This is the accepted version of the publication Ho LY, Lai CK, Ng SS. Testing the psychometric properties of the Chinese version of the Neurological Fatigue Index-Stroke. Clinical Rehabilitation. (Volume: 35 issue: 9) pp. 1329-1340. Copyright © 2021 (The Author(s)). DOI: 10.1177/02692155211001684

Testing the psychometric properties of the Chinese version of the Neurological Fatigue Index-Stroke

Abstract

Objective: To test the psychometric properties of a Chinese version of the Neurological Fatigue Index-Stroke (C-NFI-Stroke) in stroke survivors.

Design: This was a validation study. Cross-cultural adaptation of the scale was conducted according to standard guidelines. Reliability, validity, responsiveness, and interpretability were measured.

Setting: Self-help groups and a community centre

Subjects: One hundred and twelve Chinese stroke survivors and 65 healthy Chinese older people living in the community.

Interventions: Not applicable.

Main measures: The C-NFI-Stroke, Fatigue Severity Scale, Mental Fatigue Scale, General Self-Efficacy Scale, and Geriatric Depression Scale were used.

Results: Cronbach's α coefficients were 0.69-0.88; the item-level agreement was 70.3-88.9%; the weighted Kappa value was 0.47-0.79; and the intra-class correlation coefficients were 0.88-0.93. The C-NFI-Stroke had no ceiling and floor effects. It had good content validity and had two factors, 'lack of energy' and 'tiredness/weakness'. The confirmatory factor analysis showed a good fit to the model. The C-NFI-Stroke significantly correlated with existing fatigue scales (r_s=0.55-0.63), self-efficacy (r_s=-0.31 to -0.37), and depressive symptoms (r_s=0.53-0.60). The C-NFI-Stroke could discern differences between stroke survivors and healthy older people.

Conclusions: The C-NFI-Stroke is a reliable and valid tool for clinical and research use on people who have been diagnosed with stroke for a year or more, although its factor structure differs from that of the original English version.

Keywords

Depression, fatigue, Neurological Fatigue Index, outcome assessment, reliability and validity, self-efficacy, stroke

Introduction

Fatigue, which is described as ‘a tiredness in the muscles’, ‘mental tiredness’, or ‘a general feeling of tiredness’ by stroke survivors,[1] is common in stroke populations.[2] Brain lesions, inflammatory markers, neuroendocrine changes, and attentional-executive impairments are some of the factors that would trigger fatigue, while physical and behavioural factors, such as disability and reduced physical activity, have been found to be associated with fatigue following a stroke.[3] Fatigue can be sustained by psychosocial factors,[3, 4] such as self-efficacy ($r=-0.43$ to -0.50),[5, 6] which is a person’s belief in his/her ability to control life’s stressors,[7] and depressive symptoms ($r=0.21-0.61$) in stroke survivors.[8-10] Fatigue exerts negative impacts on daily life, such as on instrumental activities of daily living, mobility, and social life.[11] Thus, alleviating fatigue is an important goal in clinical stroke rehabilitation.

Different fatigue scales have been used in stroke populations. Some have not been specifically designed for stroke survivors such as the Fatigue Severity Scale.[12] Others, such as the Case Definition of Fatigue,[13] the Detection List Fatigue,[14] and the Neurological Fatigue Index-Stroke (NFI-Stroke), target stroke survivors.[15] Of these scales, only the NFI-Stroke adequately covers both physical and cognitive fatigue. It has good reliability and validity.[15] However, it is available only in English.

A validated Chinese version would facilitate assessments of fatigue in Chinese stroke populations and evaluations of interventions for alleviating fatigue in clinical settings. Therefore, the aim of this study was to test the psychometric properties of the Chinese version of the NFI-Stroke in Hong Kong, including reliability (internal consistency, test-retest reliability, standard error of measurement, ceiling and floor effects), validity (content validity, factor structure, correlations with other fatigue scales and self-efficacy as well as depressive symptoms), responsiveness (minimal detectable change), and interpretability (ability to discern differences in levels of fatigue between stroke survivors and healthy older people).

Methods

This study was conducted from January to April 2019. The Declaration of Helsinki was followed. The ethics committee of the authors' university approved this study (Reference number: HSEARS20190104001). All participants provided their written informed consent to take part in it. This work was financially supported by (1) The Hong Kong Polytechnic University; and (2) by a Departmental Research Grant from the Department of Rehabilitation Sciences, The Hong Kong Polytechnic University (Grant number: 90013897).

Translation and cultural adaptation of the Neurological Fatigue Index-Stroke

Permission to translate the NFI-Stroke into Chinese was obtained from the University of Leeds. The forward and backward translations with cultural adaptation were each conducted by two bilingual translators in accordance with the guidelines suggested by Beaton et al.[16] A content validity index, which is a method for assessing construct validity,[17] was used. A panel consisting of five experts including translators and healthcare professionals rated the experiential, conceptual, semantic, and idiomatic equivalence of each item of the scale using a 4-point scale rating (where 1, 2, 3, and 4 represented 'not relevant', 'somewhat relevant', 'quite relevant', and 'highly relevant', respectively). A rating of 1 or 2 was dichotomized as irrelevant and 3 or 4 as relevant in gauging the content validity of the index. Any disagreements over wording were discussed until a consensus was reached. This pilot Chinese version was then tested on 40 community-dwelling Chinese people who were aged ≥55, had a confirmed diagnosis of stroke for ≥1 year, no other neurological diseases, no history of transient ischemic attack, and no unstable medical conditions. They were recruited from patient self-help groups. They all agreed that the pilot version was clear, easy to read, and comprehensible. Consequently, this version, which was named the Chinese version of the Neurological Fatigue Index-Stroke (C-NFI-Stroke), underwent further testing to determine its psychometric properties.

