Supplementary Material

METHODS

Participants

The parent study was powered to examine the effects of an intervention on FOG. Power analyses showed that 62 participants would be sufficient[1]. The present analysis, based on about 50% of the participants, was not explicitly powered and was a convenient sample.

Self-Report of Questionnaires

The Characterizing Freezing of Gait Questionnaire (CFOG-Q) contains 35 self-report items, divided into four sections: 1. The presence, severity, and responsiveness of FOG to treatment. 2. Triggers of FOG. 3. The effectiveness of strategies to reduce or overcome FOG 4. The presence and severity of other types of freezing like upper limb freezing[2]. For the present analyses, part four (other types of freezing) was not included. The CFOG-Q was developed to quantify heterogeneity in freezing behavior, rather than for quantifying the severity of FOG[3].

The New Freezing of Gait Questionnaire (NFOG-Q) consists of three parts. The first part distinguishes between freezers and non-freezers, while the second part assesses the freezing severity based on the frequency and duration of FOG episodes during gait initiation and turning. The third part evaluates the impact of FOG on daily life[4]. It is an easy and quick way to differentiate between freezers and non-freezers, but it is considered insufficiently reliable to detect small changes in FOG severity[5].

Unsupervised Daily-Living Monitoring

The unsupervised daily-living activities were also monitored for seven days using a small activity monitor. The accelerometer was covered with a hypafix bandage and attached with a hydrogel adhesive positioned on the participant's lower back at the level of the fifth lumbar vertebra. The accelerometer (AX3, Newcastle, United Kingdom) is recorded with a sample frequency of 100 Hz. A previously described algorithm automatically identifies the different activities throughout the day such as walking, lying, standing and sitting[1,6].

Supervised In-Home Testing

To characterize the study cohort in the home setting, disease severity was evaluated using the Hoehn and Yahr staging scale[7] and the Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS)[8]. The MDS-UPDRS part 3 (motor functions) was evaluated OFF and ON dopaminergic medication. Quality of life was assessed using PD Questionnaire[9] and real-life functional mobility with the Life-Space Assessment[10]. To evaluate fear and fall risk, the Parkinson Anxiety Scale[11] and Falls Efficacy Scale International[12] were used. Cognitive function was assessed with the Mini-Mental State Examination (26-item version)[13] and Montreal Cognitive Assessment[14]. The Mini-Balance Evaluation Systems Test[15] assessed postural control (Table 1).

Signal Processing

First, signal windows were processed similarly during the online procedure. Next, windows identified as FOG (FOG-windows) were put together in a block if the distance between FOG-windows was up to 3 windows (thus, considering the 0.5 steps, FOG-windows could be maximally 1.5 seconds apart to be considered in the same block). Then the duration of these blocks was calculated as the number of windows in the block multiplied by 0.5 seconds (the step of overlapping windows). Blocks that were at least one second long (2 consecutive windows) were considered as probable FOG events, with the corresponding duration.

Technical Validation of the %TF in Daily Living Obtained With the DeFOG Algorithm

Sensor data and corresponding video-based labels of FOG presence from three datasets were used to provide a first technical validation (comparison with a gold standard, which is video-rating of FOG events) of the %TF obtained with the DeFOG FOG-detection algorithm.

The three datasets were:

1) *CuPiD* dataset[16]: 18 participants performing a FOG-provoking protocol, including several tasks in a laboratory setting (such as Figure of 8, passing through a door, straight walking, 180-degree and 360-degree turns while walking in circles) and a hospital tour.

2) *Leuven DeFOG pilot* dataset: 2 participants performing a FOG provoking protocol in a laboratory setting with tasks such as 360 turns in place, zig-zag walking past obstacles, passing through doors, moving a chair, turning around the chair, and sitting down.

3) *DeFOG FOG provoking* dataset in Tel Aviv and Leuven's sites[1]: 11 participants performing the FOG-provoking protocol during the assessments in the home setting at T1 (baseline) and T2 (follow-up) in two different conditions (OFF and ON medication). We considered the data from a specific person in a specific combination of time and condition as a single unit when averaging the results (e.g., the data from participant 1 in T1 in the OFF condition is one unit

and the data from participant 1 in T1 and ON condition is a different unit (as participant 1 in T2 in the OFF (or ON) condition). So, up to 4 units can be related to the same participant. The final total of units was 33 for this dataset (for some participants, not all the assessment/condition combinations were present due to technical or other issues).

By considering these datasets, there is a total of 53 units (a unit is a subject for the first two datasets and a combination of subject, assessment, and condition in the third dataset).

For each unit, we computed the total %TF based on the DeFOG algorithm, considering all the tasks related to that specific unit (i.e., the sum of all the FOG events durations identified in all tasks divided by the sum of the duration of the recorded signals in all tasks). Then, we compared this to the %TF similarly obtained by video ratings.

The first preliminary analysis was a correlation analysis, which showed, as expected, a significant positive correlation between $\%TF_{DeFOG}$ and $\%TF_{video}$ is 0.52 ($p \leq 0.001$).

