Review Article

Clinimetrics of the 9- and 19-Item Wearing-Off Questionnaire: A Systematic Review

Carlos E. Mantese,1,2 Artur Schumacher-Schuh,1 and Carlos R. M. Rieder1,3,4

1Postgraduate Program in Medical Sciences, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil
2Mãe de Deus Hospital, Porto Alegre, RS, Brazil
3Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA), Porto Alegre, RS, Brazil
4Hospital de Clínicas de Porto Alegre (HCPA), Porto Alegre, RS, Brazil

Correspondence should be addressed to Carlos R. M. Rieder; carlosrieder@gmail.com

Received 8 October 2017; Revised 17 January 2018; Accepted 19 February 2018; Published 1 April 2018

Academic Editor: Peter Hagell

Copyright © 2018 Carlos E. Mantese et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The treatment of Parkinson’s disease (PD) with dopaminergic therapy improves functionality and quality of life. However, as the disease progresses, it causes motor and nonmotor fluctuations [1]. The well-described wearing-off (WO) phenomenon is the shortening effect of levodopa, which can be managed with dosage adjustment or adjuvant therapy, such as catechol-O-methyltransferase (COMT) inhibitors [2]. Clinical evaluation has been the gold standard for diagnosing this condition. However, the WO phenomenon may not be well recognized, mainly if it is mild or involves nonmotor symptoms. Several scholars argue that recognition of WO phenomenon could change the way that it is managed and improve patient’s functionality [3].

1. Introduction

To improve the recognition of WO phenomenon, a 32-item questionnaire (WOQ-32) was developed [3]. The questionnaire consists of a checklist of symptoms that patients must identify, and they must note if these symptoms improve with medication. For practical reasons, using the same research, this questionnaire was adapted to a 19-item questionnaire (WOQ-19), which had the same properties [4]. Later, a 9-item questionnaire (WOQ-9) was developed [5], containing the most valuable questions, and it was successfully tested [6]. The WOQ-9 has been used for a number of clinical studies, translated into several languages, and adapted with several different clinimetric properties [7]. A recent review by the Movement Disorder Society set both the WOQ-19 and WOQ-9 as recommended tools for screening for WO phenomenon [8]. However, this
2 Methods

We follow the PRISMA statement.

The inclusion criterion was studies using WOQ-9 or WOQ-19 in PD patients to diagnose WO compared to the gold-standard, clinical evaluation. The studies must examine sensitivity and specificity, or they must include data that we could calculate. Also, we include studies using one of the questionnaires if they employed data regarding test-retest or responsiveness. Formal validation was not required [9], but at least a translation and face validation for the given language was applied. Reviews, abstracts, and conference meetings were excluded. Responsiveness was calculated following Cohen’s effect size [10].

The search was conducted in MEDLINE, Embase, and Web of Sciences between 01/06/2017 and 22/12/17. The terms were ((Parkinson’s OR Parkinson’s disease) AND (wearing off OR wearing-off OR motor fluctuation) AND (questionnaire)). There was also a bibliography review for the select articles and reviews already published. The articles were independently selected by title for abstract reading by two reviewers (Artur Schumacher-Schuh and Carlos E. Mantese). In case of disagreement, the articles were discussed by another author (Carlos R. M. Rieder). Later, a number of articles were selected for full reading based on abstract information.

3 Results

As shown in Figure 1, we observed 404 articles based on the title after excluding duplicates. We reviewed 194 abstracts and 43 full-text articles after excluding 4 abstracts. Ultimately, we selected 3 articles about WOQ-19 [11–13] and 5 articles about WOQ-9 [6, 14–17] for sensitivity and specificity, 1 for WOQ-19 test-retest [18], 2 for responsiveness, 1 for WOQ-9 [19], and 1 for WOQ-19 [20]. Of note, we did not include the original WOQ-19 description [4] because it was an adaptation of WOQ-32, as was WOQ-9 [5]. In addition to these articles, we exclude 17 more because they did not use the target questionnaires; among these 17, two more were from the same 32-item questionnaire [3, 21] and 15 from other means of assessment [22–36]. One article was not used because it lacked any type of language validation [37], and two articles presented post hoc analyses of the same populations outlined in other articles that we include [38, 39]. Nine articles did not present clinimetric properties [40–48], and among them, two provided data using which we could calculate only PPV [41, 43]. One was a review [7]. One paper [18] included for test-retest was not included for sensitivity and specificity because it used the same population as another article [13].

We did not identify any articles that were not included in the search from original libraries.

