



Assessing Frailty with the Tilburg Frailty Indicator (TFI): A Review of Reliability and Validity

Robbert J Gobbens ¹⁻³
Izabella Uchmanowicz ⁴

¹Faculty of Health, Sports and Social Work, Inholland University of Applied Sciences, Amsterdam, the Netherlands;

²Zonnehuisgroep Amstelland, Amstelveen, the Netherlands;

³Department Family Medicine and Population Health, Faculty of Medicine and Health Sciences, University of Antwerp, Antwerp, Belgium; ⁴Division of Nursing in Internal Medicine Procedures, Department of Clinical Nursing, The Faculty of Health Sciences, Wroclaw Medical University, Wroclaw, Poland

Objective: The Tilburg Frailty Instrument (TFI) is an instrument for assessing frailty in community-dwelling older people. Since its development, many studies have been carried out examining the psychometric properties. The aim of this study was to provide a review of the main findings with regard to the reliability and validity of the TFI.

Methods: We conducted a literature search in the PubMed and CINAHL databases on May 30, 2020. An inclusion criterion was the use of the entire TFI, part B, referring to the 15 components. No restrictions were placed on language or year of publication.

Results: In total, 27 studies reported about the psychometric properties of the TFI. By far, most of the studies ($n = 25$) were focused on community-dwelling older people. Many studies showed that the internal consistency and test-retest reliability are good, which also applies for the criterion and construct validity. In many studies, adverse outcomes of interest were disability, increased health-care utilization, lower quality of life, and mortality. Regarding disability, studies predominantly show results that are excellent, with an area under the curve (AUC) >0.80 . In addition, the TFI showed good associations with lower quality of life and the findings concerning mortality were at least acceptable. However, the association of the TFI with some indicators of health-care utilization can be indicated as poor (eg, visits to a general practitioner, hospitalization).

Conclusion: Since population aging is occurring all over the world, it is important that the TFI is available and well known that it is a user-friendly instrument for assessing frailty and its psychometric properties being qualified as good. The findings of this assessment can support health-care professionals in selecting interventions to reduce frailty and delay its adverse outcomes, such as disability and lower quality of life.

Keywords: older people, frailty, Tilburg Frailty Indicator, reliability, validity

Introduction

The Tilburg Frailty Instrument (TFI) is an instrument for assessing frailty in community-dwelling older people. It has been developed as a self-report instrument in that older people have to complete the TFI themselves.¹ The TFI is based on the following conceptual definition of frailty: ‘frailty is a dynamic state affecting an individual who experiences losses in one or more domains of human functioning (physical, psychological and social), which is caused by the influence of a range of variables and increases the risk of adverse outcomes’.² Both the conceptual definition of frailty and the TFI derived from it consider frailty as a multidimensional concept, including physical, psychological and social functioning of older people, and emphasize the importance of an integral approach to human functioning. The World Health Organization recommends this holistic approach to take care of frail,

Correspondence: Robbert J Gobbens
Faculty of Health, Sports and Social Work, Inholland University of Applied Sciences, De Boelelaan 1109, Amsterdam, 1081 HV, The Netherlands
Tel +31 6 21115578
Email robbert.gobbens@inholland.nl

older people.³ Paying sole attention to physical frailty can lead to fragmentation of care^{2,4} and possibly to a reduction of the quality of care and a decrease in the experienced quality of life in frail older people. According to Gilardi et al, a multidimensional approach to frailty can be more effective to plan and implement care services, as well as establish prevention programs for frail older people.⁵

The TFI contains two parts: part A, on 10 determinants of frailty, and part B, on 15 components of frailty.¹ The determinants are sex, age, marital status, education, income, ethnicity, lifestyle, life events, multimorbidity and living environment. The components of frailty refer to physical frailty (eight), psychological frailty (four) and social frailty (three). Physical frailty includes physically unhealthy, unexplained weight loss, difficulty walking, difficulty maintaining balance, poor hearing, poor vision, lack of strength in the hands and physical tiredness. Psychological frailty consists of the components of memory problems, feeling down, feeling nervous or anxious and unable to cope with problems. Social frailty includes living alone, lack of social relations and lack of social support. The total score of the TFI is 0–15, with a score ranging from 0 to 8 for physical frailty, 0 to 4 for psychological frailty and 0 to 4 for social frailty. Higher scores refer to greater frailty, as older persons with a total TFI score ≥ 5 are considered to be frail.¹

Originally, the TFI was developed in the Netherlands by Gobbens et al and based on an extensive literature search and opinions of an international group of frailty experts, including geriatricians, gerontologists, nurses and psychologists.⁶ Their first study examined the psychometric properties of the TFI in two Dutch samples of community-dwelling persons aged 75 years and older.¹ Subsequently, the TFI was translated into several languages, including Brazilian Portuguese,⁷ Danish,⁸ Italian,⁹ Portuguese,¹⁰ Polish,¹¹ German,¹² Chinese,¹³ Spanish,¹⁴ and Turkish.¹⁵ Until now, two systematic reviews and one narrative review have been published, indicating that the TFI is very suitable for assessing frailty among community-dwelling older people.^{5,16,17} According to Pialoux et al, both the TFI¹ and the SHARE Frailty Index¹⁸ are potentially suitable for screening frailty in older people in primary care settings.¹⁶ Sutton et al concluded that the TFI has the most robust evidence of reliability and validity of 38 frailty assessment instruments, including frequently used instruments, such as the Phenotype of Frailty¹⁹ and the Frailty Index.¹⁷ In addition, the narrative review by Gilardi et al identified the TFI as the best screening instrument to use in public health

because it was the only one of the selected instruments (including the Phenotype of Frailty,¹⁹ Vulnerable Elders Survey,²⁰ Frailty Index,²¹ and the SHARE Frailty Index)¹⁸ with three features: a multidimensional structure, quick and easy to use, and an accurate risk prediction of adverse outcomes of frailty.⁵ In addition, De Witte et al used the TFI as a gold standard for the validation of their instrument, The Comprehensive Frailty Assessment Instrument.²²

Since the TFI was developed approximately 10 years ago,¹ and many studies into its psychometric properties have been carried out since then,¹⁷ this study aims to provide a review of the main findings regarding this issue. We present the reliability and validity of these studies.

Methods

Literature Search

We conducted a literature search in the PubMed and CINAHL databases on May 30, 2020, using “Tilburg Frailty Indicator AND psychometric properties”, “TFI AND psychometric properties”, “Tilburg Frailty Indicator AND validity”, “TFI AND validity”, “Tilburg Frailty Indicator AND reliability”, and “TFI AND reliability”. An inclusion criterion was the use of the entire TFI, part B, referring to the 15 components. No restrictions were placed on language or year of publication. The studies were screened and selected for inclusion by the first author. In total, 27 studies were selected for the purpose of this review.

Reliability and Validity

Reliability

Four types of reliability were distinguished: internal consistency, test-retest, inter-rater reliability, and parallel forms reliability. Internal consistency refers to consistency across items of measurement. Statistical techniques used for this purpose were Cronbach’s alpha, Kuder–Richardson Formula 20 (KR-20) and item correlations. Cronbach’s alpha and KR-20 values >0.70 were considered acceptable.^{23,24} The higher the item correlations, the better the internal consistency of the measurement instrument.

