Manuscript File Click here to view linked References

TITLE

The Tilburg Frailty Indicator: a psychometric systematic review

ABSTRACT

Background

The Tilburg Frailty Indicator (TFI) is one of the most prominent multidimensional frailty assessment instruments. This review aimed to critically appraise and summarise its measurement properties.

Methods

Reports were eligible if they included results of studies aimed at developing the TFI or evaluating its measurement properties. We performed a literature search in MEDLINE, CINAHL, and PsycINFO databases from their inception until September 17, 2020. We also searched grey literature databases. We assessed the methodological quality of the included studies using the "COSMIN Risk of Bias". The measurement properties were evaluated using specific criteria. We graded the quality of the evidence using a GRADE approach.

Results

Fifty-five studies were included. We found moderate sufficient evidence for TFI content validity, although it is still insufficient for the comprehensiveness of its items. TFI construct validity was based on sufficient evidence from a single study of its structural validity as well as multiple hypothesis-testing for construct validity studies with inconsistent results. We did not find any studies that assessed cross-cultural validity. The TFI showed high sufficient concurrent validity with the comprehensive geriatric assessment. We identified several studies assessing its predictive validity for adverse frailty-related outcomes, although most of the evidence from these studies was insufficient. We did not find any studies that assessed the responsiveness of TFI scores.

Conclusions

The TFI is a simple measurement instrument that may be helpful in the assessment of frailty. However, more studies are needed to strengthen its usefulness as a clinical decision-making tool.

KEYWORDS

Aged, Frailty, Patient-Reported Outcome Measures, Psychometrics, Systematic Review

1. INTRODUCTION

Identifying older people who are frail or at risk of becoming frail has become the centrepiece of geriatric care in recent years (Hoogendijk et al., 2019). This identification can be approached from three perspectives: predominantly clinical, multidimensional, and holistic functional (Junius-Walker et al., 2018). Regarding the clinical perspective, frailty is understood as a clinical state determined by a series of signs and symptoms (Xue, 2011); however, this perspective may be insufficient to capture the full range of potential criteria determining frailty

(Junius-Walker et al., 2018; Sezgin et al., 2019). Moreover, other researchers suggest that frailty should not be limited to the physical domain and they advocate carrying out a multidimensional assessment that includes other factors such as psychological, cognitive, emotional, social or spiritual (Junius-Walker et al., 2018; Sezgin et al., 2019). On the other hand, World Health Organisation recommends holistic functional perspective, which includes a multidimensional approach to frailty. Nevertheless, its focus is on "total person functioning" rather than deficits and diseases (Junius-Walker et al., 2018). Because of this, it seems that the multidimensional perspective would be the most appropriate to assess all the possible factors involved in frailty (Sezgin et al., 2019).

Among the instruments available to perform a multidimensional assessment of frailty, the Tilburg Frailty Indicator (TFI) (Gobbens et al., 2010) is prominent. Existing systematic reviews show that it is one of the most robust instruments, especially to use in primary care (Pialoux et al., 2012; Sutton et al., 2016). However, these systematic reviews conclude that further in-depth evaluation of the measurement properties of TFI is needed. Moreover, these previous reviews do not cover all available evidence due to their completion date and focus on evaluating several measurement instruments rather than on a single instrument. Therefore, a specific systematic review on TFI may be more efficient in identifying all available evidence.

Developing and improving an effective individual care plan that meets the person's life goals involve a regular and multidimensional assessment of the person's needs (Dent et al., 2019). Multidimensional measures of frailty can provide more insight into these needs and enable tailored care management. The TFI is a multidimensional frailty assessment tool widely used in both clinical practice and geriatric research, so there is a need to update the available evidence to determine its suitability for this purpose. This systematic and psychometric review aimed to critically appraise and summarise the measurement properties of TFI to support evidence-based recommendations on its use and identify gaps in knowledge on its measurement properties, which can be used to design new studies.

2. METHODS

We conducted this psychometric systematic review following the COSMIN initiative (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018). The review is reporting according to the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) 2020 (Page et al., 2021) statement and the PRISMA literature search extension (Rethlefsen et al., 2021). The review protocol was registered in the International Prospective Register of Systematic Reviews (PROSPERO) on January 4, 2021 (registration number: CRD42021224427); no changes have been made to the protocol.

2.1. Search strategy and eligibility criteria

We performed a comprehensive literature search in MEDLINE (PubMed), CINAHL (EBSCOhost), and PsycINFO (EBSCOhost) databases from their inception until September 17,

2020. We also searched the grey literature on OpenGrey and Grey literature Report databases. Reports were eligible if they included results of studies aimed at developing the TFI or evaluating one or more of its measurement properties. Studies that only use the TFI as an outcome measure (e.g., clinical trials) or those used to validate another measurement instrument were excluded. Only full-text reports were included because often the minimum information about a study is found in the abstracts. Likewise, for our MEDLINE (PubMed) search, we added a highly sensitive filter to identify studies on measurement properties(Terwee et al., 2009). We imposed no language restrictions on any of the searches. The reproducible searches for all databases are available at https://doi.org/10.5281/zenodo.5513482. We manually screened reference lists of included studies to identify additional studies.

2.2. Selection process

We imported the retrieved references into the Rayyan QCRI web application program (Ouzzani et al., 2016). Two reviewers removed duplicates using the program's duplicate identification strategy and then manually. These two reviewers independently assessed the titles and abstracts of the retrieved records against the eligibility criteria. If a record seemed relevant to at least one of the reviewers, the full text of the report was independently reviewed by these two reviewers. Reviewers discussed conflicts over inclusion, and a third reviewer was consulted in case of not reaching a consensus.

2.3. Data collection process and study risk of bias assessment

We assessed the methodological quality of the included studies using the "COSMIN Risk of Bias" (available at: https://www.cosmin.nl/wp-content/uploads/COSMIN-RoB-checklist-V2-0-v17_rev3.pdf). The evaluation was carried out by each member of the review team independently. All assessments were compared, discussed, and agreed at regular meetings. Evaluation data were collected using forms designed by COSMIN (available at https://cosmin.nl/wp-content/uploads/Scoring-form-COSMIN-boxes_april_final.xlsx). Regarding criterion validity, we agreed, based on the available literature on frailty and following the COSMIN guidelines, that only comprehensive geriatric assessment (CGA) could be considered a gold standard for TFI (Hoogendijk et al., 2019; Mokkink et al., 2018; Parker et al., 2018; Prinsen et al., 2018). Predictive validity was assessed, as most of the identified studies aimed to assess this property; however, only longitudinal studies were assessed. We did not assess cross-sectional studies for this property; even though the authors described them as predictive validity studies, we considered them for construct validity.

2.4. Synthesis methods

The measurement properties were evaluated using specific criteria developed and agreed by experts (Prinsen et al., 2018). This assessment was carried out based on the number of studies available and the consistency of their results. We pooled the internal consistency results of the TFI scores by the Meta-Essentials tool (Suurmond et al., 2017) for correlational data version

1.5. We chose random-effects models based on the diversity of populations studied and in the expectation that internal consistency coefficients would differ. The meta-analyses were run with the Fisher-transformed values, which are transformed back into normal correlation coefficients for presentation (van Rhee et al., 2015). The extent and impact of between study heterogeneity were assessed by the tau^2 and the I^2 statistics, respectively. We explored possible causes of variation of results across studies by subgroup analyses based on the mode of administration. We summarised the rest of the measurement properties qualitatively.

The quality of the evidence was graded using a "Grading of Recommendations Assessment, Development and Evaluation (GRADE)" approach modified by COSMIN (Prinsen et al., 2018). As with the assessment of studies' quality, all these assessments were made independently by each review team member and subsequently compared, discussed, and agreed at regular meetings.

3. RESULTS

The literature search and study selection process is detailed in Figure 1. A total of 55 studies were included; a study had two reports (Hayajneh, 2019, 2016). The summary characteristics of the first study (TFI development) and all included studies are shown in Table 1 and Supplementary file Table 1. Twenty of the 55 studies were from The Netherlands, 17 from other European countries, ten from Asian countries, five from Brazil, one from Turkey, and two were multi-country studies. All studies assessed the measurement properties of part B of the TFI (described in Table 1). In addition, the five reports that were excluded and the reasons for exclusion are listed in Supplementary file Text 1.

The results of the methodological quality assessment of each study on a measurement property using the COSMIN Risk of Bias checklist (Mokkink et al., 2018) are shown in Supplementary file Table 2 and Table 2. Supplementary file Table 3 provides the ratings of each study against the criteria for good measurement properties (Prinsen et al., 2018). Finally, Table 3 summarises the evidence and its grade of quality by using the GRADE approach (Prinsen et al., 2018).

