Assessing measurement invariance of MSQOL-54 across Italian and English versions

(Article begins on next page)

15 July 2024

Full title: Assessing measurement invariance of MSQOL-54 across Italian and English versions

Authors

Andrea Giordano[1,2], Silvia Testa[2], Marta Bassi[3], Sabina Cilia[4], Antonio Bertolotto[5], Maria Esmeralda Quartuccio[6], Erika Pietrolongo[7], Monica Falautano[8], Monica Grobberio[9], Claudia Niccolai[10,11], Beatrice Allegri[12], Rosa Gemma Viterbo[13], Paolo Confalonieri[14], Ambra Mara Giovannetti[1,14], Eleonora Cocco[15,16], Maria Grazia Grasso[17], Alessandra Lugaresi[18,19], Elisa Ferriani[20], Ugo Nocentini[21,22], Mauro Zaffaroni[23], Alysha De Livera[24], George Jelinek[24], Alessandra Solari[1], Rosalba Rosato[2]

1. Unit of Neuroepidemiology, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milano
2. Department of Psychology, University of Turin, Turin, Italy
3. Department of Biomedical and Clinical Sciences L. Sacco, Università di Milano, Milano, Italy
4. Multiple Sclerosis Center, University Polyclinic Hospital G. Rodolico, Catania, Italy
5. Neurology & Regional Referral Multiple Sclerosis Centre (CReSM), University Hospital San Luigi Gonzaga, Orbassano, Italy
6. Department of Neuroscience, San Camillo-Forlanini Hospital, Rome, Italy
7. Department of Neurosciences, Imaging and Clinical Sciences, University G. d'Annunzio, Chieti, Italy
8. Servizio di Psicologia e Neuropsicologia, UO di Neurologia e Riabilitazione Specialistica, San Raffaele Hospital, Milan, Italy
9. Lab. of Clinical Neuropsychology, Psychology Unit, ASST Lariana, Como, Italy
10. IRCCS Fondazione Carlo Gnocchi, Florence, Italy
11.Department of NEUROFARBA, Section of Neurosciences, University of Florence, Florence, italy
12. Multiple Sclerosis Center, Neurology Unit, Hospital of Vaio, Fidenza, Italy
13. Department of Basic Medical Sciences, Neurosciences and Sense Organs, University of Bari, Bari, Italy
14. Unit of Neuroimmunology and Neuromuscular Diseases, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy
15. Department of Medical Science and Public Health, University of Cagliari, Cagliari, Italy
16. Multiple Sclerosis Center, ASSL Cagliari; ATS Sardegna, Italy
17. Multiple Sclerosis Unit, IRCCS S. Lucia Foundation, Rome, Italy
18. UOSI Riabilitazione Sclerosi Multipla, IRCCS Istituto delle Scienze Neurologiche di Bologna, Bologna, Italy
19. Dipartimento di Scienze Biomediche e Neuromotorie, Università di Bologna, Bologna, Italy
20. UOC Psicologia Ospedaliera, AUSL di Bologna, Bologna, Italy
21. Department of Systems Medicine, University of Rome "Tor Vergata", Rome, Italy
22. Neurology and Neurorehabilitation Unit, IRCCS S. Lucia Foundation, Rome, Italy
23. Multiple Sclerosis Centre, ASST Valle Olona, Gallarate, Italy
24. Neuroepidemiology Unit, Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Australia

Corresponding author

Alessandra Solari, MD

Unit of Neuroepidemiology

Fondazione IRCSS Istituto Neurologico Carlo Besta

Via Celoria 11,

20133 Milan, Italy

E-mail: alessandra.solari@istituto-besta.it

Tel. +39 02 23944664; Fax +39 02 23944056

PURPOSE. The MSQOL-54 is a specific multiple sclerosis (MS) health-related quality of life inventory consisting of 52 items organized into 12 subscales plus two single items. No study was found in literature assessing its measurement invariance across language versions. We investigated whether MSQOL-54 items provide unbiased measurements of underlying constructs across Italian and English versions.

METHODS. Three constrained levels of measurement invariance were evaluated: configural invariance where equivalent number of factor/factor patterns were required; metric invariance where equivalent factor loadings were required; and scalar invariance where equivalent item intercepts between groups were required. CFI, RMSEA and SRMR fit indices and their changes between nested models were used to assess tenability of invariance constraints.

RESULTS. Overall the dataset included 3669 MS patients: 1605 (44%) Italian, mean age 41 years, 62% women, 69% with mild level of disability; 2064 (56%) English-speaking (840 [41%] from North-America, 797 [39%] from Australasia, 427 [20%] from UK & Ireland), mean age 46 years, 83% women, 54% with mild level of disability. The configural invariance model showed acceptable fit (RMSEA = 0.052, CFI = 0.904, SRMR = 0.046); imposing loadings and intercepts equality constraints produced negligible worsening of fit (Δ RMSEA < 0.001, Δ CFI = -0.002, Δ SRMR = 0.002 for metric invariance; Δ RMSEA = 0.003, Δ CFI = -0.013, Δ SRMR = 0.003 for scalar invariance).

