

**Gender-related differences in psychometric properties of World Health Organization Disability Assessment Schedule (WHODAS 2.0)**

Short title: **Differential item functioning of WHODAS 2.0**

Niina Katajapuu MSc<sup>1,3</sup>, Katri Laimi PhD, MD<sup>2</sup>, Ari Heinonen PhD<sup>1</sup>, Mikhail Saltychev PhD, MD<sup>2</sup>

<sup>1</sup> Faculty of Health and Sport Sciences, University of Jyväskylä, Jyväskylä, Finland

<sup>2</sup> Department of Physical and Rehabilitation Medicine, Turku University Hospital and University of Turku, Turku, Finland

<sup>3</sup> Faculty of Health and Wellbeing, Turku University of Applied Sciences, Turku, Finland

**ADDRESS FOR CORRESPONDENCE**

Niina Katajapuu, Turku University of Applied Sciences, Joukahaisenkatu 3, 20540 Turku, Finland

Email: [niina.katajapuu@turkuamk.fi](mailto:niina.katajapuu@turkuamk.fi)

**CONFLICT OF INTEREST**

None to declare

**FUNDING**

None to declare

**ACKNOWLEDGEMENTS**

None to declare

## ABSTRACT

Objective of this study was to investigate the gender-related differential item function (DIF) of 12-item World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0) amongst patients with chronic musculoskeletal pain. This was a cross-sectional survey study among 1,988 consecutive patients with chronic musculoskeletal pain at a university Physical and Rehabilitation Medicine outpatient clinic. To assess a DIF, WHODAS 2.0 items were dichotomized as 'none' rated by respondents as '0' versus 'any limitation' rated as '1,2,3 or 4'. The item response theory analysis (IRT) was used to define discrimination and difficulty parameters of a questionnaire. The probit logistic regression was used to test uniformity of DIF between gender groups. The results of DIF analysis were presented and evaluated graphically as item characteristic curves based on 2-parameter IRT analysis of dichotomized responses. High to perfect discrimination ability was observed for all the items except one. Difficulty levels of eight items were shifted towards the elevated disability level, four items demonstrated a perfect difficulty property. Significant DIF between genders was observed in seven out of 12 items. All the detected DIFs were uniform. For item 'household', 'emotional affection' and 'work', men had to experience slightly worse disability than women to achieve the same score. A reverse effect was observed for items 'concentration', 'washing', 'dressing' and dealing with strangers. In this study, significant DIF between genders was found in seven of twelve items of 12-item WHODAS 2.0. amongst 1988 patients with chronic musculoskeletal pain. All the detected DIFs were uniform. Even if this study showed gender-related DIF in seven out of 12 items, we recommend using and studying 12-item WHODAS 2.0 in different populations.

## KEYWORDS

Musculoskeletal pain, WHODAS 2.0, differential item functioning, validity

## INTRODUCTION

Gender differences in the prevalence of both musculoskeletal pain and disability are well documented (Bartley. E, 2013, Binglefors and Isacson, 2004, Merrill S et al., 1997). The tools to measure pain or disabilities have to be reliable in both genders. However, the reference values of many proxy-rated, patient-reported and objective outcome measures have also shown to be gender-dependent (Bohannon, 1997, Massy-Westropp et al., 2011, Dewitt R, 2013).

The 12-item version of World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0) is a generic tool to assess health and disability across different diseases and cultures in both clinical settings and general population. The total score of WHODAS 2.0 has been reported to be age-dependent showing a 0.6 point-increase for every age year (Gomez-Olive et al., 2017). While other psychometric properties of WHODAS 2.0 have been studied (Carlozzi et al., 2015, Ustun et al., 2010, Saltychev et al., 2017), the occurrence of a gender-related differential item functioning (DIF) is not known. This means that we do not know, if WHODAS 2.0 is similarly sensitive in men and women with musculoskeletal pain. In two previous studies on patients with either major depression or myocardial infarction, no gender-related differences in any of WHODAS 2.0 items were observed (Kirchberger et al., 2014, Luciano et al., 2010). A study of the 36-item WHODAS 2.0 on patients with knee osteoarthritis, showed a gender-related DIF only in a 'getting housework done as quickly as needed' -item and gender-related DIF was observed in a modified 36 -item WHODAS functioning scale on people with mental health problems (Novak et al., 2010, Kutlay et al., 2011). The 12-item WHODAS is widely used in patients with musculoskeletal symptoms, but the gender-related DIF of this scale has not been evaluated in musculoskeletal pain yet. This information is important to justify the wide use of 12-item WHODAS in screening disability, in evaluating functioning, and in planning and reviewing rehabilitation.

