Abstract

Purpose: To investigate if the responses to the Neck Disability Index (NDI) may produce some differential item functioning (DIF) comparing men and women.

Materials and methods: Register-based study among patients undergoing cervical surgery. Item response theory (IRT) analysis including a model for detecting a DIF.

Results: Of 338 patients, 171 (51%) were women and 167 (49%) were men. Mean age was 54.0 years. For most of the items, the average level of disability in a studied sample was associated with the middle point of the scale. The ability to distinguish people with different levels of disability was high or perfect for seven out of 10 items. While the DIF could be seen for all 10 items, only three items demonstrated statistically significant DIF – "pain intensity", "headaches" and "recreation". While other seven items did not show statistically significant DIFs, better discrimination (steeper curves) for women could be graphically observed for "personal care", "lifting", "work", "driving" and "sleeping".

Conclusions: It seemed that the NDI may behave differently depending on the sex of respondents. Several items of the NDI may be more precise and more sensitive when detecting restrictions in functioning among women compared to men. This finding should be taken into account when using the NDI in research and clinical practice.

Keywords

Neck Disability Index; Neck Pain; Cervical Surgery; Differential Item Functioning

Introduction

The Neck Disability Index (NDI) has been familiar to clinicians and researchers for almost four decades (1). The NDI was developed by Howard Vernon in 1989 as the modification of a slightly older Oswestry Disability Index, which has been used to measuring disability caused by back pain. The NDI is a questionnaire containing 10 items covering pain severity, personal care, everyday life, sex life, work and recreational activity. Many journal articles and guidelines considered it a gold standard for measuring disability caused by neck pain. The properties of the NDI have broadly been studied from many points of view and using a diverse spectrum of statistical methods (2). The NDI has been found a reliable and valid scale in different settings, populations and language translations (3). However, the potential differential item functioning (DIF) of the NDI remained mostly uninvestigated and only a few studies have primarily or even secondarily focused on this topic (4-7). The DIF may, though, be very important from both perspective – scientific and clinical. E.g., two population, studied using the NDI, may not be comparable if they are predominated by groups with different properties of a scale. Also, it is also possible that the DIF is nonuniform. In other words, it behaves differently at the diverse levels of disability. Thus, the NDI estimates obtained from a population of severely disabled persons are hardly comparable with those observed among general population or people with only mild disability. Clinicians may be interested to know how much they can trust the NDI scores obtained in a particular situation.

In a simplified way, the DIF may be described as follows. Firstly, using the Item Response Theory (IRT), the average level of particular limitation of functioning in a studied sample is calculated. Secondly, the difficulty (difference between the average level and probable of each item is calculated for each choice available on a scale (e.g., “1”, “2” or “3”) .Thirdly, the discrimination property (sensitivity of an item) is calculated. Fourthly, the obtained estimates of difficulty and discrimination are compared both numerically and graphically across different

groups. When DIF is present, some groups of respondents have unequal probabilities of giving a particular response. For the NDI, the presence of DIF may mean that one group may overestimate or underestimate their disability level compared with another group. For example, older people may overestimate restrictions related to their social participation while younger may exaggerate limited physical activity. The usual grouping variables for the DIF analysis are sex and age.

The DIF is very common phenomenon. Even when using objective tests like laboratory tests or assessments of physical condition, one may expect that people of different sex or age groups may produce diverse results (8-11). The same phenomenon may be seen in subjective patient-reported outcome measures (12, 13). For example, men and women or people of different ages may grade the severity of their disability or pain differently (14-17).

Only a few previous studies have reported on the DIF of the NDI. The presence of DIF has been detected for such NDI items as “lifting”, “headaches”, “personal care”, “recreation” and “work” (4-7). As for other patient-reported outcome measures, sex and gender were the most used grouping variables. Other grouping variables used when studying the DIF of NDI were educational level, work status and clinically important change (6, 7). The presence of DIF in the NDI items might be indirectly suspected based on a few studies regarding the Oswestry Disability Index as these two scales are very alike – in fact, the NDI is a modification of the Oswestry Disability Index. One study has reported a sex-related DIF for “lifting” (18). Two previous studies have not observed any significant sex-related DIF (19, 20).

Overall, the knowledge on the presence of possible DIF in the NDI is very scarce and no strong inferences on the subject can be made. The objective of this study was to investigate if the responses to the NDI may produce some DIF comparing men and women. The hypothesis was that the NDI does not show any substantial sex-related DIF in the studied population.

