise was essential in determining the priority of these slots, to discuss and consider the range of ways in which variables could be considered, and the acceptability of the variable and method of testing the variable to both the clinical and patient community. There were a few exceptions to this process. The research team considered that it was necessary to include commonly used clinical examination procedures during the consensus stage. Ultimately, neither the systematic review nor consensus meeting identified these as important. The patchiness and limited scope of existing evidence and relatively limited sampling for the consensus group meant that the possibility of falsely excluding variables might be high, and we erred on the side of caution.

5. Development and internal validation of the SPRAINED prognostic models in the CAST dataset.

5.1 Introduction

This chapter describes the development and internal validation of the two prognostic models to identify people at risk of poor outcome after an acute ankle sprain. The development of the two models followed the same steps, using the same dataset, considered the same candidate predictors but had different definitions of outcome. Data from the Collaborative Ankle Support Trial (CAST), a randomised controlled trial on the effectiveness of three different mechanical supports compared with a double-layer tubular compression bandage for the initial management of severe ankle sprains was used to develop both models.[17]

The initial selection of variables for testing in the CAST data (prior to and for the consensus review) were guided by the systematic review (Chapter 3) and analysis of the data set. The final selection of variables for testing in internal validation was informed by the results of the consensus meeting (see Chapter 4).

5.2 Methods

5.2.1 Individual participant data used to develop the models (study population)

The CAST study was a pragmatic multicentre randomised controlled trial, with blinded assessment of the outcome, designed to estimate the clinical and cost-effectiveness of three different types of mechanical ankle support (Aircast brace, Bledsoe boot, or 10-day below-knee cast) in the treatment of severe ankle sprain (defined as an injury of grade 2 or 3, without fracture) compared with a double-layer tubular compression bandage.

The trial population comprised 584 individuals aged 16 years or older, attending emergency departments in the UK with an ankle sprain and an inability to fully bear weight on the injured ankle at the time of presentation to the emergency department (ED) and their review clinic appointment (the trial's baseline assessment). People were excluded if they present with an ankle fracture (apart from a flake fractures of less than 2mm), any other recent fracture, any contraindication to any of the

four arms of the trial, poor skin viability preventing splinting or casting, or if their injury occurred more than 7 days before the first presentation at the recruiting ED.

The different time points in the CAST study and a summary of the data collected at each point are described in Table 9.

Table 9: Definitions of time points in the CAST study.

Time point	Definition	Information collected
1. First contact with participants (Emergency Department presentation)	Individuals with an ankle sprain attending Emergency Departments (ED) recruiting for the trial were assessed for eligibility by medical staff who also completed a standard proforma with some basic clinical and socio-demographic information. Information on the trial and invitation to join the study was given to eligible individuals together with the participants' information sheet.	Initial eligibility criteria check (people aged 16 years or older, attending emergency departments no more than 7 days after injury, with sprain – not fracture – of the ankle and unable to fully weight bear at presentation); clinical examination and injury related information; socio-demographic data.
2. Follow-up clinic at 2-3 days after ED attendance (Baseline assessment)	Final eligibility check and informed consent obtained from those willing to enter the trial. Short interview performed by the research physiotherapist to ensure eligibility and, after randomisation, the participant completed the baseline questionnaire. The interventions were applied in the emergency department by an appropriately trained health professional after baseline data collection and randomisation.	Data on the main candidate predictors for the prognostic model, including age; sex; height; weight; ethnicity; pre-injury quality of life; mobility; engagement in sports activities; usual occupation and employment. Data on injury presentation; indicators of current mobility levels; pain; and weight-bearing status were also collected.

3.	Outcome measurements (Follow-up assessments)	All outcome measurements were taken at 4 weeks, 12 weeks and 9 months.	**Primary outcomes:** Foot and Ankle Outcome Score (FAOS); The Functional Limitation Profile (FLP). **Secondary outcomes:** The SF-12 scale for health related quality of life; EQ-5D for economic evaluation; VAS to estimate the pain at rest and when bearing weight; the Benefit Scale to rate the benefit received from the treatment; and health-care resource use.

5.2.2 Definition of the primary outcomes

Ankle function at nine months after ankle sprain was the primary outcome for CAST. For the SPRAINED study, our primary outcome was 'poor outcome'. We used two definitions of poor outcome based on key indicators of poor function and instability of the joint, which is typified by recurrent sprains or a significant lack of confidence in the ankle (a persistent feeling of giving way), with or without chronic pain. The selection of these outcome indicators is supported by evidence from van Rijn et al. who reported recovery was most closely associated with improvements in pain and giving way,[45] and Wikstrom et al., according to whom pain and instability are of greatest concern to patients.[46] The definitions were considered and agreed by the patient and public involvement group convened for the SPRAINED study.

Data to classify these outcomes were collected in the CAST data set as follows:

Severe persistent pain

Severe persistent pain was defined on the basis of the response given to the question "How often do you experience foot/ankle pain?" from the Foot and Ankle Outcome Score.[47] The five available response options to this question were: never, monthly, weekly, daily or always. Participants who answered "daily" or "always" were considered as having severe persistent ankle pain.

Severe functional difficulty

Severe functional difficulty was defined on the basis of the response given to the question "In general, how much difficulty do you have with your foot/ankle?" from the Foot and Ankle Outcome Score.[47] The five available response options to this question were: not at all, mildly, moderately, severely or extremely. Participants who answered "severely" or "extremely" were considered as having severe functional difficulty with the ankle.

Significant lack of confidence

Significant lack of confidence was defined on the basis of the response given to the question "How much are you troubled with lack of confidence in your foot/ankle?" from the Foot and Ankle Outcome Score.[47] The five available responses to this question are: not at all, mildly, moderately, severely or extremely. Participants who answered "severely" or "extremely" were considered as having a significant lack of confidence in the ankle.

Recurrence of injury

Recurrence of injury was defined as a new injury of the same nature (acute ankle sprain) to the same ankle, occurring after the initial assessment (baseline) and up to nine months after the date of the first injury. Data on this event was collected by using a specific question "Have you had another injury to the same ankle?".

Composite outcome generation

Two different composite outcomes were generated, focussing on self-reported recovery (outcome 1), and self-reported recovery plus whether participants had experienced a reoccurrence of their ankle sprain during the nine-month follow-up period (outcome 2). The investigation of these two different composite outcomes was conducted as reoccurrence of sprain was considered a sufficiently different clinical issue that it could potentially widen the range of patients considered as having a poor outcome and therefore warranted consideration separately.

Outcome 1

The first model was developed to predict a composite outcome (hereafter referred to as *Outcome 1*) representing the presence of at least one of the following symptoms at nine months after injury: persistent pain, functional difficulty or lack of confidence. First, individual binary outcomes (yes or no) were generated to indicate the presence of each symptom, according to the criteria described above. A single composite binary outcome (Outcome 1) was then created to indicate the presence (yes or no) of one or more of these symptoms.

Outcome 2

The second model was developed to predict a composite outcome (hereafter referred to as *Outcome 2*) representing the presence of at least one of the following symptoms or clinical events at nine months after injury: persistent pain, functional difficulty, lack of confidence or recurrence of injury. First, individual binary outcomes (yes or no) were generated to indicate the presence of each symptom or clinical event, according to the criteria described above. A single composite binary outcome (Outcome 2) was then created to indicate the presence (yes or no) of one or more of these symptoms or events.

The proportion of these outcomes observed in the CAST dataset for both outcomes and number of symptoms at nine months after injury are described in Table 10.

Table 10: Proportion of outcomes (and components) observed in the CAST dataset.

Symptoms/events	None present	1 present	2 present	3 present	4 present	Any present	% missing
Outcome 1 (Pain, lack of confidence or general difficulty)	324 (55.48%)	68 (11.64%)	19 (3.25%)	29 (4.97%)	-	116 (19.86%)	144 (24.7%)
Outcome 2 (Pain, lack of confidence, general difficulty or re-injury)	300 (51.37%)	82 (14.04%)	26 (4.45%)	23 (3.94%)	9 (1.54%)	140 (23.97%)	144 (24.7%)

5.2.3 Available candidate predictors and initial selection of variables for modelling

Complete lists of the 16 available variables in the ED proforma and 154 in the CAST baseline data sets are provided in Boxes 1 and 2, respectively. Variables available in the CAST dataset included socio-demographic indicators (e.g. age, sex, body mass index (BMI), education, employment status); pre-injury quality of life, mobility and lifestyle (e.g. engagement in sports activities); clinical data on injury presentation; indicators of current mobility levels, pain and weight-bearing status. Thirty two variables were pre-selected from these lists to form the group of candidate predictors considered to be plausibly predictive of either of the two outcomes. This initial selection was made internally by

the research team, taking into account the results from the systematic review (see Chapter 3) and the conclusions from the consensus group meeting (see Chapter 4). The pre-selected candidate predictor variables and their details (type, name, categories or units, questionnaire where the data was originally recorded and respective amount of missing data) are listed in Table 11.

Box 1: List of candidate predictor variables from the ED presentation dataset

ED Proforma

1. Date of birth
2. Sex (male / female / no response)
3. Date of ED visit
4. Date of injury
5. Location of pain
6. Anterior Drawer Test (Positive / Painful / Negative / no response)
7. Talar Tilt Test (Positive / Painful / Negative / no response)
8. Proximal Fibular tender (Positive / Painful / Negative / no response)
9. Weight bearing ability (Full / Partial / None / no response)
10. X-ray (yes / no / no response)
11. Crutches (yes / no/ no response)
12. Trial Reason for not entering trial (ankle fracture / other recent fracture / contraindication to intervention / poor skin viability / > 7 days from injury to assessment / other)
13. Additional information (if other)
14. Recruiting Centre
15. Date of Trial Clinic
16. Days from injury to assessment

Box 2: List of candidate predictor variables from the baseline assessment datasets.

Identifier variables

1. Trial Centre
2. Patient's ID
3. Date of assessment
4. Randomisation Group
5. Treatment Received
6. Calendar code
7. Calendar colour
8. Indicator of Pilot Study phase (I / II / main trial)
9. Response at baseline (yes / no)

Background Information Form

10. Age (years)
11. Sex (male / female)
12. Ethnic Group (White / Black-Caribbean / Black-African / Black-Other / Indian / Pakistani / Bangladeshi / Chinese / Other)
13. Ethnic Group details (if other)
14. First language (English / Other European / Gujarati / Hindi / Punjabi / Urdu / Bengali / Other)
15. First language additional info (if other)
16. Able to answer English questions? (yes / no)
17. Current employment status (full-time / part-time / unemployed)
18. Employment category (paid / unpaid)
19. Hours employed (Less than 10 / 10-25 / 25-40 / More than 40 hours per week)
20. Type of employment (Unskilled manual / skilled manual / unskilled non-manual / skilled non-manual / professional / other / decline d to answer)
21. Description of employment (if professional)
22. Description of employment (if other)
23. Occupation if not employed (retired / not looking for work / unable to work / looking for work / full time student / other)

Box 2: List of candidate predictor variables from the baseline assessment datasets continued.

24. Description of unemployment (if other)
25. Education (CSE / O-Level or GCSE / A-level / degree / higher degree / other)
26. Description of level of education (if other)
27. Time on feet (most of the day / > 4 hours a day / < 4 hours a day / Not much time, mostly sitting)
28. Time driving (most of the day / > 4 hours a day / < 4 hours a day / just to & from work / don't drive)
29. Current medications (since ankle injury / prior to injury / no / no answer)
30. Practice of physical activities **(11 questions)** (> once a week / < once a week / never)
41. Other physical activity (if other)
42. Height (cm)
43. Weight (kg)
44. Pain before injury (yes/no)
45. When had previous pain (exercise or heavy activities / exercise and daily activities / constantly / other)
46. Description of when had previous pain (if other)
47. Frequency of previous pain (never / monthly / weekly / daily / always)
48. Previous instability (yes / no)
49. Severity of instability (mild / moderate / severe)
50. Frequency of instability (rarely / sometimes / frequently / always)
51. Previous injury (yes / no)
52. 3+ previous injuries (yes / no)
53. Previous injury <1 year ago (yes / no)
54. Recurrent sprain – yes to all 3 above (yes / no)
55. ED attendance previously (yes / no)
56. How present injury occurred (during sport / at work / at home / outside in public place / other)
57. Description of how present injury occurred
58. Maximum weight bearable (kg)

Box 2: List of candidate predictor variables from the baseline assessment datasets continued.

Baseline questionnaire

59. FAOS components **(42 questions)**
101. Pain at rest VAS (0-100)
102. Pain bearing weight VAS (0-100)
103. FAOS Baseline symptoms (subscale)
104. FAOS Baseline pain (subscale)
105. FAOS Baseline function ADL (subscale)
106. FAOS Baseline function sport (subscale)
107. FAOS Baseline QoL (subscale)
108. FLP components **(13 questions)**
121. WORK components **(10 questions)**
131. FLP score
132. WORK score
133. 1998 SF-12 components **(12 questions)**
145. 1998 SF-12 physical score
146. 1998 SF-12 mental score
147. Baseline EQ-5D components **(5 questions)**
152. Baseline EQ-5D score
153. General level of health today (better / same / worse than the past 6 months)
154. VAS health today (0-100)

Note: Imputed scores of validated scales with specific rules for handling missing data imputation (such as FAOS, SF-12 and EQ-5D) are also present in the CAST dataset, but were not described here.

Table 11: Pre-selected candidate predictor variables from ED presentation and Baseline assessment.

Type	Variable name	Categories / units	Questionnaire	Missing values	
Binary	Sex	Male, female	Background information	0	0%
	Previous pain	Yes, No	Background information	26	4%
	Recurrent sprain	Yes, No	Background information	12	2%
Categorical (or ordinal)	Employment status	No, Part time, Full time	Background information	0	0%
	Education	CSE, GCSE, A Level, Degree, Higher Degree	Background information	20	3%
	Anterior Drawer Test	Positive, Painful, Negative, No response	ED proforma	396	68%
	Talar Tilt Test	Positive, Painful, Negative, No response	ED proforma	403	69%
	Proximal Fibular tender ligament test	Positive, Painful, Negative, No response	ED proforma	378	65%
	Able to bear weight	Full / partial / none	ED proforma	322	55%
	Treatment group	Tubular bondage, Below knee cast, Aircast brace, Bledsoe boot		0	0%
	Leisure time physical activity	None, <1 weekly, >1 weekly	Background information	7	1%
	Walking 2 miles or more	None, <1 weekly, >1 weekly	Background information	24	4%

	Previous instability	None, Mild, Moderate, Severe	Background information	27	5%
	Previous instability frequency	Never, Rarely, Sometimes, Frequently, Always	Background information	29	5%
	Injury presentation	During sport, at work, at home, outside in public	Background information	34	6%
	Ankle/foot Swelling[(1)]	Never, Rarely, Sometimes, Often, Always	Baseline questionnaire	18	3%
	Ankle/foot Grinding/clicking[(1)]	Never, Rarely, Sometimes, Often, Always	Baseline questionnaire	18	3%
	Ankle/foot catching/locking[(1)]	Never, Rarely, Sometimes, Often, Always	Baseline questionnaire	18	3%
	Ankle ROM plantar flexion[(1)]	Never, Rarely, Sometimes, Often, Always	Baseline questionnaire	18	3%
	Ankle ROM plantar dorsiflexion[(1)]	Never, Rarely, Sometimes, Often, Always	Baseline questionnaire	18	3%
	Pain at night (on bed) [(1)]	None, Mild, Moderate, Severe, Extreme	Baseline questionnaire	18	3%
	Difficulty with squatting[(1)]	None, Mild, Moderate, Severe, Extreme	Baseline questionnaire	29	5%
	Difficulty with running[(1)]	None, Mild, Moderate, Severe, Extreme	Baseline questionnaire	31	5%
	Difficulty with jumping[(1)]	None, Mild, Moderate, Severe, Extreme	Baseline questionnaire	31	5%
	Difficulty with twisting/pivoting[(1)]	None, Mild, Moderate, Severe, Extreme	Baseline questionnaire	26	4%
Continuous (or discrete)	Days from injury to assessment	0-7 days	ED proforma / Background information	312	55%
	Age	Years[(2)]	Background information	0	0%

	Body mass index[3]	kg/m^2	Background information	19	3%
	Maximum weight bearable	kg	Background information	5	1%
	Pain when resting	Visual analogue scale (0-100)	Baseline questionnaire	4	1%
	Pain when bearing weight	Visual analogue scale (0-100)	Baseline questionnaire	9	2%
	SF-12 mental component	Score (0-100)	Baseline questionnaire	5	1%

[1] Question from the Foot and Ankle Outcome Score.

[2] An inclusion criteria for the CAST study was presenting at an Emergency Department to treat the ankle sprain no more than 7 days after the injury.

[3] Calculated from height and weight (both continuous variables), as per collected with the Baseline CAST questionnaire.

In addition to the baseline predictors, a few variables from the CAST 4-weeks follow-up questionnaire were selected to be investigated as potential predictors that could add some incremental value to the developed prognostic models. The list of these variables and their characteristics are listed in Table 12.

Table 12: Selected candidate predictor variables from 4 weeks follow-up.

Type	Variable name	Categories / units	Questionnaire	Missing values	
Binary	Repeat injury to the same ankle	Yes, No	4 weeks follow-up	118	20%
	Returned to ED due to repeated injury	Yes, No	4 weeks follow-up	120	21%

Ordinal	Returned to usual sports/activities	No, partially, fully	4 weeks follow-up	121	21%
	Ankle/foot Swelling	Never, Rarely, Sometimes, Often, Always	4 weeks follow-up	102	17%
	Ankle/foot Grinding/clicking	Never, Rarely, Sometimes, Often, Always	4 weeks follow-up	102	17%
	Ankle/foot catching/locking	Never, Rarely, Sometimes, Often, Always	4 weeks follow-up	103	18%
	Able to perform ankle range of motion (ROM) plantar flexion	Never, Rarely, Sometimes, Often, Always	4 weeks follow-up	102	17%
	Able to perform ankle range of motion (ROM) dorsiflexion	Never, Rarely, Sometimes, Often, Always	4 weeks follow-up	102	17%
	Pain at night	None, Mild, Moderate, Severe, Extreme	4 weeks follow-up	101	17%
	Difficulty with squatting	None, Mild, Moderate, Severe, Extreme	4 weeks follow-up	101	17%
	Difficulty with running	None, Mild, Moderate, Severe, Extreme	4 weeks follow-up	135	22%
	Difficulty with jumping	None, Mild, Moderate, Severe, Extreme	4 weeks follow-up	137	23%
	Difficulty with twisting/pivoting	None, Mild, Moderate, Severe, Extreme	4 weeks follow-up	131	22%
Continuous	Pain at weight bearing	0-100	4 weeks follow-up	196	34%

5.2.4 Development dataset preparation

The CAST trial dataset contains individual participant information at baseline and at follow-up assessments (time-points 2 and 3, as described in Table 9) for 584 participants recruited to take part in the study. In order to include the data collected at the time of ED presentation (time-point 1, as described in Table 9) in the analysis, it was necessary to merge the CAST trial main dataset with a separate dataset which included information on 1,487 people screened during the recruitment period of the trial. The information collected at this time point was anonymised, consequently the dataset has no information on the participants' identification number. Information from these two datasets was merged by matching the cases using the individuals' information on date of birth and sex; disregarding any duplicates within each dataset to avoid mismatching.

This process added information on five of the candidate predictors collected during the first contact with the participants to 289 cases in the CAST study main dataset (see Box 1 for details on the predictors collected at the time of ED presentation). As the results of the trial have been published and all documentation archived, there was no need for further data cleaning and the only data manipulation performed with the resulting dataset included generating new variables from existing variables or re-categorisation of existing variables (see section 5.2.3 and section 5.2.5) and missing data imputation (see section 5.2.6).

5.2.5 Exploratory analysis and data transformation

Baseline and four weeks follow-up characteristics of the participants in the CAST study were summarised using means, standard deviations (SDs) and ranges for continuous variables, or counts and percentages for categorical variables. After merging the datasets, three variables from the ED dataset had more than 60% missing information (Anterior Drawer Test, Talar Tilt Test and Proximal Fibular Tender Ligament Test) and were excluded from the list of candidate predictors (see Table 11 for detailed information on the amount of missing data for each candidate predictor). Besides, it was also discussed and agreed during the consensus group meeting (Chapter 4) that it would be reasonable to exclude these variables from the pool of candidate predictors due to the variability in technique between assessors when performing the tests.

Each binary or categorical predictor was tabulated against the outcomes to check for empty or low cell counts. Where this was the case, categorical variables were re-categorised by collapsing some of their categories, providing it made clinical sense to do so. The list of manipulated variables with details on the changes performed are presented in Table 13. In cases where collapsing categories

would not solve the problem (or make clinical sense) the predictor variable was omitted from any further analysis. This was the case for the following candidate predictors: "Ankle/foot swelling" (from baseline) and "Returned to ED due to another injury to the same ankle" (from four weeks follow-up).

The distribution of continuous predictors was also assessed, first considering their empirical distributions by producing histograms and then by assessing these for normality by means of normal probability plots. The presence of any outliers was assessed based on visual examination of box plots. Extreme values were inspected to confirm whether they were clinically plausible. No individual participant information was deleted from the dataset and data transformation (normalisation) was performed as appropriate.

The correlations between candidate predictors were also examined using spearman correlation coefficient, in order to identify any highly correlations between predictors ($r \geq 0.8$). It causes unnecessary complication to include highly correlated predictors together in multivariable models. Highly correlated predictors explain the same variation in outcome, and this was found for two groups of variables: (1) "Difficulty with running", "Difficulty with jumping" and "Difficulty with twisting/pivoting" (from both baseline and four weeks follow-up); and (2) "Previous instability" and "Previous instability frequency" (from baseline).

To deal with the first group of correlated variables, a new binary variable (yes or no) was created to indicate whether a participant presented a positive answer to any of the original variables, so these individuals were characterised as presenting difficulty with running, jumping or twisting. This new composite variable was then used instead of the three highly correlated variables in the remaining analyses. The decision about which predictor should be taken to the modelling stage, between previous instability and previous instability frequency, took into account the individual predictive ability of each variable. The predictor with lower face-validity for Outcomes 1 and 2 ("Previous instability" in both cases) was then omitted from subsequent analyses.

Table 13: Format and categories/units of candidate predictor variables in the original CAST dataset and after data manipulation.

Variable Name	**In the original dataset**	**After exploratory analysis/data manipulation**

	Type	**Categories/Units**	**Type**	**Categories/Units**
Employment status	Categorical	No Part time Full time Student Retired	Categorical	None[1] Part time Full time
Injury presentation	Categorical	During sport At work At home In public Other[2]	Categorical	During sport At work At home In public
Leisure time physical activities (several types of activities)[3]	Categorical	None <1 weekly >1 weekly	Categorical	None <1 weekly >1 weekly
Ankle/foot catching/locking	Categorical	Never Rarely Sometimes Often Always	Categorical	Never Rarely/Sometimes Often/Always
Ankle/foot grinding/clicking	Categorical	Never Rarely Sometimes Often Always	Categorical	Never Rarely/Sometimes Often/Always
Previous instability frequency	Categorical	Never Rarely Sometimes Often Always	Categorical	Never Rarely/Sometimes Often/Always
Able to perform ankle range of motion (ROM) plantar flexion	Categorical	Always Often Sometimes Rarely	Categorical	Often/Always Rarely/Sometimes Never

		Never		
Able to perform ankle range of motion (ROM) dorsiflexion	Categorical	Always Often Sometimes Rarely Never	Categorical	Often/Always Rarely/Sometimes Never
Pain at night (on bed)	Categorical	None Mild Moderate Severe Extreme	Categorical	None/Mild/Moderate Severe/Extreme
Difficulty with squatting	Categorical	None Mild Moderate Severe Extreme	Categorical	None/Mild/Moderate Severe/Extreme
Difficulty with running	Categorical	None Mild Moderate Severe Extreme	Categorical	None/Mild/Moderate Severe/Extreme
Difficulty with jumping	Categorical	None Mild Moderate Severe Extreme		
Difficulty with twisting/pivoting	Categorical	None Mild Moderate Severe Extreme		
Anterior Drawer Test	Categorical	Positive Painful	Binary	Positive/painful Negative

		Negative		
Talar Tilt Test	Categorical	Positive Painful Negative	Binary	Positive/painful Negative
Proximal Fibular tender ligament test	Categorical	Positive Painful Negative	Binary	Positive/painful Negative
Weight bearing ability (ED presentation)	Categorical	Full Partial None	Binary	Yes[4] No
Days from injury to assessment[5]	Continuous	Days	Binary	1-2 days 3-7 days
Maximum weight bearable (Baseline assessment)	Continuous	kg	Binary	Unable to perform test Able to bear some weight

[1] Combination of unemployed, student and retired.

[2] The answers under the option "Other" were reviewed and regrouped with the remaining options accordingly.

[3] From a list with 11 different activities (plus an option for other activity that might not be covered) which were combined into a single variable indicating the highest frequency reported for any physical activity (except "walking 2 miles or more", explored separately).

[4] Combination of full and partial ability to bear weight.

[5] Not allowed more than 7 days in the CAST dataset.

Initial individual associations between each candidate predictor and poor recovery at nine months after ankle sprain were performed by fitting unadjusted logistic regression models for outcomes 1 and 2.

