

RESEARCH PAPER

Investigating the feasibility and reliability of the Pictorial Fit-Frail Scale

LISA MCGARRIGLE^{1,2}, EMMA SQUIRES¹, LINDSAY M. K. WALLACE^{1,2}, JUDITH GODIN^{1,2}, MARY GORMAN^{1,3}, KENNETH ROCKWOOD^{1,2}, OLGA THEOU^{1,2}

¹Geriatric Medicine Research, Centre for Health Care of the Elderly, Nova Scotia Health Authority, Halifax B3H 2E1, Nova Scotia, Canada

²Division of Geriatric Medicine, Department of Medicine, Dalhousie University, Halifax B3H 2E1, Nova Scotia, Canada

³St. Martha's Regional Hospital, Antigonish, Nova Scotia, Canada

Address correspondence to: Olga Theou, Camp Hill Veterans' Memorial Building, 5955 Veterans' Memorial Lane, Halifax B3H 2E1, Nova Scotia, Canada. Tel: (+1) 902 473 4846. Email: olga.theou@dal.ca

Abstract

Background: the Pictorial Fit-Frail Scale (PFFS) was designed as a simple and practical approach to the identification of frailty.

Objectives: To investigate the feasibility and reliability of this visual image-based tool, when used by patients, caregivers and healthcare professionals (HCPs) in clinical settings.

Design: observational study.

Setting: three outpatient geriatric healthcare settings.

Subjects: patients ($n = 132$), caregivers ($n = 84$), clinic nurses ($n = 7$) and physicians ($n = 10$).

Methods: the PFFS was administered to all patients. Where available, HCPs and caregivers completed the scale based on the patients' health. In the geriatric day hospital, the PFFS was completed on admission and administered again within 7–14 days. Time and level of assistance needed to complete the scale were recorded. Intraclass correlation coefficients (ICCs) and 95% confidence intervals (CIs) were used to assess test–retest and inter-rater reliability.

Results: mean time to complete the scale (minutes:seconds \pm SD) was 4:30 \pm 1:54 for patients, 3:13 \pm 1:34 for caregivers, 1:28 \pm 0:57 for nurses and 1:32 \pm 1:40 for physicians. Most patients were able to complete the scale unassisted (64%). Mean patient PFFS score was 11.1 \pm 5.3, mean caregiver score was 13.2 \pm 6.3, mean nurse score was 10.7 \pm 4.5 and mean physician score was 11.1 \pm 5.6; caregiver scores were significantly higher than patient ($P < 0.01$), nurse ($P < 0.001$) and physician ($P < 0.01$) scores. Test–retest reliability was good for patients (ICC = 0.78, [95%CI = 0.67–0.86]) and nurses (ICC = 0.88 [0.80–0.93]). Inter-rater reliability between HCPs was also good (ICC = 0.75 [0.63–0.83]).

Conclusion: the PFFS is a feasible and reliable tool for use with patients, caregivers and HCPs in clinical settings. Further research on the validity and responsiveness of the tool is necessary.

Keywords: frailty, reliability, feasibility, psychometric properties, assessment, older people

Key points

- The Pictorial Fit-Frail Scale (PFFS) is a recently developed frailty measure designed for use by patients, caregivers, and/or healthcare professionals (HCPs).
- Most patients were able to complete the scale unassisted.
- On average, patients and caregivers were able to complete the scale in under 5 minutes, and HCPs in under 2 minutes.
- The PFFS demonstrates good test–retest and inter-rater reliability when used in outpatient geriatric healthcare settings.

Introduction

Frailty is characterised by increased vulnerability to adverse health outcomes due to age-related decline across multiple inter-related physiological systems [1]. The proportion of older adults is increasing [2], and currently over 50% of Canadians aged 85 or older are considered frail [3]. Frailty is linked to higher mortality, poorer outcomes in relation to illness and healthcare interventions [4] and increased health and social care costs [5].

Frailty status can vary, including in response to interventions (e.g. a probiotic formulation) [6]. As such, the identification of frailty by healthcare professionals (HCPs) is important in order to protect at-risk individuals and reduce healthcare costs. Several measures are currently used to identify frailty in acute care [7]; however, many are limited by being impractical for people who are severely frail or who are experiencing communication difficulties. Many scales are developed for use with either the patient or the clinician, and as such, account for only a single perspective. Further, frailty screening measures are sometimes used as assessments, or are used without an evident relationship to care planning.