3.1. Content validity

Description and origin of the construct to be measured are clear, as well as the target population and the context of use. The TFI was developed based on previous research on frailty, but only experts were involved (Gobbens et al., 2010). Therefore, it appears that its development was not performed on a sample representing the target population. Comprehensibility and comprehensiveness are tested in a prefinal form of the instrument by representatives of professional disciplines (n=10) and by people aged 75 years and older (n=33), but these samples were not clearly described. No adjustment was necessary as this pilot test indicated that this version of the instrument was clear and comprehensive. However, the method and approach used to analyse data were not reported. For these reasons, the total quality of instrument design

and the pilot test performance were rated as inadequate and doubtful, respectively. This meant that the total rating of the instrument development was inadequate (see Supplementary file Table 2).

We identified 12 studies that analysed content validity aspects (see Table 2). Eleven of these studies assessed the comprehensibility of TFI versions translated into other languages. Among these studies, the study of the Danish version was the most salient (Andreasen et al., 2014). The quality of nine studies was assessed as doubtful mainly because the methodology's crucial aspects were not sufficiently clarified. One study on the TFI's Turkish version (Topcu et al., 2019) was assessed as inadequate because it did not describe that a cognitive interview or pilot test had been conducted in the target population. Moreover, in this study, the comprehensibility of the instrument was assessed by ten geriatricians. Only one study assessed aspects linked to the relevance and comprehensiveness of the instrument (Andreasen et al., 2015). In this study conducted with the TFI's Danish version, participants confirmed that the instrument covered most aspects of frailty. However, they identified aspects that they believed were not covered. The methodological quality of this study was rated as doubtful because it did not describe the experience with qualitative methods of the interviewers.

In summary (see Table 3), the content validity of the TFI was considered sufficient with moderate quality evidence, as the methodological quality of the studies was mostly doubtful. Evidence on comprehensiveness was considered insufficient as one content validity study indicated that the instrument did not cover important aspects related to frailty; in contrast, the evidence on the relevance and comprehensibility of the instrument was considered sufficient, although, concerning the latter aspect of content validity, the quality of evidence was considered high as the methodological quality of one of the content validity studies was adequate.

3.2. Construct validity

One study assessed the structural validity (Vrotsou et al., 2018) of the TFI, and 30 studies performed hypotheses testing for its construct validity. We did not identify any study that assessed its cross-cultural validity. Structural validity was analysed in the Spanish TFI version, and the results indicated a good fit of the three-factor model. The methodological quality of this study was rated very good, so the evidence on this measurement property was graded as high. We identified a total of 250 hypotheses for construct validity of which we considered 183 (73%) to be confirmed according to the criteria proposed by COSMIN (Prinsen et al., 2018). Nevertheless, the results were inconsistent, so we did not grade the evidence.

3.3. Criterion validity

Concurrent validity with CGA was assessed in two studies (Mazoochi et al., 2020; Si et al., 2020), for which the evidence was rated as sufficient with a high-quality grade. Several studies assessed the predictive validity of TFI for different adverse frailty-related outcomes such as higher health care use, disability, mortality, lower quality of life or falls. We also identified one

study that used a composite outcome variable with unequal importance of events such as readmission or death (Andreasen et al., 2018). The most frequent follow-up period was one year and two years. In some studies, it was six months, four years, seven years, and 12 years. One study analysed the TFI prediction for in-hospital mortality (Chong et al., 2017). The percentage lost to follow-up, in studies providing this information, was less than 20% in ten studies and higher in seven. Health care use was assessed using a wide range of variables: hospital admission, hospitalisation, unplanned readmission, receiving personal care or nursing care or informal care, general practitioner visits, contacts with health care professionals, residential care facilities, or institutionalisation. In most of these studies, the participants self-reported these variables, which contributed to the study's methodological quality being rated as doubtful. Concerning disability, all the identified studies, except one that used the Katz index (Gonzalez-Colaço Harmand et al., 2017), used the Groningen Activity Restriction Scale. Quality of life was measured in identified studies with the brief version of the World Health Organisation Quality of Life (WHOQOL-BREF) questionnaire. The evidence for TFI predictive validity was rated as sufficient for 12-year disability (Gonzalez-Colaço Harmand et al., 2017) and one-year mortality prediction (Chong et al., 2018; Daniels et al., 2012), the quality of this evidence was graded as low and moderate, respectively (see Table 3). It was rated as insufficient for all other adverse frailty-related outcomes studied.

3.4. Reliability

Nineteen studies assessed the internal consistency of TFI scores. The methodological quality of these studies was mostly rated as very good, except for three studies that only reported overall internal consistency coefficient instead of each of three TFI domains (Metzelthin et al., 2010; Topcu et al., 2019; Uchmanowicz et al., 2014). We, therefore, rated the methodological quality of these three studies as inadequate. The pooled results showed that only the physical domain had sufficient evidence in this measurement property with a moderate quality due to its inconsistency (see Table 3). Likewise, only the results of studies in which the TFI was administered by interview were shown to be consistent. Test-retest reliability was assessed in nine studies. Statistical analyses were approached by treating the response options as continuous or dichotomous, whereby the studies provide information on intraclass correlation coefficient as well as Cohen's kappa index. However, one study assessed this property using Pearson's correlation coefficient (Gobbens et al., 2010). The methodological quality of six studies was rated as doubtful mainly because these studies did not provide evidence that participants were stable, it was unclear whether the test conditions were similar, or the time interval was not appropriate. The time interval was 20 weeks in one study, so its methodological quality was considered inadequate (Freitag et al., 2016). Two studies were of adequate methodological quality (Santiago et al., 2013; Vrotsou et al., 2018). We rated the overall evidence as inconsistent, so we did not grade its quality. Only one study assessed the measurement error of

TFI scores (Vrotsou et al., 2018). In this study, the limits of agreement of scores were calculated, but there was no information about the minimal important change, so this evidence was rated indeterminate.

3.5. Responsiveness

We did not identify any study that assessed the responsiveness of a TFI change score.

4. DISCUSSION

This psychometric review was conducted to assess the quality of TFI measurement properties. We found moderate sufficient evidence for TFI content validity, although it is still insufficient for aspects such as the comprehensiveness of its items. TFI construct validity was based on sufficient evidence from a single study of its structural validity, as well as multiple hypothesis-testing for construct validity studies with inconsistent results. However, we did not find any studies that assessed cross-cultural validity. Regarding criterion validity, the TFI showed high sufficient concurrent validity with the CGA. We also identified a substantial number of studies assessing its predictive validity for adverse frailty-related outcomes, although most of the evidence from these studies was insufficient. Internal consistency of scores was the most assessed measurement property; however, only the physical domain scale showed sufficient moderate evidence. We did not find any studies that assessed the responsiveness of TFI scores.

This review has included a much larger number of studies on the TFI compared to other reviews (Pialoux et al., 2012; Sutton et al., 2016). Focusing on a single measurement instrument has allowed for a more in-depth analysis. Furthermore, the fact of not using language limits may also have contributed to the identification of a more significant number of studies. On the other hand, scoring and grading the quality of methods, the interpretation of results and the grading of evidence is a subjective process. However, all included articles were independently reviewed by the reviewers and agreed by consensus amongst the reviewer team. This process helped to resolve discrepancies and reduce variability in interpretation.

In patient-reported outcome measures such as TFI, the target population is the most appropriate assessor of the content validity of a measurement instrument. The target population's comprehensibility of the TFI items has been evaluated in different studies due to the existence of multiple language versions. However, the assessment of the relevance and comprehensiveness of its items by the target population has been scarcely evaluated. We found only two studies that evaluated these content validity aspects (Andreasen et al., 2015; Gobbens et al., 2010). Andreasen et al. (2015) find that most of the TFI items are relevant to older people; however, some items in the TFI physical domain do not seem relevant to them, such as those related to unintentional weight loss or hand strength. They also observe that the TFI does not consider some important aspects for this population, such as pain, sleep disturbances, spirituality, or meaningful activities.