Setting and sampling

Stroke survivors and healthy older people were recruited by convenience sampling from self-help groups for people with neurological diseases and from a non-governmental organization in Hong Kong. The criteria for the inclusion of stroke survivors were: (1) community-dwelling residents; (2) who are ethnic Chinese; (3) able to speak Chinese; (4) aged ≥55; and (3) who had a confirmed diagnosis of stroke for ≥1 year. People with any other neurological diseases, who had experienced a transient ischemic attack, or who had unstable medical conditions were excluded. This group of stroke survivors was recruited because the incidence of stroke doubles every decade after the age of 55[18] and those who have been diagnosed with stroke for ≥1 year are often ignored, especially when they are assumed to have fully recovered. The same inclusion and exclusion criteria, except for a diagnosis of stroke, were applied to recruit healthy older people.

Regarding the size of the sample, ≥100 stroke participants were deemed to be needed because this study involved a factor analysis.[19] Among the stroke participants, the first 27 stroke survivors were reassessed after 7-10 days to determine the test-retest reliability of the index. This sample size was established based on an intraclass correlation coefficient (ICC) of 0.7,[20] an α value of 0.05, and a β value of 0.2. To examine the ability of the C-NFI-Stroke to discern differences among groups, 64 stroke survivors and 64 healthy people would be required at α=0.05, β=0.2, and effect size=0.5. For this comparison, the required number of stroke survivors was randomly drawn from among all of the stroke participants using computer-generated random numbers.

Data collection

After obtaining the informed consent of the participants, information on their demographic characteristics was collected. Their levels of fatigue, self-efficacy, and depressive symptoms were assessed by the same researcher using Chinese versions of validated measurement scales at the university or at the premises of the participating organization.

Neurological Fatigue Index-Stroke

The C-NFI-Stroke was used to assess levels of fatigue. It has a summary scale (items 1-7, 9, 11-12), a physical subscale (items 1-8), and a cognitive subscale (items 9-12). The response choices range from 0 (strongly disagree) to 3 (strongly agree). The summary scale, and the physical and cognitive subscales of the original English version were correlated with the Fatigue Severity Scale with Spearman correlation coefficients of 0.62, 0.60, and 0.51, respectively.[15] The test-retest reliability of the summary scale, and the physical, and cognitive subscales of the original English version in terms of Spearman correlation coefficients were 0.90, 0.90, and 0.79, respectively over 2-4 weeks.[15]

Fatigue Severity Scale

Since the 9-item Fatigue Severity Scale is frequently used in stroke populations, its Chinese version[21] was chosen to quantify the convergent validity of the C-NFI-Stroke. Each item is rated from 1 (strongly disagree) to 7 (strongly agree). The internal consistency of the Fatigue Severity Scale in terms of Cronbach's α was 0.94 and its one-week test-retest reliability in terms of ICC was 0.93 in stroke survivors.[22]

Mental Fatigue Scale

As the C-NFI-Stroke considers cognitive fatigue, the Chinese version of the Mental Fatigue Scale[23] was chosen to quantify the convergent validity of the C-NFI-Stroke. The scale assesses subjective mental fatigue in the past one month through 15 questions. Each question is rated from 0 (normal function) to 3 (maximal symptom). The internal consistency of the Chinese version of the Mental Fatigue Scale in terms of Cronbach's α was 0.92 and its test-retest reliability over a week in terms of ICC was 0.97 in people with traumatic brain injury.[23]

General Self-Efficacy Scale

The Chinese version of the 10-item General Self-Efficacy Scale[24] was used to assess self-efficacy because it measures a person's confidence in handling general rather than specific situations. The total score ranges from 10-40, with higher scores representing higher levels of self-efficacy in tackling the demands of life. The internal consistency in terms of Cronbach's α was 0.91 in undergraduate students.[24]

Geriatric Depression Scale

Depressive symptoms were assessed using the Chinese version of the 15-item Geriatric Depression Scale,[25] as recommended by the American Heart Association Classification of Stroke Outcome Task Force.[26] Each item is rated either 0 or 1, with the total score ranging from 0 to 15 and higher scores indicating more depressive symptoms. The cut-off point was 8.[27] Its internal consistency was 0.78 in a stroke population.[28]

Statistical analysis

The Statistical Package for the Social Sciences (SPSS, version 25) and SPSS Amos were used to analyse the data. Tests were conducted on the psychometric properties of the instrument, which included reliability, validity, responsiveness, and interpretability, and the results were assessed.