5.2.6 Handling missing data

Some missing data in the development dataset occurred due to missed appointments and losses to follow-up during the conduct of the CAST trial, but also due to the lack of a unique patient identification in the trial's screening (ED presentation) dataset, which did not allow all the information collected at this point to be merged with the main CAST dataset (see section 5.2.4 for further details). The percentage of missing data in the final merged dataset is presented for each candidate prognostic variable in Table *11*and Table *12*. To conform to current guidelines, multiple

imputation for all participants with at least one missing value was performed.[48] Since there were several predictor variables of different types (i.e. binary, categorical, and continuous) with missing data, multiple imputation by chained equations (MICE) was carried out using the *mi impute chained* function in Stata v.14.2 (StataCorp, College Station, TX, USA) with the options *logit* (for imputation of binary variables), *mlogit* (for imputation of categorical variables) and *truncreg* (for imputation of continuous variables, setting the lower and upper limits for imputed values as 0 and 100, respectively).

In MICE, all missing values are filled in by simple random sampling with replacement from the observed values to allow the regression models to be fitted on all values. Then, the variable with the lowest amount of missing observations, for example x_1, is regressed on all other variables. Missing values are then replaced by drawing from the estimated corresponding posterior predictive distribution of x_1. Then, the next variable with the lowest amount of missing observations is regressed on all other variables including (and using the imputed values of) x_1. This process is repeated until all variables with missing values are imputed forming one cycle. Cycles are repeated to stabilize the results and the whole procedure is repeated *m* times to give *m* imputed data sets. An important characteristic of MICE is the capacity of handling different variable types (continuous, binary, unordered and ordered categorical) because each variable is imputed using its own imputation model using different types of regression analysis.

Multiple imputation was performed under the assumption that all missing data was missing at random (MAR). In other words, the probability of data being missing does not depend on the unobserved data, conditional on the observed data. Therefore, imputation models included all available observed characteristics for the predictors of interest (both from baseline and 4 weeks follow-up), predictors of predictors (e.g. weight and height for BMI) and the outcomes, as recommended by White et al.[48] The models were independently estimated for Outcomes 1 and 2, and imputations were therefore performed in separate procedures, producing two different sets of 50 complete datasets. This number of imputed datasets was chosen based on the amount of missing data for the variable with the highest rate of missing observations (312/584 for "days from injury to assessment"). No data transformation was performed on continuous predictor variables before imputing missing observations.

Despite using the augmented-regression approach,[49] some predictors were also excluded during this process due to the issue of *"perfect prediction"* when imputing categorical variables.[50] Perfect

prediction occurs whenever there is a level of a categorical explanatory variable for which the observed values of the outcome are all 1 (or all 0). Perfect prediction then leads to infinite coefficients with infinite standard errors and causes instability during estimation, which prevents the imputation model from achieving convergence. This issue was resolved by dropping the predictors causing the perfect prediction from the multiple imputation model (two from baseline: “Difficulty with running, jumping or twisting” and “Previous instability frequency”; and one from four weeks follow-up: “Able to perform ankle range of movement [ROM] plantar flexion”). A complete list of predictors excluded before the modelling process with reasons is provided in Table 14.

Table 14: Reason for exclusion of predictors before the modelling process.

Predictor	Reason for exclusion
Baseline	
Anterior Drawer Test	≥ 60% missing values, consensus agreement
Talar Tilt Test	≥ 60% missing values, consensus agreement
Proximal Fibular tender ligament test	≥ 60% missing values, consensus agreement
Ankle/foot Swelling	One or more cells with too few cases when cross-tabulated with the outcomes, regardless of re-categorisation
Difficulty with running	Highly correlated with "Difficulty with jumping" and "Difficulty with twisting/pivoting". Composite variable used instead.
Difficulty with jumping	Highly correlated with "Difficulty with running" and "Difficulty with twisting/pivoting". Composite variable used instead.
Difficulty with twisting/pivoting	Highly correlated with "Difficulty with running" and "Difficulty with jumping". Composite variable used instead.
Previous instability	Highly correlated with "Previous instability frequency"
Previous instability frequency	"Perfect prediction" during missing data multiple imputation
Difficulty with running/jumping/twisting	"Perfect prediction" during missing data multiple imputation
4 weeks follow-up	
Returned to ED due to repeated injury	One or more cells with too few cases when cross tabulated with the outcomes
Difficulty with running	Highly correlated with "Difficulty with jumping" and "Difficulty with twisting/pivoting". Composite variable used instead.

Difficulty with jumping	Highly correlated with "Difficulty with running" and "Difficulty with twisting/pivoting". Composite variable used instead.
Difficulty with twisting/pivoting	Highly correlated with "Difficulty with running" and "Difficulty with jumping". Composite variable used instead.
Ankle ROM plantar dorsiflexion	"Perfect prediction" during missing data multiple imputation

Note: Perfect prediction occurs whenever there is a level of a categorical explanatory variable for which the observed values of the outcome are all 1 (or all 0); it is often resolved by discarding the observations corresponding to offending covariate patterns or the independent variables perfectly predicting outcomes during estimation.

5.2.7 Sample size considerations

Sample size requirements for logistic regression are based on the concept of events-per-variable (EPV). It is widely recommended that to develop a prediction model, the data set should contain a minimum of 5-10 EPV.[51-56] Based on a number of at least five EPV, the outcomes rates (see Table 10) observed in the CAST dataset allowed the inclusion of up to examine up to 23 (116/5) or 28 (140/5) candidate predictor variables in the models for Outcomes 1 and 2, respectively. After the exclusion of nine pre-selected candidate predictors for the reasons described above, 23 variables from baseline remained as candidate predictors. However, some of these predictors were categorical variables with more than 2 levels, which impacts in the EPV as these predictors require the generation of indicator variables for each category (e.g.: employment status coded as "no", "part time" or "full time" will require 3 parameters to be estimated). Therefore, we ended with 35 candidate parameters, which means the EPV ratio was approximately 3 (three) and 4 (four) for Outcomes 1 and 2, respectively. It is also important to note that some of the candidate predictors were continuous variables, which could require non-linear modelling and therefore increase even more the number of regression coefficients to be estimated and affect the EPV (e.g. if using fractional polynomials and the best transformation for age was found to be age+age^2, then age would relate to two predictors instead of one). However, this was not the case.

To the best of our knowledge, this was the first project aiming to develop prediction models to assess the risk of poor recovery after an acute ankle sprain. Therefore, we have opted for relaxing the EPV rule in favour of including more potential important predictors in the analyses. However, we have

adopted several strategies to minimise bias and overfitting, including the estimation of heuristic shrinkage factors to account for possible extreme predictions resulting from overestimated associations (sections 5.2.8 to 5.2.11).

5.2.8 Data modelling

Since both outcomes were binary (poor outcome after ankle sprain – yes/no), the prognostic models were developed using a logistic regression modelling framework with the logit probability of poor outcome as the response variable. The 23 remaining candidate predictors were included together in full logistic regression models as independent variables and further selection of predictors was based on the statistical significance of their adjusted relationship with the outcomes. At this point, continuous variables were kept as continuous to avoid loss of prognostic information.[57] Therefore, the shape of the relationship between continuous predictors and the outcome should be studied and modelling performed with nonlinear functions such as fractional polynomials (FPs) where appropriate.[58]

Non-linear relationships were investigated using fractional polynomials and the "best transformation" for each continuous predictor was used when fitting the models. As more than one continuous variable were included in the full models, the multivariable fractional polynomial (MFP) algorithm was used.[59, 60] The MFP algorithm selects predictors and their respective transformations that best predict the outcome variable using a backward selection process with a nominal alpha of 0.15 used to warrant exclusion from the model to reduce the risk of over fitting. Another advantage of the MFP algorithm is that selection of predictors and transformations are done simultaneously, preserving the nominal type 1 error probability.

Since the analyses were performed in sets of 50 multiply imputed datasets, the MPF algorithm was applied using the Stata command *mfpmi* together with *logit*. The *mfpmi* allows binary, ordinal and non-ordinal categorical variables to be included together with continuous variables in the same model, and simultaneously select the appropriate FP transformation of continuous predictors and combine the estimates of multiply imputed datasets. The multivariable models are fitted in each of the 50 complete data sets and the estimated regression parameters (coefficients and variances) combined using Rubin's rule.[61, 62]

Ideally, prognostic models should be flexible, easy to understand, and parsimonious, so that they are simple and quick to apply in clinical practice. Therefore, after identifying the best transformation terms for continuous variables in the full multivariable models with all candidate predictors, the statistically significant predictors (and respective transformations of continuous variables, where applicable) were selected using the Akaike Information Criterion (AIC) as the decision rule and kept in the final model.[63] Therefore, a p-value < 0.157 (equivalent to AIC) was conservatively taken to warrant inclusion of predictors in the final model and to reduce the risk of over-fitting.

5.2.9 Model Update

After developing the prognostic models for outcomes 1 and 2 including only predictors collected at baseline (baseline variables), the additional incremental value of candidate predictors collected at 4 weeks follow-up point were investigated. First, all additional candidate predictors were included together in the final baseline models and only those predictors achieving a p<0.157 (AIC) were considered for inclusion in the updated models (i.e.: prognostic models including baseline + 4 weeks predictors). Finally, these updated models were compared with the original baseline models by decision curve analysis (DCA) plots[64, 65] to investigate whether the inclusion of additional predictors reflected in increased net benefit. The decision curve analysis was performed in Stata by using the command *dca*.

5.2.10 Assessment of model performance

After developing a prognostic model, it is important to evaluate its performance. Table *15* provides an overview of the main ways that model performance can be assessed from Thangaratinam et al..[66]

Table 15: Main methods of assessing prognostic model performance.

Terms	**Definitions**
Calibration	Calibration indicates the ability of the model to correctly estimate the absolute risks and was examined using calibration plots
Reproducibility (internal validation)	The process of determining internal validity. Internal validation assesses validity for the setting from which the development data originated

Generalisability/transportability (external validation)	The process of determining external validity of the prediction model to populations that are plausibly related
Discrimination	Discrimination describes the ability of the model to correctly distinguish those who will have an adverse outcome from those who will not
Calibration plot	In a calibration plot the predictive risk plotted against the observed incidence of the outcome. Ideally the predicted risk equals the observed incidence throughout the entire risk spectrum and the calibration plot follows the 45° line

From Thangaratinam et al.[66] Development and validation of Prediction models for Risks of complications in Early-onset Pre-eclampsia (PREP): a prospective cohort study. *Health technology assessment* 2017;**21**:1-100.

The performance of the prognostic models was characterised by evaluating calibration and discrimination.

Calibration

Calibration is the agreement between observed and predicted probabilities of poor outcome. The calibration of the developed prognostic models was assessed graphically using calibration plots, with observed risks plotted on the *y*-axis against predicted risks on the *x*-axis.[67, 68] The calibration plot is created by regressing the occurrence of the outcome on the predicted probability of the outcome using a locally weighted scatter plot smoother (lowess). This plot shows the direction and magnitude of model miscalibration across the probability range. The calibration plot was also supplemented with estimates of the calibration slope and intercept. Models with perfect calibration will have a calibration slope of 1 and intercept 0 (i.e., prediction lying on or around the 45° line).

Discrimination

Discrimination is the ability of the prognostic model to separate individuals with the event from those who do not (i.e. those with the event should have higher predicted probabilities than those without). The overall discriminatory ability was summarized by the c-statistic (or Area Under Receiver Operating Characteristic Curve [AUC ROC]) with 95% confidence interval. The c-statistic was classified as follows: 0.5 – 0.6 fail, 0.6 – 0.7 poor, 0.7 – 0.8 fair, 0.8 – 0.9 good, 0.9 – 1 excellent.

Due to complexities in the model building (e.g., a combination of variable selection, fractional polynomials and multiple imputation), we did not carry out an internal validation of the model (e.g., using bootstrapping), as not all these approaches could be replayed in the internal validation. We therefore carried out an ad-hoc hybrid of apparent performance and internal validation, whereby model performance was evaluated both on the original CAST data and also separately in each imputed data set. We calculated the model discrimination in the original CAST data, and also combined from the multiply imputed data sets using Rubin's rules. Calibration plots were created following recommendations of overlaying calibration curves from each imputed data set [69].

5.2.11 Shrinkage

Newly developed prognostic model are often optimistic due to overfitting, which leads to worse prediction in independent data. Reasons for overfitting include small EPV, selection of predictors

based on p-values, modelling non-linear relationships between predictors and the outcome, among others. To estimate the amount of overfitting likely to be present in the developed prognostic models, heuristic shrinkage factors were calculated independently for each model as:

$$(model\ \chi 2 - df)/model\ \chi 2$$

Where *model χ2* is the model likelihood ratio, or -2log likelihood of a model with only an intercept and the fitted model, and *df* the number of degrees of freedom in the fitted model. The number of degrees of freedom in the fitted model is defined by the number of degrees of freedom considered for all explored candidate predictors plus all respective transformations, where applicable.

A shrinkage factor of 1 implies no shrinkage. The regression coefficients from the prognostic models were multiplied by the shrinkage factor to adjust the models for optimism. The shrinkage of the intercept was estimated by fitting a logistic regression model for each studied outcome, including the linear predictor (log odds) calculated using the shrunk coefficients as the only independent variable, and constraining its coefficient to one (offset variable).[70]

5.3 Results

5.3.1 Baseline characteristics

Baseline characteristics of the participants in the CAST trial dataset are summarised in Table 16. Participants were on average 29.88 years, with the age range varying from 16 to 72 years. Participants had a mean BMI of 26.34 kg/m^2, lower pain sores when resting (mean 37.75/100 points) compared to when weight bearing on the injured ankle (mean 75.42/100 points), and fewer than 25% reported not being able to bear any weight on their ankles at the time of baseline assessment. The majority of participants reported not feeling pain in the ankle prior to the injury (86.56%) nor seeking treatment for a recurrent sprain (90.38%). Most of the participants were in full-time employment (61.64%), had an education level higher than GCSE (84.98%) and fewer than one quarter engaged in any leisure time physical activity more than once a week (24.09%). Among the CAST participants, injuries occurred mostly during the practice of sports (36.91%).

Table 16: Summary of baseline characteristics (candidate predictors) of the CAST trial sample.

Variable	Mean (SD)	Min - Max

Age (years)	29.88 (10.77)	16 – 72
Height (m)	1.73 (0.98)	1.47 – 2.01
Weight (kg)	78.56 (15.44)	39.92 – 133.36
BMI (kg/m^2)	26.34 (5.19)	16.07 – 53.77
Pain when resting (score)	37.75 (23.49)	0 – 100
Pain when bearing weight (score)	75.42 (19.61)	0 – 100
SF-12 Mental Component (score)	51.08 (11.26)	20.55 – 68.77
Sex	**Frequency**	**%**
Male	337	57.71
Female	247	42.29
Days from injury to assessment		
0-2	118	44.87
3 or more	145	55.13
Able to bear weight at ED presentation		
No	72	27.48
Yes	190	72.52
Able to bear weight at Baseline assessment		
No	446	77.03
Yes	133	22.97
Pain on the ankle before injury		
No	483	86.56
Yes	75	13.44
Recurrent sprain		
No	517	90.38
Yes	55	9.62
Pain on bed at night		
No	378	66.78
Yes	188	33.22
Difficulty with squatting		
None/mild/moderate	88	15.86
Severe/extreme	467	84.14
Current employment		
None	132	22.6
Part time	92	15.75
Full time	360	61.65
Treatment received for ankle sprain		
Tubular bondage	144	24.66
Below knee cast	142	24.32
Aircast brace	149	25.51
Bledsoe boot	149	25.51

Education level		
CSE level or less	84	15.02
O-level/GCSE/A-level	383	68.52
Degree/Higher degree	92	16.46
Leisure time physical activity		
None	28	4.85
< 1 time weekly	410	71.06
> 1 time weekly	139	24.09
Walking 2 miles or more per day		
None	164	29.29
< 1 time weekly	105	18.75
> 1 time weekly	291	51.96
Injury mechanism		
At home	99	18.00
Practicing sports	203	36.91
At work	79	14.36
Outside, in public	169	30.73
Ankle grinding/clicking		
Never	257	45.41
Rarely/sometimes	220	38.87
Often/Always	89	15.72
Ankle catching/locking		
Never	286	50.53
Rarely/sometimes	209	36.93
Often/Always	71	12.54
Ankle ROM plantar flexion		
Always/often	101	17.84
Sometimes/rarely	247	43.64
Never	218	38.52
Ankle ROM plantar dorsiflexion		
Always/often	81	14.31
Sometimes/rarely	227	40.11
Never	258	45.58

All continuous variables presented at least a minimal departure from a normal distribution, as evidenced in Figure 5 to Figure 11. Some outliers were observed for participants age, weight, BMI and pain score when bearing weight. However, all extreme values were clinically plausible, so no observations were dismissed.

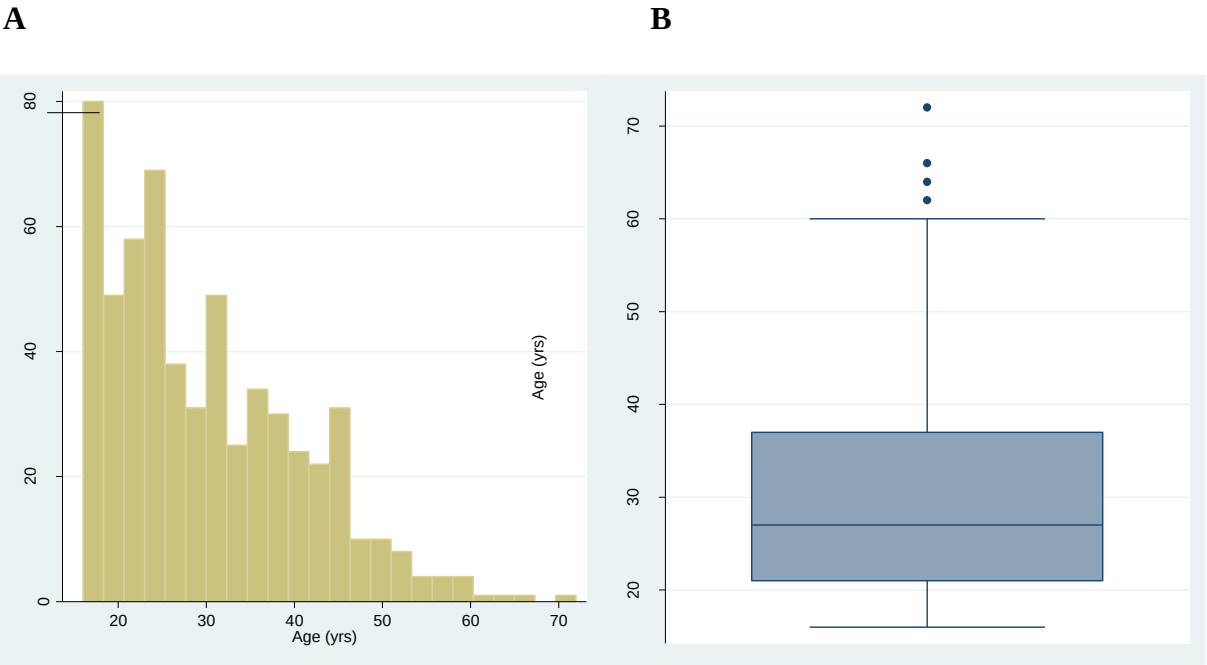


Figure 5: Participants' age histogram (A) and box plot (B).

A **B**

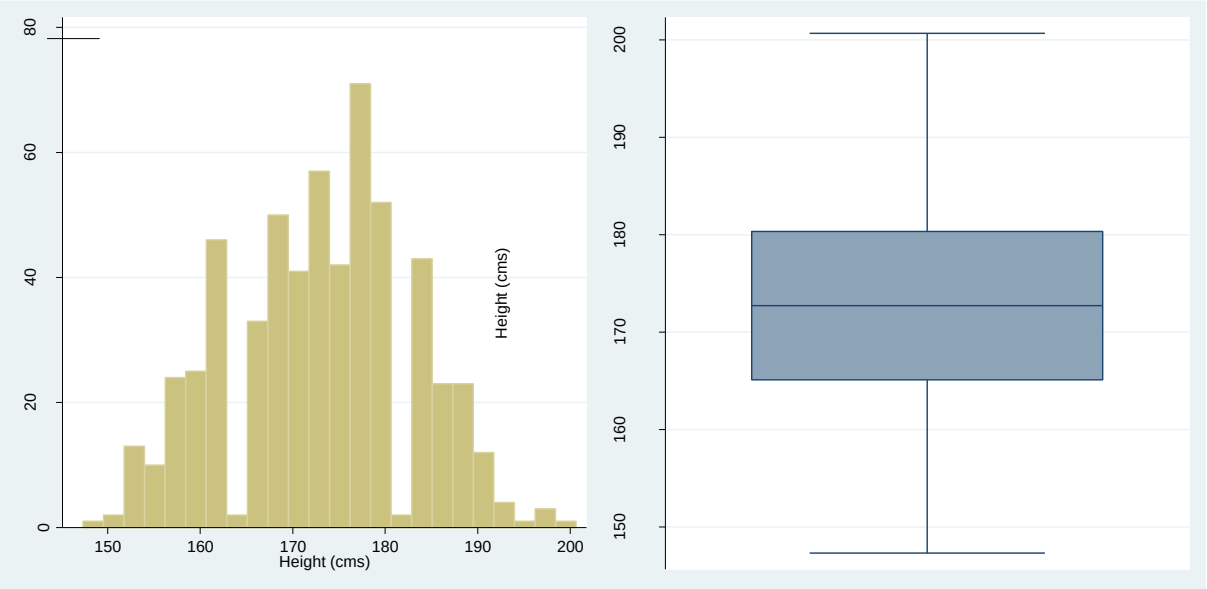


Figure 6: Participants' height histogram (A) and box plot (B).

A **B**

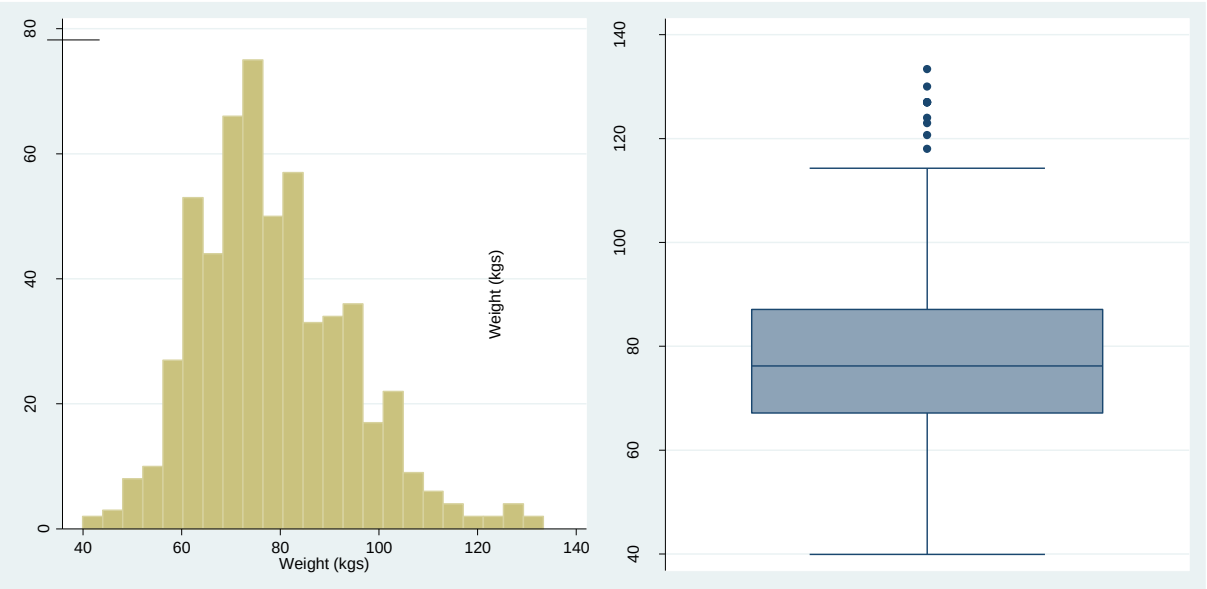


Figure 7: Participants' weight histogram (A) and box plot (B).