Test-retest indicates consistency among time (stability). Correlations, simple agreement, kappa (chance-corrected agreement) and intraclass correlation coefficient (ICC) were used to determine test–retest reliability. The higher the correlation, simple agreement, kappa value and ICC,

the higher the concordance between the two assessments will be. The correlation coefficient was evaluated using the classification of Callegari-Jacques (weak, <0.30 ; moderate, 0.30 – 0.60 ; strong, 0.60 – 0.90 ; very strong, ≥ 0.90).²⁵ For the interpretation of the Kappa value, we used the Landis and Koch evaluation (absent, <0.10 ; weak, 0.10 – 0.20 ; fair, 0.21 – 0.40 ; moderate, 0.41 – 0.60 ; substantial, 0.61 – 0.80 ; nearly perfect, 0.81 – 1.00).²⁶ The ICC was evaluated using the guideline provided by Koo and Li (poor <0.50 ; moderate, 0.50 – 0.75 ; good, 0.75 – 0.90 ; excellent >0.90).²⁷

Inter-rater reliability concerns consistency across different researchers. Frequently used techniques to establish inter-rater reliability are Kappa and ICC.

The fourth and final type of reliability, parallel forms reliability, assesses the correlation between two equivalent versions of a measurement instrument. High correlation between the two instruments indicates high forms reliability.

Validity

Six types of validity can be distinguished: criterion, construct, content, face, structural, and known-group validity. Criterion validity concerns the relation between the score on a measurement instrument and some external criterion.²⁸ If the measurement instrument corresponds to a criterion assessed simultaneously, the validity is considered concurrent. If the measurement instrument forecasts a criterion value in the future, the validity is labeled predictive.²⁸ Criterion validity can be checked by determining a correlation coefficient and conducting receiver operating characteristics (ROC) curve analyses and calculating the Area Under the Curve (AUC). An AUC <0.7 , 0.7 – 0.8 , 0.8 – 0.9 , and ≥ 0.9 , is considered poor (no discrimination), acceptable, excellent, and outstanding, respectively.²⁹

Construct validity refers to the simultaneous process of measure and theory validation.³⁰ Convergent and divergent (discriminant) validities constitute construct validity. Convergent validity involves the degree by which two measures of constructs should be or are related. By contrast, divergent validity indicates whether measures that should be unrelated, are unrelated.³¹ Correlation tests are performed to establish convergent and divergent validity and thus, construct validity.

Content validity concerns the extent in which a measurement instrument includes all necessary components of the construct to be assessed.³² According to Burns and Grove, content validity is obtained from literature,

representatives of the population concerned and experts.³³ A measurement instrument has face validity if it appears to assess what it is supposed to assess and that it will work.³⁴ As well as when determining content validity, representatives of the population concerned and experts can be involved; however, establishing face validity is more informal, compared to content validity.

Structural validity refers to the extent to which an instrument covers the hypothetical dimension of a construct.³² According to Souza et al, factorial analysis and structural equation modeling are the appropriate statistical techniques to assess structural validity.³⁵ Finally, known-group validity involves an instrument's ability to make a distinction between groups. Group differences can be determined using a chi-square test and *t*-test. When describing the findings on the reliability and validity of the TFI in the included studies, the classification described above has been used foremost.

Results

Characteristics of the Studies

Table 1 presents the characteristics of the 27 included studies. The first studies were performed in 2010.^{1,36} Thirteen studies were carried out in the Netherlands,^{1,36–47} of which two were part of a large study that also collected data in other European countries.^{46,47} Three studies were exclusively conducted in Denmark,^{8,48,49} and two in Poland^{11,50} and Brazil.^{7,51} By far, most of the studies were conducted among community-dwelling older people. Two exceptions were studies including older Danish people admitted to a hospital⁴⁹ and one study evaluating older Dutch residents of assisted living facilities.³⁹ Two-thirds of the studies used a cross-sectional design, two of which were qualitative studies conducted in Denmark,^{8,48} and the other studies were characterized by a longitudinal design. The sample size varied from 14 to 27,527 people.^{46,48} The most commonly used age groups were people aged ≥ 65 years (eight studies) and aged ≥ 70 years (seven studies). One study only showed the mean age of the sample.⁸ The highest mean age was observed among Dutch people residing at assisted living facilities (84.8 years).³⁹

Using the original TFI cut-off point of 5,¹ prevalence figures concerning the general population of community-dwelling older people ranged from 12.4% to 47.1% in samples of Chinese and Dutch individuals.^{1,13}

Table 1 General Characteristics of the Studies Included

Authors	Year	Country	Setting	Design	n	Age in years, mean ± SD	Prevalence figure
Gobbens et al ¹	2010	The Netherlands	community-dwelling	cross-sectional	479	≥ 75 years, 80.3 ± 3.8	47.1%
Mietzthin et al ³⁶	2010	The Netherlands	community-dwelling	cross-sectional	532	≥ 70 years, 77.2 ± 5.5	40.2%
Gobbens et al ⁴²	2012	The Netherlands	community-dwelling	longitudinal	484	≥ 75 years, 80.3 ± 3.8	-
Daniels et al ³⁷	2012	The Netherlands	community-dwelling	longitudinal	430	≥ 70 years, 77.2 ± 5.5	40.2%
Gobbens et al ⁴⁰	2013	The Netherlands	community-dwelling	cross-sectional	1,031	≥ 65 years, 73.4 ± 5.8	27.7%
Theou et al ⁴⁶	2013	Eleven European countries, including the Netherlands	community-dwelling	longitudinal	27,527	≥ 50 years, 65.3 ± 10.5	29.2%***
Santiago et al ⁷	2013	Brazil	community-dwelling	cross-sectional	219	≥ 60 years, sample 1: 69.8 ± 7.8 and sample 2: 71.3 ± 8.0	35.6% (sample 1); 31.7% (sample 2)
Andreasen et al ⁸	2014	Denmark	community-dwelling****	cross-sectional****	13; 21	75.9 ± 6.9; 80.6 ± 6.5	-
Uchmanowicz et al ¹¹	2014	Poland	community-dwelling	cross-sectional	100	≥ 60 years, 68.2 ± 6.5	40%
Gobbens & Van Assen ⁴¹	2014	The Netherlands	community-dwelling	longitudinal	484	≥ 75 years, 80.3 ± 3.8	-
Mulasso et al ⁹	2015	Italy	community-dwelling	cross-sectional	267	≥ 65 years, 73.4 ± 6.0	44.6%
Gobbens et al ³⁹	2015	The Netherlands	assisted living facilities	cross-sectional	221	≥ 55 years, 84.8 ± 8.9	76.5%
Coelho et al ¹⁰	2015	Portugal	community-dwelling	cross-sectional	252	≥ 65 years, 79.2 ± 7.3	54.8*
Andreasen et al ⁴⁸	2015	Denmark	community-dwelling*****	cross-sectional*****	14	≥ 65 years, 80.6	-
Uchmanowicz et al ⁵⁰	2016	Poland	community-dwelling	cross-sectional	212	≥ 60 years, 70.6 ± 7.2	44.1%
Freitag et al ¹²	2016	Germany	community-dwelling	cross-sectional	210	≥ 64 years, sample 1: 76.9 ± 5.7 and sample 2: 72.5 ± 5.5	46.7% (sample 1); 32% (sample 2)
Dong et al ¹³	2017	China	community-dwelling	cross-sectional	917	≥ 60 years, 68.6 ± 6.6	12.4%
Renne & Gobbens ⁴⁵	2018	The Netherlands	community-dwelling	cross-sectional	241	≥ 70 years, 76.5 ± 5.1	32.8%
Santiago et al ⁵¹	2018	Brazil	community-dwelling*****	longitudinal	963	≥ 60 years, 70.5 ± 8.2	44.2%
Vrotsou et al ¹⁴	2018	Spain	community-dwelling	cross-sectional	856	≥ 70 years, 78.1 ± 4.9	-
Topcu et al ¹⁵	2019	Turkey	geriatrics outpatient clinic	cross-sectional	198	≥ 70 years, 77.7 ± 5.5	63.6%
Op Het Veld et al ⁴⁴	2019	The Netherlands	community-dwelling	longitudinal	2,420	≥ 65 years, 76.3 ± 6.6	-
Op Het Veld et al ⁴³	2019	The Netherlands	community-dwelling	longitudinal	2,420	≥ 65 years, 76.3 ± 6.6	64.8%*
Alqahatani et al ⁵²	2020	Saudi Arabia	community-dwelling*****	cross-sectional	84	≥ 65 years, 72.0 ± 4.7	28.0%
Gobbens et al ³⁸	2020	The Netherlands	community-dwelling	longitudinal	180	≥ 70 years, 76.3 ± 5.1	29.4%
Gobbens & Andreasen ⁴⁹	2020	Denmark	acutely admitted patients	longitudinal	1,328	≥ 65 years, 76.9 ± 7.5	53.1%
Zhang et al ¹⁷	2020	The Netherlands, Spain, Greece, Croatia, United Kingdom	community-dwelling	cross-sectional	2,250	≥ 70 years, 79.7 ± 5.7	-