In response, our group developed a new frailty measure designed for use by patients, caregivers and/or HCPs, the Pictorial Fit-Frail Scale (PFFS) (to access or preview the scale visit www.dal.ca/sites/gmr/our-tools/pictorial-fit-frailty-scale.html). A multidisciplinary team created an initial visual frailty scale and the PFFS was then developed through a four-phase iterative process, whereby the scale was continually modified based on feedback from experts and stakeholders [8]. Changes were aimed at including the most informative items, relevant for both sexes, across cultures. The scale uses visual images to depict a range of domains associated with frailty. It was designed to be quick and simple to use, and sensitive to cultural differences and varying levels of cognition/communication ability. It allows frailty to be graded, and for areas of difficulty (to which interventions might be targeted) to be identified specifically. In this way, it allows for issues, which might otherwise go unrecognised to be documented explicitly.

The aim of this study was to investigate: (1) the feasibility of use of the PFFS in clinical settings, (2) test–retest reliability, (3) inter-rater reliability between HCPs and (4) differences in PFFS assessments between patients, caregivers, and HCPs.

Methods

Recruitment was conducted across three outpatient healthcare settings in the Nova Scotia Health Authority:

- The Geriatric Day Hospital & Falls Clinic of the Centre for Health Care of the Elderly, Halifax
- The Outpatient Geriatric Clinic at St. Martha's Regional Hospital, Antigonish
- The Geriatric Ambulatory Care/Memory Disability Clinic of the Centre for Health Care of the Elderly, Halifax

Participants were eligible for enrollment if they were English speaking without severe visual impairment (in order to see the scale and graphics). HCPs from each setting assessed patients for eligibility at their initial visit and referred to research personnel for recruitment. In the day hospital and geriatric clinic settings, patients aged 60 years or older were invited to participate. Those aged 50 years or older were invited from the memory clinic due to the broader age range of clinic attendees. A member of the research team provided the patients with a detailed overview of the study, and where available, invited their caregiver to participate, and obtained informed consent. Participants were asked to read the instructions of the PFFS (see www.dal.ca/sites/gmr/our-tools/pictorial-fit-frailty-scale.html), which advised to tick the box underneath the picture that best represented their usual level of functioning for each of the 14 health domains. To assess the feasibility of PFFS use by patients, we recorded the level of assistance required to complete the PFFS on a three-point scale: 0 = no assistance required; 1 = verbal prompt given for first domain, e.g. "Which level is closest to your usual state?"; 2 = verbal description given for each level of the first domain, e.g. "The first picture shows someone who feels happy. Would you say that you typically feel happy?" If no, the researcher described each of the adjacent levels. Time taken to complete the scale (minutes/seconds) was recorded by the researcher to assess feasibility. The nurse and/or physician who assessed the patient was asked to complete the PFFS based on the health of the patient and self-record the time taken to complete it. For all participants, we collected demographic information. Nurses were asked to rate the patients' communication capacity on a 5-point scale: 1 = Excellent; 2 = Very Good; 3 = Good; 4 = Fair; 5 = Poor. Where nurses indicated that communication capacity was "Fair" or "Poor", they were asked to identify the likely reason for this from the following options: 1 = Issues with English Proficiency; 2 = Cognitive Impairment; 3 = Other; 4 = Don't Know.

In the day hospital, patients received rehabilitation services, following a fall or a mobility impairment, twice a week over an approximate 8-week period. These participants were asked to complete the PFFS a second time 7–14 days after the initial assessment, as changes due to intervention were not expected to be apparent at this stage. The same nurses who assessed these patients at baseline, completed the PFFS a second time 7–14 days later.

Data analysis

PFFS total scores were calculated by summing the scores for each domain; level one for each domain was scored 0, level two was scored 1, etc. A standardised frailty index (FI) was constructed by dividing the PFFS score by the maximum possible score (maximum score = 43 if no data were missing) with higher scores indicating increased levels of frailty. The purpose of the FI calculation was to facilitate comparison with other studies, however, as the PFFS has not yet been concurrently validated, the FI score

should be interpreted cautiously. The FI was constructed in a similar way to previous FIs, with each level of the PFFS treated as a separate variable (e.g. Level 1 of the domain “Function” could be replaced with the question “do you have trouble shopping?”, level 2 could be replaced with the question “can you prepare your own meals?”, etc.). No participant was missing more than 20% of the data, allowing calculation of PFFS scores, including FI scores, for all participants [9].

