2018-11

Development and prospective external validation of a tool to predict poor recovery at nine months after acute ankle sprain in UK emergency departments: the SPRAINED prognostic model.

Schlussel, M

10.1136/bmjopen-2018-022802
BMJ Open
BMJ Publishing Group

DEVELOPMENT AND PROSPECTIVE EXTERNAL VALIDATION OF A TOOL TO PREDICT POOR RECOVERY AT NINE MONTHS AFTER ACUTE ANKLE SPRAIN IN UK EMERGENCY DEPARTMENTS: THE SPRAINED PROGNOSTIC MODEL

Michael M Schlüssel [1*], David J Keene [2], Gary S Collins[1], Jennifer Bostock[3], Christopher Byrne[4], Steve Goodacre[5], Stephen Gwilym[6], Daryl A Hagan[2], Kirstie Haywood[7], Jacqueline Thompson[2], Mark A Williams[8], Sarah E Lamb [1,2] for the SPRAINED study team.

[1] Centre for Statistics in Medicine, Nuffield Departments of Orthopaedic Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom

[2] Centre for Rehabilitation Research, Nuffield Departments of Orthopaedic Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom

[3] Patient and Public Involvement representative, Quality and Outcomes of Person-Centred, Care Policy Research Unit, Canterbury, Kent, UK

[4] Faculty of Health and Human Sciences, University of Plymouth, Plymouth, United Kingdom

[5] School of Health and Related Research, University of Sheffield, Sheffield, United Kingdom

[6] Oxford Trauma, Nuffield Departments of Orthopaedic Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom

[7] Warwick Research in Nursing, Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry, United Kingdom

[8] Department of Sport, Health Sciences and Social Work, Oxford Brookes University, Oxford, United Kingdom

*** Corresponding author's address:**

Dr Michael Maia Schlüssel

Centre for Statistics in Medicine

Botnar Research Centre, Old Road

OX3 7LD

Oxford, United Kingdom

Email michael.schlussel@csm.ox.ac.uk

Telephone: +44 01865 737916

Word count: 4,222 words

ABSTRACT

OBJECTIVES: To develop and externally validate a prognostic model for poor recovery after ankle sprain. **SETTING AND PARTICIPANTS:** Model development used secondary data analysis from 584 participants in a UK multicentre randomised clinical trial. External validation used data from 682 participants recruited in 10 emergency departments across the UK for a prospective observational cohort. **OUTCOME AND ANALYSIS:** Poor recovery was defined as presence of pain, functional difficulty or lack of confidence in the ankle at 9-months after injury. Twenty-three baseline candidate predictors were included together in a multivariable logistic regression model to identify the best predictors of poor recovery. Relationships between continuous variables and the outcome were modelled using fractional polynomials. Regression parameters were combined over 50 imputed datasets using Rubin's rule. To minimise over-fitting, regression coefficients were multiplied by a heuristic shrinkage factor and the intercept re-estimated. Incremental value of candidate predictors assessed at 4-weeks after injury was explored using decision curve analysis and the baseline model updated. The final models included predictors selected based on the Akaike Information Criterion ($p<0.157$). Model performance was assessed by calibration and discrimination. **RESULTS:** Outcome rate was lower in the development (6.7%) than in the external validation dataset (19.9%). Mean age (29.9 and 33.6 years), BMI (26.3 and 27.1 kg/m^2), pain when resting (37.8 and 38.5 points) or bearing weight on the ankle (75.4 and 71.3 points) were similar in both datasets. Age, BMI, pain when resting, pain bearing weight, ability to bear weight, days until assessment, and injury recurrence were the selected predictors. The baseline model had fair discriminatory ability (c-statistic 0.72; 95%CI: 0.66-0.79) but poor calibration. The updated model presented better discrimination (c-statistic 0.78; 95%CI: 0.72-0.84), but equivalent calibration. **CONCLUSIONS:** The models include predictors easy to assess clinically and show benefit when compared to not using any model. **Registry number:** ISRCTN12726986.

Keywords: prognosis, clinical prediction rule, logistic model, ankle injuries, sprains and strains

STRENGTHS AND LIMITATIONS OF THIS STUDY

- This is the first study to develop and externally validate a tool to predict poor recovery after ankle sprain, including a wide range of clinically relevant candidate predictors.
- Despite containing information on the outcomes of interest and numerous prognostic variables, the development dataset was not originally acquired to build a prognostic model.
- The number of events in the development dataset was relatively small compared to the number of candidate predictors examined.
- Yet, the prognostic models were developed using robust statistical methods, adjusted for overfitting and reported according to the most recent relevant guidelines available.
- Generalisability of findings is enhanced by the multi-centre characteristic of the datasets used in the development and external validation of the models.