The aim of this study was to investigate if there is a significant gender-related DIF in a 12-item WHODAS 2.0 amongst patients with chronic musculoskeletal pain.

## METHODS

This was a cross-sectional study of consecutive patients with chronic musculoskeletal pain in an outpatient Physical and Rehabilitation Medicine (PRM) clinic of a university hospital between April 2014 and February 2017. The survey was sent to the patients and filled up before a physician appointment. The survey included a 12-item WHODAS 2.0 questionnaire and questions on demographics, pain intensity, perceived general health, and working ability among others. The university hospital ethics committee approved the study.

The self-administered WHODAS 2.0 contains 12 items covering the most common limitations of activity and participation during the last 30 days. A Likert-like scale is used to define the severity of limitation from 0 to 4 with 0 denoting “no limitation” and 4 denoting “extreme limitation or inability to function”. The total score was calculated in our study as a sum of all 12 items divided by 48, multiplied by 100, and presented as a percentage where 100% represents the worst possible restriction.

Age was defined in full years at the time of visiting a clinic. Pain intensity was assessed using a 11-point numeric rating scale (NRS) with 0 denoting “no pain” and 10 denoting “worst possible pain”. Educational level was dichotomized “high school” vs. “no high school”. Body mass index (BMI) was calculated as a body weight divided by a squared height ( $\text{kg}/\text{m}^2$ ).

### *Statistical analysis*

Age, pain, total score of WHODAS 2.0 and BMI were presented as means and standard deviations (SDs). Gender and educational level were presented as proportions and percentages. Independent t-test and chi square test were used to investigate potential differences between men and women regarding their age, educational level, BMI, and pain intensity.

Differential item functioning (DIF) is a statistical characteristic of a scale item (here counted for each of 12 items included in WHODAS 2.0) that describes if the item is measuring an ability (here level of functioning) differently for separate subgroups (here genders) within the sample. To assess a DIF, WHODAS 2.0 items were dichotomized as ‘none’ (rated by respondents as ‘0’) versus ‘any limitation’

(rated by respondents as '1', '2', '3', or '4'). It has previously been reported that such a dichotomous version of WHODAS 2.0 is compatible with its polytomous version (World Health Organisation, 2010).

The item response theory (IRT) analysis defined discrimination and difficulty parameters of a questionnaire. A discrimination parameter describes the sensitivity of test to differentiate severity levels of symptoms. The steeper the regression curve, the more discriminative the test becomes. In this study, discrimination of 0.01 to 0.24 was considered 'none' (a totally level regression curve), 0.25 to 0.64 'low', 0.65 to 1.34 'moderate', 1.35 to 1.69 'high', and a discrimination  $>1.7$  was considered 'perfect' (a regression curve approaching a vertical line) (Baker FB, 2001). Ideally, the steepest interval corresponds to the patients who obtained average WHODAS 2.0 total scores in the studied population. In turn, difficulty is a psychometric property of a single item or an entire test, which describes how much more or less a respondent should perceive the studied ability (comparing with the average level of studied population) in order to achieve a 0.5 probability to give a particular answer.