Methods

The data were obtained from the register containing data on the patients undergoing cervical surgery of any kind between June 21, 2018 and August 17, 2021 in a university hospital. The patients responded to an online survey before the surgery. The exact number of missing responses were undetermined. To the best of our knowledge, either all or almost all the patients admitted to this particular surgery unit have responded to the survey. Of 392 patients, who have responded to a survey, 338 have responded to the NDI before the surgery. The survey contained questions on demographics and the severity of disability. A patient was included if the procedure code was one of the follows: NAG40, ABC60, ABC21, NAG41, ABC30, ABC10, NAG42 or ABC50, according to the Nordic Classification of Surgical Procedures (NCSP), version 1.15 (Table 1) The only exclusion criterion was multiple spinal procedures – all the patients were referred to their first cervical surgery.

The committee on research ethics at the Hospital District of South-West Finland allows to use register-based data, which is part of an electronic patient record system, by the researcher affiliated to the Hospital District without a separate specific approvement or a separate specific written informed consent. All the participants were informed and approved that data gathered by the register may be used for the purpose of scientific research.

Age was defined in full years at the time of surgery. Body mass index (BMI) was defined as a body weight divided by a squared height and expressed in kg/m^2. The duration of pain before surgery was defined as <6 weeks; 6-12 weeks; 3-12 months and >1 year. Pain intensity was assessed by using a visual analogue scale from 0 to 100 points with 0 indicating no pain and 100 indicating most possible pain.

The NDI is a questionnaire containing 10 items covering disability caused by neck pain: 1) pain intensity; 2) personal care; 3) lifting; 4) reading; 5) headaches; 6) concentration; 7) work;

8) driving; 9) sleeping and 10) recreation. Each item is assessed on a six-level ordinal scale with 'zero' describing 'no limitation' and 'five' describing 'extreme limitation or an inability to function'. The total score is a percentage calculated by the sum of all answers divided by 50 (the maximum possible number of points) and multiplied by 100 as follows: 'Total score = (∑item scores/50) x 100'. The equation is adjusted when the responses to one or more items are missing. A score of 0% represents the highest possible level of functioning and independence while a score of 100% represents the lowest level of functioning with total dependence.

Statistical analysis

Basic characteristics of the sample

The basic characteristics were presented as means, standard deviations (SDs), and percentage, when appropriate. Independent samples t-test and a chi-square test were used to investigate potential differences between men and women regarding their demographics.

Discrimination property

The DIF was assessed to investigate if the items of NDI behave differently when measuring disability level among men vs. women. The DIF is a subroutine of Item Response Theory (IRT). The IRT allows to define two main psychometric properties of a questionnaire – discrimination and difficulty parameters. A discrimination parameter describes the sensitivity of test to differentiate severity levels of symptoms as a regression curve. The steeper the curve, the more discriminative the test is. Ideally, the steepest interval corresponds to a disability level, which is average in a studied population. 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) (21).

Difficulty property

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 disability (comparing with the average level of studied population) in order to achieve a 0.5 probability to give a particular answer. In the ideal situation, zero difficulty can be located at the middle of scale – for the NDI with its six-level grading system (from zero to five) it places at response "two" or "three". If such an ideal situation appears, then difficulty estimates for items "zero" and "one" would carry a minus sign (less severe disability than on average in the studied population), while responses "four" or "five" would be positive (more severe disability than on average).

DIF calculation

The DIF measures the difference in two aforementioned properties between two groups, here sexes. The probit logistic regression was used to test whether an item exhibits either uniform or nonuniform DIF between sex 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 (22, 23). 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 to a midpoint and worse than women after that). A two-tailed *p*-value =<0.05 indicated a significant difference between sexes. The results of DIF analysis were also presented and evaluated graphically as item and test information curves using the graded response model (GRM) of the IRT. Test (or its single item) information is the inverse variance of test (or its item). In other words, it is 1/variance – smaller variability leads to better precision of the results. All the data analyses were performed utilizing STATA 16 (College Station, Texas, U.S.).