INTRODUCTION

Ankle sprains are one of the most common musculoskeletal injuries, representing up to 5% of all emergency department (ED) attendances in the UK.**[1]** Despite heterogeneity in sampling frame (e.g. restricted to elite athletes or excluding older people), inception, and follow-up time points, studies have indicated that approximately 30% of people have persistent problems one year after ankle sprain.**[2, 3]** In a large multi-centre randomised clinical trial conducted in the UK, a similar proportion (30%) of participants had poor outcome at 9 months.**[4]** Other studies indicate a recovery plateau at around 9 months, and residual disability after this point to be persistent.**[5]**

In the acute phase after a sprain, physical examination of the ankle is often difficult due to swelling and pain. Predicting prognosis at this stage is uncertain and based on clinical judgement. When concerned about the injury severity, clinicians operate a system of review within one week in a trauma clinic (or equivalent service), which allows some resolution of swelling and reassurance about the presence of other significant mechanical derangement.**[6]** The Ottawa ankle rule is also an alternative to reduce the requirement for imaging without missing important fractures.**[7]**

In 2008, Van Rijn et al conducted a systematic review on the clinical pathway and prognostic factors of ankle sprain recovery and found a single eligible study concluding that high levels of sports activity have prognostic value for residual symptoms.[2] In a more recent systematic review, we have identified nine studies reporting results for baseline prognostic factors of recovery after an acute ankle sprain.[8] Age, gender, swelling, range of motion, weight bearing ability, pain, injury severity, palpation/stress score, injury mechanism, self-reported recovery, re-sprain, MRI determined number of sprained ligaments and bone bruise were reported as independent predictors of poor recovery. However, almost all studies performed poorly on the risk of bias assessment, mainly due to incomplete or inadequate reporting standards for study participants, attrition, methods of assessment for predictors, confounding and statistical methods used, so results should be interpreted with caution.

To the best of our knowledge, there are no externally validated prognostic models for recovery after acute ankle sprain. Polzer et al. developed an algorithm to help clinicians with the diagnosis and treatment of acute ankle injuries, but this is considerably based on expert

judgements and do not use currently recommended methods for the development of prognostic models.[9] A robustly developed and validated prognostic model could help to target treatment better and improve outcomes for people who have an ankle sprain.[10] Therefore, the development of a new prognostic model, considering a range of plausible candidate predictors, and ideally with the evaluation of its performance on an external dataset (external validation), is indicated.

The aim of our study was to develop and externally validate the SPRAINED (Synthesising a clinical Prognostic Rule for Ankle Injuries in the Emergency Department) prognostic model, to identify people at risk of poor recovery at nine months after acute ankle sprain.

METHODS

Study populations and data collection

Data from the Collaborative Ankle Support Trial (CAST), were used to develop the prognostic model.**[11]** CAST was a pragmatic multicentre randomised controlled trial on the effectiveness of different mechanical ankle supports compared with a double-layer tubular compression bandage for managing severe ankle sprains. The trial sample comprised 584 participants aged 16 years or older, with an ankle sprain of grade 2 or 3, attending eight EDs in the UK between April 2003 and July 2005, within 7 days after their injury, and were not able to fully bear weight on the injured ankle at baseline. Further data was collected at 4 and 12 weeks, and 9 months after randomisation. The CAST methods and a CONSORT flow diagram are available elsewhere.**[11]**

To assess the model's performance in an external population, the SPRAINED prospective observational cohort was recruited. Participants were aged 16 years or above, with acute ankle sprains of any grade, attending 10 NHS EDs across England, within 7 days after their injury. Patients were excluded if they presented with an ankle fracture (except flake fractures < 2mm) or any other recent (< 3 months) lower limb fracture. Participants were not randomised, nor did they receive any interventions other than usual care at each site. The study recruited 682 participants between July 2015 and March 2016. Data collection covered clinical and socio-demographic information collected at ED presentation (baseline), with follow-up assessments at 4 weeks, 4 and 9 months after the initial injury, either by self-reported paper-based forms sent back to the study office by postal mail, electronic questionnaires, or telephone interviews. The SPRAINED questionnaires included all variables

selected as predictors in the model and the components of the outcome of interest. All participants of both studies have provided written informed consent before any data collection took place. Ethics approval was from the National Research Ethics Committee (REC) (London - Chelsea), REC number 15/LO/0538, on 10th April 2015. The study protocol was registered on 30th April 2015; registry number ISRCTN12726986.