The probit logistic regression was used to test whether an item exhibits either uniform or nonuniform DIF between gender groups that is, whether an item favors one group over the other for all values of the functioning limitation or for only some values of that (de Boeck P, 2004, Swaminathan H and Rogers HJ, 1990). A uniform DIF occurs when the difference between groups remains the same across the entire scale. In turn, a nonuniform DIF is observed when the direction of difference between groups varies at different levels of functioning limitation (e.g., if men perform better up than women to a midpoint and worse than women after that). A two-tailed  $p$ -value  $\leq 0.05$  indicated a significant difference between genders. When significant DIF was observed, the results of DIF analysis were also presented and evaluated graphically as item characteristic curves based on 2-parameter IRT analysis of dichotomized responses.

All the analyses were conducted using Stata/IC Statistical Software: Release 15. College Station (StataCorp LP, TX, USA).

## RESULTS

Of 3,150 patients visiting the clinic, 1,988 (63%) participated the study. The patients were 47.6 (SD 15.0) years old and 1,297 (65%) were women (Table 1). The average intensity of pain was 6.3 (SD 2.0) points. Most of the patients (n=1746, 88%) had a main diagnosis 'M' - 'Diseases of the musculoskeletal system and connective tissue' according to the International Classification of Diseases 10<sup>th</sup> Edition. The most frequent single diagnoses were 'M54 Dorsalgia' (n=781, 39%) and 'M79 Other soft tissue disorders' (n=202, 10%).

The total scores of WHODAS 2.0 were 27.3 (SD 19.5) points for both men and women ( $p=0.843$ ). Probably due to a large sample size, the differences between men and women in BMI ( $p<0.001$ ), pain severity ( $p=0.005$ , 95% CI 0,38 - 0.01), and educational level ( $p<0.001$ ) were statistically significant even if the absolute estimates differed only a little.

High to perfect discrimination ability was observed for all the items except for item #9 "dressing" with moderate discrimination (Table 2). Difficulty levels of eight items – #3, #4, #6, #7, #8, #9, #10, and #11 (learning, joining in community, concentrating, walking, washing, dressing, dealing with strangers, maintaining friendships) – were shifted towards the elevated disability level compared to average disability level of the entire studied population. In other words, musculoskeletal patients with mild or none disability clustered around the lowest possible scores on these items. Other four items (standing, household responsibilities, being emotionally affected, work) demonstrated a perfect difficulty property (Table 3).

Significant DIF between genders was observed in seven out of 12 items: 'household responsibilities', 'being emotionally affected', 'concentrating for 10 minutes', 'washing', 'dressing', 'dealing with strangers', and 'work'. All the detected DIFs were uniform (Table 4 and Figure 1). For items #2, #5 and #12 (household, emotional affection, work), men had to experience slightly worse disability than women to achieve the same score. A reverse effect was observed for items #6, #8, #9 and #10 (concentration, washing, dressing, dealing with strangers).

## DISCUSSION

This study amongst 1,988 patients with chronic musculoskeletal pain showed significant DIF between genders in seven of twelve items of 12-item WHODAS 2.0. All the detected DIFs were uniform meaning that the direction of gender-related differences between responses persisted across the entire spectrum of disability severity. For items 'household responsibilities', 'emotional affection', and 'work', men had to experience slightly worse disability than women to achieve the same score. A reverse effect was observed for items 'concentrating', 'washing', 'dressing' and 'dealing with strangers'.

The generalizability of the results is weakened by the fact that the sample represented a population with chronic musculoskeletal pain treated in a highly specialized health care unit (university PRM clinic). Thus, the patients might differ from those treated in e.g. primary health care. Additionally, the sample was predominated by women and there was a statistically significant difference between genders in educational level women having higher education level. This was, however, the first study on gender-related DIF of the 12-item WHODAS 2.0 with a sample large enough to achieve statistically significant results and narrow confidence intervals.