CONCLUSIONS. These findings support measurement invariance of the MSQOL-54 across the two language versions, suggesting that the questionnaire has the same meaning in the Italian and English versions.

INTRODUCTION

As many as 2.2 million people worldwide live with multiple sclerosis (MS), a chronic disabling neurological disease primarily affecting young adults [1,2]. About 80% of persons with MS (PwMS) are initially diagnosed with the relapsing-remitting form of the disease, and about 50% of them eventually develop a secondary progressive form 15 years after diagnosis. Uncertain prognosis and modest efficacy of current treatments make MS a particularly difficult disease to adjust to. MS is associated with diverse symptoms, such as fatigue, pain, depression and cognitive dysfunction, deeply affecting a variety of functioning domains in patients' lives which health care professionals often fail to detect [5,6]. Consequently, the concept of health-related quality of life (HRQOL) has increasingly received researchers' attention with the aim to incorporate in established outcome measures the assessment of those domains that are not manifest during patient-physician consultations, but are of great interest for PwMS [3,4]. In the 1990s, the first MS-specific HRQOL instruments have been published [7,8]. One of these was the Multiple Sclerosis Quality of Life-54 (MSQOL-54), which immediately gained popularity and extensive application. It was originally devised in US English, and subsequently validated in several languages [9-15]. The Italian version of MSQOL-54 was published in 1999, and is currently the most widely-used HRQOL instrument in Italy [9].

Although previous studies have investigated its reliability and validity, its measurement invariance properties have not yet been well examined.

Measurement invariance is a relevant statistical property of an instrument attesting that the same latent construct ('test structure') is measured across time or across groups [16]. Unless measurement invariance has been demonstrated, it is not possible to perform meaningful cross-group comparisons. Pooling data across samples collected in different countries with different languages may be problematic, as specific cultural beliefs and expectations may affect items

interpretation and differences in observed scores may thus not reflect actual differences in latent variables. Lack of measurement invariance across versions - as well as across cultural contexts - can be due to different reasons, such as poor translation or items that are not applicable across cultures, elicit further concepts or present ambiguous nuances [17].

Only few studies have assessed measurement invariance of instruments applied in MS [18-23]. Among these, the majority investigated measurement invariance across groups, with small sample sizes, and analyzed data using multi-group confirmatory factor analysis [19-23].

To the best of our knowledge no study has evaluated measurement invariance of MSQOL-54 across language versions. Thus, the aim of the present study was to assess the measurement invariance of MSQOL-54 across Italian and English language versions. Considering that most recent studies found evidence of partial invariance in HRQOL instruments [24-26], and that Italian and English are western languages, we expected that full or at least partial invariance would hold across the two language versions.

METHODS

Participants

Data originated from different datasets that were collected with the English and Italian versions of MSQOL-54 within ongoing or completed research projects carried out in Australia and Italy.

Data collected with the English version were obtained from the 'HOLISM study': This was an observational international study coordinated by Australian researchers (methods and results were described in [27, 28]. In brief, participants from Australasia, Europe, North America, and other countries were recruited via online platforms, including social media, websites and forums involving PwMS. The study provides a snapshot of current lifestyle and risk-modifying behaviors of a large international group of PwMS, as well as an ongoing platform for analyzing the association

5

between these variables and disease progression. In the present study we used data from English-speaking PwMS only: 840 (41%) from North-America, 797 (39%) from Australasia, and 427 (20%) from UK & Ireland.

Data collected with the Italian version were obtained from the following sources:

- The 'Care system project' [29,30], an observational research about PwMS' perceived levels of ill-being and well-being (overall, 662 PwMS from 8 MS centers).

- The study 'An abbreviated computerized version of the MSQOL-54: Development and preliminary validation using Confirmatory Factor Analysis and Item Response Theory' [31, 33, 34], which devised an abbreviated version of the MSQOL-54. We used data from 564 PwMS (5 MS centers) who participated in the retrospective phase of the study [31].

- Other research projects carried out in 5 Italian MS centers (overall 379 PwMS).

