Prognostic models for identifying risk of poor outcome in people with acute ankle sprains: the SPRAINED development and external validation study

David J Keene,[1]* Michael M Schlüssel,[1] Jacqueline Thompson,[1] Daryl A Hagan,[1] Mark A Williams,[2] Christopher Byrne,[3] Steve Goodacre,[4] Matthew Cooke,[5] Stephen Gwilym,[1,6] Philip 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, UK
[2]Department of Sport, Health Sciences and Social Work, Oxford Brookes University, Oxford, UK
[3]Faculty of Health and Human Sciences, University of Plymouth, Plymouth, UK
[4]School of Health and Related Research, University of Sheffield, Sheffield, UK
[5]Warwick Medical School, University of Warwick, Coventry, UK
[6]Emergency Department, Oxford University Hospitals NHS Foundation Trust, Oxford, UK
[7]Patient and public involvement representative

*Corresponding author david.keene@ndorms.ox.ac.uk

Declared competing interests of authors: Steve Goodacre is a member of the Health Technology Assessment (HTA) Clinical Trials Board, HTA Elective and Emergency Specialist Care 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 receive funding from the National Institute of Health Research (NIHR). She 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. Furthermore, she reports grants from the NIHR HTA programme during the conduct of this study.

Published November 2018
DOI: 10.3310/hta22640

Scientific summary

The SPRAINED development and external validation study

Health Technology Assessment 2018; Vol. 22: No. 64
DOI: 10.3310/hta22640

NIHR Journals Library www.journalslibrary.nihr.ac.uk

Scientific summary

Background

Ankle sprains are one of the most common musculoskeletal injuries. Although recovery can occur within weeks, up to one-third of patients still have problems with their ankle at 1 year post injury. 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 registered on the PROSPERO database: CRD42014014471. Electronic databases were searched [Allied and Complementary Database (AMED), EMBASE, PsycINFO, Cumulative Index to Nursing and Allied Health Literature (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 least one baseline prognostic factor were included. Eligibility assessments, data extraction and risk-of-bias assessments [using the Quality In Prognosis Studies (QUIPS) tool] were completed by two independent reviewers. A narrative synthesis was conducted.

Consensus meeting

A range of key stakeholders involved in ankle sprain care and research in the UK NHS, including patient and public representatives, health-care professionals and clinical researchers, were invited to a 1-day consensus meeting.

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

Development of the models

Data sources

Individual participant data from the existing Collaborative Ankle Support Trial (CAST) database were used to develop two prognostic models for poor outcome after ankle sprain. CAST was a pragmatic, multicentre, randomised controlled trial (RCT), with blinded assessment of the outcome, designed to estimate the clinical effectiveness and cost-effectiveness of three different types of mechanical ankle support [Aircast® ankle brace (DJO Incorporated, Vista, CA, USA), Bledsoe® boot (Bledsoe Boot Systems, Grand Prairie, TX, USA) 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 attending emergency departments (EDs) 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 ED and at 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 < 2 mm), 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 > 7 days before the first presentation at the recruiting ED. Participants were followed up at 1, 3 and 9 months after randomisation.

Candidate predictors

Twenty-three candidate predictor variables collected during the enrolment and baseline assessments of CAST were examined; all of these variables came under the following domains: 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 or not 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 post 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 post 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 9 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 model. 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 data sets created. Based on a logistic regression model, multivariable fractional polynomials (MFPs) were used to select variables and identify transformations of continuous variables that best predicted the outcome. Inclusion of predictors in the final models was based on a nominal alpha of 0.157 (equivalent to the Akaike information criterion) to reduce the risk of overfitting. 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 10 NHS EDs across England over a period of 9 months (July 2015–March 2016). There was no planned treatment allocation, as in a RCT, and EDs provided usual care in accordance with local protocols. Data collection took place at the time of a participant's presentation to any of the study recruiting sites (baseline) and subsequently at 4 weeks and 4 and 9 months after the initial injury. People aged ≥ 16 years with an acute ankle sprain (of < 7 days' duration) of any severity were invited to take part in the study. People with an ankle fracture (except a flake fracture of < 2 mm) 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 of bias and three studies at high risk of bias. Incomplete and/or inadequate reporting standards were a common issue; for example, it was difficult to determine if prognostic factors were eliminated because of statistical reasons or poor clinical utility. None of the 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 was investigated. The prognostic factors that were analysed varied considerably between studies, with no common framing across the studies. The identified studies and risk-of-bias assessments were summarised to those attending the consensus meeting.