We identified that structural validity was assessed in a single study (Vrotsou et al., 2018) which confirmed a three-factor model for three TFI domains. However, it is surprising that no further studies have been found that analyse this measurement property since such a widely studied measurement instrument. In contrast, many studies have been carried out to analyse the convergent and discriminative validity of TFI scores. Multiple measurement instruments and variables were used for this purpose. It also highlights the large number of hypotheses tested in some studies. Hypotheses for construct validity were confirmed at a high rate (73%), but inconsistent results were observed regarding the association between TFI scores and variables measuring similar or related constructs. Moreover, most studies used *P*-values rather than to assess whether the magnitude of correlations or observed differences were similar or greater than expected (Prinsen et al., 2018). We did not find any studies that analysed cross-cultural validity of the TFI despite the numerous adaptations and translations that have been carried out. This type of validity is essential to determine the equivalence of scores between the original population and the new target population (Prinsen et al., 2018).

We found two studies showing good concurrent validity of TFI with CGA (Mazoochi et al., 2020; Si et al., 2020). This finding has important implications, as CGA is a time-consuming and high-resource intensive process (Hoogendijk et al., 2019; Parker et al., 2018) that can be problematic in some healthcare settings such as primary care. In these settings, the TFI may be a simpler and more feasible tool to capture similar aspects of frailty. We identified a significant number of TFI validation studies for adverse frailty-related outcomes prediction, but their results are inconclusive. Measuring the predictive validity of a frailty measurement instrument is no easy matter. Although most people remain in their baseline frailty state at a follow-up of 1-5 years, a substantial proportion (up to 37%) experience at least one transition, including both worsening and improvement of the frailty state (Hoogendijk et al., 2019). Therefore, there is a need to carry out validation studies that consider frailty as a dynamic process.

Only the TFI physical domain subscale showed sufficient internal consistency of its scores. One possible explanation for this might be that the internal consistency coefficients are highly dependent on the number of items and the two remaining subscales have a tiny number of items. An alternative explanation is that the psychological and social domains are poorly comprehensive, as reported by Andreasen et al. (2015). We identified a substantial number of studies that administered the TFI by interview even though it was designed to be self-administered. One interesting finding is that these results were much more inconsistent when the TFI was self-administered than when it was interviewer-administered. This indicates a possible effect of the mode of administration of the TFI on their data. This influence has been described in the literature, and it may be necessary to conduct experimental studies to determine the origin, magnitude, and direction of this influence (Bowling, 2005). Results in test-retest reliability were inconsistent, probably due to the variability of different methodological aspects

of the studies, such as the lack of evidence of whether respondents were stable, the different time intervals used, and above all the fact that the instrument was not administered under similar conditions on both occasions. Test-retest reliability is an essential requirement of all measurement instruments in clinical practice and research, so this is a measurement property that needs further investigation. On the other hand, in order to evaluate the measurement error, it is necessary to have information on the minimal important change, defined as the smallest change in score that people consider important (Prinsen et al., 2018). We found no information on this issue, so there is also a need to generate evidence on it.

Responsiveness was not explored in any of the studies identified. The lack of assessment of this measurement property is common in frailty measurement instruments (Hoogendijk et al., 2019). However, it is a very relevant measurement property, as clinicians and researchers need measurement instruments that can be used to monitor changes in frailty over time.

5. CONCLUSIONS

The TFI is a simple measurement instrument that may be helpful in the assessment of frailty, especially in settings where time and resources are limited. However, there are important issues for future research, and more studies are needed to strengthen its usefulness as a clinical decision-making tool. A patient-reported outcome measure must be valid for a wide range of uses in different populations, and each use may require new evidence. However, most studies on the TFI focus on analysing the same measurement properties and other properties such as structural validity, cross-cultural validity, reliability, and responsiveness have received much less attention. Moreover, it would be interesting to establish the methodologies and evaluation criteria for testing hypotheses for construct validity. In this way, more homogeneous results would be obtained, which would give more strength to available evidence. On the other hand, it would also be desirable to conduct studies that focus on confirming fewer but strongly supported hypotheses. Finally, it is essential to highlight the importance of generating more evidence regarding the content validity of the TFI, especially in aspects related to the comprehensiveness of its items from the perspective of people who are frail.

AUTHORS STATEMENTS

Acknowledgments

The authors appreciate the review of the English text by Patryk Bialoskorski, MA.

Funding

This work was supported by the Health Department grant number from the Generalitat of Catalunya (Spain) (SLT008/18/00011); and by a grant from the Territorial Management of Barcelona of the Institut Català de la Salut (Catalan Institute of Health) to the first author (JJZS) in the 2018 and 2021 editions. The funders had no role in review design, decision to publish, or preparation of this manuscript.

Declarations of interest

None

Availability of data

Data were extracted from published research papers. All data generated or analysed during this review are included in this manuscript and its supplementary information files.

Authors' contribution

JJZS and EZO conceived and designed the review. JJZS and EZO selected the reports. All authors extracted and analysed de data. All authors accessed and verified the data. EZO wrote the first draft of the manuscript. All authors interpreted the data and contributed to the writing of the final version of the manuscript. All authors agreed with the results and conclusions of this manuscript. All authors had full access to all the data in the review and had final responsibility for the decision to submit for publication

REFERENCES

Alqahtani, B.A., Abdelbasset, W.K., Alenazi, A.M., 2020. Psychometric analysis of the Arabic (Saudi) Tilburg Frailty Indicator among Saudi community-dwelling older adults. Arch. Gerontol. Geriatr. 90, 104128. https://doi.org/10.1016/j.archger.2020.104128

Andreasen, J., Aadahl, M., Sørensen, E.E., Eriksen, H.H., Lund, H., Overvad, K., 2018. Associations and predictions of readmission or death in acutely admitted older medical patients using self-reported frailty and functional measures. A Danish cohort study. Arch. Gerontol. Geriatr. 76, 65–72. https://doi.org/10.1016/j.archger.2018.01.013

Andreasen, J., Lund, H., Aadahl, M., Gobbens, R.J.J., Sorensen, E.E., 2015. Content validation of the Tilburg Frailty Indicator from the perspective of frail elderly. A qualitative explorative study. Arch. Gerontol. Geriatr. 61, 392–399. https://doi.org/10.1016/j.archger.2015.08.017

Andreasen, J., Sørensen, E.E., Gobbens, R.J.J., Lund, H., Aadahl, M., 2014. Danish version of the Tilburg Frailty Indicator – Translation, cross-cultural adaption and validity pretest by cognitive interviewing. Arch. Gerontol. Geriatr. 59, 32–38. https://doi.org/10.1016/j.archger.2014.02.007

Bowling, A., 2005. Mode of questionnaire administration can have serious effects on data quality. J. Public Health 27, 281–291. https://doi.org/10.1093/pubmed/fdi031

Chong, E., Ho, E., Baldevarona-Llego, J., Chan, M., Wu, L., Tay, L., 2017. Frailty and Risk of Adverse Outcomes in Hospitalized Older Adults: A Comparison of Different Frailty Measures. J. Am. Med. Dir. Assoc. 18, 638.e7-638.e11. https://doi.org/10.1016/j.jamda.2017.04.011

Chong, E., Ho, E., Baldevarona-Llego, J., Chan, M., Wu, L., Tay, L., Ding, Y.Y., Lim, W.S., 2018. Frailty in Hospitalized Older Adults: Comparing Different Frailty Measures in

Predicting Short- and Long-term Patient Outcomes. J. Am. Med. Dir. Assoc. 19, 450-457.e3. https://doi.org/10.1016/j.jamda.2017.10.006

Coelho, T., Santos, R., Paúl, C., Gobbens, R.J., Fernandes, L., 2015. Portuguese version of the Tilburg Frailty Indicator: Transcultural adaptation and psychometric validation. Geriatr. Gerontol. Int. 15, 951–960. https://doi.org/10.1111/ggi.12373

Daniels, R., van Rossum, E., Beurskens, A., van den Heuvel, W., de Witte, L., 2012. The predictive validity of three self-report screening instruments for identifying frail older people in the community. BMC Public Health 12, 69. https://doi.org/10.1186/1471-2458-12-69

De Witte, N., Gobbens, R., De Donder, L., Dury, S., Buffel, T., Verté, D., Schols, J.M.G.A., 2013. Validation of the Comprehensive Frailty Assessment Instrument against the Tilburg Frailty Indicator. Eur. Geriatr. Med. 4, 248–254. https://doi.org/10.1016/j.eurger.2013.03.001

Dent, E., Martin, F.C., Bergman, H., Woo, J., Romero-Ortuno, R., Walston, J.D., 2019. Management of frailty: opportunities, challenges, and future directions. The Lancet 394, 1376–1386. https://doi.org/10.1016/S0140-6736(19)31785-4