Definition of outcome

A prognostic model was developed to predict 'poor recovery' at 9 months after an acute ankle sprain. Poor recovery was defined as the presence of pain, lack of confidence in the ankle (persistent feeling of giving way) or functional difficulty.**[12, 13]** The presence of these symptoms was assessed by patient-reported responses given to specific items (P1, Q3 and Q4) of the Foot and Ankle Outcome Score (FAOS).**[14]** Participants who answered one or more of these questions with any of the two most extreme response options ("daily" or "always" P1; "severely" or "extremely" for Q3 or Q4) were considered to have poor outcome.

Baseline candidate predictors

Thirty-two baseline candidate predictors were considered plausible predictors of poor outcome and pre-selected from a pool of 170 variables available in the CAST dataset (**Supplemental Tables 1 and 2**). This initial selection was made internally by the research team, taking into account the results from our systematic literature review **[8]** and the conclusions from a consensus group meeting convened for the SPRAINED study, which included clinicians, medical researchers, statisticians and PPI representatives. The 32 candidate predictors included socio-demographic information (e.g. age, sex, body mass index (BMI), education, employment status); pre-injury quality of life, mobility and lifestyle indicators (e.g. engagement in sports activities); clinical data on injury presentation; baseline (post-injury) mobility levels, pain and weight-bearing status (**Supplemental Table 3**)**.**

At this stage, variables were excluded or combined before statistical modelling if they had 60% or more of missing information; displayed high collinearity ($r \geq 0.8$) with another candidate predictor; presented empty or low cell counts ($n < 5$) when tabulated against the outcome; were the offending variable causing perfect prediction during the multiple imputation process (**Supplemental Table 4; Supplemental Figure 1**).

Sample size considerations

It is widely recommended that the dataset used to develop a prognostic tool should contain a minimum of 5-10 outcome events per variable (EPV) included as a predictor in the model.**[15-20]** After the exclusion of nine baseline 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 two levels, so we ended with 35 candidate parameters, meaning the EPV ratio was approximately three.

As to the best of our knowledge this was the first study aiming to develop prediction models to assess the risk of poor recovery after an acute ankle sprain, we opted for relaxing the EPV rule in favour of including more potentially important predictors. Nevertheless, we adopted several strategies to minimise bias and overfitting, as described below.

Descriptive analysis

Baseline and 4-week follow-up characteristics of the CAST and SPRAINED participants were summarised using means, standard deviations (SDs) and ranges for continuous variables, or counts and percentages for categorical variables. Inspection of extreme values (outliers) took place to confirm whether they were clinically plausible and visual assessment of data distribution for continuous predictors in both datasets was conducted. No formal statistical tests were performed to compare the values between the studies.

Prognostic model development

Using logistic regression, we developed the prognostic model to predict the probability of poor recovery. We performed multiple imputation using chained equations (MICE) **[21]** to handle missing data, with 50 imputed datasets created. Continuous variables were kept as continuous to avoid loss of prognostic information,**[22]** and the shape of their relationship with the outcome studied and modelled with nonlinear functions such as fractional polynomials (FPs) where appropriate.**[23]** As several continuous variables were included in the models, we used the multivariable fractional polynomial (MFP) algorithm.**[24, 25]**

Multiple imputation and fractional polynomials were combined using the *mfpmi* function in Stata.**[26]** The estimated regression parameters (coefficients and variances) were combined over the 50 imputed datasets using Rubin's rule.**[27, 28]** After identifying the best transformation terms for continuous variables, the final model included predictors (and respective transformations, where applicable) selected from the full multivariable model with

all candidate predictors based on the Akaike Information Criterion (equivalent to a p-value < 0.157).**[29]** To adjust for over-fitting, due to small EPV, we multiplied all regression coefficients by the heuristic shrinkage factor,**[30]** then re-estimated the intercept. All model assumptions were checked and differences between incomplete and imputed datasets inspected. Imputed data from all 584 participants were included in all analyses.