Linear mixed models for repeated measures were used to compare PFFS scores and completion times among the four types of rater (patient, caregiver, nurse and physician). This approach prevented listwise deletion due to missing data [10]. Pearson correlations were conducted to investigate if the degree of frailty, as measured by PFFS scores, was associated with time taken to complete the scale.

Intraclass correlation coefficients (ICCs) and their 95% confidence intervals (CI) were used to assess test–retest and inter-rater reliability. ICC reflects both degree of correlation and agreement between measurements [11]. Test–retest ICC estimates were based on a single measurement, absolute agreement and two-way mixed effects model. Inter-rater reliability estimates were calculated between nurse and physician scores. As this was a multi-site study, it was not possible for the same set of raters to rate all patients, and as such, inter-rater ICC estimates were based on a single measurement, one-way random effects model [12]. Recognising the arbitrariness of any cut-off point in the early stages of research, a minimum reliability of 0.70 is sufficient to conclude good reliability [13].

Analyses were conducted using SPSS version 24 [14].

Ethics

The testing protocol was reviewed and approved by the Nova Scotia Health Authority Research Ethics Board. All participants provided written informed consent.

Results

Data were collected on 150 patients across three sites. Eighteen patients did not complete the scale as their caregivers and/or HCP felt they were unable to do so ($n=5$), the patients were not interested ($n=5$) or did not have time ($n=3$), and for five patients their reason was unclear. In addition, 84 caregivers completed the PFFS, rating the patient. Seven nurses completed the PFFS relating to 146 patients, and 10 physicians completed the scale relating to 79 patients. Table 1 outlines the patient characteristics for the full sample. The mean age of patients was 78 years (SD = 8 years) and 54.7% were female. The majority of patients were white, married, with secondary level education. Twenty-nine percent of patients were rated by nurses as having fair/poor communication capacity, with cognitive impairment identified as the reason for this in 92% of cases.

Table 1. Patient characteristics ($n = 150$)

Age, mean (SD), range	77.9 (8.5), 51–99
Females, n (%)	81 (54.7)
Race, n (%)	
White	144 (97.3)
Black	3 (2)
Other	1 (0.7)
Marital status, n (%)	
Married	77 (52.7)
Divorced	21 (14.4)
Widowed	42 (28.8)
Never married	6 (4.1)
Education, n (%)	
Primary	20 (13.8)
Secondary	57 (39.3)
Post-secondary	36 (24.8)
University	32 (22.1)
Communication capacity, n (%)	
Excellent	26 (18.4)
Very good	40 (28.4)
Good	34 (24.1)
Fair	27 (19.1)
Poor	14 (9.9)

Feasibility

Of the 132 patients who completed the PFFS, 84 (64%) were able to do so unassisted, with 23 (17%) requiring level 1 assistance, and 25 (19%) requiring level 2 assistance (Figure 1a).

Mean time taken to complete the scale (minutes:seconds \pm SD) was 4:30 \pm 1:54 for patients, 3:13 \pm 1:34 for caregivers, 1:28 \pm 0:57 for nurses and 1:32 \pm 1:40 for physicians. Completion times significantly differed between raters ($P < 0.001$), with patients taking the longest to complete, followed by caregivers. HCPs had significantly shorter completion times compared to patients and caregivers, and nurses' completion times were shorter than physicians' (Figure 1b). There was a small, positive correlation between patient PFFS scores and time taken to complete the scale, with higher frailty being positively associated with longer completion times ($r = 0.22$, $P = 0.017$). Higher patient frailty was positively associated with longer completion times for physicians ($r = 0.30$, $P = 0.018$). There were no significant correlations between caregiver or nurse PFFS scores and completion times.

Reliability

Test–retest reliability was calculated for day hospital patients ($n = 66$) and nurse assessments on patients ($n = 53$); there was insufficient data at re-test for caregivers ($n = 13$) and physicians ($n = 1$). Test–retest reliability was good for patients (ICC 0.78 [95%CI = 0.67–0.86]) and nurses (ICC = 0.88 [0.80–0.93]). Using data from all sites, inter-rater reliability was found to be good between nurses and physicians (data on 77 patients) (ICC = 0.75 [0.63–0.83]) (see Appendix in the supplementary data, available at *Age and Ageing* online).

