Reliability of the new freezing of gait questionnaire: Agreement between patients with Parkinson’s disease and their carers

Alice Nieuwboera,*, Lynn Rochesterb, Talia Hermanc, Wim Vandenberghed, George Ehab Emil a, Tom Thomaesa, Nir Giladici

a Department of Rehabilitation Sciences, Katholieke Universiteit, Leuven, Belgium
b Institute for Ageing and Health, Clinical Ageing Research Unit, Newcastle University, United Kingdom
c Movement Disorders Unit, Parkinson Center, Sourasky Medical Centre Tel-Aviv, Sackler School of Medicine, Tel-Aviv University, Tel-Aviv, Israel
d Department of Neurosciences, Katholieke Universiteit, Leuven, Belgium

ARTICLE INFO

Article history:
Received 4 February 2009
Received in revised form 19 June 2009
Accepted 2 July 2009

Keywords:
Gait
Parkinson’s disease
Freezing
Validity
Reliability

ABSTRACT

Freezing of gait (FOG) is difficult to measure due to its unpredictable occurrence. This study investigated: (1) whether the new freezing of gait questionnaire (NFOG-Q) is a reliable measure of freezing by comparing patients’ ratings with those of carers’ and (2) whether adding a video improved its reliability. Non-demented people with Parkinson’s disease (PD) (N = 102) and their carers of similar age and cognitive status were recruited from movement disorders clinics in three countries. The NFOG-Q was administered to carers and patients independently before and after watching a video showing several examples of FOG. Patients had very high agreement between their pre- and post-video detection of FOG (Kappa = 0.91). However, this was less than in carers (Kappa = 0.79). The video had a significant influence (p = 0.01) on the rating of FOG severity (duration) but not on the estimation of its functional impact. Post-video freezing severity scores in the 69 freezers showed high agreement with carers’ scores (ICC = 0.78 [0.65;0.87]). We conclude that the NFOG-Q is a reliable tool to detect and evaluate the impact and severity of FOG. Adding a video does not add to the sensitivity and specificity of FOG detection but influences the estimation of FOG severity.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Freezing of gait (FOG) has recently been defined as an episodic inability to generate effective stepping in the absence of any known cause other than Parkinsonism or high level gait disorders [1]. FOG is a very disabling motor symptom of advanced Parkinson’s disease (PD) [2–4] with a significant impact on fall risk [5] and quality of life [6]. Accurate detection and rating of both the severity and impact of FOG is therefore important [7,8].

A gold standard measure of FOG is presently lacking. Its unpredictable presentation and sensitivity to medication, cues and heightened attention make it hard to trigger FOG in a clinic or research laboratory [1]. High false negative rates are likely even during the off-phase and when using specially designed gait trajectories [7]. Activity monitoring to measure FOG has shown promising results during standardized test situations in off [8], but high false positive rates may occur in the home situation. Therefore, a questionnaire has a crucial role to play in the measurement and detection of FOG.

Giladi et al. [9] developed the freezing of gait questionnaire (FOG-Q), a six-item scale (range 0–24) consisting of four items which assess FOG severity and two items which assess gait difficulties in general. The FOG-Q has satisfactory test-retest reliability, internal consistency and moderately high correlations with UPDRS-motor and ADL-scores [9,10]. Its acknowledged drawback, however, is the inclusion of general gait items, thereby reducing its FOG-specificity [10]. The recently revised MDS-UPDRS [11] includes two items on FOG (items 2.13 and 3.11). However, scoring options are based on more than one dimension, confounding an evaluation of FOG features in isolation.
4. Discussion

This study has shown the reliability of the new freezing of gait questionnaire by direct comparison of the ratings of non-demented patients with PD and their carers. Watching a teaching video altered the recognition of FOG in 4/102 patients and 9/102 carers. This suggests that patients’ self-detection may be more reliable than observation by a lay-person. However, as the true occurrence of FOG was unknown, alternative explanations may also apply. For instance, patients themselves may not have been willing to admit to their symptoms or change their opinion. Adding the video did not improve agreement between patients and carers in identifying FOG. Carers classified patients more often as freezers than patients themselves, generating a 20% false positive rate. Watching the video reduced the false positive rate only mildly to 18%. This maintained discrepancy confirms the difficulty of detecting FOG [7,10].