Incremental value analysis and model update

In addition to the baseline predictors, 14 additional variables from the CAST 4-weeks follow-up questionnaire were also selected as potential predictors that could increase the model's prognostic ability (**Supplemental Table 3**). First, all additional 4-weeks candidate predictors were included together in the final baseline model and only those achieving a p-value < 0.157 were considered for inclusion in the updated model (i.e. a model including baseline and 4-weeks predictors). Finally, the updated model was compared with the original baseline model using decision curve analysis (DCA) plots to determine whether the inclusion of additional predictors reflected in increased net benefit.**[31, 32]**

External validation: Model performance

We assessed the model performance in the prospectively collected SPRAINED cohort. Missing data in the SPRAINED cohort was handled using MICE, creating 50 imputed datasets. Performance was evaluated by assessing calibration and discrimination. Calibration is the agreement between observed and predicted probabilities of poor outcome. Calibration was assessed graphically using calibration plots, with observed risks plotted on the *y*-axis against predicted risks on the *x*-axis.**[33, 34]** The calibration plot was created by regressing the outcome on the predicted probability using a locally weighted scatter plot smoother (lowess). 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 the 45° line). Calibration plots followed recommendations of overlaying calibration curves from each imputed data set.**[35]** Discrimination reflects the ability of the model to distinguish between participants who did and did 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.**[36]** Finally, to estimate the benefit of using the developed models, 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 selected thresholds and how many of these went on to present the outcomes compared with not using the model. Individual probabilities of developing the outcomes were estimated by applying the developed prognostic models to each participant in the SPRAINED imputed datasets. We assessed the performance of both the baseline and updated models using imputed data from all 682 participants.

Patient involvement

A PPI representative was involved in the study from the beginning, providing advice on key aspects of the study design, including the definition of the research question, choice of the outcome and selection of relevant candidate predictors during the consensus group meeting. They will be consulted for the public dissemination of any product arriving from this research.

Reporting

We followed the TRIPOD Statement for the reporting of our study.**[37]**

RESULTS

Baseline characteristics for the CAST (development) and SPRAINED (validation) cohorts are summarised in **Table 1**. On average, participants were slightly older in SPRAINED than in CAST. Participants in SPRAINED had an average BMI within the overweight category, likewise those in CAST. The mean pain scores when resting or bearing weight on the ankle of SPRAINED participants were also similar to those observed for CAST participants. Differently from CAST, in SPRAINED about half of participants were female, the majority presented to an ED within 2 days from injury for assessment and were able to bear some weight on their injured ankles (**Table 1**).

Table 1. Baseline characteristics of the participants in the CAST trial and SPRAINED prospective observational cohort.

Variable	CAST Trial		SPRAINED Cohort	
	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
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
	Frequency	**%**	**Frequency**	**%**
Sex				
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
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

Table 2 shows the rates of poor recovery in the CAST trial and SPRAINED cohort datasets, as well as the number of its component symptoms, at 9 months after injury. There was a lower rate of poor recovery in the SPRAINED cohort than observed in the CAST trial, but the percentage of missing data for the outcome was similar in both studies.

Table 2. Outcome and respective symptoms components rates and proportion of missing data in the CAST trial and SPRAINED prospective observational cohort.

	Pain	Lack of confidence	Instability	Poor recovery	Missing data	TOTAL[1]
CAST	84 (14.4%)	42 (7.2%)	67 (11.5%)	116 (19.9%)	144 (24.7%)	584
SPRAINED	3 (0.4%)	23 (3.4%)	37 (5.4%)	46 (6.7%)	155 (22.7%)	682

Note: Poor recovery defined as the presence of one or more of the following symptoms: pain, lack of confidence or instability/difficulty with the ankle.

Table 3 displays the summary of the final multivariable models (predictor's coefficients, respective 95% confidence intervals and p-values). Seven of the 23 baseline candidate predictors were selected for inclusion in the baseline model: 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. The best fit for all continuous predictors were linear transformations (mean subtractions) and were later incorporated into the model by updating the intercept accordingly (**Supplemental Table 5**).

Table 3. Summary of the final baseline and updated (baseline plus 4-weeks predictors) logistic regression models and respective shrunk coefficients and intercepts.