A **B**

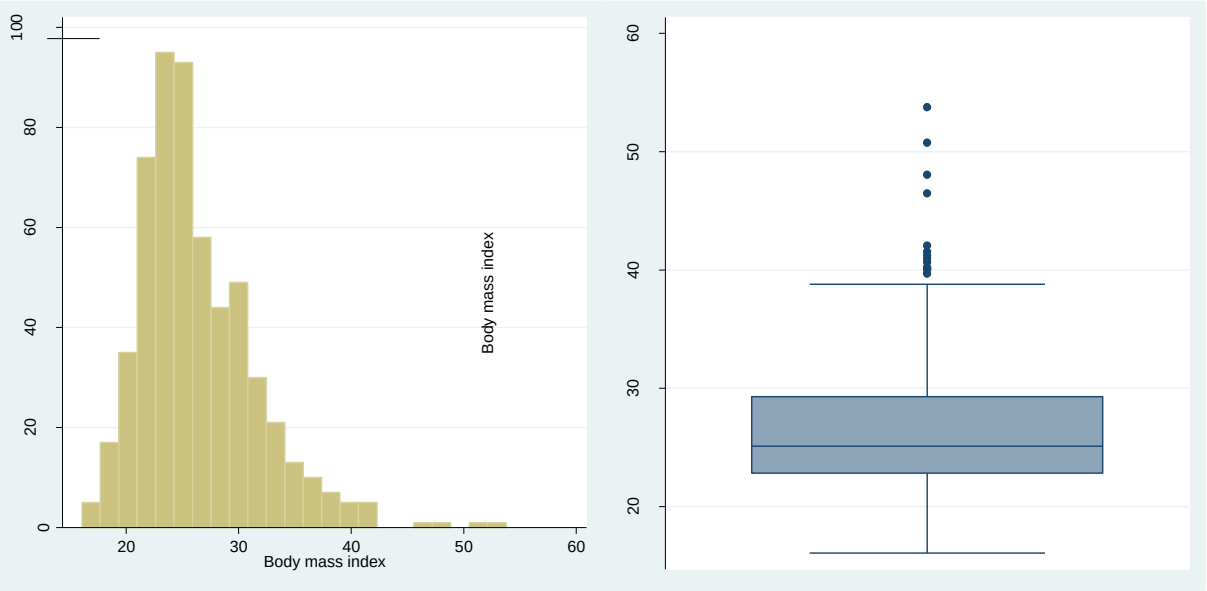


Figure 8: Participants' body mass index histogram (A) and box plot (B).

A **B**

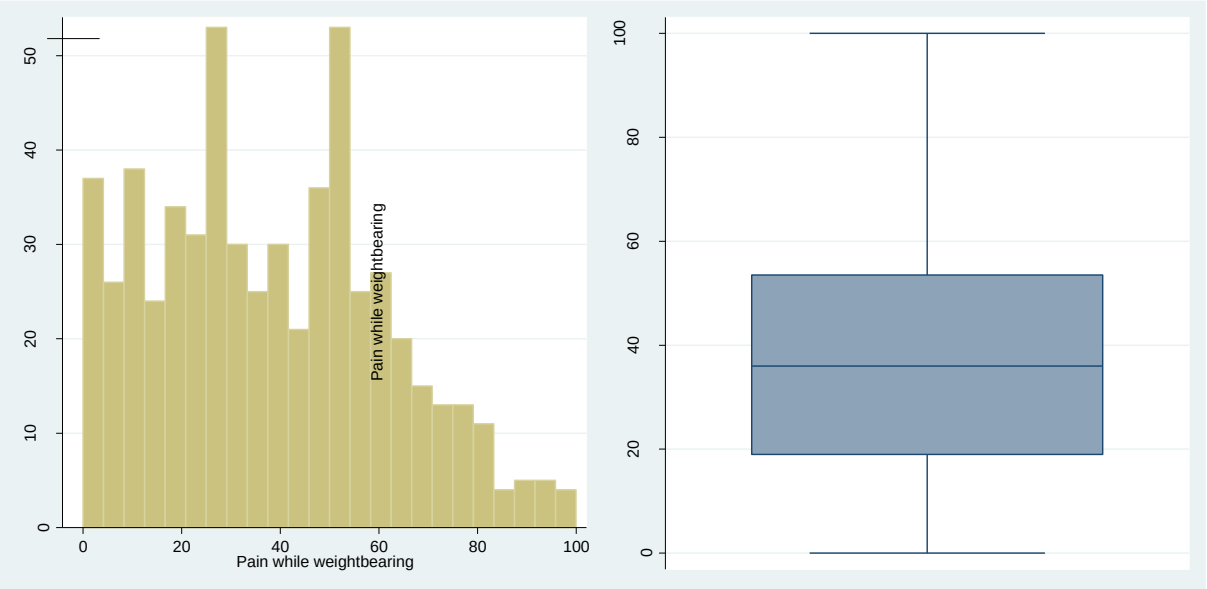


Figure 9: Pain score when resting at ED presentation histogram (A) and box plot (B).

A **B**

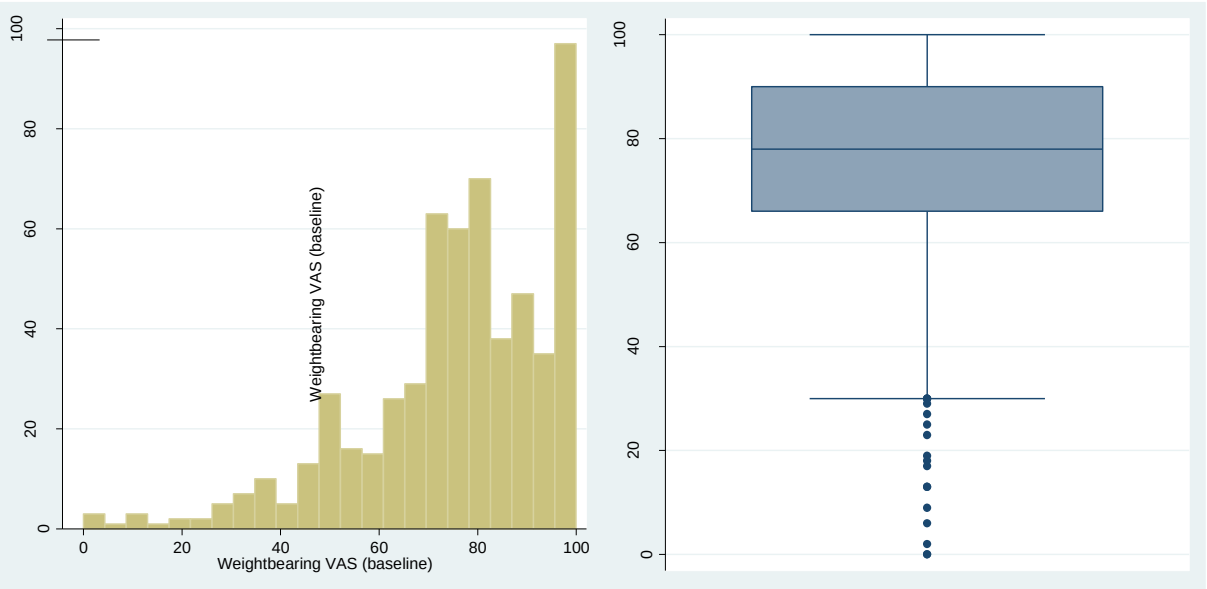


Figure 10: Pain score when bearing weight at baseline assessment histogram (A) and box plot (B).

A **B**

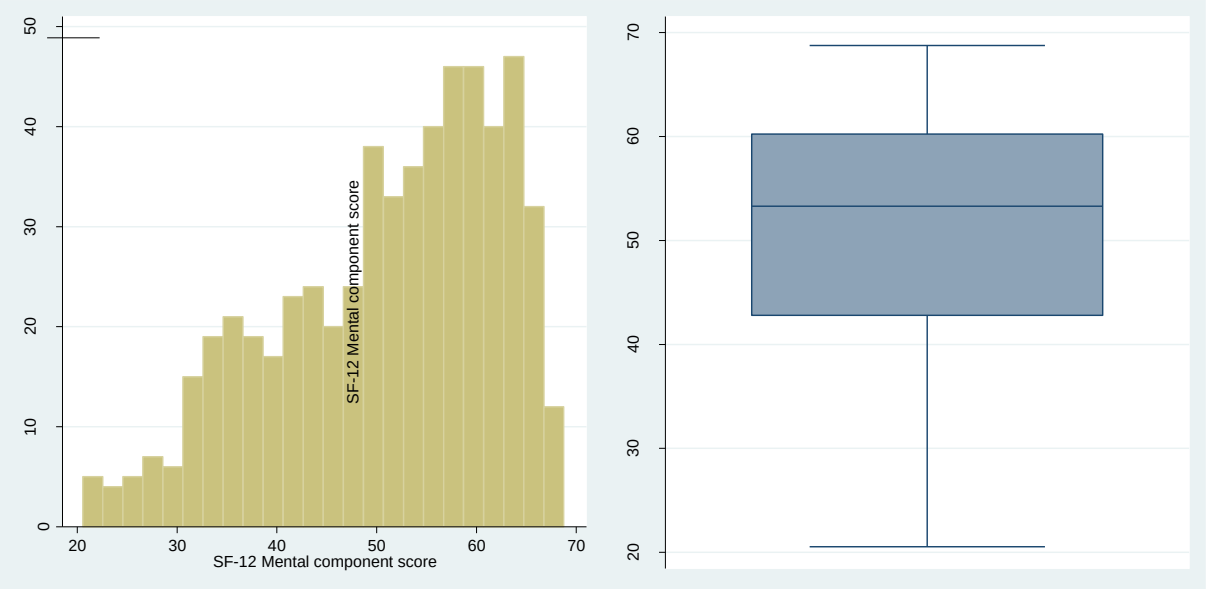


Figure 11: SF-12 mental component score histogram (A) and box plot (B).

Correlation coefficients of the baseline predictors are presented in Table 17. Highly correlated candidate predictors included “Difficulty with running” and “Difficulty with jumping” (r = 1.000); “Difficulty with running” and “Difficulty with twisting/pivoting” (r = 0.859); “Difficulty with jumping” and “Difficulty with twisting/pivoting” (r = 0.859); “Previous instability” and “Previous instability frequency (r = 0.997)”. As these variables should not be included together in the regression models, the first set of highly correlated variables was combined into a single composite variable to identify those participants with difficulties in running, jumping or twisting/pivoting. For the second pair of highly correlated variables, only previous instability frequency was included in the subsequent analysis.

5.3.2 Multivariable models

The summary of the full multivariable models estimates (predictor’s coefficients, respective 95% confidence intervals and p-values) are presented in Table 18. For outcome 1, seven of the 23 candidate predictors were selected for inclusion in the final model, based on the AIC (p<0.157): age, BMI, pain when resting, pain when bearing weight, days from injury to assessment, ability to bear weight and whether or not the injury was a recurrent sprain. For the outcome 2, almost the same set of candidate predictors were selected for inclusion in the final model, except for age and BMI. For

outcome 2 educational level was found to be a statistically important candidate predictor. However, education was identified as a low priority variable by the consensus committee. There were particular difficulties with this variable, as the criteria used in the CAST study to identify different education achievements have been superseded, and in the interim a number of new additional categories of study have become more popular (for example University of the Third Age). Given the marginal statistical significance, inability to replicate the categories in an external validation, low priority given in the consensus, and the reticence of clinicians to probe this information, we did not include this variable in the final model for outcome 2.

Table 17: Spearman correlation coefficients between pre-selected candidate predictors from baseline.

	Days from injury to assessment	Maximum bearable weight	Sex	Pain before injury	Recurrent sprain	Current employment	Education	Treatment arm	LTPA	Walking	Previous instability	Previous instability freq.	Injury mechanism	Ankle grinding	Ankle catching	Plantar ROM flexion	Plantar ROM dorsiflexion	Pain at night	Difficulty with squatting	Difficulty with running	Difficulty with jumping	Difficulty with twisting	Age	BMI	Able to bear weight	Pain when resting	Pain when bearing weight
Days from injury to assessment	-																										
Maximum bearable weight	0.050	-																									
Sex	0.014	0.019	-																								
Pain before injury	-0.039	-0.026	0.112	-																							
Recurrent sprain	-0.008	-0.008	0.042	0.389	-																						
Current employment	0.046	-0.053	-0.438	-0.065	-0.016	-																					
Education	-0.085	-0.049	-0.075	-0.070	-0.055	0.204	-																				
Treatment arm	-0.047	0.199	-0.037	-0.016	-0.024	-0.033	-0.046	-																			
LTPA	0.000	0.003	0.166	0.067	-0.068	-0.138	0.055	-0.142	-																		
Walking	0.074	0.062	0.214	0.079	0.028	-0.123	-0.086	-0.045	0.053	-																	
Previous instability	-0.056	-0.011	0.080	0.470	0.331	-0.174	-0.051	-0.115	0.015	-0.021	-																
Previous instability freq.	-0.043	-0.014	0.079	0.464	0.322	-0.179	-0.045	-0.114	0.019	-0.014	0.997	-															
Injury mechanism	0.110	0.005	0.409	0.026	0.008	-0.288	-0.189	-0.054	0.063	0.113	0.066	0.065	-														
Ankle grinding	0.079	0.033	0.049	0.267	0.136	-0.033	-0.150	-0.037	-0.108	-0.053	0.211	0.217	-0.020	-													
Ankle catching or locking	0.104	0.048	-0.005	0.192	0.013	-0.060	-0.164	0.050	0.063	-0.050	0.122	0.127	0.104	0.486	-												
Plantar ROM flexion	-0.085	0.182	-0.021	-0.020	-0.043	-0.058	-0.014	-0.002	0.015	-0.016	0.010	0.009	0.009	-0.032	0.125	-											
Plantar ROM dorsiflexion	-0.186	0.147	-0.064	0.048	0.009	0.010	0.033	-0.021	0.032	0.063	0.006	0.006	-0.075	-0.110	0.090	0.667	-										
Pain at night	-0.119	0.062	0.201	0.086	-0.041	-0.193	-0.078	-0.009	0.128	-0.024	0.035	0.040	0.176	0.224	0.175	0.169	0.119	-									
Difficulty with squatting	-0.011	0.145	0.049	-0.199	-0.110	-0.054	-0.008	0.040	-0.106	-0.042	-0.041	-0.036	0.021	0.078	0.087	0.177	0.164	0.130	-								
Difficulty with running	0.005	0.086	0.020	-0.128	-0.102	0.027	0.080	-0.007	0.072	-0.091	-0.108	-0.093	0.037	0.043	0.081	0.074	0.061	0.053	0.441	-							
Difficulty with jumping	0.005	0.086	0.020	-0.128	-0.102	0.027	0.080	-0.007	0.072	-0.091	-0.108	-0.093	0.037	0.043	0.081	0.074	0.061	0.053	0.441	1.000	-						
Difficulty with twisting	-0.050	0.178	0.066	-0.088	-0.070	0.014	0.032	0.081	-0.002	-0.091	-0.067	-0.054	0.051	0.103	0.040	0.086	0.083	0.091	0.475	0.859	0.859	-					
Age	0.101	-0.137	0.127	-0.022	-0.054	0.048	-0.004	0.004	0.124	-0.025	0.019	0.029	0.174	-0.088	-0.032	0.024	-0.012	0.021	0.124	0.068	0.068	0.082	-				
BMI	0.049	-0.123	0.268	0.113	0.010	-0.135	-0.064	-0.063	0.174	0.018	0.048	0.052	0.196	0.190	0.026	0.056	-0.076	0.059	0.006	-0.011	-0.011	0.048	0.227	-			
Able to bear weight	0.204	0.149	0.082	0.029	-0.001	-0.041	-0.042	0.007	0.001	0.205	-0.018	-0.013	-0.049	-0.031	-0.109	0.131	0.127	-0.162	-0.018	-0.100	-0.100	-0.131	0.054	-0.001	-		
Pain when resting	-0.066	0.072	0.246	0.075	0.028	-0.234	-0.157	-0.003	0.103	-0.005	0.106	0.102	0.198	0.305	0.215	0.193	0.106	0.434	0.138	-0.005	-0.005	0.047	0.065	0.158	-0.011	-	
Pain when bearing weight	-0.144	0.161	0.209	0.028	0.066	-0.196	-0.133	0.078	0.097	-0.031	0.086	0.084	0.159	0.091	0.117	0.242	0.245	0.395	0.135	0.063	0.063	0.149	0.149	0.118	-0.069	0.630	-
SF-12 Mental Component	-0.002	0.115	-0.074	-0.153	-0.108	0.106	0.131	0.016	-0.042	0.075	-0.176	-0.175	-0.102	-0.307	-0.209	0.082	0.016	-0.088	0.000	-0.071	-0.071	-0.063	-0.008	-0.126	0.238	-0.123	-0.128

Note: Variables excluded because the amount of missing data are not included in this table.

Table 18: Summary of the full multivariable logistic regression models including all 23 candidate predictors of poor outcome after ankle sprain (outcomes 1 and 2).

	Outcome 1				Outcome 2			
Variable	**Coefficient**	**95% CI**		***p***	**Coefficient**	**95% CI**		***p***
Age	**0.036**	**0.008**	**0.064**	**0.012**	0.015	-0.010	0.040	0.230
BMI	**0.039**	**-0.013**	**0.090**	**0.138**	0.012	-0.034	0.059	0.609
Pain when resting	**0.018**	**0.005**	**0.031**	**0.009**	**0.015**	**0.002**	**0.027**	**0.022**
Pain when bearing weight	**0.018**	**-0.001**	**0.037**	**0.057**	**0.013**	**-0.003**	**0.029**	**0.117**
SF-12 Mental Score	-0.006	-0.030	0.018	0.641	-0.012	-0.034	0.010	0.271
Sex (reference Male)								
Female	0.054	-0.581	0.689	0.868	-0.134	-0.734	0.466	0.661
Days from injury to assessment (reference 0-2)								
3 or more	**0.945**	**0.000**	**1.890**	**0.050**	**0.646**	**-0.129**	**1.421**	**0.101**
Able to bear weight at ED presentation (reference No)								
Yes	0.538	-0.445	1.522	0.280	0.445	-0.376	1.266	0.285
Able to bear weight at Baseline assessment (reference No)								
Yes	**-0.848**	**-1.494**	**-0.202**	**0.010**	**-0.737**	**-1.328**	**-0.147**	**0.014**
Pain on the ankle before injury (reference No)								
Yes	0.270	-0.499	1.038	0.491	0.120	-0.588	0.828	0.739
Recurrent sprain (reference No)								
Yes	**1.355**	**0.486**	**2.224**	**0.002**	**1.207**	**0.396**	**2.018**	**0.004**
Pain on bed at night (reference No)								
Yes	0.090	-0.572	0.752	0.790	-0.059	-0.647	0.528	0.843
Difficulty with squatting (reference None/mild/moderate)								
Severe/extreme	-0.223	-0.976	0.531	0.561	0.005	-0.682	0.691	0.989
Current employment (reference None)								
Part time	0.716	-0.163	1.595		0.452	-0.309	1.213	
Full time	0.685	-0.079	1.449	0.175	0.148	-0.517	0.813	0.500
Treatment received for ankle sprain (reference Tubular bondage)								
Below knee cast	-0.554	-1.287	0.179		-0.504	-1.180	0.173	
Aircast brace	-0.394	-1.115	0.326		-0.451	-1.110	0.208	
Bledsoe boot	-0.218	-0.967	0.531	0.489	-0.442	-1.125	0.242	0.443
Education level (reference CSE level or less)								
O-level/GCSE/A-level	0.433	-0.432	1.298		**0.356**	**-0.443**	**1.154**	
Degree/Higher degree	-0.217	-1.256	0.822	0.217	**-0.592**	**-1.542**	**0.358**	**0.042**
Leisure time physical activity (reference None)								
< 1 time weekly	0.007	-1.198	1.211		0.301	-0.818	1.420	
> 1 time weekly	0.206	-1.055	1.466	0.794	0.263	-0.874	1.399	0.869
Walking 2 miles or more per day (reference None)								
< 1 time weekly	-0.104	-0.867	0.659		-0.300	-1.000	0.401	
> 1 time weekly	-0.183	-0.811	0.444	0.847	-0.243	-0.788	0.303	0.626
Injury mechanism (reference At home)								
Practicing sports	0.115	-0.711	0.941		0.302	-0.457	1.062	
At work	0.444	-0.508	1.396		0.672	-0.206	1.550	
Outside, in public	-0.215	-0.966	0.535	0.524	-0.033	-0.727	0.662	0.323
Ankle grinding/clicking (reference Never)								

Rarely/sometimes	-0.226	-0.813	0.362		0.011	-0.531	0.553	
Often/Always	-0.325	-1.245	0.596	0.696	0.048	-0.772	0.869	0.993
Ankle catching/locking (reference Never)								
Rarely/sometimes	0.224	-0.383	0.832		0.021	-0.525	0.568	
Often/Always	0.487	-0.339	1.313	0.483	0.364	-0.362	1.090	0.602
Ankle ROM plantar flexion (reference Always/often)								
Sometimes/rarely	0.550	-0.380	1.479		0.474	-0.370	1.319	
Never	0.223	-0.826	1.273	0.395	-0.052	-1.002	0.897	0.185
Ankle ROM plantar dorsiflexion (reference Always/often)								
Sometimes/rarely	-0.019	-1.041	1.002		-0.127	-1.017	0.762	
Never	0.366	-0.734	1.466	0.528	0.418	-0.568	1.404	0.253
Intercept	**-3.003**	**-5.162**	**-0.845**	**0.007**	**-2.045**	**-3.892**	**-0.198**	**0.030**

95% CI: 95% confidence interval

Notes: Candidate predictors that were statistically significant according to AIC criteria are in bold. Although education level was a statistically significant predictor for outcome 2, a decision has been made not to include it in the final model (see text for details).

The best fit for all continuous predictors were found to be linear transformations (mean subtractions) which were incorporated into the model by updating the intercepts accordingly.A summary of the estimates from the final multivariable models (predictor's coefficients, respective 95% confidence intervals and p-values) are presented in Table 20. For outcome 1, BMI was not statistically significant according to AIC in the final model. Nevertheless, it was decided not to exclude this variable from the model, given its clinical importance and to reduce the risk of over-fitting. Both models were fairly simple, composed of just a few predictors that are routinely collected in the clinical setting.

Only pain when bearing weight at 4 weeks after the injury was included in the updated models (baseline + 4 weeks predictors) for both outcomes 1 and 2 (Table *19*). By inspecting the DCA plots shown in Figure 12 and Figure 13 it is possible to see a clear net benefit gain over the entire range of thresholds when using any of the developed prognostic models in comparison to considering all patients (or no patient) at risk of having poor outcome after an acute ankle sprain. Also, the inclusion of the 4 weeks predictor (pain when bearing weight) consistently improved the performance of the models for both outcomes 1 and 2.

Table 19: Summary of the full multivariable logistic regression including the predictors selected for the baseline models and the 4-weeks candidate predictors for the updated models.

	Outcome 1			**Outcome 2**		
Variable	**Coefficient**	**95% CI**	***p***	**Coefficient**	**95% CI**	***p***
Baseline predictors						

Age	0.020	-0.004	0.044	0.097	-	-	-	-
BMI	0.024	-0.027	0.074	0.356	-	-	-	-
Pain when resting	0.008	-0.005	0.021	0.228	0.004	-0.009	0.016	0.554
Pain when bearing weight	0.014	-0.002	0.031	0.090	0.010	-0.005	0.024	0.199
Days from injury to assessment (reference 0-2)								
3 or more	0.639	-0.288	1.565	0.174	0.450	-0.302	1.202	0.238
Able to bear weight at Baseline assessment (reference No)								
Yes	-0.877	-1.531	-0.223	0.009	-0.797	-1.380	-0.214	0.007
Recurrent sprain (reference No)								
Yes	1.158	0.306	2.009	0.008	1.148	0.378	1.918	0.004
4-weeks predictors								
Pain when bearing weight 4 wks. after injury	**0.019**	**0.005**	**0.033**	**0.007**	**0.026**	**0.013**	**0.039**	**< 0.001**
Another injury (reference No)								
Yes	-0.387	-1.454	0.680	0.476	0.254	-0.642	1.151	0.577
Returned to sports activities (reference Yes)								
No	-0.173	-0.785	0.440	0.580	-0.093	-0.636	0.449	0.736
Difficulty with running, jumping or twisting (pivoting) 4 wks. after injury (reference No)								
Yes	0.041	-0.801	0.882	0.924	-0.420	-1.139	0.299	0.252
Pain on bed at night 4 wks. after injury (reference No)								
Yes	0.555	-0.453	1.563	0.279	0.489	-0.481	1.459	0.322
Difficulty with squatting 4 wks. after injury (reference No)								
Yes	0.137	-0.603	0.877	0.716	0.361	-0.366	1.088	0.329
Ankle swelling 4 wks. after injury (reference Never)								
Rarely/sometimes	0.692	-0.391	1.775		0.656	-0.308	1.619	
Often/Always	0.427	-0.737	1.590	0.384	0.523	-0.501	1.546	0.404
Ankle grinding/clicking (reference Never)								
Rarely/sometimes	0.652	-0.036	1.340		0.608	0.010	1.206	
Often/Always	0.409	-0.480	1.298	0.177	0.275	-0.515	1.066	0.313
Ankle catching/locking (reference Never)								
Rarely/sometimes	-0.279	-0.982	0.423		-0.372	-1.004	0.260	
Often/Always	0.622	-0.410	1.654	0.193	0.738	-0.261	1.738	0.497
Ankle ROM plantar dorsiflexion 4 wks. after injury (reference Always/often)								
Sometimes/rarely	-0.387	-1.055	0.281		-0.229	-0.827	0.368	
Never	0.500	-0.512	1.513	0.159	0.408	-0.543	1.359	0.387
Intercept	-2.215	-3.433	-0.997	< 0.001	-1.594	-2.674	-0.515	0.004

95% CI: 95% confidence interval

Notes: Candidate predictors that were statistically significant according to AIC criteria are in bold. Although education level was a statistically significant predictor for outcome 2, a decision has been made not to include it in the final model (see text for details).

Table 20: Estimates of the final models for the prediction of outcomes 1 and 2 occurrence.

