**PROBAST**

Study:

Real-Life Gait Performance as a Digital Biomarker for Motor Fluctuations: The Parkinson@Home Validation 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?**

Logistic LASSO regression

**What is the outcome of interest?**

Distinguish pre from post medication state. Classify gait between healthy and PD.

Step 3: Assess risk of bias

**Domain 1: Participants**

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

We included a group of 25 patients with PD and 25 age-matched participants without PD (controls). Patients were recruited using various strategies, including advertisements in the Dutch Parkinson Patient Association’s magazine and on social media,visits to support groups, and through physiotherapists specialized in the treatment of PD. Controls were recruited from partners and acquaintances of the participating patients and by advertisements on social media. The inclusion criteria were (1)aged ≥30 years, (2) in possession of a smartphone running on Android 4.4 or higher, and (3) living within travelling distance from the study center. Additional inclusion criteria for the PD group were (1) Parkinson disease diagnosed by a neurologist,(2) currently using levodopa and/or a dopamine agonist, (3)experiencing at least slight motor fluctuations (Movement Disorders Society Unified Parkinson’s Disease Rating Scale[MDS-UPDRS] part IV item 4.3 ≥1), and (4) experiencing at least some Parkinson-related gait impairments (MDS-UPDRSpart II item 2.12 ≥1 and/or item 2.13 ≥1). The exclusion criteria were (1) any type of advanced treatment (deep brain stimulation or intestinal levodopa or apomorphine infusion) and (2)psychiatric or cognitive impairments that may hinder successful completion of the study protocol (based on judgement of the assessor running the recruitment). We did not exclude patients with PD or controls who used assistive devices or reported othermedical problems affecting their movements.

**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 case control format. Reasonable eligibility criteria.

**Domain 2: Predictors**

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

Accelerometer data from multiple wearable devices.

The video annotations were used to locate periods of gait(defined as five or more consecutive steps) during the free-living parts. From these, we extracted nonoverlapping gait segments of equal length (3000 samples, corresponding to 25 seconds). This length was selected because prior research showed that using shorter free-living gait segments discriminated less well between patients with PD and controls [19] and in order to achieve sufficient resolution in the frequency domain. To be included in the analyses, participants needed to have at least 10 gait segments of 25 seconds. In addition, patients were required to have at least 5 segments before and 5 segments aftermedication intake.

**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?**

PN

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

High

**Rationale of bias rating**

Features were used from annotated gait segments from video. This is hard to implement in an application setting.

**Domain 3: Outcome**

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

The clinical assessments were also conducted in the participants’ homes and included the TUG test, the Abnormal Involuntary Movement Scale, and the complete MDS-UPDRS, except for the self-reported items of part I and II, which were completed through an online survey after the visit.

we constructed continuous scores that could serve as digital biomarkers for the response to medication intake (premedication/post medication classifier) and PD gait impairment severity (PD/control classifier).

**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**

Both outcomes are independent of predictors and defined similarly for all patients.

**Domain 4: Analysis**

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

A minimum of 10 gait segments of 25 seconds were available in 18 patients with PD (median 46.5 segments, range 14-95) and 24 controls.

Amount of outcomes unclear for pre-post medication classification.

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

Next, we evaluated whether combinations of features could be used to predict whether a gait segment occurred before or after medication intake. For this, we used logistic LASSO (least absolute shrinkage and selection operator) regression with uniform prior class probabilities. To account for the varying number of gait segments per patient, we weighted each gait segment by the inverse of the number of gait segments per patient.

In addition, we evaluated the performance of logistic LASSO regression to predict whether a gait segment was from a patient with PD(premedication) or control. We used nested cross-validation forthis as well, leaving 1 patient and 1 control out in each fold.

**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)**

We evaluated the performance using leave-one-subject-out nested cross-validation (CV), with the LASSO regularization hyperparameter being selected in the inner CV loops.

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

AUC, ACC

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

Reasons for collecting an insufficient number of gait segments included rainy weather (n=4), a desire not to be filmed in the neighborhood (n=1), use of a wheelchair for longer distances (n=1), fatigue (n=1), and technical problems with the video recordings (n=1). The included patients did not differ substantially from the excluded patients in terms of disease severity. In addition, technical problems caused data loss for 1 ankle sensor in 1 participant; this participant (control) was excluded from the analyses combining multiple sensor locations

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

First, data were interpolated to a uniform sample rate of 120 Hz using piecewise cubic interpolation.

In addition, technical problems caused data loss for 1 ankle sensor in 1 participant; this participant (control) was excluded from the analyses combining multiple sensor locations

**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?**

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?**

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?**

U

**Risk of bias introduced by the analysis**

High

**Rationale of bias rating**

Low / unknown amount of outcomes. Eight patients were excluded for various reasons which may have influenced performance.

**Overall Risk of bias**

High