Predictors	Baseline model					Updated model (baseline plus 4-weeks predictors)				
	Coefficient	95% CI		p	Shrunk coefficient	Coefficient	95% CI		P	Shrunk coefficient
Age	0.027	0.006	0.048	0.014	0.019	0.018	-0.005	0.040	0.127	0.015
BMI	0.031	-0.014	0.076	0.178	0.022	0.025	-0.022	0.072	0.292	0.021
Pain when resting	0.016	0.005	0.027	0.005	0.011	0.010	-0.002	0.022	0.107	0.008
Pain when bearing weight	0.019	0.004	0.035	0.016	0.014	0.014	-0.002	0.030	0.092	0.012
Pain when bearing weight 4 wks. after injury	-	-	-	-	-	0.022	0.012	0.032	< 0.001	0.018
Days from injury to assessment (reference 0-2)										
3 or more	0.854	0.068	1.640	0.034	0.605	0.702	-0.117	1.520	0.092	0.591
Able to bear weight at Baseline (reference No)										
Yes	-0.792	-1.376	-0.207	0.008	-0.561	-0.802	-1.412	-0.192	0.010	-0.676
Recurrent sprain (reference No)										
Yes	1.180	0.417	1.944	0.003	0.836	1.170	0.386	1.953	0.004	0.985
Intercept	-1.580	-2.152	-1.008	< 0.001	-1.363	-1.543	-2.128	-0.958	< 0.001	-1. 420

95% CI: 95% confidence interval

Linear terms selected by the MFP for continuous predictors: 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.

254

Only pain when bearing weight on the sprained ankle at 4 weeks after injury was included in the updated model (baseline plus 4-week predictors) (**Table 3**). By inspecting the DCA plots shown in **Figure 1** it is possible to see a clear net benefit gain over the entire range of thresholds when using the updated prognostic model in comparison to the baseline model or considering all patients (or no patient) at risk of having poor recovery after an acute ankle sprain.

Shrinkage suggested both prognostic models (baseline and updated) had predictor-outcome associations that were too large. The heuristic shrinkage factor for the coefficients of the predictors in the baseline prognostic model was 0.71. For the updated version (baseline plus 4-weeks predictors), the estimated heuristic shrinkage factor was 0.84. The shrunk coefficients and intercepts for the final models are presented in **Table 3**.

Overall, discrimination of the baseline model was fair, with a c-statistic of 0.72 (95%CI: 0.66 to 0.79). Calibration of the baseline prognostic model in the external validation dataset was poor though, as can be evidenced by inspecting the calibration plot with overlaid calibration lines from the 50 imputed datasets (**Figure 2**). The calibration slope was 1.13 (95%CI: 0.76 to 1.5) and the calibration intercept was -0.71 (95%CI: -0.98 to -0.44). The updated model (baseline plus 4-weeks predictors) presented better discriminatory ability in the SPRAINED dataset than the baseline model (c-statistic = 0.78; 95%CI: 0.72 to 0.84), but equivalent calibration, with an intercept closer to zero (-0.51; 95%CI: -0.78 to -0.24) and slope slightly further from one (1.17; 95%CI: 0.86 to 1.48).

Table 4 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 recovery 9 months after an acute ankle sprain. There seems to be little difference between the baseline and updated models, with both identifying similar numbers of patients who would experience a poor outcome after an acute ankle sprain. However, less patients are deemed at high risk by using the updated model for (less false positives) across all thresholds of predicted probability suggesting that reassessing the patients at 4 weeks after the injury might be beneficial to a more accurate prediction of their probability of poor outcome. Using any of the models is clearly beneficial, when compared to not using any model (i.e. considering all patients – or no patients – as high risk of developing poor outcome).

Table 4. Models performance (numbers at risk and outcomes identified) at varying risk thresholds for 1000 patients.

Selected thresholds	Number of patients at risk		Number of events	
	High risk	Low risk	Identified	Not identified
Consider all high risk	1000	0	85	0
Predicted probability as per baseline model				
≥5%	971	39	85	0
≥10%	797	203	74	11
≥15%	543	457	63	22
≥20%	351	649	52	33
Predicted probability as per updated model (baseline plus 4-weeks predictors)				
≥5%	882	118	85	0
≥10%	517	483	71	14
≥15%	358	642	56	29
≥20%	259	741	41	44

DISCUSSION

We developed a prognostic model to predict a composite outcome representing the presence of at least one of the following symptoms at 9 months after an acute ankle sprain: pain, functional difficulty or lack of confidence in the ankle. The model presented fair discriminatory ability in a prospectively collected external validation cohort, but poor calibration. Including an additional variable collected at 4 weeks after the injury (pain when bearing weight on the injured ankle) improved the discriminatory ability and calibration of the model. The models included predictors that are easily collected and provided reasonable predictions of poor recovery for patients with acute ankle sprain.