Results

Basic characteristics of the sample

A total of 338 patients completed preoperative surveys including the NDI (table 1). There were 171 (51%) women with mean age of 53.7 (SD 10.8) years and 167 (49%) men with mean age of 54.2 (SD 11.0) years. Mean body mass index was 28.0 (SD 5.0) kg/m2. Of the patients, 49 (15%) had experienced neck pain for less than three months, 111 (35%) over three months and 160 (50%) over one year. The average NDI score was 44.3 (SD 17.0) preoperatively. Of 338 procedures, 242 (72%) was anterior fusion of cervical spine without fixation. The most frequent reasons for the surgery were "M50 Cervical disc disorders" (39%) and "M47 Spondylosis" (35%). Women experienced more pain than men: 58.7 (SD 27.7) vs. 48.5 (SD 28.1) points, the difference was statistically significant (p=0.012). Also, the composite score of the NDI was slightly higher among women: 47.2 (SD 16.8) vs. 41.3 (SD 16.7) points (p=0.001). Other differences between sex groups were statistically insignificant.

Difficulty parameter for the entire sample

Overall, the difficulty estimates for most of the NDI items showed a perfect situation with zero placed around responses "two" and "three", negative values for responses below "two" and positive estimates for responses over "three" (table 2). This "ideal state" also resulted in the almost perfectly shaped curve of test information function – the peak of information could be found at the average level of disability (figure 1). One item "personal care" slightly diverged from that "perfect" situation – only responses "zero" and "one" were negative. In other words, for this item, respondents with severe restrictions of functioning might grade their perceived disability as milder than an actual level compared to the average in the sample. All the estimates were statistically significant with 95% CIs not including zero.

Discrimination parameter for the entire sample

The discrimination parameter was high or perfect for seven out of 10 items. Items "lifting" (1.32 [95% CI 1.05 to 1.60]), "headaches" (0.71 [95% CI 0.50 to 0.93]) and "sleeping" (1.12 [95% CI 0.87 to 1.37]) showed only moderate discrimination abilities. All the estimates were statistically significant with 95% CIs not including zero.

Differential item functioning by sex

Some differences between sexes can be observed in tables 2 and 3. Figure 2 shows these differences in an easier form. While the differences could be seen for all 10 items, only three items demonstrated statistically significant DIF – "pain intensity", "headaches" and "recreation". The DIF curves for the item "headaches" were uniform – the distance between two curves of similar shape remained the same for the entire spectrum of disability variance. Instead, the DIF for "pain intensity" and "recreation" was nonuniform – the curves for men and women crossed each other at different levels of disability. These deviations were, however, only slight. The DIF curves for "pain intensity" were very alike regarding both difficulty and discrimination (steepness of the curve) properties. The DIF curves for "headaches" showed significantly different amount of information, which can be obtained of this NDI item. This item "headaches" was significantly more precise for women than for men. While the shapes of the DIF curves for "recreation" were alike, the difference appeared especially in the steepness of these curves – the discrimination ability of this item was better for women than for men with estimates of 2.80 vs. 2.04 (almost 40% more), as shown in table 3. While other seven items did not show statistically significant DIFs, the same trend of better discrimination (steeper curves) for women could be graphically observed for "personal care", "lifting", "work", "driving" and "sleeping". The DIF curve was steeper among men for only one item "concentration".

Discussion

This register-based study among 338 patients referred to a department of surgery for cervical operation investigated the possible DIF of the NDI questionnaire based on sexes. Overall, the difficulty estimates for most of the NDI items showed a perfect functioning of the NDI. Also, the discrimination parameter was high or perfect for seven out of 10 items. All the estimates were statistically significant. While only three items demonstrated statistically significant estimates – "pain intensity", "headaches" and "recreation", there were some DIF for all 10 items. The DIFs were uniform or almost uniform in all the cases. In general, the NDI items showed better discrimination abilities among women than among men. For some items, also the precision of information, obtainable of the NDI, was better for women.

The DIF have been studied for many other patient-reported outcome measures of disability, often resulting in the sex-related DIF of diverse magnitude (24-26). Only a few previous studies have reported on the DIF of the NDI. A study from Canada observed the DIF for a single item "lifting" based on grouping by a clinically important change (7). Similarly to the present results, a study from Netherlands reported a significant sex-related DIF for "headaches" item (4). That study has been executed on small sample of 10 patients. A study from Norway has observed a significant DIF for "work status" among 249 patients with chronic neck pain (5). Another study from Canada has reported a sex- and age-related DIF for "lifting" among 521 patients with neck pain (6).

Additionally, some indirect comparison between the present results and previous knowledge can be made of the reports on the DIF of Oswestry Disability Index (the NDI is a modification of that scale). A study from Denmark among 800 patients with low back pain has observed a sex-related DIF for "lifting", with women scoring higher on this item compared with men. The same study has also reported some other DIFs based on age and diagnoses for items "sitting",

214 "walking" and "sleeping" (18). A study from Taiwan has found no sex-related DIF for any of
215 the items of the Oswestry Disability Index. Though, there has been significant age-related DIF
216 for "sleeping" (20). A small study from Australia among 100 patients has not observed a sex-
217 related DIF, though a significant age-dependent DIF has been reported for "walking" (19).