While the video had a small impact on the accuracy of FOG-detection, the overall evaluation of FOG was rated with increased severity after watching it. The video vignettes showed real time estimates of FOG episodes and therefore it is not surprising that part II scores, assessing FOG frequency and duration, were significantly affected. The tendency to spontaneously underestimate the severity of FOG is largely in line with a recent study, showing that patients generally underate their problems compared with objective ADL performance [21]. Severely affected FRs may have become adapted and partly desensitized to its features and are therefore most likely to underestimate FOG.

On the basis of these results and given the possible impracticability of watching a video in a doctor’s clinic, its use is not recommended for routine clinical assessment. However, for accurate measurement, the video adds surplus value by calibrating patients’ evaluation of the frequency and duration of FOG-episodes. Also, in cases where patients themselves are unable to reliably identify FOG, the video may provide important visual clarification to carers.

Overall this study supports the reliability and the validity of the NFOG-Q, as it produced consistent results in both patients and carers before and after the video. Patients’ and carers’ ratings did not indicate a between-group bias and showed good inter-rater reliability. This finding reiterates an earlier study in which high agreement was found between patients, carers and neurologists on the freezing of gait item of the UPDRS-part II [22]. Between-group variability was larger when patients had less severe FOG, indicating greater difficulty when FOG is only a mild symptom.

The lack of a gold standard measure of FOG is the largest drawback of examining the validity of any FOG measure at the present time. Objective measurement or clinical observation of FOG could be carried out during complex gait tests with multiple turns and obstacles to elicit the symptom in the off-period [7,8,12]. While this method can confirm that a probable FR actually has the symptom [7], it might equally miss FOG in patients who do freeze frequently at home. Moreover, one-off assessments do not reflect true FOG severity. Trigger sensitivity, responsiveness to levodopa and ability to overcome the tendency to freeze are factors that may fluctuate, pointing to the need for ecologically valid measures involving 24 h registration in unobserved conditions. Although activity monitoring can provide reliable step parameters in controlled conditions [23], detection of FOG-episodes at home needs further development [8]. In the absence of these validated tools, carers’ continuous observation may provide an acceptable substitute.

The NFOG-Q was not directly compared with the FOG-Q due to the overlap between the scales and the limited number of repeated measures possible in one session. Comparison between FRs and NFRs revealed that in line with other studies, FRs had more severe disease, longer disease duration and more frequent falling [2–5]. Unlike the FOG-Q [10,11] no correlation was found between the UPDRS-III and NFOG-Q scores. The fact that the NFOG-Q offered a freezing-specific evaluation referring to both on and off states and that the UPDRS-III was administered during on, may have contributed to this finding. FOG is generally considered an independent disease feature with a more complex pathophysiology than other symptoms [24–26] and is not correlated with bradykinesia [24].

Similar to the previous FOGQ [10], principal component analysis revealed that the NFOG-items strongly relate to each other, likely representing the overall severity of FOG. The items had equal factor loadings and therefore contributed equally to capturing the FOG phenomenon, including the items of part III on functional impact of FOG. This argues in favor of calculating a summed NFOG-score.

Accurate assessment of FOG is an integral part of clinical decision-making and has clear relevance for research. This study supports the NFOG-Q as a FOG-specific and reliable measurement tool. Further clinimetric work is required to determine the responsiveness and validity of the NFOG-Q against objective measures.

Acknowledgements

We thank all the subjects who participated in the study. Thanks are also due to Dr. Katherine Baker, who participated in the data collection and Dr. Steffen Fieuws for his valuable statistical advice. No grant or other financial support was granted to conduct this study.

Conflict of interest statement

The authors report no conflict of interest.

Appendix A. Supplementary data


References


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.