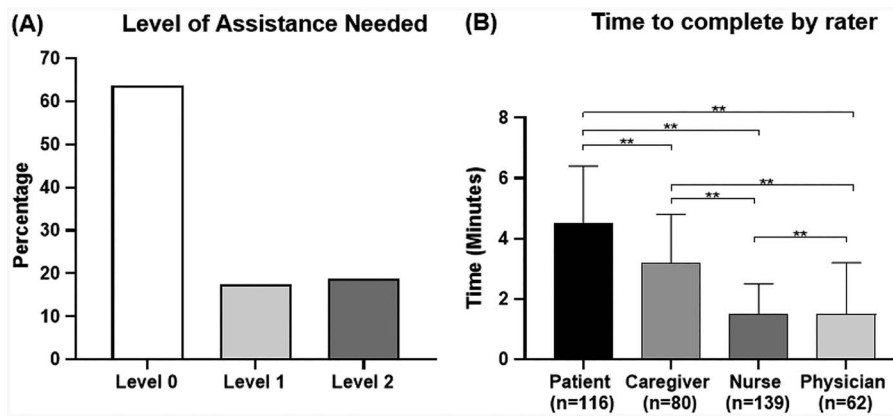


Figure 1. Level of assistance needed by patients to complete the scale and rater completion times. (A) Depicts the level of assistance needed to complete the scale. Level 0 = no assistance required; Level 1 = verbal prompt given for first domain, e.g. “Which level is closest to your usual state?”; Level 2 = verbal description given for each level of the first domain, e.g. “The first picture shows someone who feels happy. Would you say that you typically feel happy?”. If no, adjacent levels described. (B) Depicts completion times (minutes) by rater (***P* < 0.001). *n* for nurses and physicians represents the number of patients assessed.

Table 2. Mean PFFS and FI scores for each rater

Rater	<i>n</i> ^a	PFFS score (SD)	Frailty index (SD)
Patient	132	11.1 (5.3)	0.26 (0.12)
Caregiver	84	13.2 (6.3)	0.31 (0.15)
Nurse	146	10.7 (4.5)	0.25 (0.10)
Physician	79	11.1 (5.6)	0.26 (0.13)

^a*n* represents the number of patients assessed; PFFS = total PFFS score and FI = Frailty index score.

PFFS scores by rater

Mean patient PFFS score was 11.1 ± 5.3 (FI = 0.26 ± 0.12), mean caregiver score was 13.2 ± 6.3 (FI = 0.31 ± 0.15), mean nurse score was 10.7 ± 4.5 (FI = 0.25 ± 0.10), and mean physician score was 11.1 ± 5.6 (FI = 0.26 ± 0.13) (Table 2). Caregiver PFFS scores were significantly higher than patient (*P* < 0.01), nurse (*P* < 0.001) and physician (*P* < 0.01) scores. Patient, nurse and physician scores were not significantly different from each other.

Discussion

To establish the feasibility of the newly-developed PFFS, our group investigated the proportion of patients who were able to complete the scale and the time it took to complete the scale for all raters. All patients who opted to complete the scale were able to do so. Of these patients, most were able to complete the scale unassisted, however over a third of patients (36%), primarily those recruited from the memory clinic, required some level of assistance (Figure 1). On average, patients and caregivers could complete the scale in under 5 minutes, and HCPs in under 2 minutes. Test–retest reliability was good for patients and nurses and inter-rater reliability was good between HCPs. Caregiver assessments of frailty were significantly higher than patient and HCP assessments. Previous research has shown that caregiver stress

can negatively bias caregiver ratings of daily functioning and quality of life in people with dementia [15]. It may be the case that greater perceived frailty by this group can be attributed to caregiver stress, however, this was not assessed as part of this study and requires further research. As expected, time taken for patients to complete the scale was significantly longer for those with greater frailty. Higher patient frailty was also significantly associated with longer completion times for physicians, but not for nurses or caregivers. This may be due to greater time spent with the patient on the part of caregivers and nurses, resulting in increased familiarity with the patient’s abilities across domains.

