Measurement 132 (2019) 109–124

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Measuring activities of daily living in Parkinson's disease: On a road to nowhere and back again?

Peter Hagell [a,b,*]

[a] The PRO-CARE Group, Faculty of Health Science, Kristianstad University, Kristianstad, Sweden
[b] Restorative Parkinson Unit, Division of Neurology, Depratment of Clinical Sciences Lund, Lund University, Lund, Sweden

ARTICLE INFO

Article history:
Received 16 June 2018
Received in revised form 2 September 2018
Accepted 17 September 2018
Available online 18 September 2018

Keywords:
Activities of daily living
Classical test theory
Outcome measurement
Parkinson's disease
Rasch measurement theory
Rating scales

ABSTRACT

Parkinson's disease (PD) is a progressive neurodegenerative disorder associated with increasing disability and limitations in performance of activities of daily living (ADL) despite availability of effective symptomatic therapy. Following an overview of classical test theory (CTT) and Rasch measurement theory (RMT), the case of a clinical PD trial aiming to demonstrate ADL improvements by using the ADL section (part II) of the Unified PD Rating Scale (UPDRS) to measure ADL outcomes is considered and central questions related to its validity and interpretation are addressed. It is found that while CTT did not detect any issues, RMT in combination with conceptual considerations seriously challenged the role of the UPDRS II as an ADL outcome measure. Results are discussed from historical, methodological and clinical perspectives.

1. Introduction

1.1. Parkinson's disease

Parkinson's disease (PD) was first described by English physician James Parkinson in 1817 [1]. PD is a progressive neurodegenerative disorder that affects an estimated 0.3% of the population at large and 1% of people above 60 years of age [2]. The most typical and cardinal features of PD are neurological motor symptoms of bradykinesia (slowness of movement), muscle rigidity, tremor, and postural impairments. However, non-motor features (e.g., depression, anxiety, sleep disorders, fatigue, cognitive impairment, dysautonomia and pain) are also common and make significant contributions to the overall impact of PD [3,4]. The core pathology believed to cause the main motor symptoms is a striatal dopamine deficit due to progressive loss of nigrostriatal dopaminergic neurons [3]. Symptomatic dopaminergic pharmacotherapy is initially successful but a fluctuating drug response and dyskinesias often develop over time. With the occurrence and progression of both motor and non-motor symptoms, often in complex and fluctuating patterns, the disease is associated with significant consequences in terms of, e.g., deteriorating quality of life and ability to perform activities of daily living (ADL) [5–7].

Although the cause of PD remains enigmatic, there have been considerable advances in our understanding of PD over the 200 years since the disorder first was described [8]. Similarly, therapeutic advances have been significant, particularly during the past 50 to 60 years [8,9]. Up until the 1960s medical therapy was largely limited to anticholinergic drugs that only offered limited symptomatic relief. By the late 1960s, levodopa revolutionized PD therapy by offering dramatic symptomatic relief. During the following decades, additional therapeutic approaches such as dopamine agonists, enzyme inhibitors and functional neurosurgery have been introduced that together enable long-term symptom control. However, all available therapies hitherto are symptomatic and disease modifying therapies are still lacking.

1.2. Outcome measures in Parkinson's disease

The development of effective symptomatic PD therapy highlighted the need for useful and relevant outcome measures to

Abbreviations: ADL, Activities of daily living; CTT, Classical test theory; DIF, Differential item functioning; ICC, Item characteristic curve; ICF, International Classification of Functioning, disability and health; IRT, Item response theory; MDS-UPDRS, Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale; MID, Minimally important difference; MRFA, Minimum rank factor analysis; PDAQ, Penn Parkinson's Daily Activities Questionnaire; PCA, Principal component analysis; PCM, Partial credit model; PD, Parkinson's disease; PDQ-39, The 39-item Parkinson's Disease Questionnaire; PRM, Polytomous Rasch model; RCT, Randomized controlled trial; RMT, Rasch measurement theory; RSM, Rating scale model; SEM, Standard error of measurement; UPDRS, Unified Parkinson's Disease Rating Scale.

* Address: Faculty of Health Science, Kristianstad University, SE-291 88 Kristianstad, Sweden.
E-mail address: Peter.Hagell@hkr.se

https://doi.org/10.1016/j.measurement.2018.09.050
0263-2241/

determine the effectiveness and value of treatments. As in other areas of clinical medicine, such outcome measures have typically consisted of single- or multi-item instruments (or "scales") with defined response categories of various sorts [10]. The perhaps earliest instrument proposed to assess therapeutic benefit in PD was that by Duff in an early clinical trial [11]. This scale included ten "modal activities" of daily life with various degrees of movement complexities (Turning in bed, rising from and returning to bed; Dressing/undressing; Performance of the toilet, especially shaving in men; Eating; Walking; Turning; Climbing stairs; Speaking; Writing; Facial expression); each item was rated by assigning one of six ordered response categories scored 0–5 (No activity; Grossly restricted; Moderate restriction, especially by rigidity; Moderate restriction, especially by tremor; Activity approaching normal, but slow/clumsy; Activity practically normal). Other early PD rating scales include the Schwab & England activities of daily living scale [12] and the Northwestern University Disability Scale [13]. Similar to the Duff scale, these also focused on activity performance rather than symptom severity. However, it was not until following demonstration of the clinical effectiveness of levodopa [14] that instruments for outcome assessment in PD started to proliferate. These included, e.g., the Webster, King's College Hospital, Columbia University, and New York University scales [15]. Although some of these included aspects of activity performance, their primary focus tended to be on assessing the severity of various motor symptoms through standardized neurological examinations. This conceptual shift from what today is categorized as Activities to Body functions [16], was probably due to an intention to evaluate the symptomatic effects of levodopa and other emerging PD therapies.