218 Sex-related differences in responses concerning restrictions induced by neck pain or headache
219 seen in this study may be explained by a well-known difference between men and women in
220 perception of pain. Previous reports have consistently demonstrated greater pain prevalence
221 among women compared to men across multiple geographic regions (17, 27). Compared to
222 men, women tend to report more severe pain and it is more anatomically diffuse and longer-
223 lasting. Women are also more likely to visit a physician because of pain than men. Underlying
224 reasons for these differences have been searched from differences in hormonal factors related
225 to pain response, and from various psychosocial mechanisms and sociocultural beliefs as e.g.,
226 the influence of stereotypical social sex roles on pain response and threshold differences (17,
227 28-32). Similarly, several health-related, behavioral and sociodemographic factors have been
228 found to contribute to the higher prevalence of disability in women compared to men (33).

229 The reasons for the sex-related differences observed in the NDI functionality in regards
230 recreation activities can only be explained by speculations. Men and women may have different
231 needs concerning recreational activities, as well as they may favor different types of these
232 activities (34). Additionally, these differences may change over time when ageing (34).
233 According to the U.S. Bureau of Labor Statistics, men spend more time in leisure activities (5.5
234 hours) than did women (4.9 hours) (35). Also, report from Australia has suggested that women
235 are participating less than their male counterparts in leisure-time activities.

236 The generalizability of the results might be diminished by several issues. Most of the
237 respondents belonged to a particular age group around 55 years. It is, thus, possible that

populations with different age distribution might demonstrate different results. The respondents were patients referred to a single highly specialized university clinic, which suggest that the results might not stand e.g., in primary care. The exact number of missing responses were undetermined. To the best of our knowledge, either all or almost all the patients admitted to this particular surgery unit have responded to the survey. Despite the fact that the study sample was big enough to produce some statistically significant results, the sample of less than 400 might not be sufficient to detect all possible variances, especially when applying such a "size-greedy" method as the IRT. Relatively small sample size could be the reason why the DIF was not statistically significant even if its presence seemed very probable based on the visual examination of curves for several items. This study did not assess the age-related or other possible DIFs of the NDI. Due to the particular specifics of the study site, the dispersion of demographics other than sex was too narrow to form any other groups without substantial loss in a group size. However, the study sample was sufficient, at least, for this sex-related grouping, and the sex and age of the respondents were equally distributed.

Further research on the DIF of the NDI is needed. That research should be conducted on big samples of people with neck pain with different age and sex distribution, different settings and diverse language versions. Also, prospective repeated-measures studies may reveal a possible dynamic in DIF over time.

Conclusions

It seemed that the NDI may behave differently depending on the sex of respondents. Several items of the NDI may be more precise and more sensitive when detecting restrictions in functioning among women compared to men. This finding should be taken into account when using the NDI in research and clinical practice.

Data Availability

The datasets generated during and/or analyzed during the current study are not publicly available, but are available from the corresponding author on reasonable request.

References

1. Vernon H. The Neck Disability Index: state-of-the-art, 1991-2008. J Manipulative Physiol Ther. 2008;31(7):491-502. doi: 10.1016/j.jmpt.2008.08.006. PubMed PMID: 18803999.

2. Young Ia Pt D, Dunning J Pt DPT, Butts R Pt P, Mourad F Pt DPT, Cleland Ja Pt P. Reliability, construct validity, and responsiveness of the neck disability index and numeric pain rating scale in patients with mechanical neck pain without upper extremity symptoms. Physiother Theory Pract. 2019;35(12):1328-35. Epub 20180601. doi: 10.1080/09593985.2018.1471763. PubMed PMID: 29856244.

3. Pietrobon R, Coeytaux RR, Carey TS, Richardson WJ, DeVellis RF. Standard scales for measurement of functional outcome for cervical pain or dysfunction: a systematic review. Spine (Phila Pa 1976). 2002;27(5):515-22. doi: 10.1097/00007632-200203010-00012. PubMed PMID: 11880837.

