Journal of Psychosomatic Research 97 (2017) 70–81

Contents lists available at ScienceDirect

Journal of Psychosomatic Research

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

The construct validity of the Major Depression Inventory: A Rasch analysis of a self-rating scale in primary care

Marie Germund Nielsen[a,*], Eva Ørnbøl[b], Mogens Vestergaard[a,c], Per Bech[d], Kaj Sparle Christensen[a]

[a] Research Unit for General Practice, Department of Public Health, Aarhus University, Denmark
[b] Research Clinic for Functional Disorders and Psychosomatics, Aarhus University Hospital, Denmark
[c] Section for General Medical Practice, Department of Public Health, Aarhus University, Denmark
[d] Psychiatric Research Unit, Psychiatric Centre North Zealand, Copenhagen University Hospital, Denmark

ARTICLE INFO

Keywords:
Construct validity
Depression
Dimensionality
Major Depression Inventory
Primary care
Rasch model

ABSTRACT

Objective: We aimed to assess the measurement properties of the ten-item Major Depression Inventory when used on clinical suspicion in general practice by performing a Rasch analysis.
Methods: General practitioners asked consecutive persons to respond to the web-based Major Depression Inventory on clinical suspicion of depression. We included 22 practices and 245 persons. Rasch analysis was performed using RUMM2030 software. The Rasch model fit suggests that all items contribute to a single underlying trait (defined as internal construct validity). Mokken analysis was used to test dimensionality and scalability.
Results: Our Rasch analysis showed misfit concerning the sleep and appetite items (items 9 and 10). The response categories were disordered for eight items. After modifying the original six-point to a four-point scoring system for all items, we achieved ordered response categories for all ten items. The person separation reliability was acceptable (0.82) for the initial model. Dimensionality testing did not support combining the ten items to create a total score. The scale appeared to be well targeted to this clinical sample. No significant differential item functioning was observed for gender, age, work status and education. The Rasch and Mokken analyses revealed two dimensions, but the Major Depression Inventory showed fit to one scale if items 9 and 10 were excluded.
Conclusion: Our study indicated scalability problems in the current version of the Major Depression Inventory. The conducted analysis revealed better statistical fit when items 9 and 10 were excluded.

1. Introduction

Depression is one of the most common mental disorders in primary care. More than 80% of persons with depression are diagnosed by their general practitioner (GP) [1]. Reliable diagnostic tools are thus a key issue in primary care. It is of great importance to investigate the validity of applied psychometric instruments as these are prerequisites for reliable identification of mental health problems, which is the first step to ensure appropriate diagnosis and treatment of mental health problems.

Several scales have been developed for measuring depression, such as the Patient Health Questionnaire (PHQ-9), the Hospital and Anxiety Scale (HADS) and the Hopkins Symptom and Checklist (HSCL-20). In Denmark, the Major Depression Inventory (MDI) is used by many GPs.

Each year 265,000 psychometric instruments are used in general practice in Denmark. The Danish College of General Practitioners recommends using the MDI on clinical suspicion of depression [2], and the Danish GPs are reimbursed for using the MDI when diagnosing depression.

The MDI was originally developed in collaboration with the World Health Organization (WHO) for diagnosing depression in Danish general practice [3]. The MDI has later been considered as a rating scale in trials of anti-depressive interventions focusing on assessment of the total sum score to estimate symptom severity [4]. The MDI is compatible with both the ICD-10 and the DSM-IV diagnostic criteria for clinical depression. The MDI has been translated into about twenty

* Corresponding author at: Research Unit for General Practice, Department of Public Health, Aarhus University, Bartholins Allé 2, 8000 Aarhus C, Denmark.
E-mail address: mm@ph.au.dk (M.G. Nielsen).

http://dx.doi.org/10.1016/j.jpsychores.2017.04.001
Received 14 February 2017; Received in revised form 29 March 2017; Accepted 3 April 2017

different languages, including English [4], and has been widely used around the world [5–8].

The validity of the instrument has previously been investigated by classical test theory (CTT) finding satisfactory Cronbach's alpha of the MDI of respectively 0.89 and 0.90 [4,5]. Although CTT is often used to test the validity of instruments, it has some limitations. For example, the ordinal data are treated as interval level, scale evaluations are sample dependent, and the assumptions of CTT cannot be formally tested within the CTT framework [9]. Item response theory (IRT) represents a group of several distinct models which are all based on the assumption that the difficulty of each item is treated as information to be incorporated in the scaling of items [10]. In the Rasch model, the response to any particular item is a function of the difference between the estimated ability of the person (e.g. the level of depression) and the characteristics of the item; this represents the difficulty of the item (e.g. the level of depression implied by the item) [9,10].