The development and use of similar albeit different instruments by various investigators hampered the possibility to compare results from various studies [17]. This led to the development of the Unified PD Rating Scale (UPDRS), which was based on previously available scales and aimed to overcome the problem of incomparability of study results by introducing a common means of evaluating PD and therapeutic responses [15]. The UPDRS consists of four main parts intended to cover major aspects of PD: Mentation, behavior and mood (part I), Activities of daily living (part II), Motor examination (part III), and Complications of therapy (part IV). More recently, a modification of the UPDRS was conducted by the International Parkinson and Movement Disorder Society (MDS), named the MDS-UPDRS [18,19]. The basic structure remains intact compared to the original UPDRS but parts I and II were renamed as "Non-motor aspects of experiences of daily living" and "Motor aspects of experiences of daily living", respectively.

2. Psychometrics and rating scales as health outcome measures

At this point, it may be appropriate to briefly review approaches to the development and quality assurance of scales used in the health sciences, as well as in, e.g., education and psychology.

2.1. Basic principles

Most phenomena (variables) of interest in health outcome measurement (e.g., disease severity, quality of life, disability) are not directly observable. Since such latent variables cannot be observed or measured directly, one has to rely on observable manifestations or consequences of the latent variable. These manifestations are operationalized as items (questions, statements, observed behaviour or performance, etc.) that make up the instrument. Variations in the observable manifestations (item responses) are assumed to reflect variations in the latent variable. This principle is illustrated in Fig. 1.

For example, consider that we want to measure activity performance. To do so, we may ask a person to report his or her ability to perform various activities. Based on those responses, we may be able to assess the person's location on a latent activity performance continuum. With reference to Fig. 1, based on some conceptual definition of activity performance, items that express relevant activities are presented to the individual, who is asked to indicate her or his performance level on each item according to one of two or more ordered response categories that each describe a certain performance level. To quantify the qualitative information achieved in this manner, each descriptive response category (e.g., none–mild–moderate–severe) is assigned an integral numeral (e.g., 0–1–2–3) as a means of partitioning the underlying latent continuum into successively increasing (or decreasing) amounts of the variable. Item responses are then typically summed into a total score intended as the basis of locating the respondent on a continuum from less to more, in order to describe the level of individuals or groups of people on the latent variable, make comparisons and evaluate changes following therapeutic interventions.

2.2. Classical test theory

The roots of the principles reviewed in Section 2.1. can be traced to the late 19th and early 20th centuries, when behavioural scientists endeavoured to quantify latent variables such as personality, intelligence, knowledge and attitudes [10,20]. The science that grew out of this work is typically referred to as psychometrics [21], a term that appears to have been first used by Francis Galton in his 1879 paper *Psychometric experiments* [22], which he opens by stating that "Psychometry, it is hardly necessary to say, means the

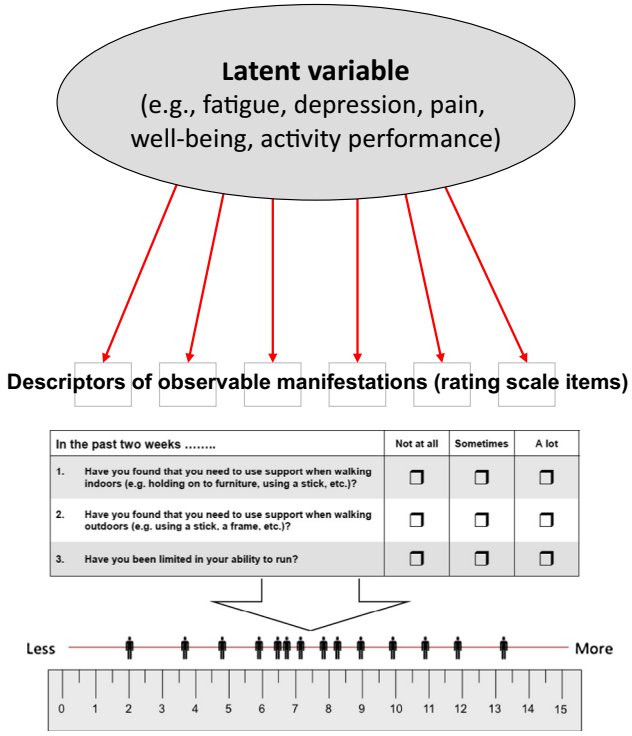

Fig. 1. Illustration of the basic instrument design and assumptions underpinning latent variable measurement. Items are observable manifestations of the unobservable latent target variable and are expected to reflect variations in the latent variable. Observed item responses form the basis in the measurement process used to locate the individual on a latent quantitative continuum intended to represent her/his position on the target variable, from less to more.