In a recent systematic review, we have reported that some of the variables selected for inclusion in our prognostic model, have been previously identified as important predictors of short, medium or long term recovery after ankle sprain.[8] According to O'Connor et al. age and weight bearing ability are predictors of ankle function, as measured by the Karlsson function score, both at 4 weeks and 4 months after injury.[38] Akacha et al. also demonstrated that age was an important predictor of slower and incomplete recovery after ankle sprain, as measure by the Foot and Ankle Outcome Score.[39] The magnitude of pain at rest at 3 months has also been shown to have prognostic value for poorer self-reported recovery at 12 months after ankle sprain by Van Middelkoop et al.[40] On the other hand, Findings regarding recurrence of ankle sprain are conflicting. McKeon et al., reported that recurrent ankle sprain was not a significant predictor of time to return-to-play after an ankle injury.[41] This is contrary to reports of an association between recurrent sprains and chronic

ankle instability reported in a systematic review conducted by Pourkazemi et al. [42] One possible explanation for these contradictory results may be the nature of the outcomes investigated in each study. When more subjective aspects of recovery (such as ankle function or instability) are considered in the definition of the endpoint, like in the present study, re-spraining the ankle seems to be an important predictor of recovery.

The inclusion of BMI in the prognostic model is another issue that deserves consideration. Although not statistically significant in the final multivariable logistic regression analysis, according to AIC ($p<0.157$), we have decided to keep BMI in the model for several reasons. First, this decision prevented another round of predictor selection, which could increase over-fitting. The model building process was not solely based on statistical rationale, and BMI was considered to be an important predictor by clinicians during our consensus group meeting. BMI is an easy to assess surrogate measure of body weight that is frequently collected at clinical routine and one that most patients know how to calculate themselves. Finally, its inclusion does not add much complexity to the models.

To the best of our knowledge, this is the first study to develop and externally validate a prognostic model to predict a clinically relevant outcome in people with acute ankle sprains exploring a wide range of clinically plausible candidate predictors. We used robust statistical methods to select the predictors and assess the model´s performance in a large external prospective cohort. 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. The observational cohort we prospectively recruited for SPRAINED is representative of patients presenting to EDs in the UK. We followed the most recent and complete guidelines available on the reporting of prognostic model development,**[37]** and applied recommended methods to minimise overfitting. 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 re-estimation of intercepts) as well as a prospective external validation phase. The amount of missing data in the external validation dataset, which is commonplace in studies of this nature, was considerably smaller than that

observed in the development dataset. Finally, we performed missing data imputation to produce a set of 50 complete datasets and enable robust analyses.

Limitations of the SPRAINED study are acknowledged. Firstly, data used to develop the 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 information on the symptoms and clinical events of interest, and a wide range of the candidate prognostic variables considered to have predictive ability. Secondly, the CAST dataset used to develop the prognostic model was relatively small compared to the number of candidate predictors.**[15-20]** As previously highlighted, the low EPV observed for the two developed models might have contributed to the optimism found for both and, therefore, to the poor calibration on the external validation dataset. Thirdly, the amount of missing data observed in the development dataset. Because of that, a number of candidate predictors were omitted before the process of data imputation, to avoid instability of the imputation models. Therefore, some important predictors could have conceivably been missed in the development phase of the SPRAINED study. Finally, the rates of poor outcome in the SPRAINED cohort were lower than in the CAST trial and those reported in previous systematic reviews.**[2, 3]** These variations in poor outcome rates and clinically important differences in baseline characteristics included in the prognostic model (such as days from injury to clinical assessment and ability to bear weight on the injured ankle) highlight the issue of different sampling frames.

Clinical examination of acute ankle sprain is challenging as tolerance of physical examination tests is often poor due to pain and swelling. Imaging is often not routinely available. A prognostic tool could enable better targeting of treatments such as immobilisation casts, which although effective can be inconvenient to patients, to those deemed at low risk of poor outcome. On the other hand, it has the potential to help clinicians targeting treatments such as surgery and physiotherapy to patients who are at highest risk of poor outcome.