Notes: *Cut-off for frailty was 6; **This study was conducted among pre-frail and frail individuals; ***This study did not use the original TFI; ****Cross-sectional qualitative research; *****Acutely admitted to hospital; *****Acutely admitted discharged to their own home; *****Users of primary health care services; *****in senior-living facilities, visiting outpatient clinic.

Abbreviation: SD, standard deviation.

Specifically, prevalence was higher among Turkish people admitted to a geriatrics outpatient clinic (63.6%),¹⁵ residents of assisted living facilities (76.5%)³⁹ and among physical pre-frail and frail community-dwelling older people (64.8%).⁴³ In the latter group, the frailty status of the participants was first assessed by the Phenotype of Frailty.¹⁹ Among community-dwelling older people, the prevalence of frailty was highest in a sample of Portuguese people (54.8%), while the cut-off point was 6.¹⁰ It should be noted that seven studies did not present a prevalence figure of frailty.^{8,14,41,42,44,47,48} A Dutch sample including 479/484 participants and a Dutch sample consisting of 2420 participants were used in three^{1,41,42} and two studies,^{43,44} respectively. Moreover, two other Dutch studies partially used the same sample.^{36,37} Additional details are displayed in Table 1.

Reliability of the TFI

In total, 16 studies report the reliability of the TFI. The four types of reliability observed were internal consistency, test-retest, inter-rater and parallel forms reliability. Fifteen studies determined the internal consistency reliability and one study failed to do so (see Table 2).⁴⁶ The Cronbach's alpha for the TFI total was 0.66 (lowest)⁹ to 0.80 (highest),⁴⁵ whereas the KR-20, calculated in three studies was 0.69, 0.70 and 0.78.^{10,14,52} Eight studies also present the Cronbach's alpha for physical, psychological and social frailty, ranging from 0.57 to 0.79,^{7,9} 0.37 to 0.63,^{1,50} and 0.25 to 0.59,^{13,50} respectively. The lowest and highest values of the KR-20, with regard to physical, psychological, and social frailty, were 0.64¹⁴ and 0.75,¹⁰ 0.48¹⁰ and 0.58,¹⁴ and 0.22¹⁴ and 0.49,¹⁰ respectively. Six studies examined the internal consistency reliability of the TFI using corrected item-total correlations.^{9,12,15,36,50,52}

Test-retest reliability was observed in nine studies, using Pearson correlations,^{1,7,10} Kappa,^{7,10,14,50} simple agreement,^{7,10,14} and ICC^{12,13,15,52} (see Table 3). Pearson correlations, with regard to frailty total were 0.88,⁷ 0.90,¹ and 0.91,¹⁰ using a period less than 3 weeks. The Pearson correlation coefficient was 0.79 for a 1-year period.¹ The correlation coefficients, with respect to the frailty domains, are detailed in Table 3. Using Kappa, the level of agreement varied greatly at item level.^{7,10,14,50} Obviously, the level of agreement was higher when the simple agreement technique was used^{7,10,14} (displayed in Table 3). For frailty total, the ICC ranged from 0.86 to 0.99^{15,52} in two studies involving a follow-up period of 1 week.

Table 3 also shows the inter-rater and parallel forms reliability of the TFI. Inter-rater reliability was identified in only one study in which, on the same day, two observers came to almost perfect agreement (ICC = 0.99).¹⁵ Finally, in two studies, parallel forms reliability was determined.^{13,46} One of these studies examined the agreement between the TFI (frailty total) and other validated frailty instruments. The highest agreement existed with the Frailty Index (FI) and the Comprehensive Geriatric Assessment and less agreement was found with the Edmonton Frail Scale and the Frail scale.⁴⁶ The other study used Kappa to establish the level of agreement between TFI items and alternative measures resulting in low and high levels of agreement (0.12 for hearing problems and 1.00 for living alone).¹³

Validity of the TFI

Criterion Validity

Tables 4–6 show an overview of the validity of the TFI, which includes 24 of the 27 studies. Criterion validity was the most frequently presented type of validity, with concurrent and predictive characteristics reported in 10 and 9 studies, respectively (see Table 4). Concurrent validity was determined using different techniques: correlations, AUC and regression analyses. Frequently occurring adverse outcomes of interest were lower quality of life,^{1,10,39,40,45} disability,^{1,9,10,13,39,47} and an increase in health-care utilization.^{1,9,10,13,39} Using different instruments for assessing quality of life (WHOQOL-BREF,⁵³ WHOQOL-OLD⁵⁴ and EUROHIS-QOL)⁵⁵, four included studies demonstrated that higher scores on the TFI were correlated with lower quality of life.^{1,10,40,45} Regarding disability and referring to limitations in performing activities of daily living (ADL) and/or instrumental activities of daily living (IADL), the AUCs were excellent in three studies,^{1,9,47} and acceptable for ADL and poor for IADL in two studies.^{10,13} Many different indicators of health-care utilization were used, eg, visits to a general practitioner, hospitalization and receiving nursing care. In most studies, the findings were poor^{9,10,13}; however, Gobbens et al observed an excellent AUC for reporting personal care, and acceptable AUCs for reporting nursing and informal care.¹ Three studies determined the discriminating ability of the TFI to identify frailty with other validated frailty measures using the AUC: the Groningen Frailty Indicator (GFI),¹⁰ the phenotype of frailty,^{10,13} Survey of Health, Ageing and Retirement in Europe-Frailty Instrument (SHARE-FI),⁴⁷ and the Frailty Index (FI).¹³ The AUCs for GFI, SHARE-FI and the FI were excellent;

however, the AUCs for the phenotype of frailty were not unequivocal. Two studies examined the correlations between the TFI and other frailty measures: the GFI,³⁶ Sherbrooke Postal Questionnaire (SPQ),³⁶ and the Phenotype of Frailty.¹⁴ The correlations between the TFI and GFI, SPQ and phenotype of frailty were 0.76, 0.42, and 0.49, respectively.^{14,36}

The predictive validity was established using regression analyses and AUC. The prediction period ranged from 6 months⁴⁹ to 5 years.⁴⁶ Most studies only used a period of 1 or 2 years.^{37,38,42–44,51} In particular, adverse outcomes of interest were disability,^{37,38,42–44,51} mortality,^{37,43,46,49,51} increased health-care utilization^{37,38,42,43,51} and lower quality of life.^{41,42}

Disability was predicted by the TFI; however, Gobbens et al^{38,42} indicated that the predictive value was excellent, while Op Het Veld et al concluded that it was poor,⁴³ presenting a positive predictive value of 42.6% and a negative predictive value of 75.2%.⁴⁴ In four studies, the findings concerning mortality were at least acceptable; only Op Het Veld et al qualified the predictive value of the TFI as poor.⁴³ As with the determination of concurrent validity, many different indicators of health-care utilization were used as outcome variables, resulting in findings that were not unanimous. For instance, the TFI predicted hospitalization in a sample of 430 Dutch people ≥ 70 years (OR = 2.59, 95% CI = 1.36–4.90),³⁷ in comparison to