Reliability and responsiveness

Internal consistency was examined using Cronbach's α, for which a value of >0.7 is acceptable.[29] Item-level agreement and a weighted Kappa value were used for item-level retest analyses. A Kappa coefficient of 0 represents poor, 0.01-0.20 slight, 0.21-0.40 fair, 0.41-0.60 moderate, 0.61-0.80 substantial, and 0.81-1 almost perfect agreement.[30] The test-retest reliability was examined using $ICC_{3,1}$, with <0.50 representing poor, 0.50-0.75 moderate, 0.75-0.90 good, and >0.90 excellent reliability.[31] The standard error of measurement was calculated by SD x $\sqrt{1-r}$,[32] and minimal detectable change was calculated at 95% confidence intervals by SD x 1.96 x $\sqrt{2(1-r)}$, where SD denoted the standard deviation of the baseline C-NFI-Stroke score and r denoted the test-retest reliability coefficient.[33] Five per cent or less, >5% to ≤10%, >10% to <20%, and 20% of the standard error of measurement relative to the total score was considered very good, good, doubtful, and negative, respectively.[34] Ceiling and floor effects were also calculated. Effects are regarded as present when >15% of the participants attain the highest or lowest possible scores.[35]

Validity

Content validity, construct validity, and convergent validity were assessed. An item-level content validity index was calculated as the number of experts giving a rating of 3 or 4 divided by 5, which was the number of experts, while the scale-level content validity index was calculated by taking the average of the sum of the item-level content validity indices.[17] An item-level content validity index and scale-level content validity index of ≥0.78 and ≥0.90, respectively are considered good.[17]

Both exploratory and confirmatory factor analyses were conducted to identify the components of the Chinese version instead of using a Rasch analysis, which is better suited for testing predetermined subscales based on theory. In exploratory analyses, items with factor loadings of over 0.30 are considered meaningful;[36] these were thus put into the confirmatory factor analysis model. The fit of the data to the model was assessed using the following criteria: $X^2/df \leq 3$,[37] root mean square error of approximation value ≤0.06, and a Comparative Fit Index score of ≥0.95.[38]

The Kolmogorov-Smirnov test showed that some data were not normally distributed. Thus, correlations between the C-NFI-Stroke and the Chinese versions of the Fatigue Severity Scale, Mental Fatigue Scale, General Self-Efficacy Scale, and Geriatric Depression Scale were examined using Spearman's correlation coefficients.

Interpretability

The C-NFI-Stroke scores of stroke survivors and healthy older people, and those of stroke survivors and healthy older people with and without depressive symptoms were compared using the Mann-Whitney U test. A *p*-value of lower than 0.05 indicated a statistically significant difference.

Results

Demographic characteristics

One hundred and twelve community-dwelling stroke survivors were recruited. Among them, 66.1% were male. Their mean (standard deviation [SD]) age was 64.15 (5.79). The mean (SD) time that had passed since their diagnosis of stroke was 73.60 (57.43) months.

Reliability and responsiveness

The internal consistency of the C-NFI-Stroke in terms of Cronbach's α coefficient was 0.69-0.88. Item-level agreement was 70.3-88.9%. The weighted Kappa value was 0.47-0.79. The ICCs were 0.88-0.93. The standard error of measurement and minimal detectable change are summarized in Table 1. No participant received the highest or lowest summary and cognitive scores. One got the highest and none got the lowest physical score. Thus, the C-NFI-Stroke had no ceiling and floor effects.

Validity

The item-level content validity index scores of the C-NFI-Stroke ranged from 0.6 to 1 and the scale-level content validity index score was 0.95. The results are presented in Table 2.

The Kaiser-Meyer-Olkin measure was 0.91, implying sufficient items for a factor analysis.[39] The Bartlett's Test of Sphericity was significant, suggesting that the factor analysis was satisfactory.[39] The exploratory factor analysis suggested a two-factor structure, 'lack of energy' (factor loadings=0.39-0.69) and 'tiredness/weakness' (factor loadings=0.60-0.77), with 56.7% of the total variance explained (Table 3). The model of the confirmatory factor analysis presented in Figure 1 shows the structure of the C-NFI-Stroke. Items 4, 5, 6, 8, 9, 10, 11, and 12 were specified to load on the lack of energy factor. Items 1, 2, 3, and 7 were specified to load on the tiredness/weakness factor. The data demonstrated a good fit to the model, with $X^2/df=1.68$, a root mean square error of approximation value of 0.06, and a Comparative Fit Index score of 0.97.

The C-NFI-Stroke correlated significantly with the Chinese versions of the Fatigue Severity Scale, Mental Fatigue Scale, General Self-Efficacy Scale, and Geriatric Depression Scale (r_s=-0.37 to 0.63, $p<0.01$). The Spearman's correlation coefficients are shown in Table 4.

Interpretability

The mean (SD) C-NFI-Stroke summary, physical, and cognitive scores of 112 stroke survivors were 15.92 (4.34), 12.52 (3.55), and 6.10 (1.84), respectively. Sixty-six stroke survivors, who were drawn by computer from the 112 stroke survivors, and 65 healthy older people were involved in the attempt to discern differences among the groups. The stroke survivors had significantly higher C-NFI-Stroke summary, physical, and cognitive scores than the healthy older people (Table 5). Both the stroke survivors and the healthy older people with depressive symptoms had significantly higher C-NFI-Stroke scores than those without depressive symptoms (Table 6).

Discussion

This study is the first to translate and culturally adapt the NFI-Stroke into Chinese. In summary, the C-NFI-Stroke had acceptable internal consistency and test-retest reliability without ceiling and floor effects. It had good scale-level content validity and two factors. All scores of the C-NFI-Stroke were correlated with the Chinese versions of the Fatigue Severity Scale, Mental Fatigue Scale, General Self-Efficacy Scale, and Geriatric Depression Scale. Stroke survivors had higher C-NFI-Stroke summary, physical, and cognitive scores than healthy older people. Both stroke survivors and healthy older people with depressive symptoms had higher C-NFI-Stroke scores than those without depressive symptoms.