Variable	Outcome 1				Outcome 2			
Baseline models	**Coefficient**	**95% CI**		**p**	**Coefficient**	**95% CI**		**p**
Age	0.027	0.006	0.048	0.014	-	-		-
BMI	0.031	-0.014	0.076	0.178	-	-		-
Pain when resting	0.016	0.005	0.027	0.005	0.014	0.004	0.024	0.008
Pain when bearing weight	0.019	0.004	0.035	0.016	0.015	0.001	0.029	0.033
Days from injury to assessment (reference 0-2)								
3 or more	0.854	0.068	1.640	0.034	0.650	0.019	1.280	0.043
Able to bear weight at Baseline (reference No)								
Yes	-0.792	-1.376	-0.207	0.008	-0.705	-1.225	-0.184	0.008
Recurrent sprain (reference No)								
Yes	1.180	0.417	1.944	0.003	1.100	0.388	1.813	0.003
Intercept	-1.580	-2.152	-1.008	< 0.001	-1.080	-1.513	-0.647	< 0.001
Updated models								
(baseline + 4 weeks predictors)	**Coefficient**	**95% CI**		**p**	**Coefficient**	**95% CI**		**p**
Age	0.018	-0.005	0.040	0.127	-	-		-
BMI	0.025	-0.022	0.072	0.292	-	-		-
Pain when resting	0.010	-0.002	0.022	0.107	0.005	-0.006	0.016	0.381
Pain when bearing weight	0.014	-0.002	0.030	0.092	0.010	-0.004	0.024	0.176
Pain when bearing weight 4 wks. after injury	0.022	0.012	0.032	< 0.001	0.026	0.016	0.035	< 0.001
Days from injury to assessment (reference 0-2)								
3 or more	0.702	-0.117	1.520	0.092	0.444	-0.230	1.118	0.194
Able to bear weight at Baseline (reference No)								
Yes	-0.802	-1.412	-0.192	0.010	-0.741	-1.288	-0.194	0.008
Recurrent sprain (reference No)								
Yes	1.170	0.386	1.953	0.004	1.168	0.416	1.919	0.002
Intercept	-1.543	-2.128	-0.958	< 0.001	-1.012	-1.468	-0.557	< 0.001

95% CI: 95% confidence interval

Linear terms selected by the MFP for continuous predictors for both outcomes 1 and 2: Age – 29.88; BMI – 26.32; Pain when resting – 37.75; Pain when bearing weight – 75.40; Pain when bearing weight at 4 weeks after injury – 36.23.

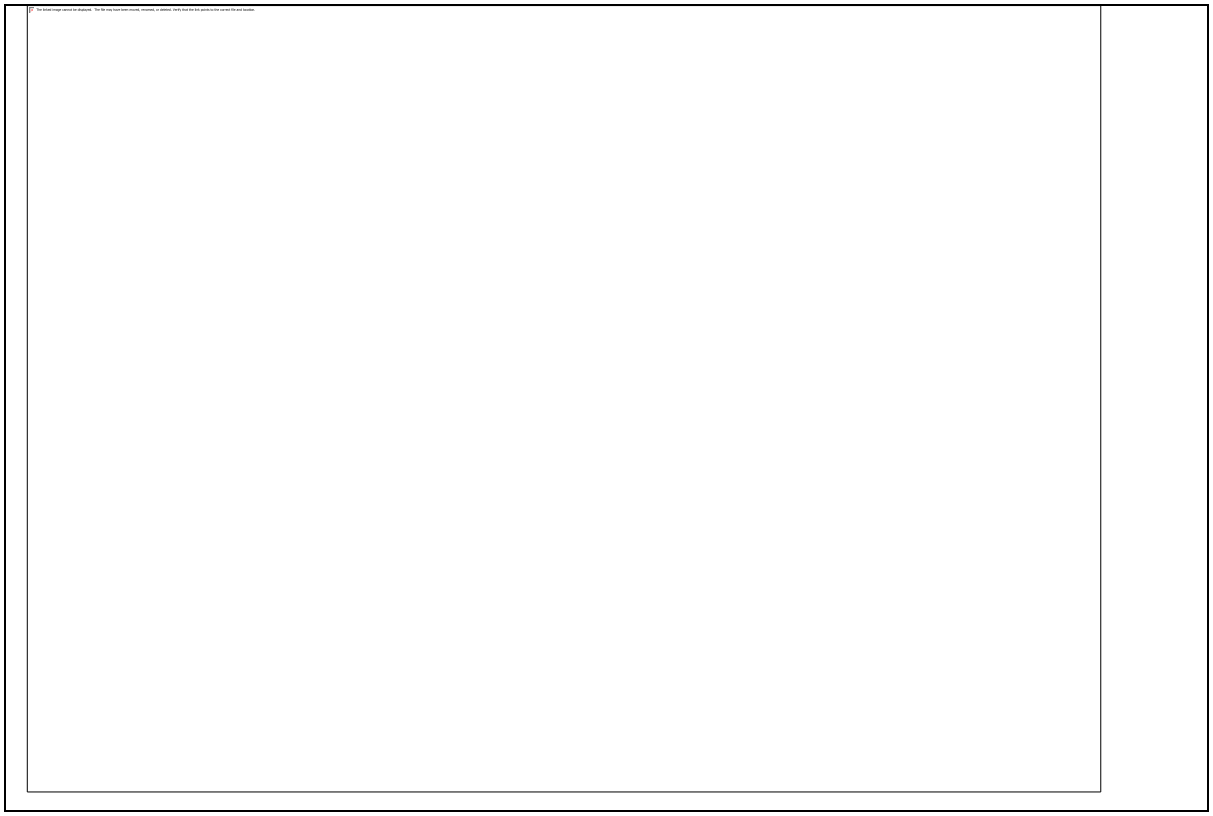


Figure 12: DCA plot for Outcome 1.

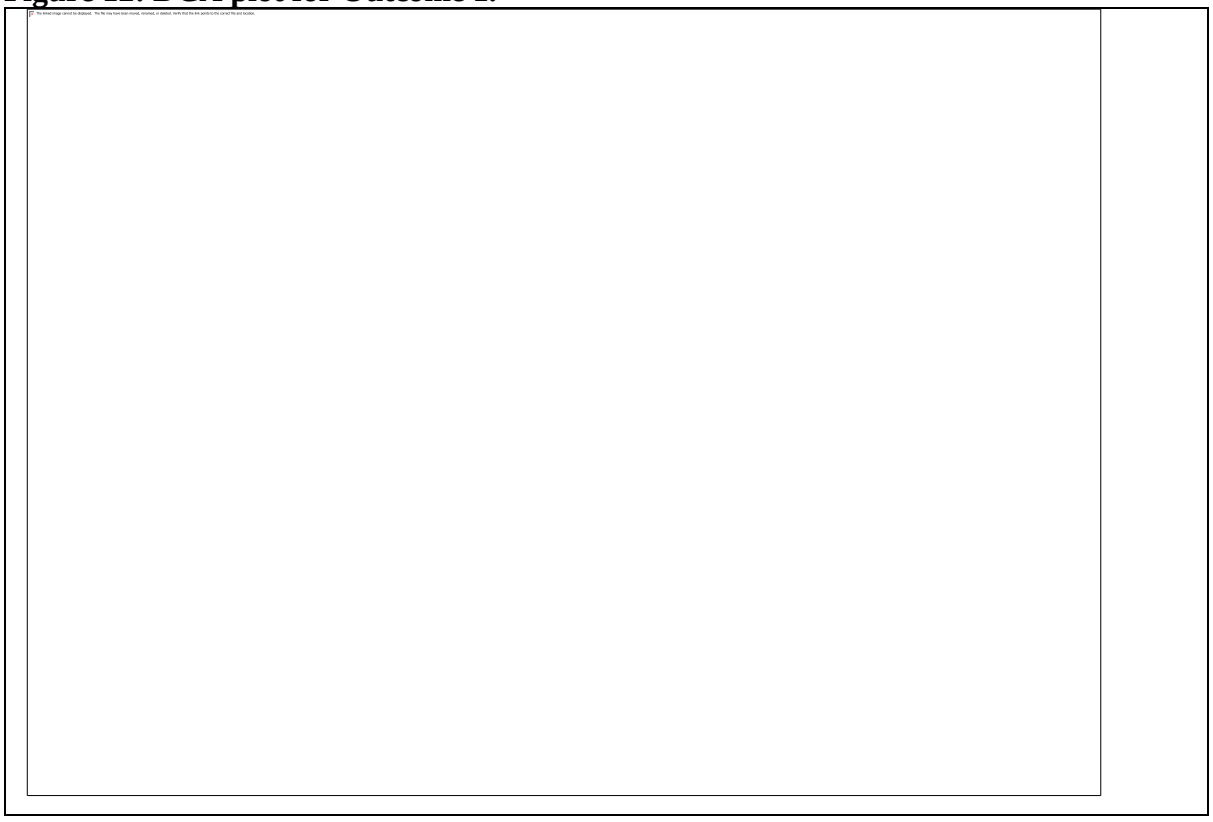
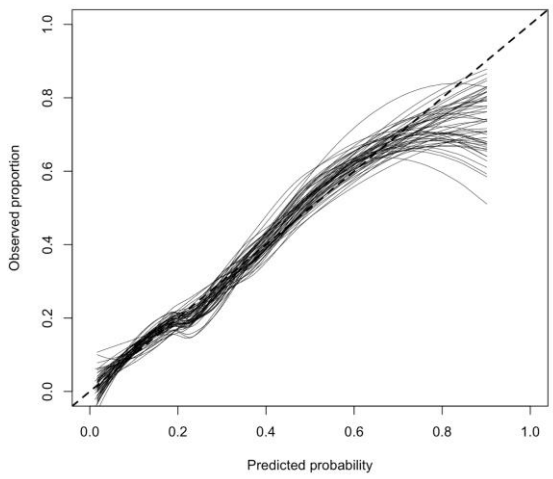


Figure 13: DCA plot for Outcome 2.

5.3.3 Model Performance

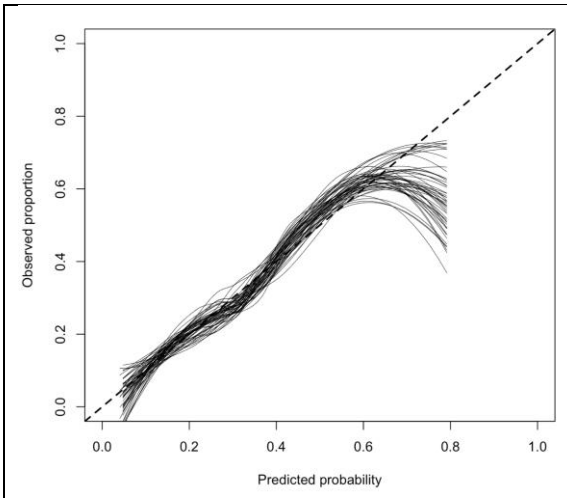
Model performance was assessed in terms of calibration and discrimination. The overall discriminatory ability of the model (apparent performance), as measured by the c-statistic estimated after regressing the predictors selected for the final model against the outcomes using the original CAST dataset (complete cases analysis; n = 194 and 200 for outcome 1 and 2, respectively), was 0.82 (95% CI 0.75 to 0.89) for the model developed to predict outcome 1 and 0.73 (95% CI 0.66 to 0.81) for the model developed to predict outcomes 2. The combined results from the analysis of the 50 imputed provided a less optimistic measure of the discriminatory ability for the two models. For the model developed to predict outcome 1, the combined c-statistic was 0.74 (95% CI 0.70 to 0.79). For the model developed to predict outcome 2, the combined c-statistic was 0.70 (95% CI 0.65 to 0.74). The addition of one variable with information on pain when bearing weight on the ankle at 4 weeks after the injury improved the discriminatory ability and apparent calibration of both models. For the updated model to predict outcome1, the c-statistic was 0.77 (95% CI 0.73 to 0.82). For the updated model to predict outcome 2, the c-statistic was 0.75 (95% CI 0.71 to 0.80).

Calibration plots overlying the results of the analysis on the 50 imputed datasets are presented in Figure 14 and Figure 15. Perfect predictions should lie on the 45-degree line for agreement with the outcome in the calibration plot. As anticipated, on average the calibration across all models was consistently strong with close agreement between observed and predicted risk of developing outcomes 1 (Figure 14) and 2 (Figure 15). Shrinkage suggested both prognostic models to be unstable, with a considerable amount of optimism. The heuristic shrinkage factor for the coefficients of the predictors in the baseline prognostic model for outcome 1 was 0.71, suggesting that 29% of the model fit was non-replicable noise. For the updated versions (baseline and 4 weeks predictors) of both prognostic models, the estimated heuristic shrinkage factor was 0.84. The shrunk coefficients and intercepts are presented in Table 21.



i. ii.

Figure 14: Calibration plot of the prognostic model for outcome 1 for (i.) the baseline model and (ii.) the updated baseline plus 4 weeks model.



i. ii.

Figure 15: Calibration plot of the prognostic model for outcome 2 for for (i.) the baseline model and (ii.) the updated baseline plus 4 weeks model.

Table 21: Intercept and regression coefficients of the prediction models for poor recovery 9 months after ankle sprain (outcomes 1 and 2), before and after correction for optimism (shrinkage).

Predictors in the baseline models	Outcome 1		Outcome 2	
	Coefficient	Shrunk coefficient	Coefficient	Shrunk coefficient
Age	0.027	0.019	-	-

BMI	0.031	0.022	-	-
Pain when resting	0.016	0.011	0.014	0.008
Pain when bearing weight	0.019	0.014	0.015	0.009
> 2 days from injury to assessment	0.854	0.605	0.650	0.396
Able to bear weight on the injured ankle	-0.792	-0.561	-0.705	-0.429
Recurrent sprain	1.180	0.836	1.100	0.670
Intercept	-1.580	-1.570	-1.080	-1.075
Predictors in the updated models (baseline + 4 weeks variables)	**Coefficient**	**Shrunk coefficient**	**Coefficient**	**Shrunk coefficient**
Age	0.018	0.015	-	-
BMI	0.025	0.021	-	-
Pain when resting	0.010	0.008	0.010	0.010
Pain when bearing weight	0.014	0.012	0.010	0.010
Pain when bearing weight 4 wks. after injury	0.022	0.018	0.026	0.022
> 2 days from injury to assessment	0.702	0.591	0.444	0.373
Able to bear weight on the injured ankle	-0.802	-0.676	-0.741	-0.623
Recurrent sprain	1.170	0.985	1.168	0.982
Intercept	-1.543	-1.527	-1.012	-1.007

5.3.4 Application of the SPRAINED model

The following section will provide example of how the internally validated SPRAINED model can be applied in practice. To make predictions with the SPRAINED prognostic models, the following equations are required (please note that all linear terms selected by the MFP for continuous predictors were incorporated into the respective models intercepts):

Baseline model for outcome 1

$$Y = -3.83 + (0.02 \times \text{age}) + (0.02 \times \text{BMI}) + (0.01 \times \text{pain when resting}) + (0.01 \times \text{pain when bearing weight}) + (0.61 \text{ if days from injury to assessment} > 2) - (0.56 \text{ if able to bear any weight on the injured ankle}) + (0.84 \text{ if the injury is a recurrent sprain})$$

Then, we need to convert the log odds (Y) into probability. This can be done by applying the following equation:

$$P = 1/(1 + \exp(-Y))$$

Where P is the probability of developing the outcome and Y is the log odds estimated with the model. To provide a practical example of how to use the SPRAINED prognostic model to predict the occurrence of outcome 1, we consider a hypothetical patient.

"Patient with ankle sprain, male, 38 years old, presenting at the emergency department 3 days after occurrence of the injury, with an estimated BMI of 25.6 kg/m^2, reporting pain when resting of 50 on the visual analogue scale, and 80 when bearing some weight on the injured ankle, willing to bear weight on the ankle and stating this is a recurrent injury, due to the practice of basketball."

To calculate the risk of having a poor recovery from this ankle sprain 9 months after the injury the information on the relevant predictors must be entered in the model.

Predictor	**Information**
Age	38
BMI	25.6
Pain when resting	50
Pain when bearing weight	80
>2 days from injury to assessment	Yes
Able to bear weight on the injured ankle	Yes
Recurrent sprain	Yes

Applying the equation,

(1) $Y = -3.83 + (0.02 \times 38) + (0.02 \times 25.6) + (0.01 \times 50) + (0.01 \times 80) + 0.61 - 0.56 + 0.84$

(2) $Y = -3.83 + 0.76 + 0.51 + 0.50 + 0.80 + 0.61 - 0.56 + 0.84$

(3) $Y = -0.37$

Applying the transformation,

(4) $P = 1/(1 + \exp(0.37))$

(5) $P = 1/(1 + 1.45)$

(6) $P = 1/2.45$

(7) $P = 0.41$ (or 41%)

The estimated probability of developing poor outcome 9 months after ankle sprain (as per the definition of outcome 1) for that patient would be 41%.
Now, if we had the chance of reassessing the patient four weeks after the injury, assessed their pain when bearing weight at this stage (say, 30 in a scale from 0-100) and applied the updated model (baseline + 4 weeks predictor), the following would need to be done.

Updated model for outcome 1 (baseline + 4 weeks predictors)

Y = -4.38 + (0.01 $\times$ age) + (0.02 $\times$ BMI) + (0.01 $\times$ pain when resting) + (0.01 $\times$ pain when bearing weight) + (0.59 if days from injury to assessment > 2) – (0.68 if able to bear any weight on the injured ankle) + (0.99 if the injury is a recurrent sprain) + (0.02 $\times$ pain when bearing weight 4 weeks after injury)

Applying the equation,

(1) $Y = -4.38 + (0.01 \times 38) + (0.02 \times 25.6) + (0.01 \times 50) + (0.01 \times 80) + 0.59 – 0.68 + 0.99 + (0.02 \times 30)$

(2) $Y = -4.38 + 0.38 + 0.51 + 0.50 + 0.80 + 0.59 – 0.68 + 0.99 + 0.60$

(3) $Y = -0.69$

Applying the transformation,

(4) $P = 1/(1 + \exp(0.69))$

(5) $P = 1/(1 + 1.99)$

(6) $P = 1/2.99$

(7) $P = 0.33$ (or 33%)

Therefore, by adding extra information on the patient follow-up, we were able to estimate a more precise probability of presenting poor outcome at 9 months after injury.

To calculate the risk of having poor recovery at 9 months after ankle sprain according to the definition of outcome 2, the following model should be applied:

Baseline model for outcome 2

Y = -2.2 + (0.01 $\underline{\times}$ pain when resting) + (0.01 $\underline{\times}$ pain when bearing weight) + (0.40 if days from injury to assessment > 2) – (0.43 if able to bear any weight on the injured ankle) + (0.67 if the injury is a recurrent sprain)

Applying the equation,

(1) $Y = -2.20 + (0.01 \times 50) + (0.01 \times 80) + 0.40 – 0.43 + 0.67$

(2) $Y = -2.20 + 0.50 + 0.80 + 0.40 – 0.43 + 0.67$

(3) $Y = -0.26$

Applying the transformation,

(4) $P = 1/(1 + \exp(0.26))$

(5) $P = 1/(1+1.30)$

(6) $P = 1/2.30$

(7) $P = 0.43$ (or 43%)

For the same patient, the probability of developing poor outcome nine months after ankle sprain (as per outcome 2 definition) would be slightly higher (43%).

To calculate the updated probability of this patient presenting poor outcome at 9 months using the model with baseline and 4 weeks predictors (considering that in the reassessment, their pain score when bearing weight was 30) the following equation should be applied:

Updated model for outcome 2 (baseline + 4 weeks predictors)

Y = -2.85 + (0.01 $\underline{\times}$ pain when resting) + (0.01 $\underline{\times}$ pain when bearing weight) + (0.37 if days from injury to assessment > 2) – (0.62 if able to bear any weight on the injured ankle) + (0.98 if the injury is a recurrent sprain) + (0.02 $\underline{\times}$ pain when bearing weight 4 weeks after injury)

Applying the equation,

(1) $Y = -2.85 + (0.01 \times 50) + (0.01 \times 80) + 0.37 – 0.62 + 0.98 + (0.02 \times 30)$

(2) $Y = -2.85 + 0.50 + 0.80 + 0.37 – 0.62 + 0.98 + 0.60$

(3) $Y = -0.22$

Applying the transformation,

(4) $P = 1/(1 + \exp(0.22))$

(5) $P = 1/(1 + 1.25)$

(6) $P = 1/2.25$

(7) $P = 0.44$ (or 44%)

Therefore, by adding extra information on the patient follow-up, the updated probability of presenting poor outcome at 9 months after injury was 44%.

The observational cohort study to enable external validation of the prognostic models presented is reported in the following chapter. The results of the prognostic model development and external validation are summarised and discussed together in Chapter 7.

6. External validation study of the SPRAINED prognostic models

6.1 Introduction

This chapter describes the external validation process of the two prognostic models (and respective updates) developed to predict the risk of poor outcome at 9 months after an acute ankle sprain. A prospective observational cohort study was conducted with the aim of obtaining data to externally validate and optimise the prognostic models for use in EDs. Before participant recruitment began the models were developed, corrected for optimism and updated with the inclusion of an additional predictor for which information was collected at 4 weeks after the injury using the CAST data set (Chapter 5), which was subsequent to a systematic literature review (Chapter 3) and a consensus process involving clinician and patient perspectives (Chapter 4).

6.2 Methods

6.2.1 Cohort design and study population

The SPRAINED cohort recruited people with acute ankle sprain attending 10 NHS EDs across England (see acknowledgements for details on recruiting centres), over a period of nine months (July 2015 to March 2016). This was an observational cohort study, therefore participants were not randomised nor did they receive any interventions other than usual care at each site. Data collection took place at the time of participant's presentation to any of the study recruiting sites (baseline) and subsequently at 4 weeks, 4 and 9 months after the initial injury.

People were invited to take part in the study if they meet the following inclusion criteria, and had no exclusion criteria:

Inclusion Criteria

1. Participant was willing and able to give informed consent for participation in the study
2. Male or Female, aged 16 years or above
3. Diagnosed with acute ankle sprain (Grade I to III, <7 days old)

Exclusion Criteria

1. Ankle fracture (apart from flake fractures <2mm);
2. Other recent (<3 months) lower limb fracture

6.2.2 Sample size

The recommended sample size estimation for an external validation of a prognostic model is that 100 outcome events are required, this being the minimum number needed to ensure accurate estimation of the calibration of the model.[67, 71] The event rates for the outcomes of interest in the CAST study were between 26% and 32%, depending on the outcomes' definition (3 symptoms and 4 symptoms/clinical events, respectively), this would require an overall sample size of between 313 and 385. Assuming a 25% loss to follow-up and a lower event rate (20%) when recruiting all grades of ankles sprains, a minimum of 675 participants were targeted for recruitment to increase the chances of achieving the required event rates. We anticipated recruiting people with a range of sprains, including Grade I to III.

6.2.3 Screening and eligibility assessment

People were screened by clinicians on admission to EDs and assessed for eligibility to take part in the SPRAINED cohort. A member of the research team at the study centres administered the study Clinical Dataset Form (CDF) and recorded responses and findings from the clinical examination (see Appendix C). The short CDF served three purposes:

1. collection of routine core clinical data set in a tick box format (reflecting the data that would be normally recorded in the course of routine clinical practice)
2. a tick box to record that clinicians had provided potential participants with the trial information pack and a brief explanation of the trial
3. a tick box to record whether the individual had given permission for a member of the research team make contact to discuss the study further and complete the informed consent process

One copy of the CDF was filed in the person's medical notes as a treatment record and a second copy, where agreement was given, passed to the local research team. The team member then contacted the individual and continued the informed consent process. Only once consent was obtained was the clinical dataset sent to the central study office. The clinical dataset of any people who did not agree to study participation remained at the site in their medical notes.

6.2.4 Informed consent and recruitment

The initial approach was made by a member of the ED clinical team. A verbal explanation of the study along with a study information leaflet were given to all potentially eligible people. Posters were displayed in all participating departments to inform participants that the study was occurring.

The informed consent process was carried out by a registered health care professional with delegated authority from the Principal Investigator (PI) at the recruiting site. Prior to consenting to participate in the study, the person was asked by a member of the local clinical team for permission to allow the local research team to speak to them, either in person or by telephone, to take forward the informed consent process. Formal consent to participation was provided either in person or by post or telephone. Before any data were provided to the study team, the participant personally signed and dated the latest approved version of the Informed Consent form (ICF), or verbal consent was recorded by a member of the local team on a form during the informed consent telephone call. The participant had the opportunity to question the clinical/research team, and to consult their GP or other independent parties to decide whether they would participate in the study.

Written informed consent was obtained by means of participant dated signature and dated signature of the person who presented and obtained the informed consent. Verbal informed consent was obtained by means of the dated signature of the local team member taking consent over the telephone. A copy of the completed written or verbal ICF was retained by the participants (or posted to the participant in the case of oral consent). One copy was sent to the study coordinating team in Oxford. The original signed Consent Form was retained in the medical notes, and a copy held in the Investigator Site File (ISF).

Participants consented to allow the study to use the CDF completed during the ED attendance and an additional questionnaire 4 weeks following this (SPRAINED study prognostic model and any additional important information), as well as follow up questionnaires at 4 and 9 months which aimed to map the recovery trajectory and final recovery status at 9 months. A questionnaire at 4 months served as reminder of the study, and as loss to follow up was likely to become larger over time, helped ensure that responses on the core components of the outcomes of interest were available for as many participants as possible.