Inter-rater reliability in this study was assessed between nurses and physicians, and as such, may have been influenced by training differences in these professions. As the study was conducted in outpatient clinics where each patient was routinely assessed by one nurse and one physician, investigating nurse–nurse or physician–physician inter-rater reliability was not possible. Even so, despite potential differences in training between raters, we found that inter-rater reliability of the PFFS was good. Further research can address whether additional rater training could enhance this, and the impact on feasible implementation, and carry-over of the result, in routine clinical care. Determining whether or not frailty assessments are consistent among different care professionals is important, particularly in geriatric care settings where nurses and physicians work in close collaboration. The utility of the PFFS in these settings is couched in its ability to rapidly gain key information about the extent to which an individual has age-related health deficits. Having established that PFFS scores are consistent between nurses and physicians trained in geriatric care, the next step will be to investigate if the PFFS may be useful in supporting HCPs who are not specifically trained in the management of frail older adults, but due to an ageing population will need to undertake complex care in this group.

A recent systematic review on the psychometric properties of multi-component frailty measures revealed that reliability was only assessed in 8/38 tools included in the study [16]. Cohen's Kappa was most frequently used to assess inter-rater reliability, with scores ranging from 0.63 (the EASY-Care Two-step Older persons Screening) to 0.72 (the Evaluative Index for Physical Frailty). Similarly to ICC, a Kappa value greater than 0.70 is considered satisfactory [17]. With an ICC of 0.75, the inter-rater reliability of the PFFS fits comfortably in the estimates for tools included in the review. Test–retest reliability was only assessed for one measure in the review (the Tilburg frailty indicator) with a Pearson correlation coefficient of 0.79 [16]. Although this indicates good test–retest reliability, the Pearson coefficient is considered a liberal measure as it often produces a value greater than true reliability, with ICC being the preferred measure in this instance [17]. PFFS test–retest reliability is again similar to the value reported in the systematic review, with ICCs of 0.78 for patients (Pearson correlation of 0.77) and 0.88 for nurses (Pearson correlation of 0.87).

Strengths and limitations

Major strengths of the PFFS are its demonstrated utility in populations who are experiencing communication difficulties and its ability to account for multiple perspectives on frailty. The PFFS aims to be a useful tool to help “geriatrize” care. Given that the ageing of the population is rapidly outpacing the supply of geriatricians, many HCPs who are not trained in the management of frail older adults will need to undertake complex care. The PFFS may offer support in that regard, a proposition we aim to test in an online, free-to-use, roll-out of the tool, which will follow a similar online multi-site validation study. Our findings suggest that the PFFS is a feasible and reliable tool for use with patients, caregivers, and HCPs, as well as for use with individuals with varying levels of communication ability; more than a quarter of patients who completed the PFFS were rated by an HCP as having poor/fair communication capacity, with cognitive impairment given as the reason for this in 92% of cases. Even so, the study had some limitations. Other aspects of feasibility, such as use in routine care and by non-geriatricians—an essential consideration if we are to “geriatrize” routine care—are subjects for future research. Communication capacity was assessed by HCPs using a five-point scale, however the health literacy of the patient was not fully tested using validated measures. This study was also limited in that it took place in Nova Scotia only and recruitment was limited to HCPs trained in geriatrics. Furthermore, there is a need for future research with a focus on determining the diagnostic test accuracy of the tool, in addition to its construct validity, predictive validity and responsiveness when compared with existing and widely used frailty measures, such as a FI derived from the Comprehensive Geriatric Assessment.

Conclusions

The PFFS is a feasible and reliable tool for use with patients, caregivers, and HCPs in outpatient geriatric healthcare settings. The tool is suitable for use with individuals with varying levels of frailty and communication ability. Future research should investigate the validity and responsiveness of the PFFS through comparison with validated frailty measures, such as the FI from a Comprehensive Geriatric Assessment. In addition, future research should further investigate the effect of health literacy and cognitive impairment on the psychometric properties of the tool, and evaluate the impact of implementing the PFFS in routine care with regard to decision-making and healthcare outcomes.

Supplementary data:

Supplementary data mentioned in the text are available to subscribers in *Age and Ageing* online.