The Cronbach's α coefficients were good in both the summary scale and physical subscale, but barely reached the level of acceptable in the cognitive subscale. This indicated that the correlations among items 9-12 and the correlation between each item and the cognitive score were not high. One possible reason for this is the small number of items (only four) in the cognitive subscale. Internal consistency is known to be a function of the number of items, and increasing the number of items can increase the value of the Cronbach's α.[40] Further testing is suggested to determine the internal consistency of the index.

The test-retest reliability of the index was satisfactory. The strength of the agreement amongst all items of the C-NFI-Stroke was moderate to substantial (0.47-0.79).[30] The relatively consistent responses from the respondents indicated that the translation of the C-NFI-Stroke was appropriate. The standard error of measurement of the C-NFI-Stroke summary scale, and the physical and cognitive subscales was 3.5%, 4.0%, and 4.7%, respectively. The C-NFI-Stroke had smaller measurement errors than other neurological fatigue indices (e.g., 8.3%-9.2% for the Dutch Neurological Fatigue Index-multiple sclerosis;[41] 5.4%-20.8% for the Brazilian Neurological Fatigue Index-multiple sclerosis;[42] and 4.8%-8.1% for the Neurological Fatigue Index-Post-Polio Syndrome).[43] Our findings support the argument that the C-NFI-Stroke is a reliable instrument for use in Chinese stroke populations.

The minimal detectable change in the C-NFI-Stroke summary score was within 10% of the total possible range of scores and was acceptable.[44] The C-NFI-Stroke (minimal detectable change relative to the total possible range of scores: summary 9.7%, physical 11.2%, cognitive 13.1%) showed a change that was not due to a measurement error. This change in score was smaller than that of the

Neurological Fatigue Index-Post-Polio Syndrome (minimal detectable change relative to the total possible range of scores: summary 18.5%, physical 13.3%, cognitive: 22.4%),[43] implying that the C-NFI-Stroke is more sensitive to change. Yet, when the C-NFI-Stroke is used to evaluate the effectiveness of interventions in clinical practice, the physical and cognitive subscales may be unable to detect changes with sensitivity.

Consistent with previous studies,[42, 43] there were no ceiling or floor effects in the C-NFI-Stroke. This study indicated that differences in fatigue could be discerned among participants and clinically meaningful data could be captured at two ends of the scale. The ability of the tool to detect changes further supports its use in clinical settings.

Concerning cultural adaptation, this study found that, with the exception of item 11, the scale-level and item-level content validity indices were good. The low scale-level content validity index of item 11 ('My coordination gets worse as the day goes on') was related to semantic inequivalence because the Chinese words did not reflect cognitive fatigue and the changes that can take place over a day. Therefore, when producing the pilot version, Chinese words that were more appropriate for describing the cognitive aspect and measuring the construct were adopted by the expert panel.

The factor structure found in this study was different from the construct determined by the original NFI-Stroke. An exploratory factor analysis of the C-NFI-Stroke identified a two-factor structure, namely 'lack of energy' and 'weakness', which was confirmed in the confirmatory factor analysis. Lack of energy (items 4, 5, 6, 8, 9, 10, 11, and 12) results in a requirement for more energy, difficulties, or even an inability to perform tasks. Tiredness/weakness (items 1, 2, 3, and 7) refers to the general feelings experienced by stroke survivors. In the original NFI-Stroke, items 1-8 and 9-12 contribute to physical and cognitive fatigue, respectively.[15] The discrepancy in the findings may be related to the methods of analysis that were employed. This study adopted exploratory and confirmatory factor analyses, while the original NFI-Stroke adopted the Rasch analysis. In addition, culture influences the perception of fatigue.[45] Our Chinese participants interpreted items under cognitive fatigue in the original version as a lack of energy, implying that language may play a role in how fatigue is perceived.

The correlation coefficients show that the construct of the C-NFI-Stroke was highly related to the Chinese versions of the Fatigue Severity Scale and the Mental Fatigue Scale, denoting convergent validity.[46] Moreover, our finding was consistent with a previous study showing a significant correlation between NFI-stroke and the 7-item Fatigue Severity Scale.[47] The C-NFI-Stroke can be said to be a valid scale for measuring fatigue.

Fatigue has been found to correlate with psychological factors in chronic stroke survivors worldwide. The findings of this study conducted in a Chinese stroke population were similar to those of previous studies, such as studies conducted in Ireland[5] and the USA,[6] in which correlations between fatigue and self-efficacy were found. Although the exact mechanism is not clear, a previous study showed that higher self-efficacy is associated with proactive coping,[48] which may enable stroke survivors to better cope with fatigue. The correlation found between fatigue and depressive symptoms in this study was similar to those of previous studies, such as those conducted in Britain,[9] the Netherlands,[10] and Japan.[8] Ormstad and Eilertsen[4] have suggested that psychosocial and behavioural factors may play a role in the development of depression over time in stroke survivors with fatigue. Douven et al.[49] found that the relationship between fatigue and depressive symptoms could be bidirectional, and a change in fatigue score was associated with a change in depression scores (B=1.47, p<0.001).