6.2.5 Data collection and management

Baseline data were collected from participants and recorded on a paper CDF. Data for the three study follow up points (4 weeks [prognostic variables], 4 and 9 months [outcome data] after baseline assessment) were collected by using paper CRFs sent to participants via post, or by telephone call where necessary. The telephone calls enabled collection of at least the core data on the outcome measures for participants that did not return the questionnaire to the trial office. Where preferred by the participant, secure online data collection took place for the 4 week timeframe.

Baseline CDFs were sent by a member of the local research team to the study coordinating office in Oxford by post. Follow up CRFs were sent by the participant to the study coordinating office in Oxford by post, using a Freepost return envelope. Where telephone follow up was used, a member of the central study team recorded data directly onto the relevant forms.

Upon receipt of data forms (CDFs and CRFs), appropriate data quality and validation checks were carried out and the data entered into a study-dedicated database which was developed and maintained by OCTRU, a UKCRN Registered Clinical Trials Unit. OpenClinica software (OpenClinica, LLC, Waltham, MA, USA) was used to develop and maintain the study database. To identify manual entry errors a 10% double entry check was carried out at regular intervals during the data collection phase of the study.

Details relating to ethics approvals and monitoring are outlined in section 2.8.

6.2.6 Study assessments

Baseline Assessments

Baseline data was collected on the clinician-completed CDF which includes:

1. Demographics (name, age, contact details)
2. Patient history
3. Clinical examination
4. Clinical Investigation
5. Clinical Management
6. Clinical Diagnosis
7. Prognostic factors

8. Agreement for research team to contact patient

Participant contact details were also collected at baseline to facilitate study follow up. This included full name, address, NHS number, mobile and/or telephone number, email address and a preferred time to be contacted. Reasons for declining the study were collected if given.

Follow up assessment 1 (prognostic variables at 4 weeks after ankle sprain)

Follow up at 4 weeks after ankle sprain was conducted by electronic, telephone or postal questionnaire. Questions included:

1. Current clinical status (recurrence of injury, swelling or pain in the ankle)
2. Return to normal activities

Follow up assessments 2 and 3 (outcome variables at 4 and 9 months after ankle sprain)

Follow up at 4 and 9 months after ankle sprain were conducted by postal or telephone questionnaire. Questions included:

1. Recurrence of injury
2. Foot and Ankle Outcome Score (FAOS)
3. Health service resource use
4. Health related quality of life (EuroQuol EQ-5D)

6.2.7 Outcome measures

For the external validation dataset (SPRAINED observational cohort study), poor outcome at 9 months after ankle sprain was defined in the same way as it was in the development study (Chapter 5). The same questions were asked to SPRAINED participants, so the same two outcomes could be constituted. Therefore, the definition of poor outcome was the presence of any (or a combination) of the following symptoms or clinical events (for further detail see section 5.2.2):

Outcome 1

- Severe persistent pain
- Severe functional difficulty
- Significant lack of confidence in the ankle

Outcome 2

- Severe persistent pain

- Severe functional difficulty
- Significant lack of confidence
- Recurrent sprain

6.2.8 Predictors of poor outcome 9 months after ankle sprain

All variables included in the prognostic models developed to predict the occurrence of poor outcome 9 months after ankle sprain (the SPRAINED prognostic models, see Chapter 5) were included in the baseline CRFs. Data collection on a few additional candidate predictors that did not make to the final models were also conducted to allow some room for model updating, if necessary. However, the data collected at baseline were kept to a minimum, prioritising the predictors included in the two developed models and those candidate predictors the consensus group considered to have most clinical importance and relevance to patients. Except for pain scores (collected as discrete variables in the SPRAINED cohort), data collection on all variables was performed respecting their original format in the CAST dataset. A complete list of the variables collected at baseline in both the CAST Trial and SPRAINED cohort, with respective formats and amount of missing data are given in Table 22**.**

6.2.9 Statistical Methods

Exploratory analysis and data transformation

Baseline characteristics of participants were summarised using means, standard deviations (SDs) and ranges for continuous variables, or counts and percentages for categorical variables. To examine differences in case-mix between the participants in the development (CAST Trial) and external validation (SPRAINED cohort), characteristics of participants included in the two studies were compared narratively (no statistical tests performed).

Categorical variables were re-categorised by collapsing some of their categories, to match the format of those included in the regression analyses during the model development stage. The distribution of the continuous predictors were also assessed, first considering their empirical distributions by producing histograms and then by assessing these for normality by means of normal probability plots, box plots and dot plots. The presence of any outliers was assessed based on visual examination of box plots. Extreme values were inspected to confirm whether they were clinically plausible.

Table 22: Predictor variables and their formats in the original CAST dataset, in the final model and SPRAINED dataset (with respective amount of missing data).

Variable Name	CAST dataset		Modelling process/Final model		SPRAINED dataset		
	Type	**Categories /Units**	**Type**	**Categories/ Units**	**Type**	**Categories /Units**	**Miss ing**
Sex	Binary	Male Female	Binary	Male Female	Binary	Male Female	-
Recurrent sprain[1]	Binary	Yes No	Binary	No Yes	Binary	No Yes	6.5%
Able to bear weight on the injured ankle	Contin uous	kg	Binary	No Yes[2]	Binary	No Yes	0.7%
Employme nt status	Catego rical	No Part time Full time Student Retired	Catego rical	None[3] Part time Full time	Catego rical	No Part time Full time Student Retired	0.3%
Injury setting	Catego rical	During sport At work At home In public Other	Catego rical	During sport At work At home In public/Othe r	Catego rical	During sport At work At home In public Other	2.1%
Ankle/foot catching/lo cking	Catego rical	Never Rarely Sometimes Often Always	Catego rical	Never Rarely/Som etimes Often/Alwa ys	Catego rical	Never Rarely Sometimes Often Always	3.1%

Ankle ROM plantar flexion	Categorical	Always Often Sometimes Rarely Never	Categorical	Often/Always Rarely/Sometimes Never	Categorical	Never Rarely Sometimes Often Always	1.5%
Ankle ROM plantar dorsiflexion	Categorical	Always Often Sometimes Rarely Never	Categorical	Often/Always Rarely/Sometimes Never	Categorical	Never Rarely Sometimes Often Always	1.6%
Age	Continuous	Years	Continuous	Years	Continuous	Years	-
Days from injury to assessment [4]	Continuous	Days	Binary	1-2 days 3-7 days	Continuous	Days	-
Body mass index[5]	Continuous	Kg/m^2	Continuous	Kg/m^2	Continuous	Kg/m^2	8.2%
Pain at rest	Continuous	0-100	Continuous	0-100	Discrete	0-10	3.4%
Pain at weight bearing	Continuous	0-100	Continuous	0-100	Discrete	0-10	4.4%
Pain at weight bearing at 4 weeks	Continuous	0-100	Continuous	0-100	Discrete	0-100	50%

[1] An ankle sprain that has happened to a previously injured ankle (at least twice), with the last time being within the past 12 months.

[2] Any value different from 0.

[3] Combination of unemployed, student and retired.

[4] Not allowed more than 7 days in the CAST dataset.

[5] Calculated from height and weight (both continuous variables), as per collected with the Baseline CAST questionnaire.

Handling missing data

As there was more than one predictor with missing data in the SPRAINED observational cohort that were needed to validate the model (up to 8%, for BMI), multiple imputation by chained equations (MICE) was used to replace missing values (percentages of missing data for predictor variables and outcomes are presented in Table 22 and Table 25, respectively). MICE uses a set of imputation equations including one for each of the predictors with missing data; all equations include all of the predictors included in the prediction model, predictors of predictors and the outcomes. It is recommended that the imputation models should take into account all predictors within the analysis model as well as the outcome (to be predicted by the prognostic model). Including more predictors within the imputation model makes the MAR assumption more plausible by potentially including factors that may explain the missingness. Multiple imputation was performed assuming that all missing variable data was MAR. This missing data mechanism assumes that the probability of an observation being missing is dependent on the observed data. To reflect the uncertainty in the imputation, 50 imputed data sets were created. The models were independently estimated for Outcomes 1 and 2, and imputations were therefore performed in separate procedures, producing two different sets of 50 complete datasets (see Chapter 5 for more details on the MICE principles, structure and commands used when handling missing data). Each of the imputed data sets was analysed separately, by calculating the model discrimination and calibration. Combined calibration plots overlaying the calibration lines of the 50 analysed datasets for each outcome. Discrimination is also presented for each model, in terms of c-statistics combined across the 50 analysed datasets for each outcome using Rubin's rules[72].

Model performance

The performance of the prognostic models were assessed in terms of calibration and discrimination. Calibration defined as ''for patients with a predicted risk of R%, on average R out of 100 should indeed suffer from the disease or event of interest". Calibration was assessed graphically by plotting the observed outcomes (on the y-axis) against the predicted probabilities from the models (on the x-axis). To produce the plots, participants were ranked from lowest predicted risk to highest predicted risk, and grouped into tenths of predicted risk

(i.e. 10 equal sized groups). For each of the ten groups, the mean predicted risk and the proportion of observed outcomes were calculated and plotted against each other. A flexible calibration curve was also fit using locally weighted scatterplot smoothing (lowess) to capture the agreement (and any miscalibration) between the observed outcomes and predicted probabilities over the entire probability range.[67]

Discrimination reflects the ability of the model to distinguish between participants who do and do not experience an event during the study period. Discrimination was assessed using the c-statistic, where a value of 0.5 represents chance and 1 represents perfect discrimination [73]. The c-statistic was classified as follows: 0.5 – 0.6 fail, 0.6 – 0.7 poor, 0.7 – 0.8 fair, 0.8 – 0.9 good, 0.9 – 1 excellent. Individual probabilities of developing the outcomes were estimated by applying the developed prognostic models to each participant in the SPRAINED observational cohort dataset. Model performance was assessed for both the baseline and updated (baseline + 4-week predictors) models.

Finally, to estimate the benefit of using the developed prognostic models, the probabilities of developing poor outcome were estimated using the models' equations and patients were ranked according to their estimated risks. These were used to calculate the number of people per 1000 identified as being at high risk according to different selected thresholds and how many of these go on to present the outcomes compared with a strategy where all individuals are deemed at high risk.

Subgroup analysis

As previously stated, a lower rate of poor outcome at 9 months after ankle sprain was expected among the patients included in the SPRAINED observational cohort. One of the inclusion criteria for the CAST trial (development dataset, Chapter 5) stated that patients would be included if they had been diagnosed with an ankle injury of grade 2 (moderate severity) or 3 (severe), and so more likely to have a poor outcome. In the SPRAINED cohort, presenting with an injury of grade 1 (mild severity) was not an exclusion criteria, as the aim was to recruit a more representative sample of the population with this type of injury seeking for medical assistance in the NHS. Therefore, a subgroup analysis was performed, with the aim of applying the prognostic models to a subsample of individuals composed of those

presenting with injury severity of grades 2 or 3 (more similar to the population in the development dataset), to check whether the models would present better performance among this specific group of patients. Model performance in the subgroup of patients with moderate or severe injuries was assessed for both the baseline and updated (baseline + 4-week predictors) models.

Model recalibration

In case of poor performance of the developed models, a strategy of recalibrating the models was planned. Recalibration methods may include adjustment of the intercept, additional adjustment of predictors coefficients (using the same method adopted during the development phase or a different approach), re-estimating predictors coefficients, and adding or removing predictors from the original model.[74] The adopted approach was to re-estimate the intercepts and predictors' coefficients (re-fit the model in the SPRAINED observation cohort dataset). The prognostic models were re-fitted using a logistic regression modelling framework with the logit probability of an adverse outcome as the response variable. The same predictors selected for the two prognostic models were included together in full logistic regression models as independent variables and no exclusion based on the statistical significance of their adjusted relationship with the outcomes was performed. Continuous variables were kept as continuous to avoid loss of prognostic information and the shape of the relationships between continuous predictors and the outcome were investigated and modelling performed using the multivariable fractional polynomial (MFP) algorithm where appropriate. The "best transformation" for each continuous predictor was used when fitting the models (see Chapter 5 for more details on the principles of modelling non-linear relationships by using fractional polynomials in logistic regression analysis). The multivariable models were fitted in each of the 50 complete data sets and the estimated regression parameters (coefficients and variances) were combined using Rubin's rules.

After re-fitting the models, the same shrinkage method used in the development phase (see Chapter 5 for details on the calculation of the heuristic shrinkage factor) was applied to correct the re-estimated intercepts and predictors' coefficients (reduce model optimism). Finally, as with any newly developed prognostic model, updated models should also be externally validated. However, that was out of the scope of the SPRAINED study.

6.3 Results

6.3.1 Exploratory analysis

The study recruited a cohort of 682 participants across 10 EDs between 20th July 2015 and 17th March 2016. The flow of participants through the cohort study is detailed in Figure 16. Baseline characteristics of the SPRAINED observational cohort study participants are summarised in Table 23. On average, participants were slightly older in the SPRAINED cohort than in CAST (33.62 vs. 29.88 years, respectively). Participants in SPRAINED cohort had an average BMI within the overweight category (27.08 kg/m^2), likewise those in the CAST trial (26.34 kg/m^2). The mean pain scores when resting (38.5 points) or bearing weight on the ankle (71.3 points) of the SPRAINED cohort participants were also very similar to those observed for the CAST participants (37.75 and 75.42 points, respectively). Differently from CAST, in the SPRAINED cohort about half of participants were female (52.05%), presented to an ED within 2 days from injury for assessment (90.03%) and were able to bear some weight on their injured ankles (73.56%).

Continuous predictor variables presented at least a minimal departure from a normal distribution, as evidenced in Figure 17 and Figure 18. Some outliers were observed for participant age and BMI. However, all extreme values were clinically plausible, so no observations were dismissed. Correlations between predictors are presented in Table 24**,** ranging from very low (r = 0.011 for body mass index and ability to bear weight on the injured ankle) to moderate (r = 0.549 for pain when resting and pain when bearing weight) values, which did not raise concerns about including them together in a multivariable model.

Events rates in the SPRAINED cohort and CAST datasets for both outcomes, and the number of symptoms, at 9 months after injury are described in Table 25. There was a lower rate of poor outcome in the SPRAINED cohort than for the CAST cohort.

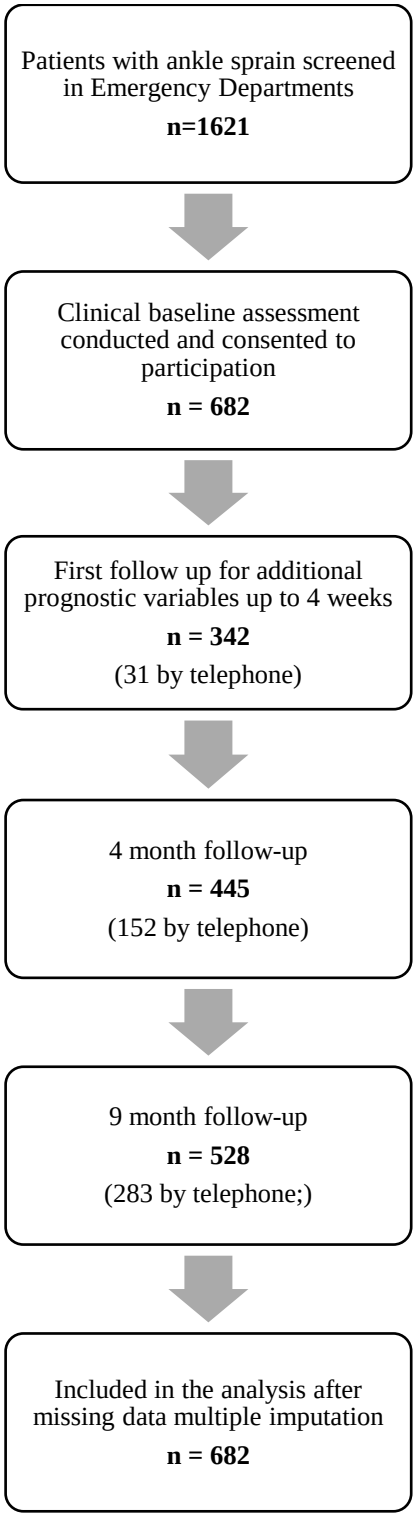


Figure 16: Flow of participants through the SPRAINED observational cohort study.

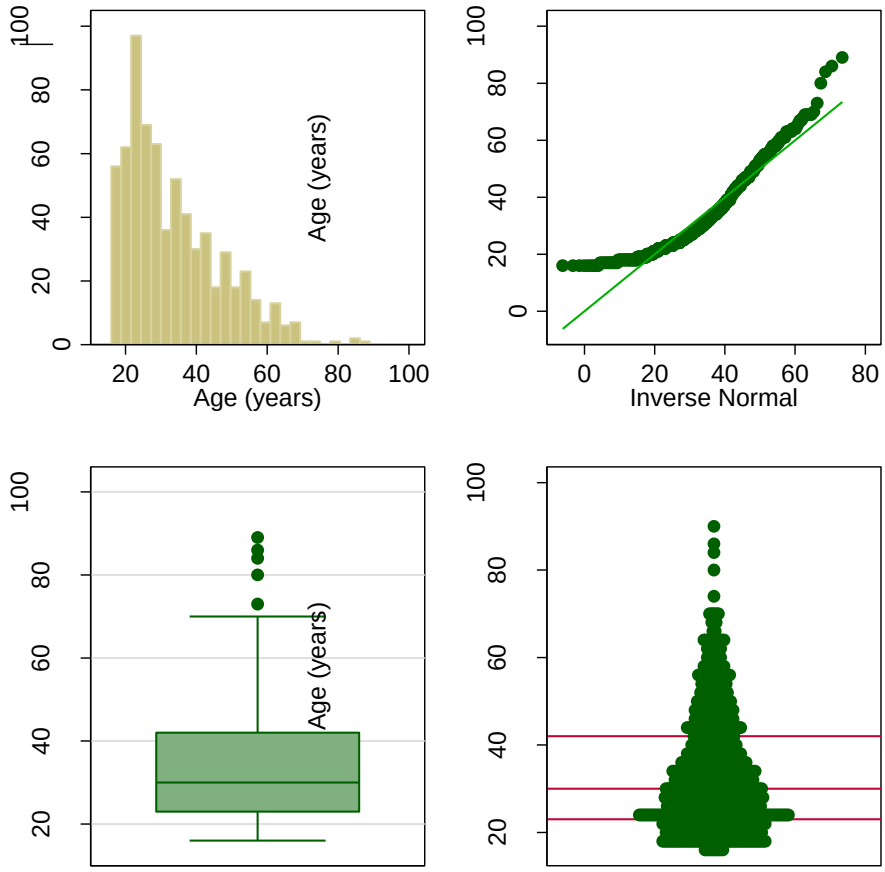


Figure 17: Histogram (top left), normal plot (top right), box plot (bottom left) and dot plot (bottom right) of the distribution of age values in the SPRAINED dataset.

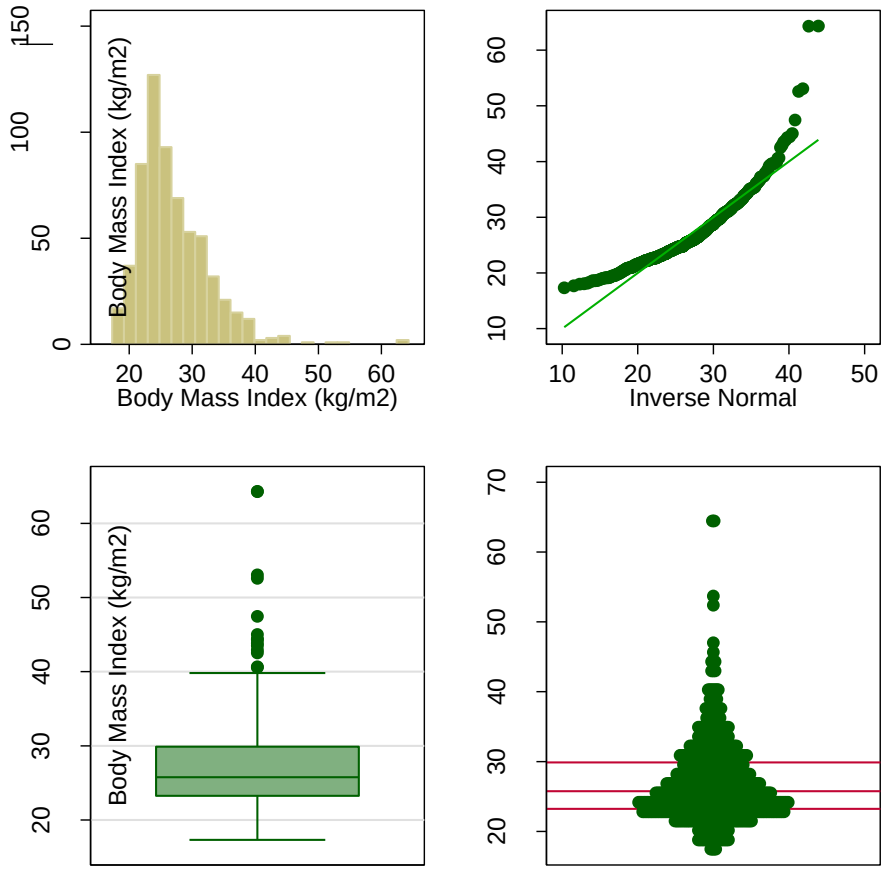


Figure 18: Histogram (top left), normal plot (top right), box plot (bottom left) and dot plot (bottom right) of the distribution of BMI values in the SPRAINED dataset.

Table 23: Baseline characteristics of the CAST Trial and SPRAINED Cohort samples.

	CAST Trial		**SPRAINED Cohort**	
Variable	**Mean (SD)**	**Min - Max**	**Mean (SD)**	**Min - Max**
Age (years)	29.88 (10.77)	16 – 72	33.62 (13.38)	16 – 89
Height (m)	1.73 (0.98)	1.47 – 2.01	1.72 (1.02)	1.50 – 2.01
Weight (kg)	78.56 (15.44)	39.92 – 133.36	80.44 (18.13)	44.50 – 180.00
Body mass index (kg/m^2)	26.34 (5.19)	16.07 – 53.77	27.08 (5.70)	17.31 – 64.30
Pain when resting (score)	37.75 (23.49)	0 – 100	38.50 (22.50)	0 – 100
Pain when bearing weight (score)	75.42 (19.61)	0 – 100	71.30 (21.00)	0 – 100
Sex	**Frequency**	**%**	**Frequency**	**%**
Male	337	57.71	327	47.95
Female	247	42.29	355	52.05
Days from injury to assessment				
0-2	118	44.87	614	90.03
3 or more	145	55.13	68	9.97
Able to bear weight at Baseline assessment				
No	446	77.03	179	26.44
Yes	133	22.97	498	73.56
Sprained the same ankle in the last 12 months				

No	197	68.40	590	87.80
Yes	91	31.60	82	12.20
Sprained the same ankle at least twice before				
No	176	61.32	472	73.63
Yes	111	38.68	169	26.37
Recurrent sprain				
No	517	90.38	583	91.38
Yes	55	9.62	55	8.62
Current employment				
None	132	22.60	161	23.68
Part time	92	15.75	92	13.53
Full time	360	61.64	427	62.79
Injury mechanism				
At home	99	18.00	144	21.56
Practicing sports	203	36.91	230	34.43
At work	79	14.36	91	13.62
Outside, in public	169	30.73	203	30.39
Ankle catching/locking				
Never	286	50.53	539	81.54
Rarely/sometimes	209	36.93	99	14.98
Often/Always	71	12.54	23	3.48

Able to perform ankle range of motion (ROM)plantar flexion				
Always/often	101	17.84	170	25.30
Sometimes/rarely	247	43.64	230	34.23
Never	218	38.52	272	40.48
Able to perform ankle range of motion (ROM) dorsiflexion				
Always/often	81	14.31	186	27.72
Sometimes/rarely	227	40.11	228	33.98
Never	258	45.58	257	38.30
Injury severity				
Grade 1	-	-	302	48.55
Grade 2	-	-	285	45.85
Grade 3	-	-	35	5.63

Table 24: Spearman correlation coefficient matrix for all predictors included in the SPRAINED prognostic model for the risk of poor outcome nine months after acute ankle injury.

	Age	Body mass index	Pain when resting	Pain when bearing weight	Days from injury to assessment	Able to bear weight on the injured ankle
Age	-					
Body mass index	0.222	-				
Pain when resting	0.021	0.060	-			
Pain when bearing weight	-0.001	0.120	0.549	-		
Days from injury to assessment	0.083	0.047	-0.084	-0.116	-	
Able to bear weight on the injured ankle	0.050	0.011	-0.258	-0.393	0.110	-
Recurrent sprain	-0.127	-0.021	0.095	-0.031	0.053	-0.009

Table 25: Frequency and rate of outcomes and component symptoms in the CAST and SPRAINED datasets.