Table 2 Internal Consistency Reliability of the TFI

Authors	Internal Consistency Reliability
Gobbens et al ¹	Cronbach's alpha: total 0.73, physical 0.70, psychological 0.63, social 0.34
Metzelthin et al ³⁶	Cronbach's alpha: total 0.79 Corrected item-total correlations: correlations ranged from 0.18 to 0.58 (mean 0.39)
Gobbens et al ⁴⁰	Cronbach's alpha: total 0.71, physical 0.67, psychological 0.54, social 0.51
Santiago et al ⁷	Cronbach's alpha: total 0.78, physical 0.79, psychological 0.53, social 0.38
Uchmanowicz et al ¹¹	Cronbach's alpha: total 0.72 Cronbach's alpha reliability coefficients after the removal of an item ranged from 0.68 (coping) to 0.73 (anxiety)
Mulasso et al ⁹	Cronbach's alpha: total 0.66, physical 0.57, psychological 0.51, social 0.36 Corrected item-total correlations for each item with the domains in general, accepted values, with some exceptions (unexplained weight loss, poor hearing)
Coelho et al ¹⁰	KR-20: total 0.78, physical 0.75, psychological 0.48, social 0.49
Uchmanowicz et al ⁵⁰	Cronbach's alpha: total 0.74, physical 0.72, psychological 0.37, social 0.59 Corrected item-total correlations: ranged from 0.12 to 0.55
Freitag et al ¹²	Cronbach's alpha: total 0.67, physical 0.66, psychological 0.43, social 0.36 Cronbach's alpha reliability coefficients after the removal of an item ranged from 0.6 (physical tiredness)–0.69 (coping) Corrected item-total correlations ranged from 0.12 (memory problems) to 0.58 (physical tiredness)
Dong et al ¹³	Cronbach's alpha: total 0.71, physical 0.71, psychological 0.51, social 0.25
Renne & Gobbens ⁴⁵	Cronbach's alpha: total 0.80, physical 0.74, psychological 0.61, social 0.51
Vrotsou et al ¹⁴	KR-20: total 0.69, physical 0.64, psychological 0.58, social 0.22
Topcu et al ¹⁵	Cronbach's alpha: total 0.68 Cronbach's alpha reliability coefficients after the removal of an item ranged from 0.62 (physical tiredness)–0.69 (lack of social relations) Corrected item-total correlations: ranged from –0.05 (living alone) to 0.57 (physical tiredness)
Alqahtani et al ⁵²	KR-20: total 0.70, physical 0.68, psychological 0.57, social 0.42; the KR-20 after the removal of an item ranged from 0.66 (coping)–0.72 (poor hearing, physical tiredness) Corrected item-total correlations ranged from 0.10 (unexplained weight loss) to 0.47 (coping)
Zhang et al ⁴⁷	Cronbach's alpha: varied among five countries involved: total 0.70 (Spain)–0.75 (Croatia), physical 0.60 (Spain)–0.73 (The Netherlands), psychological 0.38 (UK)–0.55 (Greece, Croatia), social 0.22 (Greece)–0.43 (The Netherlands)

Abbreviation: KR-20, Kuder–Richardson formula.

a poor AUC in a sample of 2420 Dutch people ≥ 65 years.⁴³ Both studies that aimed to assess the predictive value of the TFI for quality of life provided evidence that the TFI predicts lower quality of life using a follow-up period of 1, 2 and 4 years.^{41,42}

Construct Validity

The reviewed studies frequently determined the criterion validity of the TFI, as well as the construct validity. Ten studies determined the construct validity of the TFI,^{1,7,9,10,12–14,45,47,52} with nine of these studies found to address the issue of convergent and divergent validity, with exception of the study by Renne and Gobbens (see Table 5).⁴⁵ Most of the studies found the expected correlations between total frailty, domains, items, and alternative

measures, while two studies observed similar correlations between the psychological and social domains of the TFI and alternative psychological.^{10,13} Moreover, the Spanish study demonstrated a stronger correlation between social frailty and IADL, assessed with the Lawton scale, than between social frailty and physical frailty, which was expected.¹⁴ Finally, in the Brazilian study, the item coping was not correlated as expected.⁷

Content and Face Validity

Table 6 presents the content, face, structural and known-groups validity of the TFI. Three studies determined the content validity of the TFI,^{1,46,48} showing that the TFI contains the majority of important frailty items. Based on interviews with community-dwelling older people, Andreasen

Table 3 Test–Retest Reliability, Inter-Rater Reliability, and Parallel Forms Reliability of the TFI

Authors	Test-Retest Reliability
Gobbens et al ¹	Two weeks (Pearson correlation coefficient): total 0.90, physical 0.87, psychological 0.77, social 0.86 One-year period (Pearson correlation coefficient): total 0.79, physical 0.78, psychological 0.67, social 0.76
Santiago et al ⁷	7–10 days (Pearson correlation coefficient): total 0.88, physical 0.88, psychological 0.67, social 0.89. 7–10 days (simple agreement); items ranged from 0.63 (memory) to 1.00 (live alone, lack of support) 7–10 days (Kappa coefficients): five items had nearly perfect agreement (0.81–1.00), four items had substantial agreement (0.61–0.80) agreement, two items had moderate agreement (0.41–0.60), four items had fair agreement (0.21–0.40)
Coelho et al ¹⁰	12–16 days (Pearson correlation coefficient): total 0.91, physical 0.87, psychological 0.75, social 0.87. 12–16 days (simple agreement): items ranged from 0.78 (depression, anxiety) to 0.97 (living alone) 12–16 days (Kappa coefficients): ranged from 0.52 to 0.95
Uchmanowicz et al ⁵⁰	10–14 days (Kappa coefficient): a high level of agreement with regard to items was demonstrated with coefficients ranging from 0.96 to 1.00
Freitag et al ¹²	20 weeks (ICC): total 0.87, physical 0.85, psychological 0.75, social 0.84
Dong et al ¹³	10–25 days (ICC): total 0.88, physical 0.80, psychological 0.65, social 0.81
Vrotsou et al ¹⁴	7–14 days (simple agreement): item ranged from 0.77 to 0.99, except anxious (0.66) 7–14 days (Kappa coefficient): 0.98 (living alone), 0.23 to 0.34 (physical tiredness, anxious, coping, support) 0.46 to 0.57 (all other items, except unintentional weight loss)
Topcu et al ¹⁵	One week (ICC): 0.99
Alqahtani et al ⁵²	One week (ICC): 0.86
Inter-rater reliability	
Topcu et al ¹⁵	Two observers on the same day (ICC): 0.99
Parallel forms reliability	
Theou et al ⁴⁶	Kappa coefficients: TFI and Frailty Index 0.52, Frailty Index based on Comprehensive Geriatric Assessment 0.52, Clinical Frailty Scale 0.38, Frailty Phenotype 0.37, Edmonton Frail Scale 0.27, FRAIL scale 0.27, Groningen Frailty Indicator (GFI) 0.50
Dong et al ¹³	Kappa coefficients: ranged for TFI items and alternative measures from 0.12 (hearing problems)–1.00 (living alone)

Abbreviation: ICC, intraclass correlation coefficient.