Dong, L., Liu, N., Tian, X., Qiao, X., Gobbens, R.J.J., Kane, R.L., Wang, C., 2017. Reliability and validity of the Tilburg Frailty Indicator (TFI) among Chinese community-dwelling older people. Arch. Gerontol. Geriatr. 73, 21–28. https://doi.org/10.1016/j.archger.2017.07.001

Freitag, S., Schmidt, S., Gobbens, R.J.J., 2016. Tilburg frailty indicator. Z. Für Gerontol. Geriatr. 49, 86–93. https://doi.org/10.1007/s00391-015-0889-9

Giacomini, S.B.L., Fhon, J.R., Rodrigues, R.A.P., 2020. Frailty and risk of falling in the older adult living at home. Acta Paul. Enferm. 33. https://doi.org/10.37689/acta-ape/2020AO0124

Gobbens, R.J., van Assen, M.A., Luijkx, K.G., Schols, J.M., 2012c. Testing an integral conceptual model of frailty. J. Adv. Nurs. 68, 2047–2060. https://doi.org/10.1111/j.1365-2648.2011.05896.x

Gobbens, R.J.J., 2017. [Physical and mental dimensions of quality of life of frail older people]. Tijdschr. Gerontol. Geriatr. 48, 160-168. Dutch. https://doi.org/10.1007/s12439-017-0221-9

Gobbens, R.J.J., Andreasen, J., 2020. The prediction of readmission and mortality by the domains and components of the Tilburg Frailty Indicator (TFI). Arch. Gerontol. Geriatr. 89, 104077. https://doi.org/10.1016/j.archger.2020.104077

Gobbens, R.J.J., Boersma, P., Uchmanowicz, I., Santiago, L.M., 2020. The Tilburg Frailty Indicator (TFI): New Evidence for Its Validity. Clin. Interv. Aging Volume 15, 265–274. https://doi.org/10.2147/CIA.S243233

Gobbens, R.J.J., Krans, A., van Assen, M.A.L.M., 2015. Validation of an integral conceptual model of frailty in older residents of assisted living facilities. Arch. Gerontol. Geriatr. 61, 400–410. https://doi.org/10.1016/j.archger.2015.06.001

Gobbens, R.J.J., Luijkx, K.G., van Assen, M.A.L.M., 2013. Explaining quality of life of older people in the Netherlands using a multidimensional assessment of frailty. Qual. Life Res. 22, 2051–2061. https://doi.org/10.1007/s11136-012-0341-1

Gobbens, R. J. J., Luijkx, K.G., Wijnen-Sponselee, M.T., Schols, J.M.G.A., 2010. Towards an integral conceptual model of frailty. J. Nutr. Health Aging 14, 175–181. https://doi.org/10.1007/s12603-010-0045-6

Gobbens, R.J.J., van Assen, M.A.L.M., 2017. Associations between multidimensional frailty and quality of life among Dutch older people. Arch. Gerontol. Geriatr. 73, 69–76. https://doi.org/10.1016/j.archger.2017.07.007

Gobbens, R.J.J., van Assen, M.A.L.M., 2014. The prediction of quality of life by physical, psychological and social components of frailty in community-dwelling older people. Qual. Life Res. 23, 2289–2300. https://doi.org/10.1007/s11136-014-0672-1

Gobbens, R.J.J., van Assen, M.A.L.M., 2012. Frailty and its prediction of disability and health care utilization: The added value of interviews and physical measures following a self-report questionnaire. Arch. Gerontol. Geriatr. 55, 369–379. https://doi.org/10.1016/j.archger.2012.04.008

Gobbens, R.J.J., van Assen, M.A.L.M., Augustijn, H., Goumans, M., van der Ploeg, T., 2021. Prediction of Mortality by the Tilburg Frailty Indicator (TFI). J. Am. Med. Dir. Assoc. 22, 607.e1-607.e6. https://doi.org/10.1016/j.jamda.2020.07.033

Gobbens, R. J. J., van Assen, M.A.L.M., Luijkx, K.G., Schols, J.M.G.A., 2012a. The Predictive Validity of the Tilburg Frailty Indicator: Disability, Health Care Utilization, and Quality of Life in a Population at Risk. The Gerontologist 52, 619–631. https://doi.org/10.1093/geront/gnr135

Gobbens, R.J.J., van Assen, M.A.L.M., Luijkx, K.G., Wijnen-Sponselee, M.T., Schols, J.M.G.A., 2012b. [Young frail elderly: assessed using the Tilburg Frailty Indicator]. Tijdschr. Gerontol. Geriatr. 43, 296-307. Dutch. https://doi.org/10.1007/s12439-012-0043-8

Gobbens, Robbert J.J., van Assen, M.A.L.M., Luijkx, K.G., Wijnen-Sponselee, M.Th., Schols, J.M.G.A., 2010. The Tilburg Frailty Indicator: Psychometric Properties. J. Am. Med. Dir. Assoc. 11, 344–355. https://doi.org/10.1016/j.jamda.2009.11.003

Gobbens, R.J.J., van Assen, M.A.L.M., Schalk, M.J.D., 2014. The prediction of disability by self-reported physical frailty components of the Tilburg Frailty Indicator (TFI). Arch. Gerontol. Geriatr. 59, 280–287. https://doi.org/10.1016/j.archger.2014.06.008

Gonzalez-Colaço Harmand, M., Meillon, C., Bergua, V., Tabue Teguo, M., Dartigues, J.-F., Avila-Funes, J.A., Amieva, H., 2017. Comparing the predictive value of three definitions of frailty: Results from the Three-City study. Arch. Gerontol. Geriatr. 72, 153–163. https://doi.org/10.1016/j.archger.2017.06.005

Hayajneh, A.A., 2019. The Psychometric Properties of the Arabic Version of the Tilburg Frailty Indicator. Glob. J. Health Sci. 11, 123–123. https://doi.org/10.5539/gjhs.v11n9p123

Hayajneh, A.A.-R., 2016. The Arabic (Jordan) version of the Tilburg Frailty Indicator.

Hoogendijk, E.O., Afilalo, J., Ensrud, K.E., Kowal, P., Onder, G., Fried, L.P., 2019. Frailty: implications for clinical practice and public health. The Lancet 394, 1365–1375. https://doi.org/10.1016/S0140-6736(19)31786-6

Junius-Walker, U., Onder, G., Soleymani, D., Wiese, B., Albaina, O., Bernabei, R., Marzetti, E., 2018. The essence of frailty: A systematic review and qualitative synthesis on frailty concepts and definitions. Eur. J. Intern. Med. 56, 3–10. https://doi.org/10.1016/j.ejim.2018.04.023

Kendhapedi, K.K., Devasenapathy, N., 2019. Prevalence and factors associated with frailty among community-dwelling older people in rural Thanjavur district of South India: a cross-sectional study. BMJ Open 9, e032904–e032904. https://doi.org/10.1136/bmjopen-2019-032904

Mazoochi, F., Gobbens, R.J.J., Lotfi, M., Fadayevatan, R., 2020. Diagnostic accuracy of the Tilburg Frailty Indicator (TFI) for early frailty detection in elderly people in Iran. Arch. Gerontol. Geriatr. 91, 104187. https://doi.org/10.1016/j.archger.2020.104187

Metzelthin, S.F., Daniëls, R., van Rossum, E., de Witte, L., van den Heuvel, W.J., Kempen, G.I., 2010. The psychometric properties of three self-report screening instruments for identifying frail older people in the community. BMC Public Health 10, 176. https://doi.org/10.1186/1471-2458-10-176

Mokkink, L.B., de Vet, H.C.W., Prinsen, C.A.C., Patrick, D.L., Alonso, J., Bouter, L.M., Terwee, C.B., 2018. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Qual. Life Res. 27, 1171–1179. https://doi.org/10.1007/s11136-017-1765-4

Mulasso, A., Roppolo, M., Gobbens, R.J., Rabaglietti, E., 2017. Mobility, balance and frailty in community-dwelling older adults: What is the best 1-year predictor of falls? Geriatr. Gerontol. Int. 17, 1463–1469. https://doi.org/10.1111/ggi.12893

Mulasso, A., Roppolo, M., Gobbens, R.J.J., Rabaglietti, E., 2016. The Italian Version of the Tilburg Frailty Indicator. Res. Aging 38, 842–863. https://doi.org/10.1177/0164027515606192

Mulasso, A., Roppolo, M., Gobbens, R.J.J., Rabaglietti, E., 2015. [The evaluation of frailty in older adults: Translation and cross-cultural adaptation of the Tilburg Frailty Indicator to the Italian context]. G. Ital. Psicol. 42, 593-615. Italian. https://doi.org/10.1421/81168