Rasch analysis is the formal testing of an outcome scale against a mathematical measurement model that operationalizes formal measurement [11]. An important principle of the Rasch model is the specific objectivity, which is based on the assumption that the comparison of two objects should be independent of everything but the two objects and their observed relation. For a homogeneous test to conform to the Rasch model, a consequence of the principle of specific objectivity is that the estimated difference in ability between two people is independent of the difficulty of any particular test item used to compare them [12].

A former study by Olsen et al. has investigated the MDI and found the MDI to be unidimensional in a Rasch analysis [4]. A recent Rasch analysis of the MDI by Amris et al. suggested lack of unidimensionality in a clinical sample of females with chronic widespread pain [13].

The MDI scale has not been fully examined against the requirements and assumptions of the Rasch analysis. For example, differential item functioning (DIF) for gender has not previously been investigated. The presence of DIF means that individuals in different subgroups with the same level of depression will have different probabilities (or difficulties) of endorsing an item [14]. The objective of our study was to assess the measurement properties of the MDI by performing a Rasch analysis of a sample of persons who were clinically suspected of having depression in general practice.

2. Method

2.1. Procedures

We developed the webpage Sundhedsmappen.dk (in Danish) to collect web-based versions of the MDI from Danish general practice. The website is an online system intended to support diagnosis and monitoring of depression, anxiety and blood pressure in a primary care setting [15]. On clinical suspicion of depression during a consultation, the GPs asked consecutive persons to complete the web-based version of the MDI at Sundhedsmappen.dk on a tablet PC or desktop computer in the clinic. This approach ensured that the GPs received the results in their laboratory system and the data were securely saved at our database. The GP also handed out an information brochure about the study to the tested persons.

2.2. Participants

The GPs were recruited through invitations sent by postal mail to 700 practices in Denmark, newsfeeds, network practices and conference presentation. The project manager visited the GPs who showed an interest in the study and held an information meeting in the clinic. Reminders about participation in the study were sent by postal mail.

The Danish clinical guidelines recommend using the MDI on clinical suspicion of depression (i.e. presence of two or three core symptoms of depression according to the ICD-10) [2]. We included adult persons (aged 18 years or older) who were suspected of being depressed by the GP, who understood written and spoken Danish and who gave oral informed consent at their general practice. There were no exclusion criteria for this study and no missing data due to the web-based data collection.

We included 22 practices and 245 persons; mean number of persons per practice was 11 with range: 1–38. All the included persons filled out the web-based version of the MDI, and attempts were made to contact the included persons by phone regarding the diagnostic interview. In total, 150 consecutive persons were interviewed by phone by a certified Munich-Composite International Diagnostic Interview (M-CIDI) interviewer. Furthermore, 95 additional persons did not participate in the interview: 56 persons were contacted, but did not respond to repeated calls, 31 persons declined to participate in the interview and 8 did not meet the inclusion criteria. The GPs received DKK 122.57 (≈ EUR 16.50) per included person. No written informed consent was required from participants for this study. Only oral information was necessary, and no ethical permission was required according to Danish law. Permission to conduct the study was granted by the Danish Committee of Multipractice Studies in General Practice and the study was approved by the Danish Data Protection Agency, ID number: 2013–41-1756.

2.3. Measures

A web-based version of the diagnostic tool MDI was validated for use in general practice. The MDI is a self-report checklist, which includes ten items: (1) feeling sad, (2) loss of interest, (3) lack of energy, (4) lack of self-confidence, (5) feelings of guilt, (6) feeling that life is not worth living, (7) concentration problems, (8) feeling restless/slowed down, (9) sleeping too much/too little and (10) reduced/increased appetite. Included persons were asked to what extent the symptoms had been present during the last two weeks at which the frequency of each symptom are indicated from 0 (at no time), 1 (some of the time), 2 (slightly less than half of the time), 3 (slightly more than half of the time), 4 (most of the time) to 5 (all the time). The MDI can be used in two ways: a) as a diagnostic instrument covering both the DSM-IV and the ICD-10 diagnostic algorithms for depression or b) as a severity scale of depression. The ICD-10 algorithm is coded as: mild depression (at least two core symptoms + two associated symptoms), moderate depression (at least two core symptoms + four associated symptoms) or severe depression (at least three core symptoms + five associated symptoms) (see Appendix Fig. 1.A) (7,8). Core items are indicated by items 1–3. Associated symptoms are indicated by items 4–10. A core symptom is present if the score for this symptom is at least 4. To study severity of depression, a sum score of the ten items is used providing a severity rating score of 0–50. We analyzed the items according to the scoring instructions in the manual for the MDI, which states that, for items 8, 9 and 10, alternative a or b with the highest score should be considered [16]. According to the MDI manual, the cut-points for the total MDI score are: no depression (≤ 19), mild depression (20–24), moderate depression (25–29) and severe depression (≥ 30).

Eating or sleeping too much is an atypical sign of depression. Therefore, in persons reporting sleep disturbances, it is important to distinguish between too little and too much sleep. Studies suggest that