The C-NFI-Stroke was able to discern differences between stroke survivors and healthy older people. Stroke survivors with paretic limbs need additional physical energy to mobilize the limbs during daily activities and lift the body's centre of mass.[50] They also need to make additional mental effort to walk when compared with healthy people,[51] which may result in higher C-NFI-Stroke physical, cognitive, and summary scores than have been found in the healthy older. However, stroke survivors with depressive symptoms had the same or nearly the same C-NFI-Stroke summary, physical, and cognitive scores as healthy older people with depressive symptoms. This might be related to the characteristics of our healthy older comparison group. About one third of them were caregivers of stroke survivors, with one fifth of them still in the workforce. Caregiving might have contributed to low mood and fatigue,[52] while work-related stress might also add to fatigue in the working population.[53]

There are several limitations in this study. First, the participants came from only a few local self-help groups and a non-governmental organization, limiting the generalizability of this study. Second, some potential participants were unwilling to join because of the need to travel long distances to the assessment venue. Third, the sample size might not be adequate for comparisons to be made between stroke survivors and healthy older people with and without depressive symptoms. Last, since some healthy older people were caregivers of stroke survivors, they likely were not representative of the general healthy population.

The C-NFI-Stroke can usefully be applied in clinical settings. Healthcare professionals can adopt the C-NFI-Stroke to measure levels of fatigue in stroke survivors and use it to evaluate interventions for alleviating fatigue. In order to reduce fatigue, it is also necessary to screen for depressive symptoms and assess self-efficacy because these psychological factors are correlated with fatigue.

As for research applications, the C-NFI-Stroke is the first Chinese scale to measure both physical and cognitive fatigue in stroke. It can facilitate future assessments of fatigue, evaluations of fatigue interventions, and explorations of the causal relationship between fatigue and other variables. Although studies on fatigue following stroke have been increasing in Western countries, fatigue has not been extensively studied in Chinese communities. The C-NFI-Stroke can help in studies of the phenomenon in Chinese populations. It is now feasible to conduct cross-cultural comparisons of fatigue following stroke using the same tool because such a tool is now available. In future studies, a larger sample size could be used to refine the C-NFI-Stroke. The Chinese version could also be tested with people who have been diagnosed with stroke within one year, in-patients or out-patients, or stroke survivors with cognitive impairment. Replication of this study in other Chinese populations to explore what makes the factor structure of the Chinese version different from the construct determined by the original English version is recommended. Consideration can also be given in future studies to retest the internal consistency of the cognitive subscale and the minimal detectable changes in both the physical and cognitive subscales. A well-validated instrument would facilitate international comparisons and improve cross-cultural understanding of the phenomenon.

In conclusion, the C-NFI-Stroke is reliable and valid for measuring fatigue in both clinical practice and research. Fatigue was correlated with self-efficacy and depressive symptoms. The C-NFI-Stroke may help clinicians and researchers to evaluate the effectiveness of interventions designed to alleviate fatigue in stroke survivors.

Clinical Messages

- The Chinese version of the Neurological Fatigue Index-Stroke could be reliable and valid for measuring levels of fatigue in people who had been diagnosed stroke a year ago or more.
- It can be used to assess and monitor fatigue in both clinical and research settings.

References

1. Crosby GA, Munshi S, Karat AS, Worthington E and Lincoln NB. Fatigue after stroke:

Frequency and effect on daily life. *Disabil Rehabil* 2012; 34: 633-637.

2. Nadarajah M and Goh HT. Post-stroke fatigue: A review on prevalence, correlates, measurement, and management. *Top Stroke Rehabil* 2015; 22: 208-220.

3. Wu S, Mead G, Macleod M and Chalder T. Model of understanding fatigue after stroke. *Stroke* 2015; 46: 893-898.

4. Ormstad H and Eilertsen G. A biopsychosocial model of fatigue and depression following stroke. *Med Hypotheses* 2015; 85: 835-841.

5. Muina-Lopez R and Guidon M. Impact of post-stroke fatigue on self-efficacy and functional ability. *Eur J Physiother* 2013; 15: 86-92.

6. Miller KK, Combs SA, Van Puymbroeck M, et al. Fatigue and pain: Relationships with physical performance and patient beliefs after stroke. *Top Stroke Rehabil* 2013; 20: 347-355.

7. Bandura A. Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review* 1977; 84: 191-215.

8. Mutai H, Furukawa T, Houri A, Suzuki A and Hanihara T. Factors associated with multidimensional aspect of post-stroke fatigue in acute stroke period. *Asian Journal of Psychiatry* 2017; 26: 1-5.

9. Hawkins L, Lincoln NB, Sprigg N, et al. The Nottingham Fatigue After Stroke (NotFAST) study: Results from follow-up six months after stroke. *Top Stroke Rehabil* 2017; 24: 592-596.

10. de Bruijn M, Synhaeve NE, van Rijsbergen MWA, et al. Quality of life after young ischemic

stroke of mild severity is mainly influenced by psychological factors. *J Stroke Cerebrovasc* 2015; 24: 2183-2188.

11. Worthington E, Hawkins L, Lincoln N and Drummond A. The day-to-day experiences of people with fatigue after stroke: Results from the Nottingham Fatigue After Stroke study. *International Journal of Therapy & Rehabilitation* 2017; 24: 449-455.

12. Krupp LB, LaRocca NG, Muir-Nash J and Steinberg AD. The fatigue severity scale: Application to patients with multiple sclerosis and systemic lupus erythematosus. *Archives of Neurology* 1989; 46: 1121-1123.

13. Lynch J, Mead G, Greig C, Young A, Lewis S and Sharpe M. Fatigue after stroke: The development and evaluation of a case definition. *J Psychosom Res* 2007; 63: 539-544.