The SPRAINED prognostic model benefits from including predictors that are easy to measure, and usually assessed in clinical routine. Given the herby discussed limitations in its predictive performance, we suggest that its value would be in assisting the clinician to estimate the probability of a poor outcome, instead of being used as a decision making tool in isolation. Improved predictive performance of the models with the addition of information on pain

when bearing weight at 4 weeks indicates that re-assessment of prognosis after the acute phase is worth consideration for patients initially deemed to have elevated probability of delayed recovery. Besides, as it is an easy-to-use instrument, patients themselves can estimate their probability of poor outcome and gain some reassurance in their decisions to seek for further medical assistance or not.

If implemented in clinical practice, clinicians should be aware that there is a degree of uncertainty associated to the calculated risk of poor outcome when using the SPRAINED prognostic model. This uncertainty can lead to over or under referral of patients to review clinics or referral treatment such as physiotherapy. Future work could examine how well the model performs in comparison (or addition) to the clinician impression. Moreover, we recommend further research to evaluate the impact of using the SPRAINED prognostic model in clinical practice to predict patient outcomes and to assess the acceptability and uptake of the tool by clinicians in the EDs.

In conclusion, the SPRAINED prognostic models performed reasonably and despite some miscalibration show benefit in identifying patients at high risk of poor outcome after an acute ankle sprain. The models may assist clinical-decision making when assessing and advising people with ankle sprains in the ED setting and when deciding on on-going management. The models benefit from using predictors that are simple to obtain during routine clinical assessment.

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.

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.

COMPETING INTERESTS

None declared.

AUTHOR'S CONTRIBUTIONS

MMS analysed and interpreted the data, and led the writing of the manuscript. DJK had substantial contribution in data acquisition, analysis and interpretation. GSC had substantial contribution in the study conception and design, data analysis and interpretation. JB had substantial contribution in the study conception and design. CB had substantial contribution in the data acquisition. SG had substantial contribution in the study conception and design. SG had substantial contribution in the study conception and design. DAH had substantial contribution in the data acquisition. KH had substantial contribution in the study conception and design. JT had substantial contribution in the data acquisition. MAW had substantial contribution in the study conception and design, data acquisition. SEL was responsible for the study conception and design, had substantial contribution in data interpretation. All authors revised and approved the final version of the manuscript.

ACKNOWLEDGEMENTS

The SPRAINED study team members are: Sarah E Lamb (Chief Investigator); David J Keene (Study Lead); Gary S Collins, Mark Williams, Steve Goodacre, Matthew Cooke, Stephen Gwilym, Phil Hormbrey, David Wilson, Jennifer Bostock (Co-investigators); Daryl Hagan (Study administrator); Damian Haywood (Senior study manager); Jacqueline Thompson, Christopher Byrne (Research Physiotherapists); Michael Maia Schlüssel (Study statistician); Philip Hormbrey, Susan Dorrian, Steve Goodacre, Victoria Stacey, Tim Coats, Sarah Wilson, Jason Kendall, David Clarke, Antoanela Colda, Deborah Mayne (Principal Investigators and their clinical and research teams at collaborating recruitment centres); Kirstie Haywood (consultation and senior facilitation of the consensus meeting). We would like to thanks Prof. Richard Riley, Prof. Kevin Mackway-Jones and Prof. Suzanne McDonough from the Independent Study Steering Committee for their invaluable comments, support and advise throughout the study. We also acknowledge the contributions of Vicki Barber and the staff members at Oxford Clinical Trials Research Unit and Centre for Rehabilitation Research for their support in delivering the SPRAINED study; the members of the consensus group meeting, including the PPI representatives; and all those who have in some way contributed to the conception, conduction and reporting of the SPRAINED study.

REFERENCES

1. Wilson RW, Gansneder BM. Measures of functional limitation as predictors of disablement in athletes with acute ankle sprains. *J Orthop Sports Phys Ther* 2000;30:528-35.

2. van Rijn RM, van Os AG, Bernsen RM, et al. What is the clinical course of acute ankle sprains? A systematic literature review. *Am J Med* 2008;121:324-31.

3. Polzer H, Kanz KG, Prall WC, et al. Diagnosis and treatment of acute ankle injuries: development of an evidence-based algorithm. *Orthop Rev* (Pavia) 2012;4:e5.

4. Cooke MW, Marsh JL, Clark M, 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.

5. Verhagen RA, de Keizer G, van Dijk CN. Long-term follow-up of inversion trauma of the ankle. *Arch Orthop Trauma Surg* 1995;114:92-6.

6. 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.