Symptoms/events	Pain	Lack of confidence	General difficulty	Re-injury	Outcome 1	Missing	Outcome 2	Missing	TOTAL[1]
CAST dataset	84 (14.4%)	42 (7.2%)	67 (11.5%)	46 (7.9%)	116 (19.9%)	144 (24.7%)	140 (24.0%)	144 (24.7%)	584
SPRAINED dataset	3 (0.4%)	23 (3.4%)	37 (5.4%)	78 (11.4%)	46 (6.7%)	155 (22.7%)	109 (16.0%)	150 (22.0%)	682

Note: [1]Total corresponds to number of participants recruited

6.3.2 Model performance

The performance of the prediction models in the external validation dataset (SPRAINED cohort) were assessed in terms of calibration and discrimination. Calibration was graphically assessed with a calibration plots that shows calibration lines for each of the 50 imputed datasets, which was supplemented with the calibration slope and intercept. These parameters were first estimated with the original prognostic model, with poor outcome 9 months after ankle sprain (yes/no) as the outcome variable, and the linear predictor (log odds) of the original prediction model (see Chapter 5 for the equation to calculate the linear predictor) as the only covariate.

Combined performance measures (by using Rubin's rules) are presented in Table 26 and calibration plots overlaying the calibration lines from the 50 individual calibration plots are presented in Figure 19 and Figure 20.

Overall, discrimination of the models for outcome 1 stayed fairly stable, when compared with the performance of the model in the development dataset: combined c-statistic 0.72 (95% CI 0.66 to 0.79) for outcome 1. For outcome 2, a decrease in the discriminatory ability was noted, c-statistic 0.63 (95% CI 0.58 to 0.69).

Calibration of the prognostic model in the external validation dataset was poor for outcome 1, as can be evidenced by inspecting Figure 19 (a calibration plot with overlaid calibration lines from the 50 imputed datasets). Well calibrated models should produce calibration lines lying on (or at least close to) the 45 degrees dashed line of perfect prediction (observed proportion and predicted probability matching perfectly). In this scenario, the calibration slope would be equal (or very close) to 1 and the calibration intercept equal (or very close) to 0. The combined calibration slope was bigger than 1 (1.13, 95% CI 0.76 to 1.5) and the calibration intercept was smaller than 0 (-0.71, 95% CI -0.98 to -0.44).

A calibration slope bigger than 1 indicates that the regression coefficients of the original model were too close to zero; which was the case after the correction for optimism (shrinkage) of the model. A calibration intercept different from 0 indicates that the model's predicted probabilities in the validation dataset are systematically too high (intercept < 0) or too low (intercept > 0).

For the prognostic model developed to predict outcome 2, calibration was better than for the model to predict outcome 1 in terms of the calibration intercept (-0.08, 95% CI -0.27 to 0.11), and slope (1.03, 95% CI 0.65 to 1.42) (Table 26). The updated model (baseline + 4-weeks predictors) for outcome 1 presented a better discriminatory ability in the SPRAINED dataset than the baseline model (c-statistic = 0.78, 95% CI 0.72 to 0.84), but not better calibration in terms of intercept (-0.51, 95% CI -0.78 to -0.24). The same was observed for the updated model for outcome 2 (better discrimination but worse calibration) (Table 26).

Table 26: Summary of the combined performance measures (discrimination and calibration) for the prognostic models applied to the participants in the SPRAINED sample.

Model	c-statistic (95% CI)	Intercept (95% CI)	Slope (95% CI)
Outcome 1			
Baseline model	0.73 (0.66 to 0.79)	-0.71 (-0.98 to -0.44)	1.13 (0.76 to 1.50)

Updated model (baseline + 4-weeks predictors)	0.78 (0.72 to 0.84)	-0.51 (-0.78 to -0.24)	1.17 (0.86 to 1.48)
Baseline model applied to participants with moderate/severe injury (grades 2 and 3)	0.73 (0.64 to 0.81)	-0.93 (-1.33 to -0.53)	1.12 (0.55 to 1.69)
Updated model (baseline + 4-weeks predictors) applied to participants with moderate/severe injury (grades 2 and 3)	0.80 (0.72 to 0.88)	-0.74 (-1.14 to -0.33)	1.30 (0.81 to 1.78)
Outcome 2			
Baseline model	0.63 (0.58 to 0.69)	-0.08 (-0.27 to 0.11)	1.03 (0.65 to 1.42)
Updated model (Baseline + 4-weeks predictors)	0.64 (0.59 to 0.69)	0.19 (-0.01 to -0.38)	0.68 (0.46 to 0.91)
Baseline model applied to participants with moderate/severe injury (grades 2 and 3)	0.62 (0.54 to 0.69)	-0.23 (-0.51 to 0.05)	0.94 (0.36 to 0.52)
Updated model (baseline + 4-weeks predictors) applied to participants with moderate/severe injury (grades 2 and 3)	0.63 (0.54 to 0.69)	0.00 (-0.29 to 0.29)	0.65 (0.32 to 0.98

95% CI: 95% confidence interval

Note: Performance measures for the different models are a combination of the individual estimates obtained from the analyses of the 20 imputed datasets. Estimates were combined by using Rubin's rules.

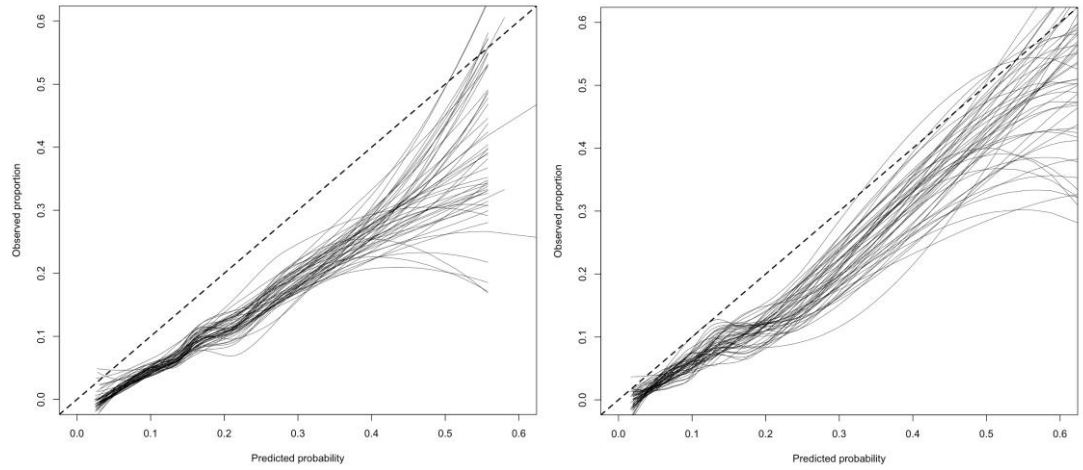


Figure 19: Calibration plot for the baseline (left) and updated (baseline + 4-weeks predictors; right) prognostic models to predict outcomes 1 overlaying the 50 calibration lines derived from the individual imputed SPRAINED datasets.

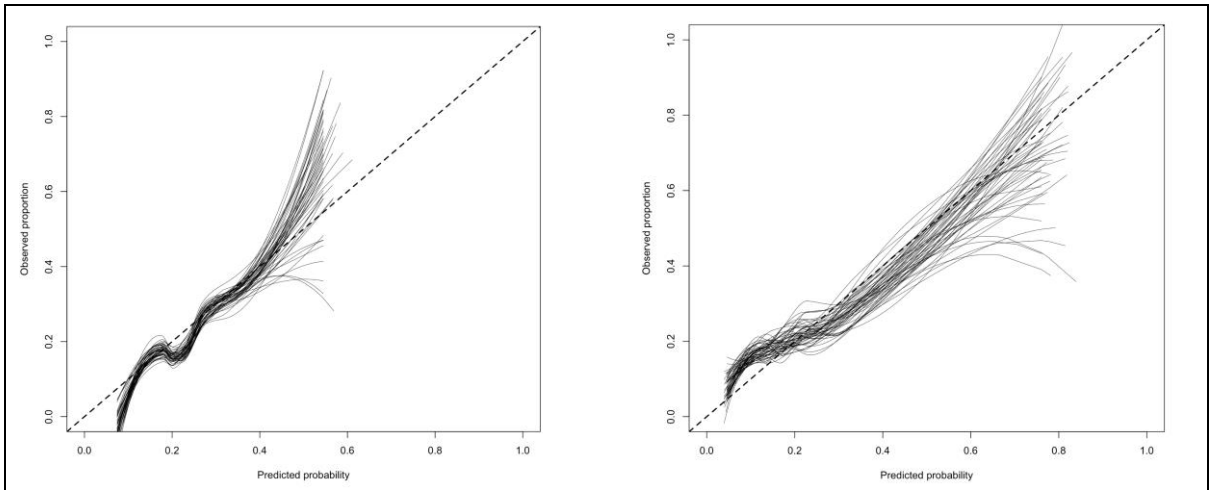


Figure 20: Calibration plot for the baseline (left) and updated (baseline + 4-weeks predictors; right) prognostic models to predict outcomes 2 overlaying the 50 calibration lines derived from the individual imputed SPRAINED datasets.

Table 27 shows how many of 1000 people would be identified as being at high risk (based on thresholds of 5, 10, 15, and 20%) using the developed prognostic models, and how many of these would actually present poor outcome 9 months after an acute ankle spain. There seems to be little difference between the baseline and updated models for outcome 1, with both identifying similar numbers of patients who experience a poor outcome after ankle sprain. However, less patients are deemed as high risk by using the updated model for outcome 1

(less false positives) across all tresholds of predicted probability as per estimated by the prognostic models. For outcome 2, the updated model misses more patients that actually develop the outcome (false negatives) when compared to the baseline model. Using any of the models seem to be beneficial, when compared to not using any model (or considering all patients as high risk of developing poor outcome).

Table 27: Models performance (numbers at risk and outcomes identified) at varying risk thresholds for 1000 patients.

	OUTCOME 1				OUTCOME 2			
	Number of patients		Number of outcomes		Number of patients		Number of outcomes	
	High risk	Low risk	Identified	Not identified	High risk	Low risk	Identified	Not identified
Consider all high risk	1000	0	85	0	1000	0	198	0
Predicted probability as per baseline model								
≥5%	959	41	85	0	1000	0	198	0
≥10%	690	310	74	11	975	25	194	4
≥15%	421	579	59	26	784	216	173	25
≥20%	280	720	41	44	488	512	110	88
Predicted probability as per updated model								
≥5%	845	155	85	0	989	11	198	0
≥10%	476	524	70	15	640	360	145	53
≥15%	328	672	55	30	424	576	102	96
≥20%	229	771	41	44	307	693	81	117

Note: Estimates based on complete cases analysis (n = 271 and 283 for outcomes 1 and 2, respectively).

6.3.3 Subgroup analyses

As the prognostic models were developed using a dataset from a clinical trial that only included participants with moderate or severe injuries (grades 2 or 3), it was decided that separate results on the models performance would also be presented for a subgroup of participants classified according to their injury severity degree (grades 2 and 3).

Overall, both the calibration (intercepts and slopes) and discrimination (c-statistics) did not show any substantial improvement in the subgroup analysis for the baseline prognostic models to predict either outcome 1 and 2 (Table 26). For the updated models (baseline + 4-week predictors), the intercept of the prognostic model to predict outcome 2 presented some improvement in terms of the calibration intercept, but not for the calibration slope (Table 26).

6.3.4 Model recalibration

Before recalibrating the models, we considered investigating the associations and predictive ability of two additional candidate predictors not included in the development phase (no data was available in the CAST dataset), but to which information was collected at baseline in the SPRAINED cohort, with the studied outcomes: sprain severity and recovery expectancy (time to recover from injury, as reported by the patients). Neither of the two variables showed statistically significant crude associations with the outcomes and presented very low predictive ability. C-statistics for sprain severity were 0.48 (95% CI 0.40 to 0.57) and 0.50 (95% CI 0.44 to 0.56), for outcomes 1 and 2 respectively. For recovery expectancy, c-statistics were 0.56 (95% CI 0.48 to 0.64) and 0.50 (95% CI 0.44 to 0.55), for outcomes 1 and 2 respectively.

Results from the model update are presented in Table 28 and Table 29. Predictor transformations were very similar to those observed for the original prognostic models developed with CAST data, apart from the fact that measures of pain were collected in a scale ranging from 0-10, and therefore an index was added indicating that values derived from assessments conducted with the VAS (which ranges from 0 to 100) should be divided by 10 before any transformation is performed when applying the model to estimate individual risks (Table 28). Coefficients obtained from the logistic regression models employed to update the models are presented in Table 29. Shrunk coefficients after applying the heuristic shrinkage factor to reduce optimism in the re-estimated model are also presented (Table 29).

Table 28: Transformations for non-categorical predictors in the recalibrated models for outcomes 1 and 2.

Variable	Outcome 1	Outcome 2
Age (years)	Age – 33.62	-
BMI (kg/m^2)	BMI – 27.05	-
Pain when resting (score, 0-100)	(Pain when resting/10) – 3.86	(Pain when resting/10) – - 3.86
Pain when bearing weight (score, 0-100)	(Pain when bearing weight/10) – 7.11	(Pain when bearing weight/10) – 7.11

Table 29: Intercept and regression coefficients of the recalibrated prediction models for poor recovery 9 months after ankle sprain (outcomes 1 and 2), before and after correction for optimism (shrinkage).

Predictor	Outcome 1		Outcome 2	
	Coefficient	Shrunk coefficient	Coefficient	Shrunk coefficient
Baseline model				
Age (years)	0.02	0.02	-	-
BMI (kg/m^2)	0.03	0.03	-	-
Pain when resting	0.19	0.17	0.07	0.06
Pain when bearing weight	0.18	0.16	0.10	0.09
>2 days from injury to assessment	-0.88	-0.78	-0.62	-0.56
Able to bear weight on the injured ankle	-0.22	-0.19	-0.05	-0.04
Recurrent sprain	1.60	1.42	2.07	1.88
Intercept	-2.60	-2.58	-1.61	-1.60
Updated model (baseline + 4-weeks predictors)				
Age (years)	0.02	0.01	-	-
BMI (kg/m^2)	0.03	0.03	-	-
Pain when resting	0.17	0.15	0.06	0.05
Pain when bearing weight	0.14	0.12	0.08	0.07
Pain when bearing weight at 4 weeks after injury	0.03	0.03	0.01	0.01
>2 days from injury to assessment	-1.23	-1.11	-0.71	-0.63
Able to bear weight on the injured ankle	-0.10	-0.09	0.07	0.06
Recurrent sprain	1.43	1.29	2.01	1.79
Intercept	-2.85	-2.82	-1.63	-1.62

The results of the prognostic development (Chapter 5) and validation are summarised and discussed together in Chapter 7.

7. Overall Discussion

The SPRAINED study research programme aimed to develop and externally validate prognostic models to aid clinical decision-making about risk of poor outcome for people attending emergency departments with an acute ankle sprain. The models were developed based on existing prognostic factor research (Chapter 3), expert consensus (Chapter 4), and using a large cohort of multicentre randomised controlled trial participants (Chapter 5). The external validation of the model was assessed in a subsequent prospective observational cohort study (Chapter 6). In this chapter, we consider the overall performance of the models, the limitations of the study, the implications for clinical practice and make recommendations for future research.

7.1 Performance of the SPRAINED prognostic models

Summary

The first prognostic model was developed to predict a composite outcome representing the presence of at least one of the following symptoms at 9 months after injury: persistent pain, functional difficulty or lack of confidence (outcome 1).

The second model was developed to predict a composite outcome representing the presence of at least one of the following symptoms or clinical events at 9 months after injury: persistent pain, functional difficulty, lack of confidence or recurrence of injury (outcome 2).

The models for outcome 1 and outcome 2 provided reasonable predictions of poor outcome for people with acute ankle sprain on the population used in their derivation (Chapter 5).

There was a slight decrease in model discrimination for both models when evaluated in prospectively collected external validation cohort (Chapter 6). The model for outcome 1 had better discrimination compared to the model for outcome 2. The variables for poor outcome used in Model 1 (persistent pain, functional difficulty or lack of confidence) were therefore easier and more reliable to predict, and appear to have good clinical utility. Hence this would be the model of choice.

The model predicting presence of either persistent pain, functional difficulty, lack of confidence or recurrence of injury (outcome 2) showed good calibration, whilst there was miscalibration of the model predicting persistent pain, functional difficulty or lack of confidence (outcome 1).

Updating these models, which used baseline data collected at the emergency department, with an additional variable at 4 weeks after the injury (pain when bearing weight on the ankle) improved the discriminatory ability and apparent calibration. However, improvements in model performance were modest. Blancing the practical challenges and resource implications of obtaining additional data at 4 weeks after presentation at the emergency department with the improvmements in prediction is likely to be an important consideration when selecting a model for use in clinical practice.

Despite some miscalibration of the models, the external validation study (Chapter 6) found that the model performance was reasonable, and showed benefit when compared to not using any model, for identifying patients at increased risk of poor outcome after acute ankle sprain. To the best of out knowledge there are no other prognostic models that have been developed and externally validated using robust methods for this patient group (see Chapter 3). The SPRAINED prognostic models may assist clinical-decision making when assessing and advising people with ankle sprains in the emergency department setting and on deciding on-going management. The models benefit from using predictors that are simple to obtain during routine clinical assessment. Recalibration of the models may be required to improve the accuracy of the predicted risks in other populations (within and outside the UK).

Differences in prognostic model performance in the development and external validation studies.

The differences in model performance between the development and external validation studies could be explained by several reasons. First, any prognostic model is expected to perform better in the dataset used in its development. Second, the very nature of the two studies can explain in part the poor calibration of the model, as the development dataset derived from a randomised clinical trial, while the external validation dataset was from a prospective observational cohort study with less restrictive eligibility criteria. The aim of the

observational cohort study was to be representative of the general population seeking medical assistance for acute ankle sprains at EDs in the UK NHS. Third, the case-mix in the two datasets might also explain the differences in model performance, as some of the most important predictors (e.g. days from injury to assessment and ability to bear weight on the injured ankle) were not equally distributed among participants in the two datasets. Finally, the differences in the outcomes rates (particularly for Outcome 1) might have also influenced the poor calibration of the models observed for the SPRAINED cohort. We recommend that the recalibrated prognostic models should be evaluated in different sets of patients.

We used an exhaustive set of predictors, and included clinical consensus to gain insight into what are factors easy to implement and acceptable. We could not include physical tests as we had insufficient data, although these have not appeared as useful tests in previous evaluations. It might be that in the future new data such as MRI or simple gait analysis will be able to add extra prognostic information. We excluded education from our considerations, but given its low priority and relatively low contribution to only one model, it is unlikely to provide much additional prognostic information.

The consensus group (Chapter 4) suggested that psychological variables may improve the prediction, and although an additional variable was collected on the particiapnts expectation about recovery, there was limited evidence that this additional variable had prognostic utility.

7.2 Strengths and limitations

To the best of our knowledge this is the first study to 1) develop a prognostic model to predict poor outcome in people with acute ankle sprains using an adequately large cohort to explore a wide range of clinically plausible candidate predictors, 2) use robust statistical methods to assess the performance of the prognostic models, and 3) conduct an a large prospective cohort study to enable external validation. We needed to conduct the observational cohort as there were no other available and sufficiently large datasets with data on a wide range of candidate predictors available for an external validation. Generalisability of the findings are enhanced by the multi-centre data from the CAST and SPRAINED cohorts that represented a range of district general and major trauma centres.

We followed the most recent guidelines available on the reporting of prognostic model development and used methods that, to the best of our knowledge, are the most widely recommended. For example, continuous variables, whenever possible, were kept as continuous to avoid loss of information. Nonlinear relationships were investigated using the best variables transformations found by multivariable fractional polynomials. The study included an internal correction for model optimism (shrinkage of regression coefficients and intercepts) as well as an external validation phase. The amount of missing data in the external validation dataset, which is almost inevitable in studies of this nature, was considerably smaller than that observed in the development dataset, and missing data imputation was also used to produce a set of 50 complete datasets, which enabled more robust analyses.

The SPRAINED project has limitations that must be considered when interpreting the results described in this report. First, the data used to develop the two proposed prognostic models were from a prior randomised controlled trial (CAST), so were not originally intended to fulfil this aim. However, the CAST cohort did represent the best dataset available, with data on the symptoms and clinical events of interest to compose the two outcomes for the SPRAINED prognostic model, and for the majority of the candidate prognostic variables considered to have predictive ability at the time of the study's conception. The CAST trial was a pragmatic randomised controlled trial, with relatively open eligibility criteria, which aimed to investigate the effect of four different interventions on a different set of (primary and secondary) outcomes. The CAST dataset was not optimally sized for developing prognostic models, and had it been larger, then it might have provided more robust estimates, resulting in models with less optimism. As previously highlighted, the low EPV observed for the two models developed might have contributed to the optimism found for both prognostic models and, therefore, to the poor calibration on the external validation dataset. Finally, another important limitation relates to the amount of missing data observed in the development dataset. Because of the extent of missing data, some of the candidate predictors had to be dropped before the process of data imputation due to the amount of missing observations (> 60%) being considered too high. Therefore, some important predictors could have conceivably been missed in the development phase of the SPRAINED study.

A key focus of the SPRAINED study was that the prognostic factor variables needed to be based on routinely collected clinical information. It is possible that information from imaging

techniques such as MRI could have resulted in a more accurate estimation of risk (see Chapter 3, Systematic Review). However, this type of investigation is not routinely used or available in the context of an emergency department consultation. We therefore restrained our investigation to prognostic factors that are or could easily be obtained during a routine assessment of a person with an acute ankle sprain in the emergency department.

The rates of poor outcome in the SPRAINED cohort study were lower than in the CAST study, 7 vs. 20% for outcome 1 and 24 vs. 16% for outcome 2, and the rates reported in previous systematic reviews of approximately 30%.[3, 4] These variations in poor outcome rates highlight the potential issue of different sampling frames. It could be argued that the observational cohort we recruited for SPRAINED was a reasonable representation of the rates of poor outcome in patients presenting to EDs in the UK as all types of adult patients with an ankle sprain were included, there was low participant burden from participation compared to most clinical trials, and we achieved good levels of follow-up.

7.3 Other prognostic models reported during the SPRAINED study

Our systematic review of the literature highlighted limitations in the evidence relating to predictive factors for recovery from ankle sprain. Since this review, Doherty and colleagues reported on movement tests performed at 2 weeks after injury as predictors of chronic ankle instability after acute ankle sprain. They found that inability to complete 2 out of 5 dynamic movement tests had a sensitivity of 83% and specificity of 55% for identifying those classified as having chronic ankle instability.[75] The need for clinical movement assessments at a later stage in recovery is not feasible in the context of most emergency department contacts in the UK, however this may indicate that other predictive factors may be appropriate for later stages in recovery.

7.4 Clinical implications of the SPRAINED study

Estimating the risk of a poor outcome for a person attending an emergency department with an ankle sprain is desirable due to the high number of individuals presenting with these injuries, and the difficulty in determining which people will struggle to recover. Many people present in the acute phase with a degree of ankle pain, swelling, loss of motion and difficulty

weight bearing on the injured leg. Clinical examination is often challenging as tolerance of physical examination tests are limited by pain, and have been found to have poorer sensitivity and specificity within the first 48 hours after injury compared to 5 days after injury.[76] As a result, it is difficult to decide which people may benefit from monitoring or rehabilitation. The value of a prognostic model is evident, but in order for it to be utilised in clinical practice, it needs to be quick and simple to use, and offer sufficiently accurate estimation of risk of poor outcome to be clinically worthwhile.

The prognostic models have the potential to assist clinicians decide whether an early review is merited and to offer some reassurance that people who are not followed up are likely to be on a positive recovery trajectory. As with other prognostic models, any potential benefits from being able to estimate an outcome should be considered in the context of the performance of the models and the potential risks of an inaccurate prediction for the person being assessed. Given some limitations in predictive performance of the SPRAINED prognostic models at the development (Chapter 5) and external validation (Chapter 6) stages, we suggest that their value would be in assisting the clinician in estimating the probability of a poor outcome, rather than being a decision making tool in isolation. If implemented in clinical practice, it should be noted that there is a degree of uncertainty in the calculated risk of poor outcome when using the SPRAINED models. This uncertainty in estimation could lead to over or under referral of patients to review clinics or treatment such as physiotherapy, and highlights the caution required when using the calculated individual risks when counselling patients about their prognosis. Further research is recommended to evaluate the impact of using the SPRAINED prognostic models on clinical practice and patient outcomes, and to assess the acceptability and uptake of use by emergency department clinicians.

Of note, 78/682 (11%) patients reported a reoccurrence of sprain within 9 months after their initial presentation in the external validation study. It could be argued that widening the classification to reoccurrence of sprain is more consistent with existing definitions of chronic ankle instability.[77] While we did not set out to predict chronic ankle instability specifically, we recognise that many with a poor outcome, as defined by the SPRAINED study, would likely include patients with this condition.

One of the important aspects of assessing the clinical usefulness of a multivariable prognostic model is that it is a better predictor of poor outcome than the overall clinical impression of clinical severity of the presenting ankle sprain. Future work could examinine how well the model performs in comparison to the clinician impression.

7.5 Implementation of the SPRAINED prognostic models

Other prediction models are in routine clinical use in the emergency department. One prediction model being used routinely is for ankle injuries, the Ottawa ankle rules, [78] which are used to help determine which patients should be considered for radiographs to rule out a fracture.[79] Patients entered into the SPRAINED study would have been assessed to rule out a fracture during their emergency department assessment. We envisage that implementation of the SPRAINED prognostic model could also be used in the assessment of this patient group, once the clinician is satisfied there is no fracture.