All these projects were approved by local ethics committees (St Vincent's Hospital Melbourne Human Research Ethics Committee [LRR 055/12]; Università di Milano; San Raffaele Hospital, Milano; University Polyclinic Hospital G. Rodolico, Catania; University of Florence; S. Anna Hospital, Como; Hospital of Vaio-Fidenza, Fidenza; University 'G. d'Annunzio', Chieti; University of Bari; San Camillo-Forlanini Hospital, Rome; University Hospital 'San Luigi Gonzaga', Orbassano; Fondazione IRCCS Istituto Neurologico 'C. Besta', Milano; IRCCS S. Lucia Foundation, Rome). Patients gave written informed consent to be included in the original projects. Additional consent was not required for this secondary analysis, for which patients' privacy and anonymity were guaranteed.

Database set up

6

Data quality check was performed. This included a search for possible multiple imputations from the same patient. Records with the same date of birth and sex were searched for, both within and across datasets, and duplicates were removed.

Database records were only eligible if MS was diagnosed according to McDonald's [35]/McDonald's revised criteria [36] (Italian version), or if MS diagnosis was confirmed by a medical doctor (English version), and if patient's age ≥18 years, gender, disease duration, Expanded Disability Status Scale (EDSS) [37] (Italian version only), and Patient Determined Disease Steps scores (PDDS [38]) were available. We included records when more than 67% of the MSQOL-54 items were completed.

Instrument

The MSQOL-54 comprises 36 generic items derived from the Short Form-36 (SF-36) [39] and 18 additional MS-specific items derived from professionals' opinion and literature review [7]. The 54 items are organized into 12 multi-item and two single-item subscales (Online Resource 1). These enquire about HRQOL over the previous month, except item 2 (Change in Health) which refers to the preceding year. As for the SF-36, two composite scores (Physical Health Composite, PHC, and Mental Health Composite, MHC) are derived by combining scores of the relevant subscales [7]. The MSQOL-54 has well documented validity in terms of content, constructs, reliability, discrimination [9, 12,15], and responsiveness [40].

Analysis

Variables were summarized using both counts and percentages, means and standard deviations (SD), or medians and minimum-maximum ranges. Categorical variables were compared using chi-square or Fisher's exact test, and continuous variables using unpaired t-test or Wilcoxon rank-sum

test (between-group comparisons), and paired t-test or Wilcoxon matched-pairs signed-ranks test (within-group comparisons), as appropriate.

We used confirmatory factor analysis with maximum likelihood estimation with robust standard errors (MLR) to separately assess whether the data from the two language versions fitted the (original) MSQOL-54 12-factor model, and then to assess measurement invariance across the two language versions.

Three increasingly constrained levels of measurement invariance were assessed via multi-group confirmatory factor analysis. First, we tested configural invariance which tests if the same pattern of loadings exists across the groups under investigation (i.e. Italian and English language versions), requiring that the same items have non-zero loadings on the same factors. Second, we tested metric invariance which requires that unstandardized factor loadings be the same across groups. Finally, we tested scalar invariance which requires meeting the assumptions of configural and metric invariance, and that unstandardized item intercepts be invariant across groups [41].

We considered the model fit acceptable if the following criteria were met: root mean square error of approximation (RMSEA) <0.08; comparative fit index (CFI) >0.90; and standardized root mean square residual (SRMR) <0.08 [42, 43]. According to Chen [44], a worsening of CFI that exceeds the threshold of 0.010, supplemented by a change of ≥ 0.015 in RMSEA or a change of ≥ 0.030 in SRMR was considered as indication of absence of metric invariance; when testing scalar invariance, the cut-off value for CFI and RMSEA were the same as for metric invariance, while it was 0.010 for SRMR. We did not rely on the χ2 difference test that is typically used to compare the fit of two nested models, as it is sensitive to sample size and thus tends to give significant results with moderate to large sample sizes [45].

Inspired by Bebber et al. [46], we compared individual factor scores obtained from the unconstrained model (i.e. configural invariance) with those obtained from the more constrained

model (i.e. scalar invariance), in order to investigate the practical consequences of imposing equality constrains on model parameters across the two groups. More in detail, within each group we transformed the factor scores obtained under the two models (configural and scalar) into T-scores having a mean of 50 and a SD of 10, and for each respondent we calculated the absolute difference between the two T-scores. We expected that all the absolute differences be close to zero if the constrained model fitted the data well.

In addition, a sensitivity analysis was conducted to account for possible selection biases by assessing measurement invariance: a) across English-speaking geographic areas (Australasia/North-America/UK & Ireland); b) across two sub-samples (N=985 each) matched for gender, age (18-30 years, 31-40, 41-50, 51-60, 61+), level of disability, and disease duration (0-11 years, 12-23, 24+), by using 1:1 coarsened exact matching [47]. These stratification variables were selected because previous research indicated they are associated with differences in the conceptualization of HRQOL and other patient-reported outcomes [48].

All analyses were performed with Stata Statistical Software, release 12.0 (Stata Corp LP, College Station, USA), and Mplus software 7.0 [49].