Consensus meeting

The consensus meeting was attended by 30 participants. The final consensus voting identified 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 data set

The first model predicted the presence of persistent pain, functional difficulty or lack of confidence at 9 months and comprised age, body mass index, pain when resting, pain when bearing weight, number of days from injury to assessment, whether or not 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% confidence interval (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.

The second model predicted the presence of either persistent pain, functional difficulty, lack of confidence or recurrence of injury at 9 months and comprised pain when resting, pain when bearing weight, days from injury to assessment, ability to bear any weight on the injured ankle and whether or not 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.

Updating these models, which used baseline data collected at the ED, with an additional variable at 4 weeks after the injury (pain when bearing weight on the ankle), improved the predictions of the models when compared, using decision curve analysis plots.

A substudy to pilot dynamic consent recruited 22 participants in the later phase of the prospective cohort study. Eight participants accessed their dynamic consent online web page and none changed his/her consent decisions during the study.

Performance of the models in the external data set

Discrimination of the model for outcome 1 was similar to that observed in the development data set (combined *c*-statistic across the 50 imputed data sets = 0.73, 95% CI 0.66 to 0.79), but calibration was poor (combined calibration plot intercept = –0.91, 95% CI –1.18 to –0.65, and slope = 1.13, 95% CI 0.76 to 1.50). For the outcome 2 model, the combined *c*-statistic across the 50 imputed data sets was 0.63 (95% CI 0.58 to 0.69), the calibration plot intercept was –0.25 (95% CI –0.44 to –0.06) and the slope was 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 = –0.62, 95% CI –0.89 to –0.34, 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), the calibration plot intercept was 0.12 (95% CI –0.07 to –0.32) and slope was 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 grade 2 or 3). All models were recalibrated (i.e. had their regression coefficients and intercepts re-estimated) using the external validation data set.

A substudy to pilot dynamic consent recruited 22 participants in the later phase of the prospective cohort study. Eight participants accessed their dynamic consent online web page and none changed his/her consent decisions during the study.

Conclusions

Both models and their 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 a prospectively collected external validation cohort. The models predicting presence of persistent pain, functional difficulty, lack of confidence or recurrence of injury showed good calibration, whereas 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 health care

The SPRAINED (Synthesising a clinical Prognostic Rule for Ankle Injuries in the Emergency Department) study prognostic models performed reasonably well and showed benefit when compared with not using any model (i.e. consider all patients to be at a high risk of poor outcome); therefore, the models may assist clinical decision-making when assessing and advising people with ankle sprains in the ED setting and when deciding on ongoing 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 is recommended. Further refinement of the models, including external validation of the recalibrated 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 ED staff, and their impact on patient outcomes, should be investigated.

Trial registration

This trial is registered as ISRCTN12726986.

Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research. Funding was also recieved from the NIHR Collaboration for Leadership in Applied Health Research, Care Oxford at Oxford Health NHS Foundation Trust, NIHR Biomedical Research Centre, Oxford and the NIHR Fellowship programme.

Health Technology Assessment

HTA/HTA TAR

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 4.513

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the Clarivate Analytics Science Citation Index.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: http://www.nets.nihr.ac.uk/programmes/hta

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 13/19/06. The contractual start date was in November 2014. The draft report began editorial review in August 2017 and was accepted for publication in February 2018. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).