Table 4 Criterion Validity of the TFI

Authors	Criterion Validity
Gobbens et al ¹	Concurrent using correlations: large for total frailty and quality of life domains physical, psychological, environmental, and medium to large for quality of life domain social assessed with the WHOQOL-BREF Concurrent using AUC: excellent for disability and reporting personal care, acceptable for reporting nursing and informal care, poor for reporting visits general practitioner and hospitalization
Metzelthin et al ³⁶	Concurrent using correlations: correlation between TFI and Groningen Frailty Indicator (GFI) was 0.76; the correlation between TFI and Sherbrooke Postal Questionnaire (SPQ) was 0.42
Gobbens et al ⁴²	Predictive, one and two years later, using multiple regression analyses: an increase in predictive accuracy of most adverse outcomes (disability, indicators of health care utilization, and quality of life) Predictive, one and two years later, using AUC: excellent for disability and reporting personal care, acceptable for reporting nursing, informal care, and facilities in residential care, poor for contacts with health care professionals and hospitalization, not significant for visits to a general practitioner
Gobbens et al ⁴⁰	Concurrent using sequential regression analyses: all components of the TFI together explained the scores on quality of life domains physical health, psychological, social relations, environmental assessed with the WHOQOL-BREF
Daniels et al ³⁷	Predictive, one year later, using OR unadjusted: disability 3.96, 95% CI = 2.48–6.30, mortality 3.08, 95% CI = 1.04–9.13, hospitalization 2.59, 95% CI = 1.36–4.90 Predictive, one year later, using AUC: poor for disability, mortality, and hospitalization
Theou et al ⁴⁶	Predictive, two and five years later, using AUC: acceptable for mortality
Gobbens and Van Assen ⁴¹	Predictive, two and four years later, using sequential regression analyses: the items physical unhealthy, difficulty in walking, difficulty in maintaining balance, physical tiredness, feeling down, and lack of social support predicted quality of life scores assessed with the WHOQOL-BREF
Mulasso et al ⁹	Concurrent using AUC: excellent for disability, poor for falls and visits to general practitioner
Gobbens et al ³⁹	Concurrent using regression analyses: all three domains (physical, psychological, social) together had an effect on disability, quality of life (physical health, psychological, social relationships, environmental), visits to a general practitioner, and falls; no effects were observed with contacts with health care professionals, hospitalization, receiving personal care, receiving nursing care, receiving informal care, and facilities in nursing home/rehabilitation center
Coelho et al ¹⁰	Concurrent using multiple regression analyses: the TFI domains predicted 38.7% and 42.1% of quality of life variance, assessed with EUROHIS-QOL, and WHOQOL-OLD, respectively Concurrent using AUC: acceptable for disability in ADL, poor for disability in IADL, and health care utilization Concurrent using AUC: discriminating ability was excellent regarding identifying frailty by the Groningen Frailty Indicator (GFI) (0.86, 95% CI = 0.85–0.93) and acceptable for frailty assessed with the Frailty Phenotype by Fried et al (0.75, 95% CI = 0.68–0.81)
Dong et al ¹³	Concurrent using AUC: excellent for depression; acceptable for disability in ADL, and low social support; poor for disability in IADL, and for health care utilization (hospitalization, emergency use) Concurrent using AUC: discriminating ability was excellent regarding identifying frailty by the Frailty Phenotype by Fried et al (0.87, 95% CI = 0.87–0.93) and the Frailty Index (0.86, 95% CI = 0.82–0.91)
Renne and Gobbens ⁴⁵	Concurrent using sequential multiple linear regression analyses: all fifteen items together explained 36.5% of the variance of the score of quality of life
Santiago et al ⁵¹	Predictive, 1 year later, using sequential logistic regression analyses: total frailty predicted mortality, adjusted for sex and age (HR = 2.72, 95% CI = 1.01–7.31); after controlling for sociodemographic variables the frailty domains (physical, psychological, social) improved the prediction of hospitalization (OR = 1.83, 95% CI = 1.10–3.06), falls (OR = 2.08, 95% CI = 1.21–3.58), disability in ADL (OR = 3.03, 95% CI = 1.45–6.29), disability in IADL (OR = 1.51, 95% CI = 1.05–2.17)
Vrotsou et al ¹⁴	Concurrent using correlations: the correlation between total and the Frailty Phenotype by Fried et al was 0.49
Op Het Veld et al ⁴⁴	Predictive, 2 years later: positive predictive value 42.6% and negative predictive value 75.2% for disability in IADL

(Continued)

Table 4 (Continued).

Authors	Criterion Validity
Op Het Veld et al ⁴³	Predictive, 2 years later, using AUC: poor for mortality, hospitalization, and disability in IADL
Gobbens et al ³⁸	Predictive, 1 year later, using linear and logistic regression analyses: the three frailty domains together predicted disability, visits general practitioner, contacts with health care professionals, receiving nursing; no effects were found on hospitalization, receiving personal care, falls (after controlling for sociodemographic characteristics and multimorbidity) Predictive, 1 year later, using AUC: excellent for total frailty with respect to disability and receiving personal care; poor for receiving nursing, falls, and hospitalization
Gobbens and Andreasen ⁴⁹	Predictive, 6 months later, using sequential logistic regression analyses: physical and social frailty predicted readmission and mortality; psychological frailty predicted only readmission
Zhang et al ⁴⁷	Concurrent using AUC: all AUC were excellent for SHARE-FI, and disability; all AUC were acceptable for limited function, poor mental health, and feeling lonely

Abbreviations: AUC, area under the curve; CI, confidence interval; ADL, activities of daily living; IADL, instrumental activities of daily living; OR, odds ratio; SHARE-FI, SHARE Frailty Instrument.

et al argue that items referring sleep quality, pain, spirituality, and meaningful activities should be included in the TFI.⁴⁸

Face-validity was established in only two studies.^{1,8} In the first study, the TFI was checked by participants at geriatric meetings and in the second study, conducted in Denmark, after translating the TFI, a pretest was performed by cognitive interviewing concluding that the TFI could be further tested in practice.

Structural Validity and Known-Groups Validity

Both structural validity and known-groups validity were only determined by Vrotsou et al (see Table 6).¹⁴ Fit indexes of a second-order model of three factors (frailty domains) were acceptable and the TFI differentiated well between frail and non-frail people, as defined by the Gérontopôle Frailty Screening Tool (GFST)⁵⁶ and the Short Physical Performance Battery (SPPB).⁵⁷

Table 5 Construct Validity of the TFI

Authors	Construct Validity
Gobbens et al ¹	Convergent and divergent validity using correlations: the 15 single components and three domains correlated as expected with validated measures
Santiago et al ⁷	Convergent and divergent validity using correlations: the correlations between the items and their corresponding measures were as expected, except the item 'coping'
Mulasso et al ⁹	Convergent and divergent validity using correlations: all items correlated with single corresponding frailty measures
Coelho et al ¹⁰	Convergent and divergent validity using correlations: physical and social domains correlated as expected with alternative measures; psychological measures showed similar correlations with the psychological and physical domains of the TFI
Freitag et al ¹²	Convergent and divergent validity using correlations: total frailty was correlated with all alternative measures of frailty. In addition, the domains correlated good with corresponding alternative measures
Dong et al ¹³	Convergent and divergent validity using correlations: the three domains correlated with alternative measures; however, psychological measures had similar correlations with the psychological and physical domain
Renne & Gobbens ⁴⁵	Construct validity using correlations: total and all three domains correlated with all six quality of life domains assessed with the WHOQOL-OLD
Vrotsou et al ¹⁴	Convergent and divergent validity using correlations: the three frailty domains correlated as expected, except social frailty that had a stronger correlation with the Lawton scale than physical frailty
Alqahtani et al ⁵²	Convergent and divergent validity: correlations as expected between total and six frailty-related measures
Zhang et al ⁴⁷	Convergent and divergent validity using correlations: validity of physical, psychological and social frailty was supported by all the alternative measures in all five countries

Table 6 Content Validity, Face Validity, Structural Validity, and Known-Groups Validity of the TFI

Authors	Content Validity
Gobbens et al ¹	Determined by representatives of professional disciplines and people aged ≥ 75 years
Theou et al ⁴⁶	The TFI records items referring to limitations in self-rated health, nutrition, mobility, energy, cognition, mood
Andreasen et al ⁴⁸	Determined by interviewing frail community-dwelling older people: the majority of important frailty items were covered by the TFI; pain, sleep quality, meaningful activities and spirituality are not present in the TFI
Face Validity	
Gobbens et al ¹	Checked by participants at geriatric meetings
Andreasen et al ⁸	A pretest was performed by cognitive interviewing. The TFI was translated and adapted in such a manner that it can be implemented and further tested in clinical practice
Structural Validity	
Vrotsou et al ¹⁴	Confirmatory factor analysis (CFA) showed that fit indexes of a second-order model of three factors (frailty domains) were acceptable
Known-Groups Validity	
Vrotsou et al ¹⁴	Total and physical frailty scores differentiated well between frail and non-frail people defined by the GFST and the SPPB

Abbreviations: GFST, Gérontopôle Frailty Screening Tool; SPPB, Short Physical Performance Battery.