RESULTS

Descriptive analysis

The original dataset (including the two language versions) comprised 3877 PwMS. Of those, 37 were excluded as they were duplicates, 96 because they did not complete any MSQOL-54 item, and 75 because they completed less than 67% of the items. Out of the 3669 included PwMS, 1605 (44%) were Italian (mean age 41 years, 62% women, 69% with a mild disability level) and 2064 (56%) were English-speaking (840 [41%] from North-America, 797 [39%] from Australasia, 427 [20%] from UK and Ireland), mean age 46 years, 83% women, 54% with a mild disability level).

Compared to Italians, English-speaking participants were older and had longer disease duration (p< 0.001) (Table 1).

Measurement invariance

The (original) 12-factor model of the MSQOL-54 was estimated separately in the two language versions, obtaining good fit indices for RMSEA and SRMR (Italian: RMSEA=0.050; SRMR=0.045; English: RMSEA=0.054; SRMR=0.047), and an acceptable value for CFI (Italian: CFI=0.906; English: CFI=0.903). As shown in Table 2, the model assessing the first level of measurement invariance (i.e. configural) produced analogous results to those in the separate samples: good fit indices for RMSEA and SRMR and a less satisfactory, but still acceptable, value for CFI. As for the model in which the loadings were constrained to be equal across groups, the fit indices were acceptable and the worsening with respect to the unconstrained model (configural) was negligible (ΔRMSEA<0.001; ΔCFI=-0.002, ΔSRMR=0.002), thus supporting the metric invariance of the instrument. Finally, when both loadings and intercepts were constrained to be equal across groups (scalar invariance), the model fitted the data well in terms of RMSEA and SRMR, and CFI was slightly under the cut-off of 0.90. Concerning the changes in fit indices as compared to the metric invariance model, the cut-off values were reached, except for ΔCFI (ΔRMSEA=0.003; ΔCFI=-0.013, ΔSRMR=0.003), thus supporting scalar invariance.

Further evidence of measurement invariance across the two language versions was obtained by comparing the factor scores derived from the unconstrained model (configural invariance) with those derived from the more constrained model (scalar invariance). As shown in Table 3, the absolute difference between individuals' T scores never exceeded 2 T-points; the maximum difference (1.72 T points) was observed for the Pain subscale in the Italian sample. Furthermore, the number of participants with an absolute difference greater than 1 T-point was very low,

ranging from 1 (Sexual function and Overall Quality of Life subscales in the English-speaking sample, and Emotional Wellbeing and Overall Quality of Life subscales in the Italian-speaking sample) to 19 (the Pain subscale in the Italian sample).

Sensitivity analysis

Measurement invariance was also assessed across English-speaking geographic areas (North-America/Australasia/UK & Ireland). Results supported configural, metric, and scalar invariance across the three subgroups, indicating that the loadings and intercepts of the MSQOL-54 items can be considered equal across the different English-speaking areas (Online Resource 2).

Results from the matched-pairs subgroup analysis supported configural, metric, and scalar measurement invariance, indicating that the results of the main analysis reported in Table 2 were not biased by the demographic and clinical differences across the language version samples (Online Resource 3).

DISCUSSION

Measurement invariance is an important prerequisite for meaningful group comparisons. To the best of our knowledge, this is the first study investigating the measurement invariance of the MSQOL-54 across language versions. Findings support the measurement invariance of the English and Italian MSQOL-54, suggesting that the questionnaire has the same meaning across languages, and also that individuals who have the same score on a MSQOL-54 domain would obtain the same value on the observed variable, irrespective of the language version.

In the sensitivity analysis we found that measurement invariance was further supported across English-speaking countries, which is quite important considering that the original US English

version of the MSQOL-54 was used in all these countries. Further, measurement invariance was supported across subgroups matched for age, sex, level of disability, and disease duration.

All in all, these findings indicate that the MSQOL-54 can be used to assess HRQOL among both Italian- and English-speaking PwMS. They further demonstrate that it is possible to pool data together or compare scores between these two language groups (and within English-speaking groups) obtaining meaningful interpretations. Any perceived similarities or differences in HRQOL levels between Italian- and English-speaking PwMS would therefore indicate true similarities or differences. Notably, the (original) US English version (used with English-speaking participants from the 'HOLISM study') and the Italian version, which has been linguistically validated according to international guidelines, can be considered as culturally equivalent. The UK English version of the questionnaire has not yet been validated. However, it is not always feasible to validate an instrument in each target language group, so its validity in our populations is encouraging and produces evidence to support using the MSQOL-54 in other English-speaking populations.