The results were similar with two previous studies on the subject (Kutlay et al., 2011, Novak et al., 2010). Novak et al. have found gender-related DIF in many daily activities in people with mental health problems. In turn, Kutlay et al. have reported gender-related DIF in 'life activities' in patients with knee osteoarthritis. Both studies have also observed a significant DIF in 'taking care of household responsibilities' as seen in the present study: comparing to women, men had to experience worse disability to reach a similar score in this item. Neither of these studies used the 12-item version of WHODAS 2.0. The populations of interest and statistical methods used by both previous reports have been different from the present study by. To assess DIF, Novak et al. have used odds ratio. The participants in that study were asked to evaluate their functioning level based on the worse month in previous year while the 12-item WHODAS 2.0 is based on responses concerning the previous 30 days. The study by Kutlay et al. have employed the 36-item WHODAS 2.0. When studying the 36-item WHODAS 2.0, Silva et al. have observed some gender-dependent differences in limitations in household activities – women had had more limitations than men

had (Silva et al., 2013a). Another study by the same team has reported that men had stated more often on not doing housework marking the respective question as 'not applicable' (Silva et al., 2013b). In the present study, this phenomenon was not observed probably due to cultural differences between studied populations. Both studies by Silva C et al. and Silva AG et al. have employed the 36-item version of WHODAS 2.0 (Silva et al., 2013a, Silva et al., 2013b) using different statistical methods without assessing DIF.

The findings in this study differed from the results of two previous reports that did not observe DIF in any of WHODAS items (Luciano et al., 2010, Kirchberger et al., 2014). Differently to present study, they employed samples of patients with acute health conditions like myocardial infarction and major depressive episode and the respondents were older when compared to present study population.

The design used in this study does not provide any explanations to the gender differences in psychometric properties of 12-item WHODAS 2.0. As this is the first study on patients with chronic musculoskeletal pain, the results cannot be straightly reflected to previous studies either. If the found differences between genders in this study population were only due to the gender-specific way of reporting disability with one gender over-estimating and other underestimating functioning limitations, the gender effect would probably not change across items as in our study. Women had to experience slightly worse disability to get the same score than male in four items, while male had to have more disability to achieve the score of women in three items. This finding supports the results of a previous study in elderly (Merrill et al 1997), where self-reported disability was highly associated with measured difficulties in both genders.

Further research is needed to reveal possible gender-related DIFs in other settings and patient groups with different levels of functioning. Repeated measures design may reveal potential fluctuations in DIFs over time. In the light of our study results, men and women might answer differently to part of WHODAS 2.0 items.

Even if this study showed gender-related DIF in seven out of 12 items of the self-administered WHODAS 2.0 in musculoskeletal pain, these differences were uniform across the whole scale of severity, and we still recommend using and studying 12-item WHODAS 2.0 in different populations.