In WOQ-19, there are 3 trials selected (Table 1). One of them used 1-item cutoff [12], while the others [11, 13] used 2-item cutoff. The sensitivity ranged from 0.81 to 0.90, and the specificity was 0.39–0.80. PPV was 0.62–0.88, and NPV was 0.64–0.84. The wide range of specificity seems secondary to one study that used 1-item cutoff. This trial does not exhibit better aggregate sensitivity but has shown worse specificity.

In WOQ-9 (Table 2), all studies had 1-item cutoff. Sensibility was 0.87–1, specificity was 0.10–0.69, PPV was 0.48–0.86, and NPV was 0.71–1. All studies showed excellent sensitivity but lacked specificity.

Test-retest stability was assessed in one paper [18], two weeks apart from each test, which showed an intraclass correlation of number of positive items of 0.858. It was applied to stable patients; however, it did not mention the clinical stability or the type of intraclass correlation.

For responsiveness, two studies were analyzed [19, 20]. Both were clinical trials with COMT inhibitors. One of them [19] used WOQ-9 and reported improvement in most items in proportion of patients with improvements; however, it did not provide data to calculate the effect size. In the other trial [20], Cohen’s effect size was 0.5.

4 Discussion

This report describes the first systematic review of quantitative clinimetric properties of WOQ-19 and WOQ-9.
Systematic reviews are fundamental to summarize important data for research and clinical practices. Additionally, this report describes the first independent review of the clinimetric properties of these questionnaires. The WOQ-19 seems to have good accuracy, which is an excellent tool in both research and clinical practice, when a 2-item cutoff is used. However, most of the trials used WOQ-9, which has excellent sensitivity but poor specificity. Thus, the WOQ-9 could be used as a screening tool to identify certain at-risk individuals, but it would need a clinical evaluation to confirm the diagnosis, as several trials have done [42, 46]. Stacy [7] has argued that office visits could fail to recognize WO, and its position as the gold standard of care may need reevaluation [6]. This hypothesis seems difficult to prove. Moreover, most clinical trials for Parkinson’s disease treatment use wearing-off outcomes diaries or UPDRS (Unified Parkinson’s Disease Rating Scale) wearing-off subitems. Raciti et al. [49] showed that UPDRS has 0.87 sensitivity and 0.43 specificity compared to clinical stability. WOQ-19 exhibits reliability and therefore lost specificity. In a ROC curve plotted by Martinez-Martin et al. [11], the questionnaire showed better accuracy with no change in clinical status. However, this article involves data from a different questionnaire and WOQ-9 questionnaires. Several articles were post hoc analyses of primary data, which we had already included. We excluded several important articles, such as Stacy et al.’s [3] description of the WOQ-32 questionnaire and its transformation into the WOQ-19 [4] and WOQ-9 [5]. However, this article involves data from a different questionnaire, which was later transformed into the WOQ-19 and WOQ-9 questionnaires. Several articles were post hoc analyses of primary data, which we had already included. One study did not have any type of validation and did not meet our inclusion criteria. Most of the excluded trials were with no comparator; therefore, we could not address clinimetric properties.

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>N</th>
<th>Cutoff</th>
<th>Sn</th>
<th>Sp</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Martinez-Martin et al.</td>
<td>2008</td>
<td>222</td>
<td>2</td>
<td>0.88</td>
<td>0.8</td>
<td>0.88</td>
<td>0.79</td>
</tr>
<tr>
<td>Seki et al. [12]</td>
<td>2013</td>
<td>464</td>
<td>1</td>
<td>0.81</td>
<td>0.39</td>
<td>0.62</td>
<td>0.64</td>
</tr>
<tr>
<td>Stocchi et al. [13]</td>
<td>2014</td>
<td>617</td>
<td>2</td>
<td>0.90</td>
<td>0.63</td>
<td>0.76</td>
<td>0.84</td>
</tr>
</tbody>
</table>

Sn: sensitivity; Sp: specificity; PPV: positive predictive value; NPV: negative predictive value.

5. Conclusions

We conducted the first systematic review of WOQ-19 and WOQ-9, an important tool for screening and diagnosing WO. The lack of certain data suggests caution when using the WOQ-9. However, the WOQ-19 exhibits reliability and...
was validated to use as a diagnostic tool. Moreover, we suggest that the authors report complete properties when they are publishing papers validating their methods.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

Carlos R. M. Rieder was supported by the Brazilian government research grant agency CNPq (no. 311970/2014-1). The authors also thank the Fundo de Incentivo a Pesquisa e Eventos (FYPE), Hospital de Clnicas de Porto Alegre (HCPA), for article processing charges.

References