As far as the methods of analysis are concerned, we chose multi-group confirmatory factor analysis because it is one of the most powerful analytical approach in cross-cultural research. Given the response structure of some MSQOL-54 items (i.e. 2/3/4/5/6 response options), an estimation method for ordered response categories (e.g. weighted least square mean and variance adjusted estimator [WLSMV] using the polychoric correlation) would have been more appropriate [41]. However, no statistical methods other than χ^2 are currently available to assess the measurement invariance between nested models when WLSMV is employed. Criteria for changes in CFI and other goodness of fit (GFI) indices have not yet been set, and the few studies addressing this issue suggest the users avoid interpreting the changes in GFI, especially for mis-specified models [49]. Moreover, in the present study we used a large dataset and it is known that the χ^2 test statistic is sensitive to sample size, such that it tends to yield significant results [45].

This study has some limitations. First, differences must be acknowledged in the recruitment strategies adopted to gather Italian and English data. Particularly, Italian data stem from research projects where clinical information was provided by investigators. By contrast, English data were derived from an online survey requiring a high level of literacy of participants. Moreover, higher levels of physical disability may have prevented some PwMS from participating and completing the survey without support. Further, some PwMS were directly recruited through a website and associated forums promoting lifestyle changes: This may have facilitated the participation of individuals with a specific interest in this topic. In spite of these differences, our results globally support the robustness of the questionnaire.

Second, in the two datasets disability level was assessed using different scales, the EDSS in Italy and the PDSS in the English-speaking population. To overcome this issue, EDSS scores were transformed into PDDS levels [38, 51,52], improving the completeness of the data collected.

Third, other potential variables (such as level of education, employment, and disease form) were not available in the two original datasets; we therefore could not take them into account in data analysis.

To conclude, results from this study further support the inclusion of the MSQOL-54 as a patient-reported outcome measure (PROM) in clinical practice and research involving both Italian- and English-speaking PwMS. Moreover, findings show that data gathered with these language versions can be suitable for group comparisons and can be pooled together to obtain large international datasets needed to conduct analysis such as the application of the multidimensional computerized adaptive testing to the MSQOL-54.

Future studies should be conducted to further assess measurement invariance across language version groups matching the samples by a broad set of individual and clinical variables, such as

levels of education, employment, and disease forms. Taking into account those variables would increase confidence that comparisons across language versions are meaningful.

Finally, researchers have recently shown substantial interest in using electronic PROMs to routinely monitor patients with long-term conditions. One step forward could be to assess measurement invariance across the modes of MSQOL-54 administration (paper vs. electronic) in both Italian and English versions of the instrument.

COMPLIANCE WITH ETHICAL STANDARDS

Conflict of interest

The authors declare that they have no conflict of interest.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent

Patients gave written informed consent to being included in the original projects. Additional consent was not required for this secondary analysis, for which patients' privacy and anonymity was guaranteed.

REFERENCES

1. Compston, A., McDonald, I., Noseworthy, J., Lassmann, H., Miller, D., Smith, K., et al. (2006). *McAlpine's Multiple Sclerosis*, 4th edn. Edinburgh, UK: Churchill Livingstone Elsevier.

2. GBD 2016 Multiple Sclerosis Collaborators. (2019). Global, regional, and national burden of multiple sclerosis 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurology*, 18(3), 269-285.

3. Rothwell, P.M., McDowell, Z., Wong, C.K., Dorman, P.J. (1997). Doctors and patients don't agree: cross sectional study of patients and doctors perceptions and assessments of disability in multiple sclerosis. *BMJ,* 314, 1580-1583.

4. Solari A. (2005). Role of health-related quality of life measures in the routine care of people with multiple sclerosis. *Health Quality of Life Outcomes,* 3, 16.

5. Miller, D.M., Allen, R. (2010). Quality of life in multiple sclerosis: determinants, measurement, and use in clinical practice. *Current Neurology and Neuroscience Reports*, 10, 397–406.

6. Mitchell, A.J., Benito-León, J., González, J.M., Rivera-Navarro, J. (2005). Quality of life and its assessment in multiple sclerosis: integrating physical and psychological components of wellbeing. *Lancet Neurology*, 4, 556-566.

7. Vickrey, B.G., Hays, R.D., Harooni, R., Myers, L.W., Ellison, G.W. (1995) A health-related quality of life measure for multiple sclerosis. *Quality Life Research* 1995; 4: 187–206.

8. Cella, D.F., Dineen, M.A., Arnason, B., Heeley, E., Rostgaard, I., Løvendahl, B., et al. (1996). Validation of the Functional Assessment of Multiple Sclerosis quality of life instrument. *Neurology*, 47, 129–139.

9. Solari, A., Filippini, G., Mendozzi, L., Ghezzi, A., Cifani, S., Barbieri, E., et al. (1999). Validation of Italian multiple sclerosis quality of life 54 questionnaire. *Journal Neurology Neurosurgery Psychiatry*, 67, 158–162.