4. Ailliet L, Knol DL, Rubinstein SM, de Vet HC, van Tulder MW, Terwee CB. Definition of the construct to be measured is a prerequisite for the assessment of validity. The Neck Disability Index as an example. J Clin Epidemiol. 2013;66(7):775-82; quiz 82 e1-2. Epub 2013/04/27. doi: 10.1016/j.jclinepi.2013.02.005. PubMed PMID: 23618795.

5. Johansen JB, Andelic N, Bakke E, Holter EB, Mengshoel AM, Roe C. Measurement properties of the norwegian version of the neck disability index in chronic neck pain. Spine (Phila Pa 1976). 2013;38(10):851-6. Epub 2012/12/04. doi: 10.1097/BRS.0b013e31827fc3e9. PubMed PMID: 23202354.

6. van der Velde G, Beaton D, Hogg-Johnston S, Hurwitz E, Tennant A. Rasch analysis provides new insights into the measurement properties of the neck disability index. Arthritis Rheum. 2009;61(4):544-51. Epub 2009/04/01. doi: 10.1002/art.24399. PubMed PMID: 19333989.

7. Walton DM, MacDermid JC. A brief 5-item version of the Neck Disability Index shows good psychometric properties. Health Qual Life Outcomes. 2013;11:108. Epub 2013/07/03. doi: 10.1186/1477-7525-11-108. PubMed PMID: 23816395; PubMed Central PMCID: PMCPMC3718697.

8. Bohannon RW. Comfortable and maximum walking speed of adults aged 20—79 years: reference values and determinants. Age and Ageing. 1997;26(1):15-9. doi: 10.1093/ageing/26.1.15.

9. Massy-Westropp NM, Gill TK, Taylor AW, Bohannon RW, Hill CL. Hand Grip Strength: age and gender stratified normative data in a population-based study. BMC Research Notes. 2011;4(1):127. doi: 10.1186/1756-0500-4-127.

10. Nunn A.J GI. New regression equations for predicting peak expiratory flow in adults. BMJ 1989(298).

11. Dewitt R. 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. 2013;15(1):4. Epub 22 Mar 2013. doi: 0.1080/10671188.1944.106248151.

12. Gelin M and Zumbo B. Differential item functioning results may change depending on how item is scored: An illustration with the center for epidemiologicstudies depression scale. Educational and Psychological Measurement. 2003;63(1):10. doi: 10.1177/0013164402239317.

13. Fleishman J SWaAB. Impact of Differential Item Functioning on Age and Gender Differences in Functional Disability. Journal of Gerontology: SOCIAL SCIENCES. 2002;57B(5):10.

14. Merril S ST, Kasl S and Bergman L,. Gender Differences in the Comparison of Self-Reported Disability and Performance Measures. Journal of Gerontology: MEDICAL SCIENCES. 1997;52A(1):8.

15. Bingefors K, Isacson D. Epidemiology, co-morbidity, and impact on health-related quality of life of self-reported headache and musculoskeletal pain--a gender perspective. Eur J Pain. 2004;8(5):435-50. Epub 2004/08/25. doi: 10.1016/j.ejpain.2004.01.005. PubMed PMID: 15324775.

16. Pickering G DJ, Eschalier A, Dubray C,. Impact of Age, Gender and Cognitive Functioning on Pain Perception. Gerontology. 2002;48:7.

17. Bartley E. FR. Sex differences in pain: a brief review of clinical and experimental findings. Brittish Journal of Anaesthesia. 2013;111(1):52-8. Epub 2013/06/26. doi: 10.1093/bja/aet127. PubMed PMID: 23794645; PubMed Central PMCID: PMCPMC3690315.

18. Comins J, Brodersen J, Wedderkopp N, Lassen MR, Shakir H, Specht K, et al. Psychometric Validation of the Danish Version of the Oswestry Disability Index in Patients With Chronic Low Back Pain. Spine (Phila Pa 1976). 2020;45(16):1143-50. Epub 2020/03/25. doi: 10.1097/BRS.0000000000003486. PubMed PMID: 32205707.

19. Davidson M. Rasch analysis of three versions of the Oswestry Disability Questionnaire. Man Ther. 2008;13(3):222-31. Epub 2007/03/17. doi: 10.1016/j.math.2007.01.008. PubMed PMID: 17363319.

20. Lu YM, Wu YY, Hsieh CL, Lin CL, Hwang SL, Cheng KI, et al. Measurement precision of the disability for back pain scale-by applying Rasch analysis. Health Qual Life Outcomes. 2013;11:119. Epub 2013/07/23. doi: 10.1186/1477-7525-11-119. PubMed PMID: 23866814; PubMed Central PMCID: PMCPMC3717282.