14. Kruithof N, Van Cleef MHM, Rasquin SMC and Bovend'Eerdt TJH. Screening poststroke fatigue; feasibility and validation of an instrument for the screening of poststroke fatigue throughout the rehabilitation process. *J Stroke Cerebrovasc* 2016; 25: 188-196.

15. Mills RJ, Pallant JF, Koufali M, et al. Validation of the neurological fatigue index for stroke (NFI-Stroke). *Health and Quality of Life Outcomes* 2012; 10: 51.

16. Beaton DE, Bombardier C, Guillemin F and Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 2000; 25: 3186-3191.

17. Polit DF, Beck CT and Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Res Nurs Health* 2007; 30: 459-467.

18. Yousufuddin M and Young N. Aging and ischemic stroke. *Aging (Albany NY)* 2019; 11: 2542-2544.

19. Arafat SMY, Chowdhury HR, Qusar M and Hafez M. Cross cultural adaptation & psychometric validation of research instruments: A methodological review. *Journal of Behavioral Health* 2016; 5: 129-136.

20. Polit DF. Getting serious about test–retest reliability: A critique of retest research and some recommendations. *Qual Life Res* 2014; 23: 1713-1720.

21. Wang MY, Liu IC, Chiu CH and Tsai PS. Cultural adaptation and validation of the Chinese version of the Fatigue Severity Scale in patients with major depressive disorder and nondepressive people. *Qual Life Res* 2016; 25: 89-99.

22. NadaraJah M, Mazlan M, Abdul-Latif L and Goh HT. Test-retest reliability, internal consistency and concurrent validity of fatigue severity scale in measuring post-stroke fatigue. *Eur J Phys Rehab Med* 2017; 53: 703-709.

23. Chiu HY, Li W, Lin JH, Su YK, Lin EY and Tsai PS. Measurement properties of the Chinese version of the Mental Fatigue Scale for patients with traumatic brain injury. *Brain Injury* 2018; 32: 652-664.

24. Zhang JX and Schwarzer R. Measuring optimistic self-beliefs: A Chinese adaptation of the General Self-Efficacy Scale. *Psychologia* 1995; 38: 174-181.

25. Lee HCB, Chiu HF, Kwok WY, Kwong PK, Leung CM and Chung DWS. Chinese elderly and

the GDS short form: A preliminary study. *Clin Gerontologist* 1993; 14: 37-42.

26. Kelly-Hayes PM, Robertson JT, Broderick JP, et al. The American Heart Association stroke outcome classification. *Stroke* 1998; 29: 1274-1280.

27. Lee HCB, Chiu HFK and Kwong PPK. Cross-validation of the Geriatric Depression Scale short form in the Hong Kong elderly. *Bulletin of the Hong Kong Psychological Society* 1994; 32/33: 72-77.

28. Chau J, Martin CR, Thompson DR, Chang AM and Woo J. Factor structure of the Chinese version of the Geriatric Depression Scale. *Psychology, Health & Medicine* 2006; 11: 48-59.

29. Polit DF and Beck CT. *Nursing research: Generating and assessing evidence for nursing practice*. 8th ed. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins, 2008.

30. Landis JR and Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159-174.

31. Koo TK and Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine* 2016; 15: 155-163.

32. Stratford PW. Getting more from the literature: Estimating the standard error of measurement from reliability studies. *Physioth Can* 2004; 56: 27-30.

33. Haley SM and Fragala-Pinkham MA. Interpreting change scores of tests and measures used in physical therapy. *Phys Ther* 2006; 86: 735-743.

34. Ostelo RWJG, de Vet HCW, Knol DL and van den Brandt PA. 24-item Roland-Morris Disability

Questionnaire was preferred out of six functional status questionnaires for post-lumbar disc surgery. *J Clin Epidemiol* 2004; 57: 268-276.

35. McHorney CA and Tarlov AR. Individual-patient monitoring in clinical practice: Are available health status surveys adequate? *Qual Life Res* 1995; 4: 293-307.

36. Floyd FJ and Widaman KF. Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment* 1995; 7: 286-299.

37. Iacobucci D. Structural equations modeling: Fit indices, sample size, and advanced topics. *Journal of Consumer Psychology* 2010; 20: 90-98.

38. Hu Lt and Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal* 1999; 6: 1-55.

39. Leech NL, Barrett KC and Morgan GA. *IBM SPSS for intermediate statistics: Use and interpretation*. 5th ed. New York, NY: Routledge, 2015.

40. Dilorio C. The evaluation of self-report instruments for use in research. *Journal of Neuroscience Nursing* 1992; 24: 350-353.

41. Derksen A, Mokkink LB, Rietberg MB, Knol DL, Ostelo RW and Uitdehaag BM. Validation of a Dutch version of the Neurological Fatigue Index (NFI-MS) for patients with multiple sclerosis in the Netherlands. *Qual Life Res* 2013; 22: 2435-2441.

42. Lopes J, Lavado EL and Kaimen-Maciel DR. Validation of the Brazilian version of the

neurological fatigue index for multiple sclerosis. *Arq Neuro-Psiquiat* 2016; 74: 320-328.

43. Young CA, Wong SM, Quincey A-MC and Tennant A. Measuring Physical and Cognitive Fatigue in People With Post-Polio Syndrome: Development of the Neurological Fatigue Index for Post-Polio Syndrome (NFI-PP). *PM&R* 2018; 10: 129-136.