10. Acquadro, C., Lafortune, L., Mear, I. (2003). Quality of life in multiple sclerosis: translation in French Canadian of the MSQoL-54. *Health Quality Life Outcomes*, 1, 70.

11. Yamamoto, T., Ogata, K., Katagishi, M., Shimizu, H., Ogawa, M., Yamamura, T., et al. (2004). Validation of the Japanese-translated version Multiple Sclerosis Quality of Life-54 instrument. *Rinsho Shinkeigaku*, 44, 417–421.

12. Idiman, E., Uzunel, F., Ozakbas, S., Yozbatiran, N., Oguz, M., Callioglu, B., et al. (2006). Cross-cultural adaptation and validation of multiple sclerosis quality of life questionnaire (MSQOL-54) in a Turkish multiple sclerosis sample. *Journal Neurological Sciences*, 240, 77-80.

13. Pekmezovic, T., Kisic Tepavcevic, D., Kostic, J., Drulovic, J. Validation and cross-cultural adaptation of the disease-specific questionnaire MSQOL-54 in Serbian multiple sclerosis patients sample. *Quality Life Research*, 16, 1383–1387.

14. Füvesi, J., Bencsik, K., Benedek, K., Mátyás, K., Mészáros, E., Rajda, C., et al. (2008). Cross-cultural adaptation and validation of the 'Multiple Sclerosis Quality of Life Instrument' in Hungarian. *Multiple Sclerosis*, 14, 391-398.

15. El Alaoui Taoussi, K., Ait Ben Haddou, E., Benomar, A., Abouqal, R., Yahyaoui, M. (2012). Quality of life and multiple sclerosis: Arabic language translation and transcultural. adaptation of MSQOL-54. *Revue Neurologique,* 168, 444-449.

16. Putnick, D.L., Bornstein, M.H. (2016). Measurement Invariance Conventions and Reporting: The State of the Art and Future Directions for Psychological Research. *Developmental Review,* 41, 71-90.

17. Boer, D., Hanke, K., He, J. (2018). On Detecting Systematic Measurement Error in Cross-Cultural Research: A Review and Critical Reflection on Equivalence and Invariance Tests. *Journal Cross-Cultural Psychology,* 49(5), 713–734.

18. Motl, R.W., McAuley, E., Suh, Y. (2010). Validity, invariance and responsiveness of a self-report measure of functional limitations and disability in multiple sclerosis. *Disability Rehabilitation*, 32, 1260–1271.

19. Motl, R.W., McAuley, E., Mullen, S. (2011). Longitudinal measurement invariance of the Multiple Sclerosis Walking Scale-12. *Journal Neurological Sciences*, 305, 75-79.

20. Motl, R.W , Mullen, S., McAuley, E. (2012) Multi-group measurement invariance of the multiple sclerosis walking scale-12? *Neurological Research,* 34(2), 149-152.

21. Cox, S.D, Pakenham, K.I. (2014) Confirmatory factor analysis and invariance testing of the Young Carer of Parents Inventory (YCOPI). *Rehabilitation Psychology,* 59, 439-452.

22. Chung, H., Kim, J., Askew, R.L., Jones, S.M., Cook, K.F., Amtmann, D. (2015). Assessing measurement invariance of three depression scales between neurologic samples and community samples. *Quality Life Research*, 24, 1829-1834.

23. Chung, H., Kim, J., Park, R., Bamer, A.M., Bocell, F.D., Amtmann, D. (2016). Testing the measurement invariance of the University of Washington Self-Efficacy Scale short form across four diagnostic subgroups. *Quality Life Research,* 25, 2559-2564.

24. Byrne, B. M., Shavelson, R. J., Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, 105(3), 456.

25. Santos, D., Abad, F.J., Miret, M., Chatterji, S., Olaya, B., Zawisza, K., et al. (2017). Measurement invariance of the WHOQOL-AGE questionnaire across three European countries. *Quality of Life Research*, 27, 1015–1025.

26. Geyh, S., Fellinghauer, B.A., Kirchberger, I., Post, M.W. (2010). Cross-cultural validity of four quality of life scales in persons with spinal cord injury. *Health Qual Life Outcomes,* 3, 8, 94. Https://doi: 10.1186/1477-7525-8-94.

27. Hadgkiss, E.J., Jelinek, G.A., Weiland, T.T., Pereira, N.G., Marck, C.H., van derMeer, D.M. (2013). Methodology of an International Study of People with Multiple Sclerosis Recruited through Web 2.0 Platforms: Demographics, Lifestyle, and Disease Characteristics. *Neurology Research International*, 580-596.