Op het Veld, Linda P. M., Beurskens, A.J.H.M., de Vet, H.C.W., van Kuijk, S.M.J., Hajema, K., Kempen, G.I.J.M., van Rossum, E., 2019a. The ability of four frailty screening instruments to predict mortality, hospitalization and dependency in (instrumental) activities of daily living. Eur. J. Ageing 16, 387–394. https://doi.org/10.1007/s10433-019-00502-4

Op het Veld, L.P.M., van Rossum, E., Kempen, G.I.J.M., Beurskens, A.J.H.M., Hajema, K.J., de Vet, H.C.W., 2019b. 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 1–6. https://doi.org/10.14283/jfa.2019.17

Ouzzani, M., Hammady, H., Fedorowicz, Z., Elmagarmid, A., 2016. Rayyan-a web and mobile app for systematic reviews. Syst. Rev. 5, 210. https://doi.org/10.1186/s13643-016-0384-4

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., McGuinness, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P., Moher, D., 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 372, n71. https://doi.org/10.1136/bmj.n71

Parker, S.G., McCue, P., Phelps, K., McCleod, A., Arora, S., Nockels, K., Kennedy, S., Roberts, H., Conroy, S., 2018. What is Comprehensive Geriatric Assessment (CGA)? An umbrella review. Age Ageing 47, 149–155. https://doi.org/10.1093/ageing/afx166

Pialoux, T., Goyard, J., Lesourd, B., 2012. Screening tools for frailty in primary health care: A systematic review: Screening tools for frailty. Geriatr. Gerontol. Int. 12, 189–197. https://doi.org/10.1111/j.1447-0594.2011.00797.x

Prinsen, C.A.C., Mokkink, L.B., Bouter, L.M., Alonso, J., Patrick, D.L., de Vet, H.C.W., Terwee, C.B., 2018. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual. Life Res. 27, 1147–1157. https://doi.org/10.1007/s11136-018-1798-3

Renne, I., Gobbens, R.J., 2018. Effects of frailty and chronic diseases on quality of life in Dutch community-dwelling older adults: a cross-sectional study. Clin. Interv. Aging 13, 325–334. https://doi.org/10.2147/CIA.S156116

Rethlefsen, M.L., Kirtley, S., Waffenschmidt, S., Ayala, A.P., Moher, D., Page, M.J., Koffel, J.B., 2021. PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. Syst. Rev. 10, 39–39. https://doi.org/10.1186/s13643-020-01542-z

Roppolo, M., Mulasso, A., Gobbens, R.J., Mosso, C.O., Rabaglietti, E., 2015. A comparison between uni- and multidimensional frailty measures: prevalence, functional status, and relationships with disability. Clin. Interv. Aging 10, 1669–78. https://doi.org/10.2147/CIA.S92328

Santiago, L.M., Gobbens, R.J.J., Mattos, I.E., Ferreira, D.B., 2019. A comparison between physical and biopsychosocial measures of frailty: Prevalence and associated factors in Brazilian older adults. Arch. Gerontol. Geriatr. 81, 111–118. https://doi.org/10.1016/j.archger.2018.12.003

Santiago, L.M., Gobbens, R.J.J., van Assen, M.A.L.M., Carmo, C.N., Ferreira, D.B., Mattos, I.E., 2018. Predictive validity of the Brazilian version of the Tilburg Frailty Indicator for adverse health outcomes in older adults. Arch. Gerontol. Geriatr. 76, 114–119. https://doi.org/10.1016/j.archger.2018.02.013

Santiago, L.M., Luz, L.L., Mattos, I.E., Gobbens, R.J.J., 2012. [Cross-cultural adaptation of the Tilburg Frailty Indicator (TFI) for use in the Brazilian population]. Cad. Saúde Pública 28, 1795-1801. Portuguese. https://doi.org/10.1590/S0102-311X2012000900018

Santiago, L.M., Luz, L.L., Mattos, I.E., Gobbens, R.J.J., van Assen, M.A.L.M., 2013. Psychometric properties of the Brazilian version of the Tilburg frailty indicator (TFI). Arch. Gerontol. Geriatr. 57, 39–45. https://doi.org/10.1016/j.archger.2013.03.001

Sezgin, D., O'Donovan, M., Cornally, N., Liew, A., O'Caoimh, R., 2019. Defining frailty for healthcare practice and research: A qualitative systematic review with thematic analysis. Int. J. Nurs. Stud. 92, 16–26. https://doi.org/10.1016/j.ijnurstu.2018.12.014

Si, H., Jin, Y., Qiao, X., Tian, X., Liu, X., Wang, C., 2020. Comparing Diagnostic Properties of the FRAIL-NH Scale and 4 Frailty Screening Instruments among Chinese Institutionalized Older Adults. J. Nutr. Health Aging 24, 188–193. https://doi.org/10.1007/s12603-019-1301-z

Sutton, J.L., Gould, R.L., Daley, S., Coulson, M.C., Ward, E.V., Butler, A.M., Nunn, S.P., Howard, R.J., 2016. Psychometric properties of multicomponent tools designed to assess frailty in older adults: A systematic review. BMC Geriatr. 16, 55. https://doi.org/10.1186/s12877-016-0225-2

Suurmond, R., van Rhee, H., Hak, T., 2017. Introduction, comparison, and validation of Meta-Essentials: A free and simple tool for meta-analysis. Res. Synth. Methods 8, 537–553. https://doi.org/10.1002/jrsm.1260

Terwee, C.B., Jansma, E.P., Riphagen, I.I., de Vet, H.C.W., 2009. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual. Life Res. 18, 1115–1123. https://doi.org/10.1007/s11136-009-9528-5

Terwee, C.B., Prinsen, C.A.C., Chiarotto, A., Westerman, M.J., Patrick, D.L., Alonso, J., Bouter, L.M., de Vet, H.C.W., Mokkink, L.B., 2018. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual. Life Res. 27, 1159–1170. https://doi.org/10.1007/s11136-018-1829-0

Topcu, Y., Tufan, F., Kılıç, C., 2019. Turkish version of the Tilburg Frailty Indicator. Clin. Interv. Aging 14, 615–620. https://doi.org/10.2147/CIA.S197512

Uchmanowicz, I., Gobbens, R., Jankowska-Polanska, B., Loboz-Rudnicka, M., Manulik, S., Loboz-Grudzien, K., 2014. Cross-cultural adaptation and reliability testing of the Tilburg Frailty Indicator for optimizing care of Polish patients with frailty syndrome. Clin. Interv. Aging 997. https://doi.org/10.2147/CIA.S64853

Uchmanowicz, I., Gobbens, R.J., 2015. The relationship between frailty, anxiety and depression, and health-related quality of life in elderly patients with heart failure. Clin. Interv. Aging 1595. https://doi.org/10.2147/CIA.S90077

Uchmanowicz, I., Jankowska-Polańska, B., Uchmanowicz, B., Kowalczuk, K., Gobbens, R.J.J., 2016. Validity and Reliability of the Polish Version of the Tilburg Frailty Indicator (TFI). J. Frailty Aging 5, 27–32. https://doi.org/10.14283/jfa.2015.66

Uchmanowicz, I., Kuśnierz, M., Wleklik, M., Jankowska-Polańska, B., Jaroch, J., Łoboz-Grudzień, K., 2018. Frailty syndrome and rehospitalizations in elderly heart failure patients. Aging Clin. Exp. Res. 30, 617–623. https://doi.org/10.1007/s40520-017-0824-6

Uchmanowicz, I., Lisiak, M., Wontor, R., Łoboz-Grudzień, K., 2015. Frailty in patients with acute coronary syndrome: comparison between tools for comprehensive geriatric assessment and the Tilburg Frailty Indicator. Clin. Interv. Aging 521. https://doi.org/10.2147/CIA.S78365

van der Vorst, A., Op het Veld, L.P.M., De Witte, N., Schols, J.M.G.A., Kempen, G.I.J.M., Zijlstra, G.A.R., 2018. The impact of multidimensional frailty on dependency in activities of daily living and the moderating effects of protective factors. Arch. Gerontol. Geriatr. 78, 255–260. https://doi.org/10.1016/j.archger.2018.06.017

van Rhee, H., Suurmond, R, Hak, T, 2015. User manual for Meta-Essentials: Workbooks for meta-analysis (Version 1.4).