Declaration of Conflicts of Interest:

K.R. is the President and Chief Science Officer of DGI Clinical, which in the last 5 years has contracts with pharma and device manufacturers (Baxter, Baxalta, Shire, Hollister, Nutricia, Roche, Otsuka) on individualized outcome measurement. In 2017, he attended an advisory board meeting with Lundbeck. Otherwise, any personal fees are for invited guest lectures and academic symposia, received directly from event organizers, chiefly for presentations on frailty. He is Associate Director of the Canadian Consortium on Neurodegeneration in Aging, which is funded by the Canadian Institutes of Health Research, and with additional funding from the Alzheimer Society of Canada and several other charities, as well as, in its first phase (2013–2018), from Pfizer Canada and Sanofi Canada. He receives career support from the Dalhousie Medical Research Foundation as the Kathryn Allen Weldon Professor of Alzheimer Research, and research support from the Canadian Institutes of Health Research, the QEII Health Science Centre Foundation, the Capital Health Research Fund and the Fountain Family Innovation Fund of the QEII Health Science Centre Foundation. O.T. and K.R. have asserted copyright of the Pictorial Fit-Frail Scale. It will be made freely available for research, education and not-for-profit care.

Declaration of Funding:

L.M. received support from the Mitacs Elevate Program as the recipient of a Postdoctoral Fellowship award (IT09427) to complete this project. L.M.K.W. is supported by a Canadian Institutes of Health Research Doctoral Scholarship (GSD-154170). O.T. was supported by the Canadian Frailty Network (Transformative Grant Pilot Study Award

TG2015-24PC) and the Nova Scotia Health Authority Research Fund (2016-Theou). K.R.'s work on frailty is supported by the Canadian Institutes of Health Research, the Dalhousie Medical Research Foundation Kathryn Allen Weldon Chair of Alzheimer Disease Research, and the Fountain Family Innovation Fund of the Queen Elizabeth II Health Research Foundation.

References

1. Clegg A, Young J, Iliffe S, Rikkert MO, Rockwood K. Frailty in elderly people. *Lancet* 2013; 381: 752–62.
2. Ferrucci L, Giallauria F, Guralnik JM. Epidemiology of aging. *Radiol Clin N Am* 2008; 46: 643–52.
3. Hoover M, Rotermann M, Sanmartin C, Bernier J. Validation of an index to estimate the prevalence of frailty among community-dwelling seniors. *Health Rep* 2013; 24: 10–7.
4. Muscedere J, Waters B, Varambally A *et al.* The impact of frailty on intensive care unit outcomes: a systematic review and meta-analysis. *Intensive Care Med* 2017; 43: 1105–22.
5. Walters K, Frost R, Kharicha K *et al.* Home-based health promotion for older people with mild frailty: the HomeHealth intervention development and feasibility RCT. *Health Technol Assess* 2017; 21: 1–128.
6. Theou O, Jayanama K, Fernández-Garrido J *et al.* Can a prebiotic formulation reduce frailty levels in older people? *J Frailty Aging* 2019; 8: 1–5.
7. Theou O, Squires E, Mallery K *et al.* What do we know about frailty in the acute care setting? A scoping review. *BMC Geriatr* 2018; 18: 139.
8. Theou O, Andrew M, Ahip SS *et al.* The pictorial fit-frail scale: developing a visual scale to assess frailty. *Can Geriatr J* 2019; 22: 64–74.
9. Searle SD, Mitnitski A, Gahbauer EA, Gill TM, Rockwood K. A standard procedure for creating a frailty index. *BMC Geriatr* 2008; 8: 24.
10. Weinger K, Beverly EA, Lee Y, Sitnikov L, Ganda OP, Caballero AE. The effect of a structured behavioral intervention on poorly controlled diabetes: a randomized controlled trial. *Arch Intern Med* 2011; 171: 1990–9.
11. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016; 15: 155–63.
12. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1996; 1: 30–46.
13. Nunnally JC. *Psychometric Theory*. 2nd edition. New York, NY: McGraw-Hill, 1978.
14. Corp IBM. *IBM SPSS Statistics For Windows*. Armonk, NY: IBM Corp., 2016.
15. Pfeifer L, Drobetz R, Fankhauser S, Mortby ME, Maercker A, Forstmeier S. Caregiver rating bias in mild cognitive impairment and mild Alzheimer's disease: impact of caregiver burden and depression on dyadic rating discrepancy across domains. *Int Psychogeriatr* 2013; 25: 1345–55.
16. Sutton JL, Gould RL, Daley S *et al.* Psychometric properties of multicomponent tools designed to assess frailty in older adults: a systematic review. *BMC Geriatr* 2016; 16: 55.
17. Streiner DL, Norman GR, Cairney J. *Health Measurement Scales: A Practical Guide to their Development and Use*. 5th edition. United Kingdom: Oxford University Press, 2015.

Received 4 March 2019; editorial decision 30 June 2019