## REFERENCES

- BAKER FB 2001. *The basics of item response theory*, USA, ERIC Clearinghouse on Assessment and Evaluation.
- BARTLEY, E, F. R. 2013. Sex differences in pain: a brief review of clinical and experimental findings. *British Journal of Anaesthesia*, 111, 52-8.
- BINGEFORS, K. & ISACSON, D. 2004. Epidemiology, co-morbidity, and impact on health-related quality of life of self-reported headache and musculoskeletal pain--a gender perspective. *Eur J Pain*, 8, 435-50.
- BOHANNON, R. W. 1997. Comfortable and maximum walking speed of adults aged 20—79 years: reference values and determinants. *Age and Ageing*, 26, 15-19.
- CARLOZZI, N. E., KRATZ, A. L., DOWNING, N. R., GOODNIGHT, S., MINER, J. A., MIGLIORE, N. & PAULSEN, J. S. 2015. Validity of the 12-item World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0) in individuals with Huntington disease (HD). *Qual Life Res*, 24, 1963-71.
- DE BOECK P, W. M. 2004. Explanatory Item Response Models. In: DE BOECK P, W. M. (ed.) *A Generalized Linear and Nonlinear Approach*. Springer-Verlag New York.
- DEWITT R 2013. A Study of the Sit-up Type of Test as a Means of Measuring Strength and Endurance of the Abdominal Muscles. *Research Quarterly. American Association for Health, Physical Education and Recreation*, 15, 4.
- GOMEZ-OLIVE, F., SCHRÖDERS, J. & ABODERIN, I. 2017. Variations in disability and quality of life with age and sex between eight lower income and middle-income countries: data from the INDEPTH WHO-SAGE collaboration. *BMJ Glob Health*, 2.
- KIRCHBERGER, I., BRAITMAYER, K. & COENEN, M. 2014. Feasibility and psychometric properties of the German 12-item WHO Disability Assessment Schedule (WHODAS 2.0) in a population-based sample of patients with myocardial infarction from the MONICA/KORA myocardial infarction registry. *Population Health Metrics*, 12, 13.
- KUTLAY, Ş., KÜÇÜKDEVECİ, A. A., ELHAN, A. H., ÖZTUNA, D., KOÇ, N. & TENNANT, A. 2011. Validation of the World Health Organization disability assessment schedule II (WHODAS-II) in patients with osteoarthritis. *Rheumatology International*, 31, 339-346.
- LUCIANO, J. V., AYUSO-MATEOS, J. L., AGUADO, J., FERNANDEZ, A., SERRANO-BLANCO, A., ROCA, M. & HARO, J. M. 2010. The 12-item World Health Organization Disability Assessment Schedule II (WHO-DAS II): a nonparametric item response analysis. *BMC Med Res Methodol*, 10, 45.
- MASSY-WESTROPP, N., GILL, T., TAYLOR, A., BOHANNON, R. & HILL, C. 2011. Hand Grip Strength: age and gender stratified normative data in a population-based study. *BMC Research Notes*, 4, 127.
- MERRILL S, SEEMAN T, KASL S & L, B. 1997. Gender Differences in the Comparison of Self-Reported Disability and Performance Measures. *Journal of Gerontology: MEDICAL SCIENCES*, 52A, 8.
- NOVAK, S., COLPE, L., BARKER, P. & GFROERER, J. 2010. Development of a brief mental health impairment scale using a nationally representative sample in the USA. *International Journal of Methods in Psychiatric Research*, 11.
- SALTYCHEV, M., BARLUND, E., MATTIE, R., MCCORMICK, Z., PALTAMAA, J. & LAIMI, K. 2017. A study of the psychometric properties of 12-item World Health Organization Disability Assessment Schedule 2.0 in a large population of people with chronic musculoskeletal pain. *Clin Rehabil*, 31, 262-272.
- SILVA, A. G., ALVARELHÃO, J., QUEIRÓS, A. & ROCHA, N. P. 2013a. Pain intensity is associated with self-reported disability for several domains of life in a sample of patients with musculoskeletal pain aged 50 or more. *Disability and Health Journal*, 6, 369-376.
- SILVA, C., COLETA, I., SILVA, A. G., AMARO, A., ALVARELHÃO, J., QUEIROS, A. & ROCHA, N. 2013b. Adaptation and validation of WHODAS 2.0 in patients with musculoskeletal pain. *Rev Saude Publica*, 47, 752-8.
- SWAMINATHAN H AND ROGERS HJ 1990. Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement* 27, 361-370.

- USTUN, T. B., CHATTERJI, S., KOSTANJSEK, N., REHM, J., KENNEDY, C., EPPING-JORDAN, J., SAXENA, S., VON KORFF, M., PULL, C. & PROJECT, W. N. J. 2010. Developing the World Health Organization Disability Assessment Schedule 2.0. *Bull World Health Organ*, 88, 815-23.
- WORLD HEALTH ORGANISATION 2010. *Measuring Health and Disability: Manual for WHO Disability Assessment Schedule (WHODAS 2.0)*, WHO Press, World Health Organization.

Figure 1. Item characteristics curves of WHODAS 2.0 items 1 to 12

Estimates for men are presented in dash lines, estimates for women (or the entire sample) – in solid lines. Y-axis presenting the probability of endorsing the item, X-axis presenting the ability level

S1 = standing S2 = household responsibilities S3 = learning S4 = joining community S5 = emotional affection, S6 = concentrating, S7 = walking, S8 = washing, S9 = dressing, S10 = dealing with strangers, S11 = maintaining a friendship, S12 = work