Vrotsou, K., Machón, M., Rivas-Ruíz, F., Carrasco, E., Contreras-Fernández, E., Mateo-Abad, M., Güell, C., Vergara, I., 2018. Psychometric properties of the Tilburg Frailty Indicator

in older Spanish people. Arch. Gerontol. Geriatr. 78, 203–212. https://doi.org/10.1016/j.archger.2018.05.024

Xie, J., Yu, Q., Yang, L., Sun, Y., 2020. A Preliminary Study of the Effects of the Tilburg Frailty Indicator, Frailty Phenotype, and Silver Nanoparticle-Silver Needle Therapy in Senile Inpatients with Frailty. J. Nanosci. Nanotechnol. 20, 6518–6524. https://doi.org/10.1166/jnn.2020.18602

Xue, Q.-L., 2011. The Frailty Syndrome: Definition and Natural History. Clin. Geriatr. Med. 27, 1–15. https://doi.org/10.1016/j.cger.2010.08.009

Yang, Y., Liao, L., Deng, Q., 2020. [Comparison of FRAIL Scale and Tilburg Frailty Indicator Scale in assessment of frailty status of elderly patients with diabetes in the community]. Nurs. Integr. Tradit. Chines West. Med. 6, 1-4. Chinese. https://doi.org/10.11997/nitcwm.2020001001

Zhang, X., Tan, S.S., Bilajac, L., Alhambra-Borrás, T., Garcés-Ferrer, J., Verma, A., Koppelaar, E., Markaki, A., Mattace-Raso, F., Franse, C.B., Raat, H., 2020. Reliability and Validity of the Tilburg Frailty Indicator in 5 European Countries. J. Am. Med. Dir. Assoc. 21, 772-779.e6. https://doi.org/10.1016/j.jamda.2020.03.019

Zhang, X., Tan, S.S., Franse, C.B., Alhambra-Borrás, T., Durá-Ferrandis, E., Bilajac, L., Markaki, A., Verma, A., Mattace-Raso, F., Voorham, A.J.J., Raat, H., 2019. Association between physical, psychological and social frailty and health-related quality of life among older people. Eur. J. Public Health 29, 936–942. https://doi.org/10.1093/eurpub/ckz099

Figure 1. PRISMA 2020 Flow-Chart of Study Selection (Page et al., 2021)

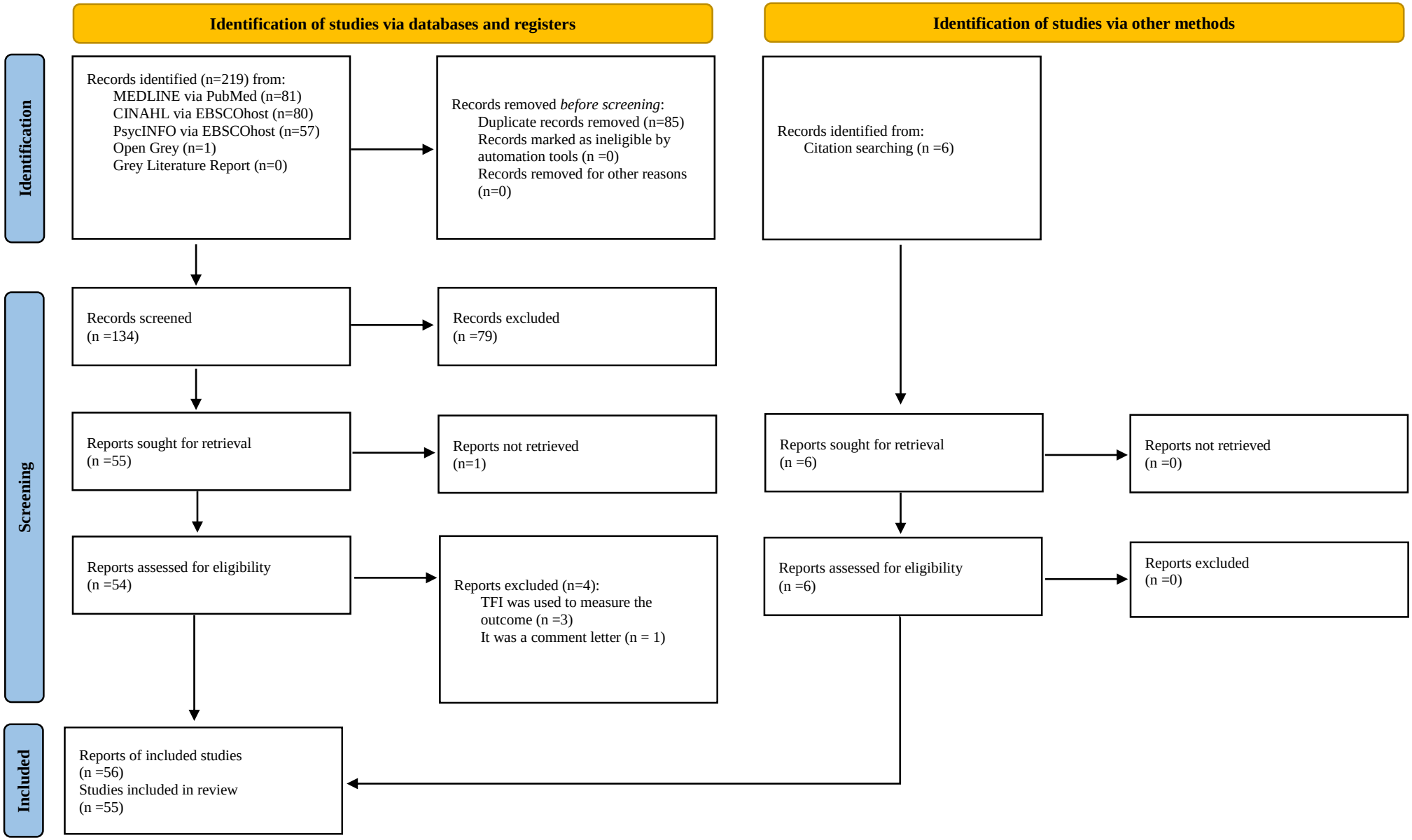

Table 1. Main characteristics of the Tilburg Frailty Indicator (Gobbens et al., 2010)

Target population	Mode of administration	Recall period	(Sub)scale (s) (number of items)	Response options	Range of scores/scoring	Original language	Available translations
Community-dwelling older people	Self-administered	Now	It consists of 2 parts: - Part A contains ten questions on determinants of frailty and diseases (multimorbidity) - Part B contains 15 questions divided into three domains of frailty: physical (8 questions), psychological (4 questions), and social (3 questions)	- Yes/No (11 questions) - Yes/ Sometimes /No (4 questions)	- The score for frailty and the three domains of frailty are determined by adding items' responses belonging to each scale. The response options "sometimes" and "yes" are grouped into a single score. - Score ranges from 0 to 15. - Frailty cut off point≥5 - Higher score=Higher degree of frailty	Dutch	-Arabic (Jordan & Saudi Arabia) -Chinese (China) -Croatian -Danish -English (UK) -French (France) -German -Greek -Italian -Persian -Polish -Portuguese (Brazil & Portugal) -Spanish (Spain) -Turkish

Table 2. Quality of studies on measurement properties of Tilburg Frailty Indicator

	Instrument development	Content validity					Structural validity	Internal consistency	Cross-cultural validity	Reliability	Measurement error	Criterion validity	Construct validity		Responsiveness			
		Asking patients			Asking experts													
Study		**Relevance**	**Comprehensiveness**	**Comprehensibility**	**Relevance**	**Comprehensiveness**							**Convergent validity**	**Known groups validity**	**Comparison with gold standard**	**Comparison with other instruments**	**Comparison between subgroups**	**Comparison before and after intervention**
Gobbens et al., 2010	I							V		D		D	V					
Metzelthin et al., 2010								I					V					
Daniels et al., 2012												D						
Gobbens and van Assen, 2012												D						
Gobbens et al., 2012a												D	D					
Gobbens et al., 2012b								V										
Gobbens et al., 2012c													D					
Santiago et al., 2012				D														
De Witte et al., 2013													V	D				
Gobbens et al., 2013								V					V					
Santiago et al., 2013								V		A			D					
Andreasen et al., 2014				A														
Gobbens and van Assen, 2014								V				D						
Gobbens et al., 2014								V				D						