21. Baker FB. The basics of item reponse theory. 2nd ed. USA: ERIC Clearinghouse on Assessment and Evaluation; 2001.

22. de Boeck P WM. Explanatory Item Response Models: Springer-Verlag New York; 2004.

23. Swaminathan H and Rogers HJ. Detecting differential item functioning using logistic regression procedures. Journal of Educational Measurement 1990;27:361-70.

24. Kalkers NF, Galan I, Kerbrat A, Tacchino A, Kamm CP, O'Connell K, et al. Differential item functioning of the Arm function in Multiple Sclerosis Questionnaire (AMSQ) by language, a study in six countries. Mult Scler. 2021;27(1):90-6. Epub 2019/12/18. doi: 10.1177/1352458519895450. PubMed PMID: 31845614.

25. Parker DJ, Werth PM, Christensen DD, Jevsevar DS. Differential item functioning to validate setting of delivery compatibility in PROMIS-global health. Qual Life Res. 2022;31(7):2189-200. Epub 2022/01/21. doi: 10.1007/s11136-022-03084-4. PubMed PMID: 35050447.

26. Penton H, Dayson C, Hulme C, Young T. An Investigation of Age-Related Differential Item Functioning in the EQ-5D-5L Using Item Response Theory and Logistic Regression. Value Health. 2022;25(9):1566-74. Epub 2022/04/30. doi: 10.1016/j.jval.2022.03.009. PubMed PMID: 35487819.

27. Wijnhoven HA, de Vet HC, Picavet HS. Explaining sex differences in chronic musculoskeletal pain in a general population. Pain. 2006;124(1-2):158-66. Epub 20060522. doi: 10.1016/j.pain.2006.04.012. PubMed PMID: 16716517.

28. Samulowitz A, Gremyr I, Eriksson E, Hensing G. "Brave Men" and "Emotional Women": A Theory-Guided Literature Review on Gender Bias in Health Care and Gendered Norms towards Patients with Chronic Pain. Pain Res Manag. 2018;2018:6358624. Epub 20180225. doi: 10.1155/2018/6358624. PubMed PMID: 29682130; PubMed Central PMCID: PMCPMC5845507.

29. Hurley RW, Adams MC. Sex, gender, and pain: an overview of a complex field. Anesth Analg. 2008;107(1):309-17. doi: 10.1213/01.ane.0b013e31816ba437. PubMed PMID: 18635502; PubMed Central PMCID: PMCPMC2715547.

30. Pieretti S, Di Giannuario A, Di Giovannandrea R, Marzoli F, Piccaro G, Minosi P, et al. Gender differences in pain and its relief. Ann Ist Super Sanita. 2016;52(2):184-9. doi: 10.4415/ann_16_02_09. PubMed PMID: 27364392.

31. Bernardes SF, Keogh E, Lima ML. Bridging the gap between pain and gender research: a selective literature review. Eur J Pain. 2008;12(4):427-40. Epub 20071023. doi: 10.1016/j.ejpain.2007.08.007. PubMed PMID: 17936655.

32. Keogh E. The gender context of pain. Health Psychol Rev. 2021;15(3):454-81. Epub 20200908. doi: 10.1080/17437199.2020.1813602. PubMed PMID: 32875959.

33. Leveille SG, Resnick HE, Balfour J. Gender differences in disability: evidence and underlying reasons. Aging (Milano). 2000;12(2):106-12. doi: 10.1007/bf03339897. PubMed PMID: 10902052.

34. Sjogren K, Stjernberg L. A gender perspective on factors that influence outdoor recreational physical activity among the elderly. BMC Geriatr. 2010;10:34. Epub 2010/06/10. doi: 10.1186/1471-2318-10-34. PubMed PMID: 20529318; PubMed Central PMCID: PMCPMC2891794.

35. The U.S. Bureau of Labor Statistics. TED: The Economics Daily image 2020 [January 18, 2022]. Available from: https://www.bls.gov/opub/ted/2020/men-spent-5-point-5-hours-per-day-in-leisure-activities-women-4-point-9-hours-in-2019.htm.

Figure legends

Figure 1. Test information function for the NDI composite score among all 338 respondents included. Test information is an inverse variance representing the preciseness of the scores.

Figure 2. Test information functions for the NDI items by sex. Test information is an inverse variance representing the preciseness of the scores.

* Statistically significant difference between sexes (p<=0.005)