44. Smidt N, van der Windt DA, Assendelft WJ, et al. Interobserver reproducibility of the assessment of severity of complaints, grip strength, and pressure pain threshold in patients with lateral epicondylitis. *Arch Phys Med Rehab* 2002; 83: 1145-1150.

45. Karasz A and Mckinley PS. Cultural differences in conceptual models of everyday fatigue: A vignette study. *J Health Psychol* 2007; 12: 613-626.

46. Swank JM and Mullen PR. Evaluating evidence for conceptually related constructs using bivariate correlations. *Meas Eval Couns Dev* 2017; 50: 270-274.

47. Kuppuswamy A, Clark EV, Turner IF, Rothwell JC and Ward NS. Post-stroke fatigue: A deficit in corticomotor excitability? *Brain* 2015; 138: 136-148.

48. Tielemans NS, Schepers VP, Visser-Meily JM, Post MW and van Heugten CM. Associations of proactive coping and self-efficacy with psychosocial outcomes in individuals after stroke. *Arch Phys Med Rehab* 2015; 96: 1484-1491.

49. Douven E, Kohler S, Schievink SHJ, et al. Temporal associations between fatigue, depression, and apathy after stroke: Results of the cognition and affect after stroke, a prospective evaluation of risks study. *Cerebrovascular Diseases* 2017; 44: 330-337.

50. Stoquart G, Detrembleur C and Lejeune TM. The reasons why stroke patients expend so much energy to walk slowly. *Gait Posture* 2012; 36: 409-413.

51. Smulders K, van Swigchem R, de Swart BJM, Geurts ACH and Weerdesteyn V. Community-dwelling people with chronic stroke need disproportionate attention while walking and negotiating obstacles. *Gait Posture* 2012; 36: 127-132.

52. Woodford J, Farrand P, Watkins ER and LLewellyn DJ. ”I don’t believe in leading a life of my own, I lead his life”: A qualitative investigation of difficulties experienced by informal caregivers of stroke survivors experiencing depressive and anxious symptoms. *Clin Gerontologist* 2018; 41: 293-307.

53. Rose D, Seidler A, Nübling M, et al. Associations of fatigue to work-related stress, mental and physical health in an employed community sample. *BMC Psychiatry* 2017; 17: 167.

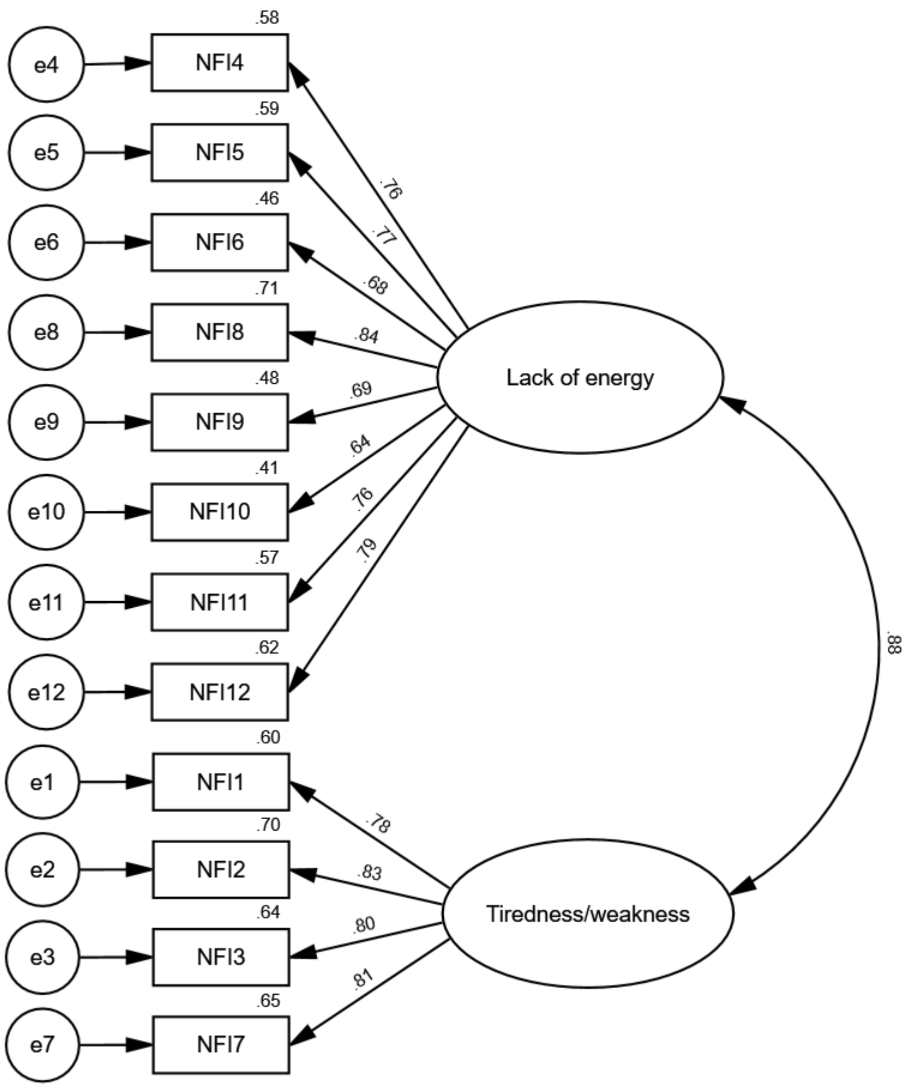

Figure 1 Confirmatory factor analysis of the Chinese version of the Neurological Fatigue Index-Stroke: the 2-factor model

Rectangles represent items in the Chinese version of the Neurological Fatigue Index-Stroke, with NFI1 representing item 1 of the scale. Circles represent measurement errors, with e1 representing a measurement error of item 1 of the scale. Ellipses are the latent factors of the scale. The figure on top of the rectangle indicates a squared multiple correlation. The figure on the straight line indicates a standardized regression weight. The figure on the curved line indicates the correlation between the two latent factors of the scale.