An application of the SPRAINED prognostic models we recommend for further investigation is whether they can be used to stratify patients to post-injury interventions that are matched to the level of risk of poor outcome. There have been inconsistencies in the findings of trials investigating the effectiveness of physiotherapy rehabilitation after acute ankle sprain.[80, 81] We hypothesise that as most patients attending the emergency department have a good prognosis, better targeting of higher intensity interventions to those at greater risk of poor outcome may enhance the clinical and cost-effectiveness of rehabilitation, however this required formal evaluation.

The prognostic model requires a calculation too complex for easy use in the clinical setting so it would require a computer application to facilitate the calculation of probability for poor outcome for the person being examined in the emergency department. A web-based calculator or application could be developed specifically for the SPRAINED prognostic models model and this will be an area of work taken forward by the SPRAINED investigators. Due to limitations in the performance of the models, an issue to address when presenting the calculated risks to clinicians will be to concurrently make users aware of the prediction accuracy.

7.6 Recommendations for future research

Further research is recommended to:

- determine appropriate cut points or score ranges from the prognostic model for identifying patients more likely to benefit from different clinical pathways
- to assess whether the prognostic model can make an impact in improving decision making and targeting of treatment, and ultimately patient outcomes
- evaluate the acceptability and uptake of use by emergency department clinicians
- examine how well the model performs in comparison to clinician impression on prognosis and assessnment of clinical severity of the presenting ankle sprain
- investigate whether a wider range of psychological, or other types of variables that were not included in the SPRAINED study, improve prediction

We also note that recalibration of the models may be required to improve the accuracy of the predicted risks in other populations (within and outside the UK).

7.7 Conclusions

The SPRAINED study research programme aimed to develop and externally validate prognostic models to aid clinical decision-making about risk of poor outcome for people attending emergency departments with an acute ankle sprain. The models were developed based on existing prognostic factor research, expert consensus, and using a large cohort of multicentre randomised controlled trial participants. The external validation of the model was assessed in a subsequent prospective observational cohort study.

The SPRAINED study prognostic models performed reasonably and show benefit in identidying patients at high risk of poor outcome after acute ankle sprain when compared to not using any model (consider all patients high risk of poor outcome), so may assist clinical-decision making when assessing and advising people with ankle sprains in the emergency department setting and when deciding on on-going management. The models benefit from using predictors that are simple to obtain during routine clinical assessment.

Further research to evaluate the performance of the models in other settings is recommended. Further refinement of the models, including external validation of the re-calibrated models or identifying additional predictors may be required. The impact of implementing and using either model in clinical practice, in terms of acceptability and uptake by emergency department staff and their impact on patient outcomes should also be investigated.

Acknowledgements

Contribution of authors

David J Keene (NDORMS Research Fellow in Trauma Rehabilitation/NIHR Postdoctoral Research Fellow) – study lead, led the development and authorship of the report, responsible for overall management of the project.

Michael M. Schüssel (Medical Statistician) – developed and carried out data analysis, co-authorship of report, and provided statistical input throughout the study.

Jacqueline Thompson (Research Physiotherapist) – led the systematic review, provided training and clinical support to collaborating sites, facilitated follow-up, co-production of consensus meeting chapter.

Daryl A. Hagan (Study Co-ordinator and Administrator) – day to day co-ordination of project, data collection, queries and data cleaning, collation and editing of report chapters.

Mark A. Williams (Senior Lecturer in Physiotherapy and Rehabilitation, Oxford Brookes University) – previous study lead, led the consensus meeting process and co-production of consensus meeting chapter.

Christopher Byrne (Lecturer in Physiotherapy, University of Plymouth) – produced the systematic review, provided training and clinical support to collaborating sites.

Steve Goodacre (Professor of Emergency Medicine, University of Sheffield) – provided academic expertise and advice at key points, recruiting site principal investigator, reviewed report.

Matthew Cooke (Professor of Emergency Medicine, University of Warwick) – provided academic expertise and advice at key points, reviewed report.

Stephen Gwilym (Consultant Surgeon & Honorary Senior Lecturer, Oxford University Hospital NHS Foundation Trust) – provided academic expertise and advice at key points,

reviewed report.

Phillip Hormbrey (Consultant in Emergency Medicine, Oxford University Hospital NHS Foundation Trust) – provided academic expertise and advice at key points, recruiting site principal investigator, review of report.

Jennifer Bostock (PPI Representative) – provided consultation and key input of patient and public perspective throughout study, reviewed report.

Kirstie Haywood (Senior Research Fellow in Patient Reported Outcomes, University of Warwick) – provided consultation and senior facilitation of the consensus meeting process.

David Wilson (Honorary Consultant Radiologist, Oxford University Hospital NHS Foundation Trust) – provided clinical expertise and advice at key points, reviewed report.

Gary S Collins (Professor of Medical Statistics) – responsible for study design, supervised data analysis, provided academic expertise and advice throughout project, co-authorship of report.

Sarah E Lamb (Professor of Trauma Rehabilitation/Director of Oxford Clinical Trials Research Unit) – chief investigator, overall responsibility of study, design, academic leadership, authorship of report.

SPRAINED study team

Chief Investigator: Sarah E Lamb

Study Lead: David J Keene

Co-investigators: Gary S Collins, Mark Williams, Steve Goodacre, Matthew Cooke, Steve Gwilym, Phil Hormbrey, David Wilson, Jennifer Bostock

Study co-ordinator and administrator: Daryl Hagan

Senior study manager: Damian Haywood

Research Physiotherapists: Jacqueline Thompson, Christopher Byrne

Study statisticians: Michael Maia Schlüssel, Gary S Collins

Table 30: Principal Investigators / research fellows, nurses or therapists by hospital site in order of date starting recruitment

Principal Investigator	Research nurses, therapists and associates	Hospital name	NHS Trust name
Dr Philip Hormbrey	Sally Beer, Amanda Budden, Alexis Espinosa, Dominique Georgiou, Louise Findlay	John Radcliffe Hospital	Oxford University Hospital NHS Foundation Trust
Dr Susan Dorrian	Samantha Stafford, Nathan Humphries, David Hunt	Heartlands Hospital & Solihull Hospital	Heart of England NHS Foundation Trust
Prof Steve Goodacre	Rachel Walker, Anna Wilson, Nicola Hindmarch, Craig Jones, Zoe Dutton, John Parry, Charlotte Green	Northern General Hospital	Sheffield Teaching Hospitals NHS Foundation Trust
Dr Victoria Stacey	Claire Hunt, Natalie Bynorth, Pauline Brown, Kayleigh Collins, Estelle Nambela	Cheltenham General Hospital & Gloucester Royal Hospital	Gloucestershire Hospital NHS Foundation Trust
Prof Tim Coats	Lisa McClelland, Elisabeth Cadman-Moore	Leicester Royal Infirmary	University Hospitals of Leicester NHS Trust
Dr Sarah Wilson	Louise Chandler, Louise Foster, Vikki Diduca, Joana Da Rocha	Wexham Park Hospital	Frimley Health NHS Foundation Trust
Dr Jason Kendall	Lee Cameron, Rachel Ozanne, Sue Kempson, Ruth Worner, Beverley Faulkner, Caroline Ellis	Southmead Hospital	North Bristol NHS Trust
Dr David Clarke	Nicola Jacques, Dariusz Pabianczyk, Ria Diel, Andrzej Adamowicz, Abby Brown, Claire Burnett, Daniel Sedgewick, Claire Sayner, Jane Macpherson, Elizabeth Oastler, Mitzi Baylis, Caroline Lewis, Helen Ingolfsrid, Rikki Davies, Carys Davies, Teresa Hobbs	Royal Berkshire Hospital	Royal Berkshire NHS Foundation Trust

Ms Antoanela Colda	Gill Ritchie, Seema Chavda	Milton Keynes University Hospital	Milton Keynes University Hospital NHS Foundation Trust
Dr Deborah Mayne	Jackie Berry, Sarah Patch, Julie Camsooksai, Lee Tbaily	Poole Hospital	Poole Hospital NHS Foundation Trust

Study Steering Committee

Prof Richard Riley (Chair), Prof Kevin Mackway-Jones, Prof Suzanne McDonough

Other acknowledgements

Special thanks to the Centre for Health, Law and Emerging Technologies (HeLEX) for their collaboration on the Dynamic Consent pilot study, in particular Prof Jane Kaye, Harriet Teare and Jeremy Holland.

We also recognise the contributions of the following individuals at the Oxford Clinical Trials Research Unit and Centre for Rehabilitation Research for their support in delivering the SPRAINED study: Vicki Barber, Lesley Morgan, Emma Roberts, Scott Parsons, Sue Davolls, Katie Chegwin, Emma Haines, Oliver Conway, Hannah Ashby, Asima Qayyum, Tim Cranston, Patrick Julier, Lucy Eldridge, Simon Shayler, Joanna Black, Deborah Brown and others who have provided advice and support throughout the course of the study.

SPRAINED consensus meeting participants

Chrissy Aimes (Paramedic, South Central Ambulance Service), Emma Batchelor (Physiotherapist, University Hospital Birmingham NHS Foundation Trust), Emma Bolton (ED Extended Nurse Practitioner, Gloucestershire Hospitals NHS Foundation Trust), Mrs Jennifer Bostock (PPI lead/co-applicant), Ms Lucy Cameron Paramedic (Paramedic, South Central Ambulance Service), Dr David Clarke (ED Consultant, Royal Berkshire NHS Trust), Prof Matthew Cooke (ED Consultant, Heart of England Foundation Trust), Mr Jason Franks (Paramedic, South Central Ambulance Service), Prof Steve Goodacre (ED consultant, Sheffield Teaching Hospitals NHS Foundation Trust), Dr Phil Hormbrey (ED consultant, Oxford University Hospital NHS Trust), Mr Nathan Humphries (ESP Physiotherapist, Heart of England Foundation Trust), Mr David Hunt (ESP Physiotherapist, Heart of England

Foundation Trust), Mrs Claire Hunt (ESP Physiotherapist, Gloucestershire Hospitals NHS Foundation Trust), Dr Liza Keating (ED Consultant, Royal Berkshire NHS Trust), Dr David Keene (Clinical Researcher, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford), Mrs Sue Kempson (ESP Physiotherapist, North Bristol NHS Trust), Prof Sallie Lamb (Clinical Researcher, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford), Dr David Wilson (Consultant Radiologist, St Luke's Radiology).

Funding

The SPRAINED study was funded by the National Institute of Health Research (NIHR) Health Technology Assessment programme (project number 13/19/06). Supported by the NIHR Biomedical Research Centre, Oxford, and the NIHR Fellowship programme (Dr David Keene, PDF-2016-09-056). Sarah Lamb receives funding from the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care Oxford at Oxford Health NHS Foundation Trust. The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the Health Technology Assessment programme, NIHR, NHS or the Department of Health.

Publications

Thompson, J.Y., Byrne, C., Williams, M.A., Keene, D.J., Schlussel, M.M, Lamb, S.E. (2017) Prognostic Factors for Outcome Following Acute Lateral Ankle Ligament Sprain. A Systematic Review. BMC Musculoskeletal Disorders, 18:421.

Data sharing

All data requests should be submitted to the corresponding author for consideration. Access to anonymised data may be granted following review. Exclusive use will be retained until the publication of major outputs.

Word Count

27,360

References

1. Cooke MW, Lamb SE, Marsh J, Dale J. A survey of current consultant practice of treatment of severe ankle sprains in emergency departments in the United Kingdom. *Emergency medicine journal : EMJ* 2003;**20**:505-7.

2. Martin RL, Davenport TE, Paulseth S, Wukich DK, Godges JJ. Ankle stability and movement coordination impairments: ankle ligament sprains. *The Journal of orthopaedic and sports physical therapy* 2013;**43**:A1-40. https://doi.org/10.2519/jospt.2013.0305

3. van Rijn RM, van Os AG, Bernsen RM, Luijsterburg PA, Koes BW, Bierma-Zeinstra SM. What is the clinical course of acute ankle sprains? A systematic literature review. *The American journal of medicine* 2008;**121**:324-31 e6. https://doi.org/10.1016/j.amjmed.2007.11.018

4. Polzer H, Kanz KG, Prall WC, Haasters F, Ockert B, Mutschler W, *et al.* Diagnosis and treatment of acute ankle injuries: development of an evidence-based algorithm. *Orthopedic reviews* 2012;**4**:e5. https://doi.org/10.4081/or.2012.e5

5. Cooke MW, Marsh JL, Clark M, Nakash R, Jarvis RM, Hutton JL, *et al.* Treatment of severe ankle sprain: a pragmatic randomised controlled trial comparing the clinical effectiveness and cost-effectiveness of three types of mechanical ankle support with tubular bandage. The CAST trial. *Health technology assessment* 2009;**13**:iii, ix-x, 1-121. https://doi.org/10.3310/hta13130

6. Verhagen RA, de Keizer G, van Dijk CN. Long-term follow-up of inversion trauma of the ankle. *Archives of orthopaedic and trauma surgery* 1995;**114**:92-6.

7. Wikstrom EA, Hubbard-Turner T, McKeon PO. Understanding and treating lateral ankle sprains and their consequences: a constraints-based approach. *Sports medicine* 2013;**43**:385-93. https://doi.org/10.1007/s40279-013-0043-z

8. Stiell I, Wells G, Laupacis A, Brison R, Verbeek R, Vandemheen K, *et al.* Multicentre trial to introduce the Ottawa ankle rules for use of radiography in acute ankle injuries. Multicentre Ankle Rule Study Group. *Bmj* 1995;**311**:594-7.

9. van Dijk CN, Mol BW, Lim LS, Marti RK, Bossuyt PM. Diagnosis of ligament rupture of the ankle joint. Physical examination, arthrography, stress radiography and sonography compared in 160 patients after inversion trauma. *Acta orthopaedica Scandinavica* 1996;**67**:566-70.

10. Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A, *et al.* Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes. *Bmj* 2013;**346**:e5595. https://doi.org/10.1136/bmj.e5595

11. Hingorani AD, Windt DA, Riley RD, Abrams K, Moons KG, Steyerberg EW, *et al.* Prognosis research strategy (PROGRESS) 4: stratified medicine research. *Bmj* 2013;**346**:e5793. https://doi.org/10.1136/bmj.e5793

12. Riley RD, Hayden JA, Steyerberg EW, Moons KG, Abrams K, Kyzas PA, *et al.* Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *PLoS medicine* 2013;**10**:e1001380. https://doi.org/10.1371/journal.pmed.1001380

13. Hiller CE, Ightingale EJN, Lin CWC, Coughlan GF, Caulfield B, Delahunt E. Characteristics of people with recurrent ankle sprains: a systematic review with meta-analysis. *Brit J Sport Med* 2011;**45**:660-72. https://doi.org/10.1136/bjsm.2010.077404

14. Linde F, Hvass I, Jurgensen U, Madsen F. Early mobilizing treatment in lateral ankle sprains. Course and risk factors for chronic painful or function-limiting ankle. *Scandinavian journal of rehabilitation medicine* 1986;**18**:17-21.

15. Kerkhoffs GM, Rowe BH, Assendelft WJ, Kelly K, Struijs PA, van Dijk CN. Immobilisation and functional treatment for acute lateral ankle ligament injuries in adults. *The Cochrane database of systematic reviews* 2002; 10.1002/14651858.CD003762:CD003762. https://doi.org/10.1002/14651858.CD003762

16. Kerkhoffs GM, Handoll HH, de Bie R, Rowe BH, Struijs PA. Surgical versus conservative treatment for acute injuries of the lateral ligament complex of the ankle in adults. *The Cochrane database of systematic reviews* 2007; 10.1002/14651858.CD000380.pub2:CD000380. https://doi.org/10.1002/14651858.CD000380.pub2

17. Lamb SE, Marsh JL, Hutton JL, Nakash R, Cooke MW, Collaborative Ankle Support Trial Team. Mechanical supports for acute, severe ankle sprain: a pragmatic, multicentre, randomised controlled trial. *Lancet* 2009;**373**:575-81. https://doi.org/10.1016/S0140-6736(09)60206-3

18. Roos EM, Brandsson S, Karlsson J. Validation of the foot and ankle outcome score for ankle ligament reconstruction. *Foot & ankle international* 2001;**22**:788-94.

19. Brooks R. EuroQol: the current state of play. *Health policy* 1996;**37**:53-72.

20. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, *et al.* Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. *Ann Intern Med* 2015;**162**:W1-W73.

21. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. *Systematic reviews* 2016;**5**:210. https://doi.org/10.1186/s13643-016-0384-4

22. Hayden JA, van der Windt DA, Cartwright JL, Cote P, Bombardier C. Assessing bias in studies of prognostic factors. *Ann Intern Med* 2013;**158**:280-6. https://doi.org/10.7326/0003-4819-158-4-201302190-00009

23. de Bie RA, de Vet HC, van den Wildenberg FA, Lenssen T, Knipschild PG. The prognosis of ankle sprains. *International journal of sports medicine* 1997;**18**:285-9. https://doi.org/10.1055/s-2007-972635

24. Wilson RW, Gansneder BM. Measures of functional limitation as predictors of disablement in athletes with acute ankle sprains. *The Journal of orthopaedic and sports physical therapy* 2000;**30**:528-35. https://doi.org/10.2519/jospt.2000.30.9.528

25. Cross KM, Worrell TW, Leslie JE, Van Veld KR. The relationship between self-reported and clinical measures and the number of days to return to sport following acute lateral ankle sprains. *The Journal of orthopaedic and sports physical therapy* 2002;**32**:16-23. https://doi.org/10.2519/jospt.2002.32.1.16

26. Akacha KH, J.L. Lamb, S.E. Modelling treatment, age- and gender-specific recovery in acute injury studies. In: The University of Warwick Centre for Research in Statistical Methodology; 2010: 1-12.

27. Langner I, Frank M, Kuehn JP, Hinz P, Ekkernkamp A, Hosten N, *et al.* Acute inversion injury of the ankle without radiological abnormalities: assessment with high-field MR imaging and correlation of findings with clinical outcome. *Skeletal radiology* 2011;**40**:423-30. https://doi.org/10.1007/s00256-010-1017-y

28. van Middelkoop M, van Rijn RM, Verhaar JA, Koes BW, Bierma-Zeinstra SM. Re-sprains during the first 3 months after initial ankle sprain are related to incomplete recovery: an observational study. *Journal of physiotherapy* 2012;**58**:181-8. https://doi.org/10.1016/s1836-9553(12)70109-1

29. van der Wees P, Hendriks E, van Beers H, van Rijn R, Dekker J, de Bie R. Validity and responsiveness of the ankle function score after acute ankle injury. *Scandinavian journal of medicine & science in sports* 2012;**22**:170-4. https://doi.org/10.1111/j.1600-0838.2010.01243.x

30. O'Connor SR, Bleakley CM, Tully MA, McDonough SM. Predicting functional recovery after acute ankle sprain. *PloS one* 2013;**8**:e72124. https://doi.org/10.1371/journal.pone.0072124

31. Medina McKeon JM, Bush HM, Reed A, Whittington A, Uhl TL, McKeon PO. Return-to-play probabilities following new versus recurrent ankle sprains in high school athletes. *Journal of science and medicine in sport* 2014;**17**:23-8. https://doi.org/10.1016/j.jsams.2013.04.006

32. van Ochten JM, Mos MCE, van Putte-Katier N, Oei EHG, Bindels PJE, Bierma-Zeinstra SMA, *et al.* Structural abnormalities and persistent complaints after an ankle sprain are not associated: an observational case control study in primary care. *Br J Gen Pract* 2014;**64**:e545-e53. https://doi.org/10.3399/bjgp14X681349

33. Royston P, Moons KGM, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *Bmj* 2009;**338**. https://doi.org/10.1136/bmj.b604

34. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD statement. *Journal of Clinical Epidemiology* 2015;**68**:112-21. https://doi.org/http://dx.doi.org/10.1016/j.jclinepi.2014.11.010

35. Black N, Murphy M, Lamping D, McKee M, Sanderson C, Askham J, *et al.* Consensus Development Methods: A Review of Best Practice in Creating Clinical Guidelines. *J Health Serv Res Policy* 1999;**4**:236-48. https://doi.org/10.1177/135581969900400410

36. Delbecq AL, Van de Ven AH, Gustafson DH. *Group techniques for program planning: a guide to nominal group and Delphi processes*: Scott, Foresman; 1975.

37. Boers M, Kirwan JR, Wells G, Beaton D, Gossec L, d'Agostino M-A, *et al.* Developing Core Outcome Measurement Sets for Clinical Trials: OMERACT Filter 2.0. *J Clin Epidemiol* 2014;**67**:745-53. https://doi.org/https://doi.org/10.1016/j.jclinepi.2013.11.013

38. Harvey N, Holmes CA. Nominal group technique: An effective method for obtaining group consensus. *Int J Nurs Pract* 2012;**18**:188-94. https://doi.org/10.1111/j.1440-172X.2012.02017.x

39. Haywood KL, Griffin XL, Achten J, Costa ML. Developing a core outcome set for hip fracture trials. *Bone Joint J* 2014;**96-B**:1016-23. https://doi.org/10.1302/0301-620x.96b8.33766

40. Schünemann HJ BJ, Guyatt G, Oxman A, The GRADE Working Group. *GRADE Handbook*; 2013. URL: http://gdt.guidelinedevelopment.org/app/handbook/handbook.html (accessed 14 July 2017).

41. Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, *et al.* Developing core outcome sets for clinical trials: issues to consider. *Trials* 2012;**13**:132-. https://doi.org/10.1186/1745-6215-13-132

42. Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CF, Askham J, *et al.* Consensus development methods, and their use in clinical guideline development. *Health technology assessment* 1998;**2**:i-iv, 1-88.

43. Tugwell P, Boers M, Brooks P, Simon L, Strand V, Idzerda L. OMERACT: An international initiative to improve outcome measurement in rheumatology. *Trials* 2007;**8**:38-. https://doi.org/10.1186/1745-6215-8-38

44. Laisne F, Lecomte C, Corbiere M. Biopsychosocial predictors of prognosis in musculoskeletal disorders: a systematic review of the literature (corrected and republished) *. *Disability and rehabilitation* 2012;**34**:1912-41. https://doi.org/10.3109/09638288.2012.729362

45. van Rijn RM, Willemsen SP, Verhagen AP, Koes BW, Bierma-Zeinstra SM. Explanatory variables for adult patients' self-reported recovery after acute lateral ankle sprain. *Phys Ther* 2011;**91**:77-84.

46. Wikstrom EA, Hubbard-Turner T, McKeon PO. Understanding and treating lateral ankle sprains and their consequences: a constraints-based approach. *Sports Med* 2013;**43**:385-93.

47. Roos EM, Brandsson S, Karlsson J. Validation of the foot and ankle outcome score for ankle ligament reconstruction. *Foot & Ankle International* 2001;**22**:788-94.

48. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med* 2011;**30**:377-99.

49. White IR, Daniel R, Royston P. Avoiding bias due to perfect prediction in multiple imputation of incomplete categorical data. *Comp Stat Data Anal* 2010;**54**:2267-75.

50. Albert A, Anderson JA. On the existence of maximum likelihood estimates in logistic regression models. *Biometrika* 1984;**71**:1-10.

51. Harrell Jr FE, Lee KL, Califf RM, Pryor DB, Rosati RA. Regression modelling strategies for improved prognostic prediction. *Stat Med* 1984;**3**:143-52.

52. Harrell Jr FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;**15**:361-87.

53. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996;**49**:1373-9.

54. Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, *et al.* Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. *PLoS medicine* 2014;**11**:e1001744. https://doi.org/10.1371/journal.pmed.1001744

55. Pavlou M, Ambler G, Seaman SR, Guttmann O, Elliott P, King M, *et al.* How to develop a more accurate risk prediction model when there are few events. *Bmj* 2015;**351**:h3868. https://doi.org/10.1136/bmj.h3868

56. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. *Am J Epidemiol* 2007;**165**:710-8.

57. Collins GS, Ogundimu EO, Cook JA, Manach YL, Altman DG. Quantifying the impact of different approaches for handling continuous predictors on the performance of a prognostic model. *Stat Med* 2016;**35**:4124-35. https://doi.org/10.1002/sim.6986

58. Royston P, Altman DG. Regression Using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modelling. *J R Stat Soc Ser C Appl Stat* 1994;**43**:429-67. https://doi.org/10.2307/2986270

59. Royston P, Sauerbrei W. MFP: Multivariable Model-Building with Fractional Polynomials. In: *Multivariable Model-Building*: John Wiley & Sons, Ltd; 2008:115-50. https://doi.org/10.1002/9780470770771.ch6

60. Sauerbrei W, Royston P, Binder H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Stat Med* 2007;**26**:5512-28. https://doi.org/10.1002/sim.3148

61. Marshall A, Altman DG, Holder RL, Royston P. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. *BMC medical research methodology* 2009;**9**:57. https://doi.org/10.1186/1471-2288-9-57

62. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*: Wiley; 2004.

