**PROBAST**

Study:

Digital health technologies and machine learning augment patient reported outcomes to remotely characterise rheumatoid arthritis

Step 2: Type of prediction study

**Is the study a diagnostic or a prognostic study?**

Diagnostic

**Is the study a development only, development and validation or validation only study?**

Development only

**What is the model of interest?**

Logistic regression model

**What is the outcome of interest?**

RA disease detection, RA disease severity estimation

Step 3: Assess risk of bias

**Domain 1: Participants**

**Describe the sources of data and criteria for participant selection**

Digital wearable devices—a wrist-worn Apple Watch for passive monitoring and an iPhone, integrated with a bespoke mobile app. which prescribed daily guided assessments—collected high-frequency, objective sensor data in 30 RA patients and 30matched Healthy Controls (HCs).

**1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?**

Y

**1.2 Were all inclusions and exclusions of participants appropriate?**

Y

**Risk of bias introduced by selection of participants:**

**Low**

**Rationale of bias rating**

Case control useful for differentiating RA vs healthy control. No further eligibility criteria given that may limit generalizability.

**Domain 2: Predictors**

**List and describe predictors included in the final model, e.g. definition and timing of assessment**

The daily physical activity of RA participants and healthy controls were estimated with a deep convolutional neural network (DCNN)that was first pre-trained on 100,000 participants in the publicly available UK Biobank, following a multi-task self-supervised learning (SSL) methodology27, which was subsequently fine-tuned on the free-living Capture-24 dataset 28 of < 150 participants to determine broad activity patterns of interest {sleep, sedentary, light physical activity, moderate-to-vigorous physical activity(MVPA)}29,30 and fine-grained activity prediction labels {sleep, sitting/standing, mixed, vehicle, walking, bicycling}

**2.1 Were predictors defined and assessed in a similar way for all participants?**

Y

**2.2 Were predictor assessments made without knowledge of outcome data?**

Y

**2.3 Are all predictors available at the time the model intended to be used?**

Y

**Risk of bias introduced by predictors or their assessment**

Low

**Rationale of bias rating**

Predictors are easy to implement and independent of outcome.

**Domain 3: Outcome**

**Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination:**

1. Distinguish RA status: prior diagnosis
2. Estimate severity levels: , participants were denoted as having moderate or severe RA based on baseline clinician-assessedRAPID-3 score

**3.1 Was the outcome determined appropriately?**

Y

**3.2 Was a pre-specified or standard outcome definition used?**

Y

**3.3 Were predictors excluded from the outcome definition?**

Y

**3.4 Was the outcome defined and determined in a similar way for all participants?**

Y

**3.5 Was the outcome determined without knowledge of predictor information?**

Y

**3.6 Was the time interval between predictor assessment and outcome determination appropriate?**

Y

**Risk of bias introduced by the outcome or its determination**

Low

**Rationale of bias rating**

Diagnosis or gold standard test were taken for outcome labeling.

**Domain 4: Analysis**

**Describe number of participants, number of candidate predictors, outcome events and events per candidate predictor**

Two RA participants withdrew immediately after enroling in the study. Data from these participants were not collected, leaving 28 RA participants, 28 matched HCs, and 2 unmatched HCs for a total of58 participant

**Describe how the model was developed, predictor selection and risk group definition**

**Describe whether and how the model was validated, either internally (cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants)**

To determine the generalisability of our models, a stratified subject-wise k-fold cross-validation (CV) was employed. This consisted of randomly partitioning the dataset into k=5 folds, which was stratified with equal class proportions where possible. Participant data remained independent between training, validation, and testing splits.

**Describe the performance measures of the model, e.g. calibration, discrimination, classification, net benefit, and whether they were adjusted for optimism**

CK, F1, R2, MAE, RMSE

**Describe any participants who were excluded from the analysis**

Two RA participants withdrew immediately after enroling in the study. Data from these participants were not collected, leaving 28 RA participants, 28 matched HCs, and 2 unmatched HCs for a total of58 participant

**Describe missing data on predictors and outcomes as well as methods used for missing data**

Not described

**4.1 Were there a reasonable number of participants with the outcome?**

N

**4.2 Were continuous and categorical predictors handled appropriately?**

Y

**4.3 Were all enrolled participants included in the analysis?**

N

**4.4 Were participants with missing data handled appropriately?**

U

**4.5 Was selection of predictors based on univariable analysis avoided?**

Y

**4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls)**

**accounted for appropriately?**

Y

**4.7 Were relevant model performance measures evaluated appropriately?**

Y

**4.8 Were model overfitting and optimism in model performance accounted for?**

Y

**4.9 Do predictors and their assigned weights in the final model correspond to the results**

**from multivariable analysis?**

Y

**Risk of bias introduced by the analysis**

High

**Rationale of bias rating**

Small amount of outcomes

**Overall Risk of bias**

High