7. 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.

8. Thompson JY, Byrne C, Williams MA, et al. Prognostic factors for recovery following acute lateral ankle ligament sprain: a systematic review. *BMC Musculoskelet Disord*. 2017;18:421.

9. 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.

458 10. Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A, et al. Prognosis
459 research strategy (PROGRESS) 1: A framework for researching clinical outcomes. Bmj
460 2013;346:e5595.

461 11. Lamb SE, Marsh JL, Hutton JL, et al. Mechanical supports for acute, severe ankle sprain:
462 a pragmatic, multicentre, randomised controlled trial. *Lancet* 2009;373:575-81.

463 12. van Rijn RM, Willemsen SP, Verhagen AP, et al. Explanatory variables for adult
464 patients' self-reported recovery after acute lateral ankle sprain. *Phys Ther* 2011;91:77-84.

465 13. Wikstrom EA, Hubbard-Turner T, McKeon PO. Understanding and treating lateral
466 ankle sprains and their consequences: a constraints-based approach. *Sports Med*
467 2013;43:385-93.

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

470 15. Harrell Jr FE, Lee KL, Califf RM, et al. Regression modelling strategies for improved
471 prognostic prediction. *Stat Med* 1984;3:143-52.

472 16. Harrell Jr FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing
473 models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*
474 1996;15:361-87.

475 17. Peduzzi P, Concato J, Kemper E, et al. A simulation study of the number of events per
476 variable in logistic regression analysis. *J Clin Epidemiol* 1996;49:1373-9.

477 18. Moons KG, de Groot JA, Bouwmeester W, et al. Critical Appraisal and Data Extraction
478 for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. *PLoS med*
479 2014;11:e1001744.

480 19. Pavlou M, Ambler G, Seaman SR, et al. How to develop a more accurate risk prediction
481 model when there are few events. *BMJ* 2015;351:h3868.

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

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

22. Collins GS, Ogundimu EO, Cook JA, et al. Quantifying the impact of different approaches for handling continuous predictors on the performance of a prognostic model. *Stat Med* 2016;35:4124-35.

23. 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.

24. Royston P, Sauerbrei W. MFP: Multivariable Model-Building with Fractional Polynomials. In: Multivariable Model-Building: John Wiley & Sons, Ltd; 2008:115-50.

25. 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.

26. Morris TP, White IR, Carpenter JR, et al. Combining fractional polynomial model building with multiple imputation. *Stat Med* 2015;34:3298-317.

27. Marshall A, Altman DG, Holder RL, et al. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. *BMC Med Res Methodol* 2009;9:57.

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

29. Atkinson AC. A note on the generalized information criterion for choice of a model. *Biometrika* 1980;67:413-8.

30. van Houwelingen JC, LeCessie S. Predictive value of statistical models. *Stat Med* 1990;9:1303-25.

31. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006;26:565-74.

32. 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.

33. 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.

34. 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.

35. Janssen KJ, Moons KG, Kalkman CJ, et al. Updating methods improved the performance of a clinical prediction model in new patients. *J Clin Epidemiol* 2008;61:76-86.

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

37. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 2015;162:55-63.

38. O'Connor SR, Bleakley CM, Tully MA, McDonough SM. Predicting functional recovery after acute ankle sprain. PLoS One. 2013;8(Suppl 8):e72124.

39. Akacha M, Hutton JS, Lamb SE. Modelling treatment, age- and genderspecific recovery in acute injury studies. In: The University of Warwick Centre for Research in Statistical Methodology. 2010;10:11-12. http://www2.warwick.ac.uk/fac/sci/statistics/crism/research/2010/paper10-09. Accessed 17 Feb 2015.

40. 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. J Phys. 2012;58(Suppl 3):181–8.

41. 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. J Sci Med Sport. 2014;17(Suppl 1):23–8.

42. Pourkazemi F, Hiller CE, Raymond J, Nightingale EJ, Refshauge KM. Predictors of chronic ankle instability after an index lateral ankle sprain: a systematic review. J Sci Med Sport / Sports Med Aust. 2014;17(Suppl 6):568–73.

FIGURE LEGENDS

Figure 1. Decision curve analysis for the baseline and updated (baseline plus 4-weeks predictors) prognostic models.

Figure 2. Calibration plots for the baseline (left) and updated (right) SPRAINED prognostic models, overlaying calibration lines derived from the analyses of 50 imputed datasets.