Discussion

The TFI is a questionnaire that is increasingly used to determine frailty in older people. Since the introduction of the TFI in 2010, many studies on its psychometric properties have been conducted. In this study, we aimed to present a review of findings regarding the reliability and validity of the TFI. The present study is the first to assess these psychometric properties of the TFI.

The literature search performed through May 30, 2020, showed that 27 studies reported on the psychometric properties of the TFI, as related to reliability, validity or both. Most of the studies ($n = 25$) were focused on community-dwelling older people. Since the TFI was developed in the Netherlands,¹ it is not surprising that 13 of the included studies were conducted in that country and there appeared to be large differences in the prevalence of frailty. The lowest and highest prevalence figures were 12.4%¹³ and 76.5%,³⁹ respectively. Our review shows higher prevalence figures of frailty are closely related to greater age. This finding is supported by a systematic review containing cross-sectional data from community-based cohorts.⁵⁸ In addition, we found higher prevalence figures among people residing in settings other than the community, eg, acutely admitted patients,⁴⁹ residents of assisted living facilities,³⁹ and people admitted to a geriatrics outpatient clinic.¹⁵ Zhang et al emphasized that the mean score on the TFI is influenced by the country of residence of the

participants.⁴⁷ For example, the mean score for people with an average age of 75.3 years from Greece was 5.80, while this score was 4.25 in a sample of Dutch people with an average age of 81.5 years.⁴⁷ The country of residence of the participants may also be the possible explanation for the high prevalence among older Portuguese people (54.8%);¹⁰ this prevalence is exceptionally high if we consider that the TFI cut-off point of 6¹⁰ was taken, not the established cut-off point of 5.¹

Our study shows that the reliability of the TFI has been comprehensively assessed; 15 and nine studies examined the internal consistency and test-retest reliability, respectively. In many cases, the reliability of the TFI, reflected by Cronbach's alpha, was >0.70 , which indicates satisfactory reliability.²⁴ The internal consistency of the individual domains of the TFI was worse, particularly for the psychological and social domains, which can be explained by the fact that these domains only contain four and three items. Test-retest reliability, reflected by correlations (>0.60) and ICC (>0.75), was good.^{25,27} Moreover, kappa coefficients showed a substantial or nearly perfect level of agreement concerning many individual TFI items. The two other types of reliability, inter-rater and parallel forms, have only been examined to a limited extent.^{13,15,46}

The validity of the TFI has been established in 24 studies. In many studies focusing on criterion validity (concurrent and predictive), adverse outcomes of interest

were disability, increased health-care utilization, lower quality of life and mortality. Regarding disability, studies predominantly show results that are excellent,^{1,9,38,42,47} with AUCs >0.80.²⁹ The findings pertaining to an increase in health-care utilization present a less unambiguous picture and seem to depend strongly on the indicator used (eg, personal care, informal care, visits to a general practitioner or hospitalization). The TFI, however, clearly has a poor association with visits to a general practitioner^{1,9,42} and hospitalization.^{1,13,37–39,42,43} In all six studies using quality of life as the outcome, the TFI showed good associations with lower quality of life, independent of the quality of life instrument that has been used.^{1,10,40–42,45} Moreover, the TFI predicted mortality, with only Op Het Veld et al⁴³ qualifying the predicted value for this outcome as low. It should be noted that follow-up periods were short, consisting of 1 or 2 years, except the study by Theou et al⁴⁶ that included a follow-up of 5 years. Unfortunately, Theou et al did not use the original TFI for assessing frailty.⁴⁶ Therefore, we recommend determining the prediction value of the original TFI for mortality using a follow-up period >5 years. Finally, concerning criterion validity, the TFI showed discriminating ability in regards to GFI,¹⁰ SHARE-FI,⁴⁷ and FI,¹³ reflected by excellent AUCs (>0.80).

Construct validity of the TFI, distinguishing convergent and divergent validity, was established in nine studies. In many cases, the domains and components of the TFI correlated as expected with alternative measures providing evidence for good construct validity. Content and face validity were only determined in three^{1,46,48} and two studies,^{1,8} respectively. Two of these studies were qualitative in nature,^{8,48} which seems ideally suited to establish these types of validity. In our opinion, to validating an instrument like the TFI, the involvement of preferably older people, is necessary, whom, from their perspective, can indicate what needs to be questioned in the context of frailty. Only one study assessed the structural and known groups validity.¹⁴ These findings were satisfactory.

Some limitations of our study should be noted. First, different instruments and questions have been used to assess disability, eg, Groningen Activity Restriction Scale (GARS)⁵⁹ and Katz scale,⁶⁰ and indicators of health-care utilization. Differences concerning variables controlled in regression analyses also exist, which could have influenced the findings. Secondly, many of the included studies aimed to determine the reliability and validity of the TFI were conducted in samples of community-dwelling older people.

More studies are needed in order to establish good reliability and validity of the TFI in other samples of older people (eg, residents of assisted living facilities, nursing homes, mental health institutions and hospitalized patients). Thirdly, many data on reliability and validity were available and it was impossible to present all these data. A selection based on the most relevant data for the present study was therefore utilized and the individual studies were referred to for more detailed information. Finally, the majority of studies have been performed in Europe and especially in the Netherlands, with less in Brazil,^{7,51} Saudi Arabia,⁵² and China.¹³ The determination of the psychometric properties of the TFI in the other continents of the world, such as Africa, North America and Australia, is recommended.

The TFI is specifically designed for screening of multidimensional frailty among community-dwelling older people; our review provided much evidence that the TFI is ideally suited to that target group. Well-known determinants of frailty, assessed with the TFI, are higher age, being a woman, and a low socio-economic status, expressed by low-educational and low-income level.^{36,61,62} To prevent frailty, it is recommended to start screening among people that meet those criteria. Based on this screening, primary health-care professionals (eg, general practitioners, nurses, physiotherapists, occupational therapists) can determine, preferably in a multidisciplinary consultation, whether these people need a comprehensive geriatric assessment (CGA) and a care needs assessment. In addition, the findings of the screening of frailty using the TFI can provide primary health-care professionals a first direction to the interventions that should be conducted next. Evidence of beneficial effects of multidomain interventions compared to unidomain interventions on frailty status or score is limited but increasing.⁶³

In conclusion, our literature search revealed 27 studies examining the reliability and/or validity of the TFI. Many studies showed that the internal consistency and test–retest reliability are noteworthy, as well as the criterion and construct validities. In contrast, the association of the TFI with some indicators of health-care utilization can be indicated as poor (eg, visits to a general practitioner or hospitalization). Knowing that population aging is occurring all over the world, the availability of the TFI is critical. In addition to the qualification of its psychometric properties as good, it is well known that the TFI is a user-friendly instrument for assessing frailty. The findings of this assessment can support health-care professionals in selecting interventions to reduce frailty and delay its adverse outcomes, such as disability and lower quality of life.

