**PROBAST**

Study:

Evaluation of walking activity and gait to identify physical and mental fatigue in neurodegenerative and immune disorders: preliminary insights from the IDEA-FAST feasibility study.

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?**

General linear mixed model

**What is the outcome of interest?**

Patient fatigue

Step 3: Assess risk of bias

**Domain 1: Participants**

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

Participants with IMIDs and NDDs (Parkinson’s disease (PD), Huntington’s disease (HD), rheumatoid arthritis (RA), systemic lupus erythematosus (SLE), primary Sjogren’s syndrome (PSS), and inflammatory bowel disease (IBD)) wore a lower‑back IMU continuously for up to 10 days at home.

The inclusion criteria ensured the participants were over 18 and were willing and able to participate. The exclusion criteria ensured participants had no diagnoses or symptoms relating to their disorder that could interfere with the aims of the study (e.g., sleep disorder, chronic fatigue, etc.)

**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**

Reasonable eligibility criteria.

**Domain 2: Predictors**

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

The current study analysed data from a lower-back IMU approximately at the level of the L5 vertebra for two periods of five consecutive days in a free-living environment. The IMU was the McRoberts [41] Dynaport device, from which only the triaxial accelerometer data was used, with a sampling rate of 100 Hz and a range of ± 8 g (1 g is equivalent to 9.81 m/s2).

**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**

Accelerometer data are independent and assessed the same way for everyone.

**Domain 3: Outcome**

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

Whilst wearing the device, participants were asked to complete short PROs up to four times a day: morning (09:00–12:00), early afternoon (13:00–16:00), late afternoon (17:00–20:00), and evening (21:00–24:00) on a smart phone provided. This included two questions asking how the participants were feeling with regards to their physical and mental fatigue (Likert items on a scale of 0–6).

For the intersubject method, the PROs were bina-rised by setting 0–2 as low fatigue and 3–6 as high fatigue [64]. For the intrasubject method, the PROs were bina-rised to low or high after splitting the data into the train-ing/testing sets (five-fold CV) by setting threshold for binarisation as the mean of the training data.

**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**

Standard questionnaire was used to determine patient patient fatigue.

**Domain 4: Analysis**

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

71 patients. Amount of outcomes unclear but > 1000.

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

The relationships between the PRO scores and the digital measures gait described above were assessed using a generalised linear mixed effect model (GLMM) and machine learning classifiers.

**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)**

two methods of splitting the data were used: intersubject cross validation (CV) (five-fold leave-subjects-out CV where 20% of the subjects were excluded as the test data) and intrasubject CV (five-fold cross validation within the data of each individual subject).

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

Correlation coefficient

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

Therefore, participants whose entire data were excluded based on these criteria were also excluded from the analysis, giving 72 participants. Initially 131.

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

For both methods, any missing features (macros: 6.5% of variability of the bout lengths; and micros: 0.5% of the step length and step velocity and 10.8% of the step length asymmetry) were replaced with the median of the training data.

one or two PROs from three participants had to be excluded since there was not enough range in the PRO scores to create two classes

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

Y

**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?**

Y

**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?**

PN

**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**

Almost half of patients excluded from analysis due to incomplete data. Also insufficient reporting of metrics for imbalanced classes.

**Overall Risk of bias**

High