28. Jelinek, G.A., De Livera, A.M., Marck, C.H., Brown, C.R., Neate, S.L., Keryn, L., et al. (2016). Lifestyle, medication and socio-demographic determinants of mental and physical health-related quality of life in people with multiple sclerosis. *BMC Neurology*, 16, 235. https://doi.org/10.1186/s12883-016-0763-4.

29. Bassi, M., Falautano, M., Cilia, S., Goretti, B., Grobberio, M., Pattini, M., et al. (2016). Illness Perception and Well-Being Among Persons with Multiple Sclerosis and their caregivers. *Journal Clinical Psychology Medical Settings*, 23, 33-52.

30. Bassi, M., Falautano, M., Cilia, S., Goretti, B., Grobberio, M., Pattini, M., et al. (2014). The coexistence of well- and ill-being in persons with multiple sclerosis, their caregivers and health professionals. *Journal Neurological Sciences*, 337, 67-73.

31. Rosato, R., Testa, S., Bertolotto, A., Confalonieri, P., Patti, F., Lugaresi, A., et al. (2016). Development of a short version of MSQOL-54 using factor analysis and item response theory. *PLoS One*, 11, e0153466.

32. Rosato, R., Testa, S., Bertolotto, A., Scavelli, F., Giovannetti, A.M., Confalonieri, P., et al. (2018). Prospective validation of the abbreviated, electronic version of the MSQOL-54. *Multlple Sclerosis,* 25(6), 856-866.

33. Solari, A., Motta, A., Mendozzi, L., Pucci, E., Forni, M., Mancardi, G., et al. (2004). Computer-aided retraining of memory and attention in people with multiple sclerosis: a randomized, double-blind controlled trial. *Journal Neurological Sciences, 222*, 99-104.

34. Massacesi, L., Tramacere, I., Amoroso, S., Battaglia, M.A., Benedetti, M.D., et al. (2014). Azathioprine versus beta interferons for relapsing-remitting multiple sclerosis: a multicentre randomized non-inferiority trial. *PLoS ONE*, 9, e113371.

35. Polman, C., Reingold, S., Edan, G., Filippi, M., Hartung, H., Kappos, L., et al. (2005). Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria". *Annals Neurology*, 58, 840–846.

36. Polman, C.H., Reingold, S.C., Banwell, B., Clanet, M., Cohen, .JA., Filippi, M., et al. (2011). Diagnostic criteria for multiple sclerosis: 2010 Revisions to the McDonald criteria. *Annals Neurology*, 69, 292–302.

37. Kurtzke, J.F. (1983). Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology*, 33, 1444-1452.

38. Hohol, M.J., Hohol, M.J., Orav, E.J., Weiner, H.L. (1995). Disease steps in multiple sclerosis: a simple approach to evaluate disease progression. *Neurology,* 45, 251-255.

39. Ware, J.E., Snow, K.K., Kosinski, M., Gandek, B. (1993). *SF-36 Health survey manual and interpretation guide*. Boston, MA: The Health Institute.

40. Giordano, A., Pucci, E., Naldi, P., Mendozzi, L., Milanese, C., Tronci, F., et al. (2009). Responsiveness of patient- reported outcome measures in multiple sclerosis relapses: the REMS study. *Journal Neurology Neurosurgery Psychiatry,* 80, 1023-1028.

41. Millsap, R. E., Yun-Tein, J. (2004). Assessing Factorial Invariance in Ordered-Categorical Measures. *Multivariate Behavioral Research*, 39(3), 479–515.

42. Hu, L., Bentler, P.M. (1999) Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling,* 6, 1-55.

43. Hu, L., Bentler, P.M. (1995). Evaluating model fit. In: R.H. Hoyle (Ed.), *Structural equation modeling. Concepts, issues, and applications* (pp.76-99). London, UK: Sage.

44. Chen, F.F. (2007) Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 464–504.

45. Hays, R.D., Revicki, D., Coyne, K.S. (2005). Application of structural equation modeling to health outcomes research. *Evaluation Health Professions*, 28, 295–309.

46. van Bebber, J., Flens, G., Wigman, J. T. W., de Beurs, E., Sytema, S., Wunderink, L., & Meijer, R. R. (2018). Application of the Patient-Reported Outcomes Measurement Information System (PROMIS) item parameters for Anxiety and Depression in the Netherlands. *International Journal of Methods in Psychiatric Research*, 27(4), e1744. https://doi.org/10.1002/mpr.1744.

47. Iacus, S.M., King, G., Porro, G. (2009). cem: Software for Coarsened Exact Matching. *Journal Statistical Software*, 30 (9). https://doi.org/10.18637/jss.v030.i09.

48. Lix, L.M., Osman, B.A., Adachi, J.D., Towheed, T., Hopman, W., Davison, K.S.. Leslie, W.D. (2012). Measurement equivalence of the SF-36 in the canadian multicentre osteoporosis study. *Health Quality of Life Outcomes*, 10, 29.