Table 1 Summary of reliability and responsiveness of the Chinese version of the Neurological Fatigue Index-Stroke

Item	Exact agreement (%)	Weighted Kappa value (95% confidence interval)
1	88.9	0.79 (0.57-1.01)
2	70.4	0.47 (0.17-0.77)
3	70.4	0.48 (0.20-0.77)
4	81.5	0.58 (0.27-0.89)
5	77.8	0.58 (0.28-0.87)
6	77.8	0.60 (0.33-0.87)
7	85.2	0.69 (0.44-0.95)
8	77.8	0.47 (0.14-0.80)
9	77.8	0.53 (0.21-0.86)
10	85.2	0.73 (0.49-0.97)
11	81.5	0.65 (0.39-0.91)
12	74.1	0.52 (0.22-0.81)

	Intra-class correlation coefficient	Cronbach's α	Standard error of measurement	Minimal detectable change
Summary scale	0.93*	0.88	1.05	2.92
Physical subscale	0.92*	0.87	0.97	2.68
Cognitive subscale	0.88*	0.69	0.57	1.57

*P<0.001

Table 2 Item-level and scale-level content validity index of the Chinese version of the Neurological Fatigue Index-Stroke

Item	Expert A	Expert B	Expert C	Expert D	Expert E	Item-level content validity index
1	4	4	4	4	4	1
2	4	4	2	4	4	0.8
3	4	4	4	4	4	1
4	4	4	4	4	4	1
5	4	4	3	4	4	1
6	4	4	3	4	4	1
7	4	4	4	4	4	1
8	4	4	4	4	4	1
9	4	4	3	4	4	1
10	4	4	4	4	4	1
11	4	3	3	2	2	0.6
12	4	4	4	4	4	1
				Scale-level content validity index		0.95

Table 3 Rotated factor matrix of the Chinese version of the Neurological Fatigue Index-Stroke (N=112)

	Factor	
	1	2
Item	(Lack of energy)	(Tiredness/weakness)
5	0.69	
8	0.65	
12	0.62	
6	0.60	
4	0.58	
10	0.54	
11	0.43	
9	0.39	
2		0.77
1		0.69
7		0.68
3		0.60
Eigenvalues	5.64	1.16
Variance explained (%)	47.0	9.65

Note: Items 1-8 and 9-12 contribute to physical and cognitive fatigue, respectively, in the original Neurological Fatigue Index-Stroke[15]

Table 4 Spearman's correlation coefficients of the Chinese version of the Neurological Fatigue Index-Stroke

	Fatigue Severity Scale	Mental Fatigue Scale	General Self-Efficacy Scale	Geriatric Depression Scale
C-NFI-Stroke summary scale	0.62[a]	0.63[a]	-0.35[a]	0.60[a]
C-NFI-Stroke physical subscale	0.60[a]	0.61[a]	-0.31[a]	0.58[a]
C-NFI-Stroke cognitive subscale	0.55[a]	0.61[a]	-0.37[a]	0.54[a]

C-NFI-Stroke = Chinese version of the Neurological Fatigue Index-Stroke.

[a] $P<0.01$

Table 5 Comparison of fatigue scores between stroke survivors (N=66) and healthy older people (N=65)

	Median (interquartile range)	Mann-Whitney U	Z	*p*
Summary score		1133.50	-4.67	<0.001
Stroke survivors	16.00 (6.50)			
Healthy older people	12.00 (10.50)			
Physical score		1128.50	-4.70	<0.001
Stroke survivors	13.00 (5.00)			
Healthy older people	9.00 (8.00)			
Cognitive score		987.00	-5.39	<0.001
Stroke survivors	6.00 (2.25)			
Healthy older people	4.00 (4.50)			

Table 6 Comparison of fatigue scores between participants with and without depressive symptoms

	Stroke survivors (N=66)				Healthy older people (N=65)			
	Median (IQR)	**Mann-Whitney U**	**Z**	***p***	**Median (IQR)**	**Mann-Whitney U**	**Z**	***p***
Summary score		180.50	-4.03	<0.001		98.50	-3.22	0.001
GDS score ≥8	19.00 (7.00)				19.00 (4.00)			
GDS score <8	15.00 (6.00)				11.00 (10.00)			
Physical score		172.50	-4.16	<0.001		115.50	-2.91	0.004
GDS score ≥8	15.00 (5.50)				14.50 (3.75)			
GDS score <8	11.00 (4.00)				8.00 (7.00)			
Cognitive score		211.00	-3.65	<0.001		109.50	-3.05	0.002
GDS score ≥8	7.00 (2.50)				7.00 (2.75)			
GDS score <8	6.00 (3.00)				4.00 (4.00)			

GDS = Geriatric Depression Scale; IQR = interquartile range.