63. Atkinson AC. A note on the generalized information criterion for choice of a model. *Biometrika* 1980;**67**:413-8. https://doi.org/10.1093/biomet/67.2.413

64. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006;**26**:565-74. https://doi.org/10.1177/0272989X06295361

65. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *Bmj* 2016;**352**:i6. https://doi.org/10.1136/bmj.i6

66. Thangaratinam S, Allotey J, Marlin N, Mol BW, Von Dadelszen P, Ganzevoort W, *et al.* Development and validation of Prediction models for Risks of complications in Early-onset Pre-eclampsia (PREP): a prospective cohort study. *Health technology assessment* 2017;**21**:1-100. https://doi.org/10.3310/hta21180

67. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. *J Clin Epidemiol* 2016;**74**:167-76. https://doi.org/10.1016/j.jclinepi.2015.12.005

68. Austin PC, Steyerberg EW. Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. *Stat Med* 2014;**33**:517-35. https://doi.org/10.1002/sim.5941

69. Wood AM, Royston P, White IR. The estimation and use of predictions for the assessment of model performance using large samples with multiply imputed data. *Biom J* 2015; 10.1002/bimj.201400004. https://doi.org/10.1002/bimj.201400004

70. Janssen KJ, Moons KG, Kalkman CJ, Grobbee DE, Vergouwe Y. Updating methods improved the performance of a clinical prediction model in new patients. *J Clin Epidemiol* 2008;**61**:76-86. https://doi.org/10.1016/j.jclinepi.2007.04.018

71. Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stat Med* 2016;**35**:214-26. https://doi.org/10.1002/sim.6787

72. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York: J. Wiley & Sons; 1987.

73. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, *et al.* Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;**21**:128-38.

74. Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, *et al.* Risk prediction models: II. External validation, model updating, and impact assessment. *Heart (British Cardiac Society)* 2012;**98**:691-8. https://doi.org/10.1136/heartjnl-2011-301247

75. Doherty C, Bleakley C, Hertel J, Caulfield B, Ryan J, Delahunt E. Recovery From a First-Time Lateral Ankle Sprain and the Predictors of Chronic Ankle Instability: A

Prospective Cohort Analysis. *The American journal of sports medicine* 2016;**44**:995-1003. https://doi.org/10.1177/0363546516628870

76. van Dijk CN, Lim LS, Bossuyt PM, Marti RK. Physical examination is sufficient for the diagnosis of sprained ankles. *The Journal of bone and joint surgery British volume* 1996;**78**:958-62.

77. Guillo S, Bauer T, Lee JW, Takao M, Kong SW, Stone JW, *et al.* Consensus in chronic ankle instability: Aetiology, assessment, surgical indications and place for arthroscopy. *Orthop Traumatol Surg Res* 2013;**99**:S411-S9. https://doi.org/http://dx.doi.org/10.1016/j.otsr.2013.10.009

78. Stiell IG, Greenberg GH, McKnight RD, Nair RC, McDowell I, Reardon M, *et al.* Decision rules for the use of radiography in acute ankle injuries. Refinement and prospective validation. *Jama* 1993;**269**:1127-32.

79. Bachmann LM, Kolb E, Koller MT, Steurer J, Riet Gt. Accuracy of Ottawa ankle rules to exclude fractures of the ankle and mid-foot: systematic review. *Bmj* 2003;**326**:417. https://doi.org/10.1136/bmj.326.7386.417

80. Doherty C, Bleakley C, Delahunt E, Holden S. Treatment and prevention of acute and recurrent ankle sprain: an overview of systematic reviews with meta-analysis. *Brit J Sport Med* 2017;**51**:113-25. https://doi.org/10.1136/bjsports-2016-096178

81. Brison RJ, Day AG, Pelland L, Pickett W, Johnson AP, Aiken A, *et al.* Effect of early supervised physiotherapy on recovery from acute ankle sprain: randomised controlled trial. *Bmj* 2016;**355**. https://doi.org/10.1136/bmj.i5650

Appendices

Appendix A - Systematic review search strategy

AMED (Allied and Complementary Medicine) via OVID 1985 to July 2015

1. exp Ankle/
2. ankle.ti,ab.
3. Calcaneus/
4. calcane$.ti,ab.
5. Talus/
6. talus.ti,ab.
7. talocrural.ti,ab.
8. talofibular.ti,ab.
9. calcaneofibular.ti,ab.
10. Ankle Joint/
11. (ankle adj joint$).ti,ab.
12. Tarsal Joint/
13. (tarsal adj joint$).ti,ab.
14. Tarsal bones/
15. (tarsal adj bone$).ti,ab.
16. (lateral adj1 ligament$).ti,ab.
17. OR/1-16
18. Ankle Injury/
19. (ankle adj injur$).ti,ab.
20. Sprains and Strains/
21. (sprain$ or strain$).ti,ab.
22. inversion.ti,ab.
23. OR/18-22
24. exp Prognosis/
25. prognos$.ti,ab.
26. predict$.tw.
27. exp Follow Up Studies/
28. (follow adj up adj stud$).ti,ab.

29. incidence.ti,ab.

30. course.ti,ab.

31. exp Longitudinal Studies/

32. longitudinal.ti,ab.

33. Prospective Studies/

34. prospect$.ti,ab.

35. Risk factors/

36. (risk adj factor$).ti,ab.

37. Cohort Studies/

38. (cohort adj stud$).ti,ab.

39. OR/24-38

40. 17 AND 23 AND 39

CENTRAL Updated Search Strategy

#1 Ankle :MH 1364

#2 ankle :TI,AB,KY 4530

#3 (Ankle Joint):MH 505

#4 (ankle joint*):TI,AB,KY 814

#5 (Tarsal Bones):MH 16

#6 (tarsal bones):TI,AB,KY 19

#7 (tarsal joint*):TI,AB,KY 12

#8 (Tarsal Joints):MH 10

#9 Calcaneus:MH115

#10 calcane*:TI,AB,KY 353

#11 Talus:MH 20

#12 talocrural:TI,AB,KY 25

#13 talofibular:TI,AB,KY 9

#14 calcaneofibular:TI,AB,KY 10

#15 (Lateral Ligament, Ankle):MH 0

#16 (lateral ligament*):TI,AB,KY96

#17 #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16 4835

#18 (Ankle Injury):MH 0

#19 (ankle injur*):TI,AB,KY 561

#20 (Ankle Sprain):MH 0

#21 (ankle sprain):TI,AB,KY 245

#22 (Sprains and Strains):MH 267

#23 (sprain* or strain*):TI,AB,KY 7127

#24 inversion:TI,AB,KY 582

#25 #18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 7923

#26 Prognosis:MH 10961

#27 prognos*:TI,AB,KY 23331

#28 Forecasting:MH 463

#29 predict*:TI,AB,KY 51680

#30 (Follow Up):MH 48086

#31 follow?up*:TI,AB,KY 2075

#32 Incidence:MH 7849

#33 incidence:TI,AB,KY 59777

#34 (Cohort Studies):MH 6214

#35 (cohort stud*):TI,AB,KY 9473

#36 (Prospective Studies):MH 73954

#37 (prospect* stud*):TI,AB,KY 97763

#38 (Retrospective Studies):MH 6414

#39 (retrospect* stud*):TI,AB,KY 8809

#40 (Longitudinal Studies):MH 4966

#41 (longitudinal stud*):TI,AB,KY 6982

#42 (Risk Factors):MH 19329

#43 (risk factor*):TI,AB,KY 35375

#44 (Decision Support Techniques):MH 469

#45 #26 OR #27 OR #28 OR #29 OR #30 OR #31 OR #32 OR #33 OR #34 OR #35 OR #36 OR #37 OR #38 OR #39 OR #40 OR #41 OR #42 OR #43 OR #44 251623

#46 #17 AND #25 AND #45 324

#47 fracture:TI,AB,KY 7565

#48 #17 AND #25 AND #45 NOT 47 302

#49 01/01/2015 TO 27/07/2016:CD 118692

#50 #48 AND #49 33

CINAHL via EBSCOHost – 1982 to september 2015

MH Ankle

TI ankle* OR AB ankle*

TI calcaneofibular OR AB calcaneofibular

TI talofibular OR AB talofibular

TI talocrural OR AB talocrural

TI (ankle N1 joint*) OR AB (ankle N1 joint*)

TI “tarsal joint*” OR AB “tarsal joint*”

TI “tarsal bone*” OR AB “tarsal bone*”

MH Calcaneus

MH Talus

MH Tarsal Bones+

MH Lateral Ligament, Ankle

TI (lateral N1 ligament) OR AB (lateral N1 ligament)

MH Ankle Sprain

MH Sprains and Strains

TI sprain* OR AB sprain*

TI strain* OR AB strain*

MH Ankle Injuries

TI (injur* N1 ankle) OR AB (injur* N1 ankle)

TI (inversion N1 sprain*) OR AB (inversion N1 sprain*)

MH Incidence

TI predict* OR AB predict*

TI “cohort stud*” OR AB “cohort stud*”

TI course OR AB course

MH Predictive research

MH Prognosis

TI prognos* OR AB prognos*

TI “follow up stud*” OR AB “follow up stud*”

TI “follow-up stud*” OR AB “follow-up stud*”

MH Prospective studies+

TI “longitudinal stud*” OR AB “longitudinal stud*”

MH Risk Factors

TI recovery OR AB recovery

TI (treatment N1 outcome*) OR AB (treatment N1 outcome*)

S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13

S14 OR S15 OR S16 OR S17 OR S18 OR S19 OR S20

S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28 OR S29 OR S30 OR S31 OR S32 OR S33 OR S34

S35 AND S36 AND S37 retrieved 194 articles / 204 articles on the 26th july 2016

EMBASE via Ovid - 1974 to 2016 week 30 (July)

exp Ankle/

ankle.ti,ab.

Ankle Lateral Ligament/

(ankle adj lateral adj ligament).ti,ab.

Calcaneus/

calcane$.ti,ab.

Talus/

talus.ti,ab.

calcaneofibular.ti,ab.

talofibular.ti,ab.

talocrural.ti,ab.

(ankle adj joint$).ti,ab.

Tarsal Joint/

(tarsal adj joint$).ti,ab.

OR/1-14

Ankle Sprain/

Sprain/

sprain$.ti,ab.

strain$.ti,ab.

(inversion adj sprain$).ti,ab.

Ankle Injury/

OR/16-21

follow-up.mp.

prognos:.tw.

ep.fs.

OR/23-25

15 AND 22 AND 26

OpenGREY search strategy

Simple search in titles and abstracts for “ankle sprain or ankle”

PEDro search strategy

Simple search in titles and abstracts for “ankle sprains”

PsycINFO via Ovid – 1806 to July week 3 2016

1. exp Ankle/
2. ankle.ti,ab.
3. (ankle adj lateral adj ligament).ti,ab.
4. calcane$.ti,ab.
5. talus.ti,ab.
6. calcaneofibular.ti,ab.
7. talofibular.ti,ab.
8. talocrural.ti,ab.
9. (ankle adj joint$).ti,ab.
10. (tarsal adj joint$).ti,ab.
11. OR/1-10
12. sprain$.ti,ab.
13. strain$.ti,ab.
14. inversion.ti,ab.
15. OR/12-14
16. Prognosis/
17. prognos$.ti,ab.
18. predict$.ti,ab.

19. Followup Studies/
20. (follow?up adj stud$).ti,ab.
21. incidence.ti,ab.
22. course.ti,ab.
23. Longitudinal Studies/
24. (longitudinal adj stud$).ti,ab.
25. Prospective Studies/
26. (prospective adj stud$).ti,ab.
27. Risk Factors/
28. (risk adj factor$).ti,ab.
29. Cohort Analysis/
30. (cohort adj stud$).ti,ab.
31. Disease course/
32. OR/16-32

PubMed search strategy – 26th July 2016

1. Ankle [mh]
2. ankle* [tiab]
3. Lateral Ligament, Ankle [mh]
4. calcane* [tiab]
5. Ankle Joint [mh]
6. ankle joint* [tiab]
7. tarsal joint* [tiab]
8. calcaneofibular [tiab]
9. talofibular [tiab]
10. talocrural [tiab]
11. talus [tiab]
12. #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11
13. Ankle Injuries [mh]
14. sprain* [tiab]
15. strain* [tiab]
16. Sprains and Strains [mh]

17. inversion [tiab]
18. #14 OR #15 OR #16 OR #17
19. Prognosis [MeSH:noexp]
20. diagnosed [tiab]
21. cohort* [tiab]
22. Cohort effect [mh]
23. Cohort studies [MeSH:noexp]
24. predictor* [tiab]
25. death [tiab]
26. "models, statistical" [mh]
27. #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25 OR #26
28. #12 AND #18 AND #27

SportDiscus via EBSCOHost 1966 -2016.

1. SU Ankle
2. TI ankle* OR AB ankle*
3. TI calcaneofibular OR AB calcaneofibular
4. TI talofibular OR AB talofibular
5. TI talocrural OR AB talocrural
6. TI “ankle joint*” OR AB “ankle joint*”
7. TI “tarsal joint*” OR AB “tarsal joint*”
8. TI “tarsal bones” OR AB “tarsal bones”
9. TI calcane* OR AB calcane*
10. TI talus OR AB talus
11. SU Ankle Lateral Ligament
12. TI “lateral ligament” OR AB “lateral ligament”
13. S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12
14. SU Sprains
15. SU Strain
16. TI sprain* OR AB sprain*
17. TI strain* OR AB strain*
18. TI (injur* N1 ankle) OR AB (injur* N1 ankle)
19. TI (inversion N1 sprain*) OR AB (inversion N1 sprain*)
20. S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19

21. TI incidence OR AB incidence
22. TI predict* OR AB predict*
23. TI course OR AB course
24. TI cohort* OR AB cohort*
25. TI “cohort stud*” OR AB “cohort stud*”
26. SU Prognosis
27. TI prognos* OR AB prognos*
28. TI “follow up stud*” OR AB “follow up stud*”
29. TI “follow-up stud*” OR AB “follow-up stud*”
30. TI “longitudinal stud*” OR AB “longitudinal stud*”
31. TI “risk factor*” OR AB “risk factor*”
32. TI forecasting OR AB forecasting
33. TI “decision making” OR AB “decision making”
34. TI predict* and AB predict*
35. SU Cohort analysis
36. S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28 OR S29 OR S30 OR S31 OR S32 OR S33 OR S34 OR S35
36. S11 AND S20 AND S36

Appendix B – Consensus Meeting Pre-meeting Questionnaire

Please enter your Name here:__

Below is a questionnaire we would like you to complete and return prior to the Consensus Meeting on March the 27th. The results will inform our discussions during the meeting. **THEREFORE PLEASE COMPLETE AND RETURN BY WEDNESDAY 25TH MARCH**

We have formatted the questionnaire so it is easiest to complete electronically. Once you have completed it, please save the file and include your surname in the file name and then email it back to us at sprained@ndorms.ox.ac.uk .

The questionnaire asks about different factors that may help predict recovery following an ankle sprain. Before you complete this you should look at the information provided in the summary pack that accompanies this questionnaire.

Your responses will be collated with those from other people attending the Consensus Meeting. During the meeting the overall group ratings will be summarised and you will have your own results provided to you in confidence for you to compare.
Please note there are no right or wrong answers.

You will be asked to respond to the questions using a nine point scale. In all cases please mark your response clearly in one box only. If you are completing this electronically, you just need to click on one box. An example is shown below:

1	2	3	4	5	6	7	8	9
☐	☐	☐	☐	☒	☐	☐	☐	☐
	Not important				Important			Critical

In some cases you may feel you are unable to answer the question. In those cases please mark the “Don’t know” box.

At the time of assessment in A&E, how important are the following factors in predicting recovery from an ankle sprain:

1. The time between injury and visiting A&E

1	2	3	4	5	6	7	8	9
☐	☐	☐	☐	☐	☐	☐	☐	☐

Not important Important Critical

☐

Don't know

---

2. The amount of ankle pain a person has

1	2	3	4	5	6	7	8	9
☐	☐	☐	☐	☐	☐	☐	☐	☐

Not important Important Critical

☐

Don't know

---

3. The amount of ankle pain a person has when putting weight on their injured ankle

1	2	3	4	5	6	7	8	9
☐	☐	☐	☐	☐	☐	☐	☐	☐

Not important Important Critical

☐

Don't know

---

4. The ability to put full weight on their ankle

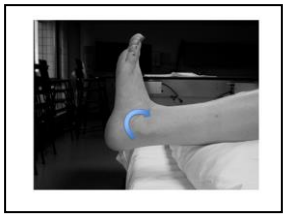
1	2	3	4	5	6	7	8	9
☐	☐	☐	☐	☐	☐	☐	☐	☐

Not important Important Critical

☐

Don't know

5. The amount of ankle movement a person has pulling their toes up towards their head (dorsiflexion)



1 2 3 4 5 6 7 8 9

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

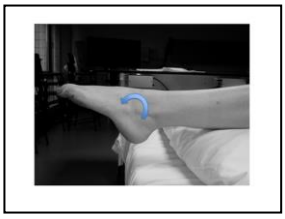
Not important Important Critical

☐

Don't know

---

6. The amount of ankle movement a person has pointing their toes away from their head (plantarflexion)



1 2 3 4 5 6 7 8 9

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

Not important Important Critical

☐

Don't know

---

7. Abnormal imaging findings (for example ultrasound or MRI scans)

1 2 3 4 5 6 7 8 9

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

Not important Important Critical

☐

Don't know

---

8. A person's age

1 2 3 4 5 6 7 8 9

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

Not important Important Critical

☐

Don't know

9. A person's Body Mass Index (combination of their weight and height)

1	2	3	4	5	6	7	8	9
☐	☐	☐	☐	☐	☐	☐	☐	☐

Not important Important Critical

☐

Don't know

---

10. A person's working status (unemployed or working part-time or full time)

1	2	3	4	5	6	7	8	9
☐	☐	☐	☐	☐	☐	☐	☐	☐

Not important Important Critical

☐

Don't know

---

11. A person's level of education

1	2	3	4	5	6	7	8	9
☐	☐	☐	☐	☐	☐	☐	☐	☐

Not important Important Critical

☐

Don't know

---

12. *How* a person injured their ankle

1	2	3	4	5	6	7	8	9
☐	☐	☐	☐	☐	☐	☐	☐	☐

Not important Important Critical

☐

Don't know

13. That a person has repeatedly sprained their ankle before

1	2	3	4	5	6	7	8	9
☐	☐	☐	☐	☐	☐	☐	☐	☐

Not important Important Critical

☐

Don’t know

14. Whether a person’s ankle is catching or locking

1	2	3	4	5	6	7	8	9
☐	☐	☐	☐	☐	☐	☐	☐	☐

Not important Important Critical

☐

Don’t know

We would be interested to hear about other factors that you think are important in predicting recovery from an ankle sprain. Please type/write the most important factors below (maximum 2) and rate their importance.

15. Extra Factor A. _

1	2	3	4	5	6	7	8	9
☐	☐	☐	☐	☐	☐	☐	☐	☐

Not important Important Critical

☐

Don’t know

16. Extra Factor B. _

1	2	3	4	5	6	7	8	9
☐	☐	☐	☐	☐	☐	☐	☐	☐

Not important Important Critical

☐

Don’t know

Some research studies have shown that it is beneficial to collect information after the initial visit to A&E. Collecting delayed information can often improve the accuracy of the prediction of how people will recover following an ankle sprain.

---

17. If we were to collect further information like this, how many weeks after the initial visit do you think we should collect this information?

1 week	2 weeks	3 weeks	4 weeks	5 weeks	6 weeks
☒	☐	☐	☐	☐	☐

☐

Don't know

---

18. How should we collect this information?

Hospital visit	Postal Questionnaire	Online Questionnaire	Telephone Questionnaire
☐	☐	☐	☐

☐

Don't know

If you have any additional comments, please add them below:

MANY THANKS FOR TAKING TIME TO COMPLETE THIS

Appendix C – Emergency Department Clinical Dataset Form

BASELINE CLINICAL DATA SET
REC Reference: 15/LO/0538

SPRAINED

SITE

ATTACH PATIENT STICKER

Tel No. _____

_____

Best time: _____

Email add: _____

(Research nurse: If patient is suitable for the study and has been registered, please tear/fold here to anonymise prior to sending a copy to the sprained study office)

SPRAINED BASELINE CLINICAL DATA SET: ANKLE INJURY (AGE 16 +YEARS)

HISTORY

1. Age: ___ ___ 2. Sex (tick): ☐ Male ☐ Female

3. Patient's reported Height: Feet ___ ___ Inches ___ ___ OR ___ ___ ___cms

4. Patient's reported Weight: Stone ___ ___ lbs ___ ___ OR ___ ___ ___ kgs

5. Currently employed (tick): ☐ None ☐ Part-time ☐ Full-time ☐ Student ☐ Retired

6. Days since injury ___ ___ Date of injury (DD/MM/YYYY) ___ ___/___ ___ ___/___ ___ ___ ___

7. Injury setting (tick): ☐ At home ☐ At work/uni/school ☐ Playing sport ☐ Outside in public ☐ Other _____

8. Sprained this ankle in last 12 months (tick) ☐ Y ☐ N 9. Sprained this ankle at least twice before (tick) ☐ Y ☐ N

EXAMINATION

10. Ankle side (tick): ☐ L ☐ R 11. Patient able to weight bear (tick)? ☐ Y ☐ N

12. Pain at rest on 0-10 scale? (0= no pain, 10 = worst pain imaginable) ___ ___

13. Pain on weight bearing on 0-10 scale? (0= no pain, 10 = worst pain imaginable) ___ ___

14. Ankle movement limited (tick)? ☐ Y ☐ N

15. Since injury can patient dorsiflex fully? (circle one) Always / Often / Sometimes / Rarely / Never

16. Since injury can patient plantarflex fully? (circle one) Always / Often / Sometimes / Rarely / Never

17. Since injury has pt experienced catching/ locking when moving? (circle one) Always / Often / Sometimes / Rarely / Never

18. How long does patient expect recovery will take? (circle one)

< 2 wks / 2-8 wks / 2-6 mths / 6-12 mths / > 1 year / Not sure will recover / don't know

INVESTIGATION

Xray of ankle/foot: (tick) ☐ Y ☐ N

SPRAIN SEVERITY (tick)

Mild (Gd I) / Moderate (Gd II) / Severe (Gd III)

☐ ☐ ☐

SUITABLE FOR SPRAINED STUDY?

PATIENT CONTACT DETAILS (mobile) CONFIRMED AS CORRECT ☐

Is patient suitable for the SPRAINED Study(tick)? ☐ Y ☐ N If YES, trial information and invite given? ☐ Y ☐ N

If patient does not wish to be contacted about SPRAINED please tick here ☐

Record reason here if patient declines SPRAINED study _____

Signature of clinician: _____ Date form completed: ___/___/_____

PLEASE NOW HAND THIS FORM TO YOUR SPRAINED RESEARCH CLINICIAN / PLACE IN SPRAINED BOX FILE

When registered on to the SPRAINED Study: Enter study number and initials below, anonymise this sheet and post to the SPRAINED Study Office in Freepost envelope.

Study code:		Site ID code:		Participants Study Number:					Initials:		
S	P										

Appendix D – Dynamic Consent

SPRAINED study pilot of Dynamic Consent

Aim: to pilot Dynamic Consent in the SPRAINED study to explore how it might improve the consent procedure, and whether or not it influences trial adherence.

Objective: to determine whether Dynamic Consent can be introduced to a clinical study and integrated appropriately with study management software and existing recruitment processes.

Background: Dynamic Consent is an approach to informed consent that is designed to allow participants to have greater control over how their samples and data are used, to interact with the study team more easily, and to receive updates on how the research is progressing. Participants receive access to a personal profile that allows them to review their consent decisions, to change their mind, and to receive relevant information about the study.

Researchers at the HeLEX centre have developed software to support a dynamic consent approach, and worked with members of the SPRAINED research team to trial the software (tailored to the study) in the SPRAINED study, to see if it would influence trial retention rates. If participants were reminded of their involvement in SPRAINED, received notifications of up-coming questionnaires, and were informed of the value of their continued involvement even if they'd fully recovered, it was hoped that this would help study retention.

It was important to ensure that Dynamic Consent did not adversely affect the SPRAINED study. On this basis it was introduced in the later stages of recruitment, once the centres had initiated recruitment processes and were familiar with the study. Ethics approval for the amendment to the study protocol was received, allowing dynamic consent to be implemented. Participants were consented if they visited the emergency department with a sprained ankle. For the participants that were asked to trial dynamic consent, the consent process was the same as for those following a traditional consent pathway, with an additional question included on the form asking whether they would be happy to use dynamic consent. They then signed a paper consent form, providing an email address, and were sent a web-link to their secure Dynamic Consent page, where they could review their consent decisions or make any changes at any stage in the study. They also received notifications of any updates to the pages, including articles reminding them to complete the follow-up questions at 4 weeks, 4 months and 9 months.

Challenges: Dynamic consent presented a minor change to the recruitment process, and as such implementing the change took longer than anticipated, as recruitment teams had to update their paperwork and remember to ask about involvement in the additional aspect of the study. Not all participants provided email addresses, which limited the opportunity to set up dynamic consent accounts.

Results: 22 participants were recruited to use Dynamic Consent, out of a total of 682 participants in the SPRAINED study. Of these 21 users, 8 accessed their dynamic consent pages during the study (none of the participants changed their consent decisions during the study). It is not possible to determine from this whether Dynamic Consent improved response rates or study adherence, however it was successful in demonstrating the possibility for Dynamic Consent software to integrate with clinical trial management software, and confirmed that the process for consent using the dynamic consent software worked within a clinical setting.

Future work: Having confirmed the viability of the software, it is now important to apply it to a larger study, with a greater number of participants to further explore user experience, and to demonstrate how dynamic consent influences study experience and adherence.