Disclosure

The authors report no conflicts of interest in this work.

References

- Gobbens RJ, van Assen MA, Luijkx KG, Wijnen-Sponselee MT, Schols JM. The Tilburg Frailty Indicator: Psychometric properties. *J Am Med Dir Assoc.* 2010;11(5):344–355.
- Gobbens RJ, Luijkx KG, Wijnen-Sponselee MT, Schols JM. Toward a conceptual definition of frail community dwelling older people. *Nurs Outlook.* 2010;58(2):76–86. doi:10.1016/j.outlook.2009.09.005
- World Health Organization. World report on ageing and health. Geneva (Switzerland): World Health Organization; 2015. Available from: <https://www.who.int/ageing/events/world-report-2015-launch/en/>.
- Markle-Reid M, Browne G. Conceptualizations of frailty in relation to older adults. *J Adv Nurs.* 2003;44(1):58–68. doi:10.1046/j.1365-2648.2003.02767.x
- Gilardi F, Capanna A, Ferraro M, et al. Frailty screening and assessment tools: a review of characteristics and use in Public Health. *Ann Ig.* 2018;30(2):128–139. doi:10.7416/ai.2018.2204
- Gobbens RJ, Luijkx KG, Wijnen-Sponselee MT, Schols JM. In search of an integral conceptual definition of frailty: opinions of experts. *J Am Med Dir Assoc.* 2010;11(5):338–343. doi:10.1016/j.jamda.2009.09.015
- Santiago LM, Luz LL, Mattos IE, Gobbens RJ, van Assen MA. Psychometric properties of the Brazilian version of the Tilburg frailty indicator (TFI). *Arch Gerontol Geriatr.* 2013;57(1):39–45. doi:10.1016/j.archger.2013.03.001
- Andreasen J, Sorensen EE, Gobbens RJ, Lund H, Aadahl M. Danish version of the Tilburg frailty indicator - translation, cross-cultural adaptation and validity pretest by cognitive interviewing. *Arch Gerontol Geriatr.* 2014;59(1):32–38. doi:10.1016/j.archger.2014.02.007
- Mulasso A, Roppolo M, Gobbens RJ, Rabaglietti E. The Italian version of the Tilburg Frailty Indicator: analysis of psychometric properties. *Res Aging.* 2016;38(8):842–863. doi:10.1177/0164027515606192
- Coelho T, Santos R, Paul C, Gobbens RJ, Fernandes L. Portuguese version of the Tilburg Frailty Indicator: transcultural adaptation and psychometric validation. *Geriatr Gerontol Int.* 2015;15(8):951–960. doi:10.1111/ggi.12373
- Uchmanowicz I, Jankowska-Polanska B, Loboz-Rudnicka M, Manulik S, Loboz-Grudzien K, Gobbens RJ. Cross-cultural adaptation and reliability testing of the Tilburg Frailty Indicator for optimizing care of Polish patients with frailty syndrome. *Clin Interv Aging.* 2014;9:997–1001. doi:10.2147/CIA.S64853
- Freitag S, Schmidt S, Gobbens RJ. Tilburg frailty indicator. German translation and psychometric testing. *Z Gerontol Geriatr.* 2016;49(2):86–93. doi:10.1007/s00391-015-0889-9
- Dong L, Liu N, Tian X, et al. Reliability and validity of the Tilburg Frailty Indicator (TFI) among Chinese community-dwelling older people. *Arch Gerontol Geriatr.* 2017;73:21–28. doi:10.1016/j.archger.2017.07.001
- Vrotsou K, Machon M, Rivas-Ruiz F, et al. Psychometric properties of the Tilburg Frailty Indicator in older Spanish people. *Arch Gerontol Geriatr.* 2018;78:203–212. doi:10.1016/j.archger.2018.05.024
- Topcu Y, Tufan F, Kilic C. Turkish version of the Tilburg Frailty Indicator. *Clin Interv Aging.* 2019;14:615–620. doi:10.2147/CIA.S197512
- Pialoux T, Goyard J, Lesourd B. Screening tools for frailty in primary health care: a systematic review. *Geriatr Gerontol Int.* 2012;12(2):189–197. doi:10.1111/j.1447-0594.2011.00797.x
- 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(1):55. doi:10.1186/s12877-016-0225-2
- Romero-Ortuno R, Walsh CD, Lawlor BA, Kenny RA. A frailty instrument for primary care: findings from the Survey of Health, Ageing and Retirement in Europe (SHARE). *BMC Geriatr.* 2010;10:57. doi:10.1186/1471-2318-10-57
- Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. *J Gerontol a Biol Sci Med Sci.* 2001;56(3):M146–56. doi:10.1093/gerona/56.3.M146
- Saliba D, Elliott M, Rubenstein LZ, et al. The Vulnerable Elders Survey: a tool for identifying vulnerable older people in the community. *J Am Geriatr Soc.* 2001;49(12):1691–1699. doi:10.1046/j.1532-5415.2001.49281.x
- Mitnitski AB, Mogilner AJ, Rockwood K. Accumulation of deficits as a proxy measure of aging. *ScientificWorldJournal.* 2001;1:323–336. doi:10.1100/tsw.2001.58
- De Witte N, Gobbens R, De Donder L, et al. Validation of the comprehensive frailty assessment instrument against the Tilburg Frailty Indicator. *Eur Geriatr Med.* 2013;4(4):248–254. doi:10.1016/j.eurger.2013.03.001
- McDowell I. *Measuring Health: A Guide to Rating Scales and Questionnaires.* 3rd ed. New York: Oxford University Press; 2006.
- Taber ST. The use of Cronbach's alpha when developing and reporting research instruments in science education. *Res Sci Educ.* 2016;48(1):1–24.
- Callegari-Jacques SM. *Biostatistics: Principles and Applications.* Porto Alegre: Artmed; 2003.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159–174. doi:10.2307/2529310
- Koo TK, Li MYA. Guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* 2016;15(2):155–163. doi:10.1016/j.jcm.2016.02.012
- Kimberlin CL, Winterstein AG. Validity and reliability of measurement instruments used in research. *Am J Health Syst Pharm.* 2008;65(23):2276–2284. doi:10.2146/ajhp070364
- Hosmer DW, Lemeshow S. *Applied Logistic Regression.* 2nd ed. New York: John Wiley and Sons; 2000.
- Strauss ME, Smith GT. Construct validity: advances in theory and methodology. *Ann Rev Clin Psychol.* 2009;5:1–25. doi:10.1146/annurev.clinpsy.032408.153639
- Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull.* 1959;56(2):81–105. doi:10.1037/h0046016
- Polit DF. Assessing measurement in health: beyond reliability and validity. *Int J Nurs Stud.* 2015;52(11):1746–1753. doi:10.1016/j.ijnurstu.2015.07.002
- Burns N, Grove SK. *The Practice of Nursing Research: Conduct, Critique & Utilization.* 2nd ed. Philadelphia: WB Saunders Company; 1993.
- Bolarinwa OA. Principles and methods of validity and reliability testing of questionnaires used in social and health science researches. *Niger Postgrad Med J.* 2015;22(4):195–201. doi:10.4103/1117-1936.173959
- Souza AC, Alexandre NMC, Guirardello EB. Psychometric properties in instruments evaluation of reliability and validity. *Epidemiol Serv Saude.* 2017;26(3):649–659. doi:10.5123/S1679-49742017000300022
- Metzelthin SF, Daniels R, van Rossum E, de Witte L, van den Heuvel WJ, Kempen GI. The psychometric properties of three self-report screening instruments for identifying frail older people in the community. *BMC Public Health.* 2010;10:176. doi:10.1186/1471-2458-10-176
- Daniels R, van Rossum E, Beurskens A, van den Heuvel W, de Witte L. The predictive validity of three self-report screening instruments for identifying frail older people in the community. *BMC Public Health.* 2012;12:69. doi:10.1186/1471-2458-12-69
- Gobbens RJ, Boersma P, Uchmanowicz I, Santiago LM. The Tilburg Frailty Indicator (TFI): new evidence for its validity. *Clin Interv Aging.* 2020;15:265–274. doi:10.2147/CIA.S243233