	Instrument development	Content validity					Structural validity	Internal consistency	Cross-cultural validity	Reliability	Measurement error	Criterion validity	Construct validity		Responsiveness			
		Asking patients			Asking experts													
Study		**Relevance**	**Comprehensiveness**	**Comprehensibility**	**Relevance**	**Comprehensiveness**							**Convergent validity**	**Known groups validity**	**Comparison with gold standard**	**Comparison with other instruments**	**Comparison between subgroups**	**Comparison before and after intervention**
Uchmanowicz et al., 2014				D				I										
Andreasen et al., 2015		D	D															
Coelho et al., 2015				D				V		D			V					
Gobbens et al., 2015													D					
Mulasso et al., 2015				D														
Roppolo et al., 2015													V					
Uchmanowicz and Gobbens, 2015													V	D				
Uchmanowicz et al., 2015													V	D				
Freitag et al., 2016				D				V		I			V					
Mulasso et al., 2016								V					V					
Uchmanowicz et al., 2016								V		D								
Chong et al., 2017												I						
Dong et al., 2017				D				V		D			V					
Gobbens, 2017													V	D				

	Instrument development	Content validity					Structural validity	Internal consistency	Cross-cultural validity	Reliability	Measurement error	Criterion validity	Construct validity		Responsiveness			
		Asking patients			Asking experts													
Study		**Relevance**	**Comprehensiveness**	**Comprehensibility**	**Relevance**	**Comprehensiveness**							**Convergent validity**	**Known groups validity**	**Comparison with gold standard**	**Comparison with other instruments**	**Comparison between subgroups**	**Comparison before and after intervention**
Gobbens and van Assen, 2017													V					
Gonzalez-Colaço Harmand et al., 2017												D	V					
Mulasso et al., 2017												I		V				
Andreasen et al., 2018												D						
Chong et al., 2018												D						
Renne and Gobbens, 2018								V					V	D				
Santiago et al., 2018												I						
Uchmanowicz et al., 2018												V						
van der Vorst et al., 2018												I						
Vrotsou et al., 2018				I			V	V		A	A		V	D				
Hayajneh, 2016 Hayajneh, 2019				D				V					V	D				
Kendhapedi and Devasenapathy, 2019													V					
Op het Veld et al., 2019a												V						

	Instrument development	Content validity					Structural validity	Internal consistency	Cross-cultural validity	Reliability	Measurement error	Criterion validity	Construct validity		Responsiveness			
		Asking patients			Asking experts													
Study		**Relevance**	**Comprehensiveness**	**Comprehensibility**	**Relevance**	**Comprehensiveness**							**Convergent validity**	**Known groups validity**	**Comparison with gold standard**	**Comparison with other instruments**	**Comparison between subgroups**	**Comparison before and after intervention**
Op het Veld et al., 2019b												V						
Santiago et al., 2019													V					
Topcu et al., 2019				I				I		D								
Zhang et al., 2019														V				
Alqahtani et al., 2020				D				V		D			V					
Giacomini et al., 2020													V	D				
Gobbens and Andreasen, 2020												I						
Gobbens et al., 2020												D						
Mazoochi et al., 2020												V						
Si et al., 2020												V						
Xie et al., 2020													V					
Yang et al., 2020													V					
Zhang et al., 2020								V					V					
Gobbens et al., 2021												V						

V, Very good; A, Adequate; D, Doubtful; I, Inadequate. Empty cells indicate that this measurement property had not been assessed.

Table 3. Summary of findings

Measurement property	Summary or pooled result	Overall rating	Quality of evidence
Content validity	Not applicable	Sufficient	Moderate: TFI development study inadequate quality, and there are multiple content validity studies of doubtful quality available
Relevance	Not applicable	Sufficient	Moderate: TFI development study inadequate quality, and there is one content validity study of doubtful quality available
Comprehensiveness	Not applicable	Insufficient	Moderate: TFI development study inadequate quality, and there is one content validity study of doubtful quality available
Comprehensibility	Not applicable	Sufficient	High: TFI development study inadequate quality, and there is at least one content validity study of adequate quality available
Structural validity	3 factors: CFI=0.961 and RMSEA=0.035	Sufficient	High: There is one study of very good quality available
Internal consistency			
Physical domain	Global: coefficient 0.69 (95% CI, 0.67 to 0.72), n= 7736, I^2=61.5%	Sufficient	Moderate: There are multiple studies of very good quality available, and there is moderate inconsistency
	Self-administered mode: coefficient 0.68 (95% CI, 0.65 to 0.71), n=6961, I^2=62.4%	Sufficient	Moderate: There is multiple studies of very good quality available, and there is moderate inconsistency
	Interviewer-administered mode: coefficient 0.74 (95% CI, 0.69 to 0.78), n=775, I^2=0.0%	Sufficient	High: There is multiple studies of very good quality available
Psychological domain	Global: coefficient 0.54 (95% CI, 0.49 to 0.58), n=7307, I^2=68.1%	Insufficient	Moderate: There is multiple studies of very good quality available, and there is moderate inconsistency

Measurement property	Summary or pooled result	Overall rating	Quality of evidence
	Self-administered mode: coefficient 0.55 (95% CI, 0.51 to 0.60), n=6532, I^2=68.1%	Insufficient	Moderate: There is multiple studies of very good quality available, and there is moderate inconsistency
	Interviewer-administered mode: coefficient 0.47 (95% CI, 0.37 to 0.56), n=775, I^2=26.0%	Insufficient	Moderate: There is multiple studies of very good quality available, and there is low inconsistency
Social domain	Global: coefficient 0.41 (95% CI, 0.34 to 0.48), n=7503, I^2=89.9%	Insufficient	Low: There is multiple studies of very good quality available, and there is high inconsistency
	Self-administered mode: coefficient 0.39 (95% CI, 0.30 to 0.48), n=6728, I^2=91.9%	Insufficient	Low: There is multiple studies of very good quality available, and there is high inconsistency
	Interviewer-administered mode: coefficient 0.47 (95% CI, 0.34 to 0.57), n=775, I^2=53.1%	Insufficient	Moderate: There is multiple studies of very good quality available, and there is moderate inconsistency
Cross-cultural validity\Measurement invariance	No info available		
Reliability	ICC range 0.23 to 0.99	Inconsistent	
Measurement error	LoA were calculated but MIC was not defined	Indeterminate	
Criterion validity			
- *Concurrent validity*	2 out of 2 hypotheses confirmed (100%)	Sufficient	High: There are studies of very good quality available
- *Predictive validity*	Health care use: 17 out of 45 hypotheses confirmed (38%)	Insufficient	High: There are studies of very good quality available
	Disability (all studies): 5 out of 12 hypotheses confirmed (42%)	Insufficient	High: There are studies of very good quality available
	Disability (one year follow-up): 4 out of 7 hypotheses confirmed (57%)	Insufficient	Moderate: There are studies of doubtful quality available
	Disability (two-year follow-up): 1 out of 2 hypotheses confirmed (50%)	Insufficient	High: There are studies of very good quality available
	Disability (12-year follow-up): 1 out of 1 hypothesis confirmed (100%)	Sufficient	Low: There is one study of doubtful quality available
	Mortality (all studies): 6 out of 9 hypotheses confirmed (67%)	Insufficient	High: There are studies of very good quality available

Measurement property	Summary or pooled result	Overall rating	Quality of evidence
	Mortality (one-year follow-up): 2 out of 2 hypotheses confirmed (100%)	Sufficient	Moderate: There are studies of doubtful quality available
	Mortality (two-year follow-up): 1 out of 1 hypothesis unconfirmed (0%)	Insufficient	High: There are studies of very good quality available
	Mortality (≥four-year follow-up): 3 out of 5 hypotheses confirmed (60%)	Insufficient	High: There is one study of very good quality available
	QoL: 1 out of 4 hypotheses confirmed (25%)	Insufficient	Moderate: There are studies of doubtful quality available
	Falls: 1 out of 4 hypotheses confirmed (25%)	Insufficient	Moderate: There are studies of doubtful quality available
Hypotheses testing for construct validity	183 out of 250 hypotheses confirmed (73%)	Inconsistent	
Responsiveness	No info available		

CFI, Comparative Fit Index; CI, Confidence Interval; ICC, Intraclass Correlation Coefficient; LoA, Limits of Agreement; MIC, Minimal Important Change; QoL, Quality of life; RMSEA, Root Mean Square Error of Approximation; TFI, Tilburg Frailty Indicator

Supplementary file Text 1 Click here to access/download;Table;Supplementary file Text 1.docx

Supplementary file Text 1.

