developed to prepare for Phase 3, which is currently validating the ClinRO. Training included an online tutorial that described the clinical FOG definition, scoring explanation, and representative FOG severity examples. Raters independently scored examples, and discussion with experts refined their approach. Raters were certified using a test with curated videos of various FOG severities. Training evaluation provided feedback on the burden, confidence, difficulty, and feasibility of ClinRO use. Pilot and training scores were evaluated for inter-rater reliability using intraclass correlation coefficients (ICCs). Results: The ClinRO was designed to maximally provoke and accurately rate FOG severity for use in clinical trials to reduce FOG. As such, the tool needed to take under 10 minutes to complete, with minimal setup, is scored live and is suitable for clinical settings. The 13 rater's ICC in Phase 2 ICC was 0.81 (95% CI: 0.72-0.88). The 34 raters in Phase 3 (clinicians & researchers) across the 7 sites completed the ClinRO training program. Experience with PwP and FOG varied. 35% of raters had less than 2 yrs of experience working with PwP, and 44% observed FOG in < 20 PwP. Of the group, 25 (78%) passed the certification testing on the first attempt. Failed attempts only had an error of 1.25% greater than the acceptable score range and all passed on the second attempt. ICC of ClinRO scores was 0.925 (95% CI: 0.868-0.967; $F_{15,472} = 460$, $p<0.001$), showing excellent inter-rater reliability in the trained raters. Feedback from raters revealed appropriate burden, the implementation somewhat to very feasible and the confidence somewhat to very high after training. Conclusion: The training of ClinRO scoring improved the inter-rater reliability of the ClinRO and it was beneficial to align FOG severity scoring even in a relatively inexperienced group of raters. Rater feedback provided useful insights for the ongoing validation of the FOG ClinRO.

P03-N-98 - Can a deep learning algorithm assess freezing of gait during a turning task in people with Parkinson's disease?

Goris Maaike [1], Po-Kai Yang [2], Pieter Ginis [2], Benjamin Filtjens [3], Clint Hansen [4], Christian Schlenstedt [5], Jeffrey Hausdorff [6], Walter Maetzler [7], Wim Vandenberghe [8], Bart Vanrumste [9], Alice Nieuwboer [2], Moran Gilat [2]

[1] Ku Leuven, [2] Katholieke Universiteit Leuven, [3] University Health Network, [4] Kiel University, [5] Medical School Hamburg, [6] Tel Aviv University; Sourasky Medical Center, [7] University Hospital Schleswig-Holstein, [8] University Hospitals Leuven, [9] KU Leuven

Background: Freezing of gait (FOG) is a disabling symptom of Parkinson's disease (PD), often triggered by turning tasks such as one-minute 360° dual-task turning (360 turn). Accurately classifying patients as freezers and assessing their FOG severity during this task is proven challenging and algorithms based on threshold-based methods show limited accuracy for FOG detection. Deep learning (DL) algorithms have been developed that offer potential for improved FOG detection. This study aims to evaluate the accuracy of a wearable sensor-based DL algorithm to classify freezers and assess its validity to quantify FOG severity.Methods: 154 people with PD (Age:63.21±9.71 years, Disease duration:7.81±4.9 years, MDS-UPDRS-III:30.27±12.50) performed the 360 turn in their

ON medication state while wearing IMU sensors at the feet, shins, and lower back. Self-reported FOG was determined via the NFOGQ. Expert FOG assessment was based on post-hoc video analysis by two independent raters. The Algorithm was trained to detect FOG on another dataset of 29 people with PD with severe FOG and 9 without FOG, performing several gait tasks and voluntary stops (1). Afterwards, it was tested on the 154 included participants. Agreement of the algorithm's classification (FOG/non-FOG) with expert ratings and self-reports was assessed using Cohen's kappa statistics. Absolute agreement in percentage time frozen (%TF) scored by experts versus algorithm-extracted %TF was evaluated using the intraclass correlation coefficient (ICC). ROC curves were used to evaluate sensitivity and specificity of the DL algorithm.Results: The algorithm showed moderate agreement with expert ratings for FOG classification (κ=0.428, p<0.001) and weak agreement with self-reports (κ=0.112, p=0.036). For all participants, the ICC between expert and algorithm for %TF was 0.628 (95% CI=0.522–0.715, p<0.001). Among the 26 self-reported freezers, the ICC for %TF was 0.591 (95% CI=0.280–0.790, p<0.001), and among the 98 clinically rated freezers, it was 0.576 (95%CI=0.427–0.694, p<0.001). The algorithm provided a sensitivity of 65.3% and specificity of 80.4% when compared to expert FOG classification (AUC=0.78, p<0.001, Threshold=1.24). Against self-report, the algorithm achieved a sensitivity of 57.5% and specificity of 83.6% (AUC=0.70, p=0.001, cutoff=5.59). Examples of agreement in Figure 1. Conclusion: Automated FOG classification remains a challenge, with moderate agreement of the DL algorithm with clinical ratings and poor agreement with self-reports. However, when considering FOG severity as a continuum and expressed as %TF, the model's performance improved. The study participants ranged from non-FOG to severe FOG, while the model was trained mostly on severe FOG cases. This may explain the lower performance for less severe cases compared to previous work. For clinical implementation, it is crucial that DL algorithms perform effectively across the full spectrum of freezing severity.Funding: Mobilise-D Funding from the Innovative Medicines Initiative 2 JointUndertaking (JU) under grant agreement No. 820820. KU Leuven internally funded FOG-ITproject (C3/20/109). M. Goris is supported by a fellowship of the Research Foundation Flanders (FWO) (1SHEK24N). The other authors do not have funding to declare.1. 1. Yang PK, Filtjens B, Ginis P, Goris M, Nieuwboer A, Gilat M, et al. Freezing of gait assessment with inertial measurement units and deep learning: effect of tasks, medication states, and stops. J Neuroeng Rehabil. 2024 Dec 1;21(1).

P03-N-99 - Turning performance in Parkinson's disease: classification of prospective fallers and freezers using wearable inertial sensors

Sonia Pianzi [1], Luigi Vinciguerra [1], Ilaria D'ascanio [1], Rana Zia UR Rehman [2], Riona Mc Ardle [2], Silvia Del Din [2], Brook Galna [3], Thomas Alan [2], Lynn Rochester [2], Jose Albites-Sanabria [1], Pierpaolo Palumbo [1], Stefania Bandinelli [4], Lorenzo Chiari [1], Luca

[1] University of Bologna, [2] Newcastle University, [3] Murdoch University, [4] Azienda Sanitaria Toscana Centro