The other two evaluation methods, were the analysis of agreement (Bland-Altman plot) and reliability (IntraClass Correlation Coefficient, ICC).

In the first, the Bland-Altman plot (see Supplementary Figure 2), reported below, shows a low bias (~1%) and a minimal detectable change around 13%. As a note, it can be seen that for low values of %TF, it seems that there are smaller differences (the algorithm works better).

In the second, we found a moderate absolute and consistency ICC ($0.4<ICC<0.7$), with $ICC_{consistency}= 0.4703$ and $ICC_{absolute}=0.4649$.

These results provide a first validation of the %TF obtained from the DeFOG algorithm. As a limitation, the datasets used for validation are mainly made of structured or semi-structured (e.g. hospital tour) tests, which cannot fully mimic the unsupervised daily activities of people. Also, the first two datasets are related to data collected in a laboratory setting and transferability of results between the laboratory and home/real-life might not be guaranteed. These limitations are inherent to the necessity of having video-based labels to use as the gold standard for comparison and are common to all current state-of-the art tools.

When analyzing the presented results, one should also consider that the DeFOG algorithm was designed for daily living and therefore was not optimized or tuned for specific tasks that are present in these datasets but may be very infrequent in daily life (e.g., full 360 degree turns).

Statistical Analyses

$$\%\text{TF single task} = \frac{\text{time spent frozen during the task x 100}}{\text{duration task}}$$

$$\%\text{TF provoking protocol (all tasks)} = \frac{\text{time spent frozen during all tasks x 100}}{\text{duration all tasks}}$$

$$\text{FOG episodes per hour} = \frac{\text{total number of FOG episodes}}{\text{observation time}}$$

References

1 Zoetewei D, Herman T, Brozgol M *et al.* Protocol for the DeFOG trial: A randomized controlled trial on the effects of smartphone-based, on-demand cueing for freezing of gait in Parkinson's disease. *Contemporary Clinical Trials Communications* 2021:100817.

2 Ehgoetz Martens KA. Development of the Characterizing Freezing of Gait Questionnaire [Questionnaire]; 2018 2018.

3 Ehgoetz Martens KA, Shine JM, Walton CC *et al.* Evidence for subtypes of freezing of gait in Parkinson's disease. *Mov Disord* 2018;**33(7)**:1174–8.

4 Herman T, Dagan M, Shema-Shiratzky S *et al.* Advantages of timing the duration of a freezing of gait-provoking test in individuals with Parkinson's disease. *J Neurol* 2020.

5 Hulzinga F, Nieuwboer A, Dijkstra BW *et al.* The New Freezing of Gait Questionnaire: Unsuitable as an Outcome in Clinical Trials? *Mov Disord Clin Pract* 2020;**7(2)**:199–205.

6 Galperin I, Hillel I, Del Din S *et al.* Associations between daily-living physical activity and laboratory-based assessments of motor severity in patients with falls and Parkinson's disease. *Parkinsonism Relat Disord* 2019;**62**:85–90.

7 Goetz CG, Poewe W, Rascol O *et al.* Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: status and recommendations. *Mov Disord* 2004;**19(9)**:1020–8.

8 Goetz CG, Tilley BC, Shaftman SR *et al.* Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Mov Disord* 2008;**23(15)**:2129–70.

9 Jenkinson C, Fitzpatrick R, Peto V, Greenhall R, Hyman N. The PDQ-8: Development and validation of a short-form parkinson's disease questionnaire. *Psychology & Health* 1997;**12(6)**:805–14.

10 Baker PS, Bodner EV, Allman RM (2003) Measuring life-space mobility in community-dwelling older adults.

11 Leentjens AFG, Dujardin K, Pontone GM, Starkstein SE, Weintraub D, Martinez-Martin P. The Parkinson Anxiety Scale (PAS): development and validation of a new anxiety scale. *Mov Disord* 2014;**29(8)**:1035–43.

12 Kempen GIJM, Yardley L, van Haastregt JCM *et al.* The Short FES-I: a shortened version of the falls efficacy scale-international to assess fear of falling. *Age Ageing* 2008;**37(1)**:45–50.

13 Newkirk LA, Kim JM, Thompson JM, Tinklenberg JR, Yesavage JA, Taylor JL. Validation of a 26-point telephone version of the Mini-Mental State Examination. *J Geriatr Psychiatry Neurol* 2004;**17(2)**:81–7.

14 Nasreddine ZS, Phillips NA, Bédirian V *et al.* The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc* 2005;**53(4)**:695–9.

15 Franchignoni F, Horak F, Godi M, Nardone A, Giordano A. Using psychometric techniques to improve the Balance Evaluation Systems Test: the mini-BESTest. *J Rehabil Med* 2010;**42(4)**:323–31.

16 Palmerini L, Rocchi L, Mazilu S, Gazit E, Hausdorff JM, Chiari L. Identification of Characteristic Motor Patterns Preceding Freezing of Gait in Parkinson's Disease Using Wearable Sensors. *Front Neurol* 2017;**8**:394.

17 Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J Neurol Neurosurg Psychiatry* 1992;**55(3)**:181–4.