39. Gobbens RJ, Krans A, van Assen MA. Validation of an integral conceptual model of frailty in older residents of assisted living facilities. *Arch Gerontol Geriatr.* 2015;61(3):400–410. doi:10.1016/j.archger.2015.06.001
40. Gobbens RJ, Luijckx KG, van Assen MA. Explaining quality of life of older people in the Netherlands using a multidimensional assessment of frailty. *Qual Life Res.* 2013;22(8):2051–2061. doi:10.1007/s11136-012-0341-1
41. Gobbens RJ, van Assen MA. The prediction of quality of life by physical, psychological and social components of frailty in community-dwelling older people. *Qual Life Res.* 2014;23(8):2289–2300. doi:10.1007/s11136-014-0672-1
42. Gobbens RJ, van Assen MA, Luijckx KG, Schols JM. The predictive validity of the Tilburg Frailty Indicator: disability, health care utilization, and quality of life in a population at risk. *Gerontologist.* 2012;52(5):619–631. doi:10.1093/geront/gnr135
43. Op Het Veld LPM, Beurskens A, de Vet HCW, et al. The ability of four frailty screening instruments to predict mortality, hospitalization and dependency in (instrumental) activities of daily living. *Eur J Ageing.* 2019;16(3):387–394. doi:10.1007/s10433-019-00502-4
44. Op Het Veld LPM, van Rossum E, Kempen G, Beurskens A, Hajema KJ, de Vet HCW. Can the Combined Use of Two Screening Instruments Improve the Predictive Power of Dependency in (Instrumental) Activities of Daily Living, Mortality and Hospitalization in Old Age? *J Frailty Aging.* 2019;8(4):180–5.
45. Renne I, Gobbens RJJ. Effects of frailty and chronic diseases on quality of life in Dutch community-dwelling older adults: a cross-sectional study. *Clin Interv Aging.* 2018;13:325–334. doi:10.2147/CIA.S156116
46. Theou O, Brothers TD, Mitnitski A, Rockwood K. Operationalization of frailty using eight commonly used scales and comparison of their ability to predict all-cause mortality. *J Am Geriatr Soc.* 2013;61(9):1537–1551. doi:10.1111/jgs.12420
47. Zhang X, Tan SS, Bilajac L, et al. Reliability and Validity of the Tilburg Frailty Indicator in 5 European Countries. *J Am Med Dir Assoc.* 2020;21(6):772–9.e6. doi:10.1016/j.jamda.2020.03.019
48. Andreasen J, Lund H, Aadahl M, Gobbens RJ, Sorensen EE. Content validation of the Tilburg Frailty Indicator from the perspective of frail elderly. A qualitative explorative study. *Arch Gerontol Geriatr.* 2015;61(3):392–399. doi:10.1016/j.archger.2015.08.017
49. Gobbens RJJ, Andreasen J. The prediction of readmission and mortality by the domains and components of the Tilburg Frailty Indicator (TFI): a prospective cohort study among acutely admitted older patients. *Arch Gerontol Geriatr.* 2020;89:104077. doi:10.1016/j.archger.2020.104077
50. Uchmanowicz I, Jankowska-Polanska B, Uchmanowicz B, Kowalczyk K, Gobbens RJ. Validity and reliability of the Polish version of the Tilburg Frailty Indicator (TFI). *J Frailty Aging.* 2016;5(1):27–32. doi:10.14283/jfa.2015.66
51. Santiago LM, Gobbens RJJ, van Assen M, Carmo CN, Ferreira DB, Mattos IE. Predictive validity of the Brazilian version of the Tilburg Frailty Indicator for adverse health outcomes in older adults. *Arch Gerontol Geriatr.* 2018;76:114–119.
52. Alqahtani BA, Abdelbasset WK, Alenazi AM. Psychometric analysis of the Arabic (Saudi) Tilburg Frailty Indicator among Saudi community-dwelling older adults. *Arch Gerontol Geriatr.* 2020;90:104128. doi:10.1016/j.archger.2020.104128
53. WHOQOL Group. Development of the World Health Organization WHOQOL-BREF quality of life assessment. *Psychol Med.* 1998;28(3):551–558. doi:10.1017/s0033291798006667
54. Power M, Quinn K, Schmidt S. Development of the WHOQOL-old module. *Qual Life Res.* 2005;14(10):2197–2214. doi:10.1007/s11136-005-7380-9
55. Schmidt S, Mühlan H, Power M. The EUROHIS-QOL 8-item index: psychometric results of a cross-cultural field study. *Eur J Public Health.* 2006;16(4):420–428. doi:10.1093/eurpub/cki155
56. Vellas B, Balardy L, Gillette-Guyonnet S, et al. Looking for frailty in community-dwelling older persons: the Gérontopôle Frailty Screening Tool (GFST). *J Nutr Health Aging.* 2013;17(7):629–631. doi:10.1007/s12603-013-0363-6
57. Guralnik JM, Ferrucci L, Pieper CF, et al. Lower extremity function and subsequent disability: consistency across studies, predictive models, and value of gait speed alone compared with the short physical performance battery. *J Gerontol a Biol Sci Med Sci.* 2000;55(4):M221–31. doi:10.1093/gerona/55.4.M221
58. Collard RM, Boter H, Schoevers RA, Oude Voshaar RC. Prevalence of frailty in community-dwelling older persons: a systematic review. *J Am Geriatr Soc.* 2012;60(8):1487–1492. doi:10.1111/j.1532-5415.2012.04054.x
59. Kempen GI, Suurmeijer TP. The development of a hierarchical polychotomous ADL-IADL scale for noninstitutionalized elders. *Gerontologist.* 1990;30(4):497–502. doi:10.1093/geront/30.4.497
60. Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW. Studies of illness in the aged. The index of ADL: a standardized measure of biological and psychosocial function. *JAMA.* 1963;185:914–919. doi:10.1001/jama.1963.03060120024016
61. Gobbens RJ, van Assen MA, Luijckx KG, Wijnen-Sponselee MT, Schols JM. Determinants of frailty. *J Am Med Dir Assoc.* 2010;11(5):356–364.
62. van Assen MA, Pallast E, Fakiri FE, Gobbens RJ. Measuring frailty in Dutch community-dwelling older people: reference values of the Tilburg Frailty Indicator (TFI). *Arch Gerontol Geriatr.* 2016;67:120–129. doi:10.1016/j.archger.2016.07.005
63. Dedeyne L, Deschodt M, Verschueren S, Tournoy J, Gielen E. Effects of multi-domain interventions in (pre) frail elderly on frailty, functional, and cognitive status: a systematic review. *Clin Interv Aging.* 2017;12:873–896. doi:10.2147/CIA.S130794

Clinical Interventions in Aging

Dovepress

Publish your work in this journal

Clinical Interventions in Aging is an international, peer-reviewed journal focusing on evidence-based reports on the value or lack thereof of treatments intended to prevent or delay the onset of maladaptive correlates of aging in human beings. This journal is indexed on PubMed Central, MedLine, CAS, Scopus and the Elsevier

Bibliographic databases. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/clinical-interventions-in-aging-journal>