49. Muthén, L. K., Muthén, B. O. (1998-2011). *Mplus User's Guide*. Sixth Edition. Los Angeles, CA: Muthén & Muthén.

50. Sass, D. A., Schmitt, T. A., Marsh, H. W. (2014). Evaluating Model Fit With Ordered Categorical Data Within a Measurement Invariance Framework: A Comparison of Estimators. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(2), 167–180).

51. Marrie, R.A., Cutter, G., Tyry, T., Hadjimichael, O., Campagnolo, D., et al. (2005). Changes in the ascertainment of multiple sclerosis. *Neurology*, 65, 1066–1070.

Table 1. Characteristics of the entire dataset (N=3669 patients) by MSQOL-54 language version.

	English-speaking (N=2064)	Italian (N=1605)	P value
Women (%)[1]	1704 (83)	996 (62)	<0.001
Mean age in years, SD (range)[2]	46.1, 10.5 (18–87)	40.9, 10.8 (18–79)	<0.001
Mean years from MS diagnosis, SD (range)[3]	9.0, 7.3 (1–42)	4.9, 7.8 (0-48)	<0.001
Median EDSS score (range)[4]	-	2.5 (0–9.5)	-
Patient Determined Disease Steps (%)[5]			
Mild disability	1110 (54)	1097 (69)	
Moderate disability	722 (35)	308 (19)	
Severe disability	219 (11)	194 (12)	<0.001
Mean MSQOL-54 PHC, SD (range)	57.7, 21.5 (3–100)	61.1, 20.2 (2–100)	<0.001

Mean MSQOL-54 MHC, SD (range)	66.6, 21.3 (1-100)	62.9, 20.7 (2–100)	<0.001

EDSS, Expanded Disability Status Scale; MSQOL-54, Multiple Sclerosis Quality of Life-54; PHC/MHC, Physical and Mental Health Composite; SD standard deviation.

1. Missing replies for sex: N=21 (English-speaking).

2. Missing replies for age: N=62 (English-speaking); N=53 (Italy)

3. Missing replies for disease duration: N=11 (English-speaking); N=227 (Italy)

4. Missing replies for EDSS: N=6 (Italy).

5. Missing replies for PDDS: N=13 (English-speaking); N=6 (Italy).

Table 2. Measurement invariance of the MSQOL-54.

	χ^2(df)	χ^2 p-value	RMSEA	CFI	SRMR	Δ RMSEA	Δ CFI	Δ SRMR
Italian (N=1605)	5987.5 (1208)	<0.0001	0.050	0.906	0.045	-	-	-
English-speaking (N= 2064)	8596.3 (1208)	<0.0001	0.054	0.903	0.047	-	-	-
Configural invariance	14508.0 (2416)	<0.0001	0.052	0.904	0.046	-	-	-
Metric invariance	14829.6 (2456)	<0.0001	0.052	0.902	0.048	0.000	-0.002	0.002
Scalar invariance	16551.8 (2496)	<0.0001	0.055	0.889	0.051	0.003	-0.013	0.003

CFI, comparative fit index; RMSEA, root mean square error of approximation; SRMR, standardized root mean square residual.

Table 3. Summary statistics for the absolute difference between individuals' T scores as resulted from the unconstrained model (configural invariance) and the most constrained model (scalar invariance).

	Physical Health	Role Limitations- Physical	Role Limitations- Emotional	Bodily Pain	Emotional Wellbeing	Energy	Health Perceptions	Social Function	Cognitive Function	Health Distress	Sexual Function	Overall Quality of Life
Italian sample												
Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Max	0.76	0.49	0.67	**1.72**	1.14	1.05	1.29	0.39	0.41	1.17	1.24	1.08
Mean	0.11	0.09	0.14	0.22	0.18	0.21	0.26	0.07	0.07	0.11	0.17	0.14
SD	0.12	0.09	0.16	0.23	0.15	0.18	0.20	0.06	0.07	0.11	0.23	0.12
% ≥ 1*	0	0	0	1.18 (19)	0.06 (1)	0.19(3)	0.19(3)	0	0	0.06(1)	0.81 (13)	0.06 (1)
English-speaking sample												
Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Max	0.88	0.44	0.51	0.52	0.56	0.61	0.90	0.26	0.29	0.43	1.02	**1.11**
Mean	0.10	0.08	0.06	0.10	0.09	0.12	0.14	0.04	0.03	0.03	0.13	0.14
SD	0.09	0.10	0.08	0.07	0.08	0.10	0.12	0.04	0.03	0.03	0.14	0.11
% ≥ 1*	0	0	0	0	0	0	0	0	0	0	0.05 (1)	0.05 (1)

SD, standard deviation.

*Number of participants in brackets.