List of excluded reports according to reasons of exclusion

Reason: measurement instrument was used to measure the outcome

Gobbens, R.J., van Assen, M.A., Luijkx, K.G., Wijnen-Sponselee, M.T., Schols, J.M., 2010. Determinants of frailty. J. Am. Med. Dir. Assoc. 11, 356-364. https://doir.org/10.1016/j.jamda.2009.11.008

Ntanasi, E., Yannakoulia, M., Mourtzi, N., Vlachos, G. S., Kosmidis, M. H., Anastasiou, C. A., Dardiotis, E., Hadjigeorgiou, G., Megalou, M., Sakka, P., & Scarmeas, N., 2020. Prevalence and Risk Factors of Frailty in a Community-Dwelling Population: The HELIAD Study. J. Aging. Health. 32, 14–24. https://doi.org/10.1177/0898264318801735

Verver, D., Merten, H., de Blok, C., & Wagner, C., 2019. A cross sectional study on the different domains of frailty for independent living older adults. BMC Geriatr.19, 61. https://doi.org/10.1186/s12877-019-1077-3

Reason: Report was a comment letter

Freitag S., 2016. Advancing the diversity of frailty measurements in Germany. Z. Gerontol. Geriatr. 49, 263–264. https://doi.org/10.1007/s00391-016-1069-2

Reason: Report not retrieved

Xiaoqian L., 2017. Application of Tilburg Frailty Indicator scale in evaluation of senile frailty. Chinese Nursing Research. 31, 4186-4189.

Supplementary file Table 1 Click here to access/download;Table;Supplementary file Table 1.docx

Supplementary file Table 1. Characteristics of the included studies

	Population					Health status/Determinants of frailty				Instrument administration				
Study	**N**	**Age Mean (SD, range) yr**	**Female**	**Education level**	**Health care use**	**Health behaviours**	**Physical**	**Psychological**	**Socioeconomic**	**Mode**	**Setting**	**Country**	**Language**	**Response rate**
Gobbens et al., 2010	245	80.3 (3.9, 75-93)	54.7%	- No or primary education, 37.9% - Secondary, 45.3% - Higher, 16.9%	-Visits GP ≥5 last yr, 29.7% - Hospitalisation, 21.3% -Receiving: personal care, 15.3%; nursing care, 13.9%, and informal care, 26.6%	Healthy lifestyle, 74.3%	- Comorbidity, 49.2% - Poor physical health, 28.5% - Difficulty in walking, 49.0% - Poor hearing, 36.5% - Poor vision, 21.9%	- Problems with memory, 10.2% - Feeling down, 39.6% - Feeling nervous or anxious, 29.1% - Death loved one, 34.0%	- Monthly income: ≤900€, 17.1%; 901-2100€, 65.0%; >2100€, 17.9% - Live alone, 47.8% - Social relations, 57.0% - Social support, 15.5% - Satisfaction residence, 95.9%	Self-administered	Community-based	The Netherlands	Dutch	54.0%
Metzelthin et al., 2010	532	77.2 (5.5, 70-97)	58.5 %	- No or primary education, 35,7% - Secondary, 48,6% - Higher, 15,7%	NIA	NIA	NIA	NIA	- Monthly income: ≤900€, 18.7%; 901-1500€, 39.0%; >1500€, 42.4% - Live in an urban area, 64.0%	Self-administered	Community-based	The Netherlands	Dutch	77.4%
Daniels et al., 2012	430	77.2 (5.5)	60.2%	- No or primary education, 32.9% - Secondary, 50.1%	Hospitalisation, 17.4%	NIA	NIA	NIA	- Monthly income: ≤900€, 15.6%; 901-1500€, 38.4%;	Self-administered	Community-based	The Netherlands	Dutch	86.0%

	Population					Health status/Determinants of frailty				Instrument administration				
Study	**N**	**Age Mean (SD, range) yr**	**Female**	**Education level**	**Health care use**	**Health behaviours**	**Physical**	**Psychological**	**Socioeconomic**	**Mode**	**Setting**	**Country**	**Language**	**Response rate**
				- Higher, 16,9%					>1500€, 46.0%					
Gobbens and van Assen, 2012	245	80.3 (3.9, 75-93)	54.7%	NIA	NIA	Healthy lifestyle, 74.3%	- Multimorbidity, 49.2% - Poor balance, 42.3%	NIA	NIA	Self-administered	Community-based	The Netherlands	Dutch	53.0%
Gobbens et al., 2012a	479	80.3 (3.8)	56.8%	- No or primary education, 38.1% - Secondary, 46.5% - Higher, 15.4%	-Visits GP ≥5 last yr, 26.7% - Hospitalisation last yr, 21.4% -Receiving: personal care, 13.5%; nursing care, 9.5%, and informal care, 30.6%	Healthy lifestyle, 73.6%	Multimorbidity, 48.5%	Death loved one, 33.0%	- Monthly income: ≤900€, 18.9**%;** 901-2100€, 63.5%; >2100€, 17.6% - Satisfaction residence, 96.4%	Self-administered	Community-based	The Netherlands	Dutch	42.0%
Gobbens et al., 2012b	308	61.4 (1.9, 58-64)	63.0%	- No or primary education, 23.3% - Secondary, 65.2% - Higher, 11.5%	NIA	Healthy lifestyle, 76.6%	- Multimorbidity, 22.4% - Severe illness, 4.2% - Poor physical health, 15.6% - Difficulty in walking, 14.8% - Poor hearing, 14.1% - Poor vision, 5.9%	- Problems with memory, 3.2% - Feeling nervous or anxious, 31.2% - Death loved one, 24.7% - Sadness, 35.2% - Loneliness, 26.7%	- Monthly income: ≤900€, 10.6%; 901-2100€, 54.9%; >2100€, 34,5% - Living alone, 14.9% - Insufficient social support, 8.6% - Satisfaction residence, 92.8%	Self-administered	Community-based	The Netherlands	Dutch	15.0%

	Population					Health status/Determinants of frailty				Instrument administration				
Study	**N**	**Age Mean (SD, range) yr**	**Female**	**Education level**	**Health care use**	**Health behaviours**	**Physical**	**Psychological**	**Socioeconomic**	**Mode**	**Setting**	**Country**	**Language**	**Response rate**
Gobbens et al., 2012c	213	80.3 (3.7)	59.6%	- No or primary education, 37.4% - Secondary, 48.4% - Higher, 14.2%	-Visits GP last yr: 1-2, 27.1%; 3-4, 36,2%; and ≥5, 28.6% - Hospitalisation last yr, 23.3% -Receiving: personal care, 15.9%; nursing care, 14.9%; and informal care, 25.8%	Healthy lifestyle, 74.4%	Multimorbidity, 47.6%	- Death loved one, 31.6%	- Monthly income: ≤900€, 20.6%; 901-2100€, 61.5%; >2100€, 17.9% - Satisfaction residence, 97.1%	Self-administered	Community-based	The Netherlands	Dutch	34.0%
Santiago et al., 2012	30	NIA	NIA	NIA	NIA	NIA	NIA	NIA	NIA	Interview-administered	NIA	Brazil	Portuguese	NIA
De Witte et al., 2013	178	- Median 74 yr - 47.5% were aged 75 or older	67.2%	NIA	NIA	NIA	NIA	NIA	NIA	Self-administered	Community-based	Belgium	Dutch	92.2%
Gobbens et al., 2013	1031	73.4 (5.8, 65-95)	33.2%	- No or primary education, 16.3% - Secondary, 43.0% - Higher, 40.7%	NIA	NIA	Multimorbidity, 39.3 %	NIA	- Monthly income: <1500€, 14.8%; 1500- 3000€, 57.3%; >3000€, 27.9%	Self-administered	Community–based	The Netherlands	Dutch	77.9%
Santiago et al., 2013	219	70.5 (7.9)	52.5%	<4 yr or illiterate, 67.7% 5-8 yr, 17.1%	NIA	NIA	- Poor physical health, 21.0%	- Problems with memory, 30.5% - Feeling or sadness or	- Living alone, 17.3% - Social relations, 45.2%	Interview-administered	Community-based	Brazil	Portuguese	NIA

	Population					Health status/Determinants of frailty				Instrument administration				
Study	**N**	**Age Mean (SD, range) yr**	**Female**	**Education level**	**Health care use**	**Health behaviours**	**Physical**	**Psychological**	**Socioeconomic**	**Mode**	**Setting**	**Country**	**Language**	**Response rate**
				> 8 yr, 15.3%			- Difficulty in walking, 25.8% - Poor hearing, 33.7% - Poor vision, 52.6%	depression, 21.2% - Feeling nervous or anxious, 34.2% - Feeling down, 32.2%	- Social support, 11.0%					
Andreasen et al., 2014	34	78.8 (6.9, 65-94)	62.0%	Primary school 38.0% had a craftsman education	Acutely admitted hospital, 61.8%	Healthy lifestyle, 79.0%	Comorbidity, 38.0%	NIA	Satisfied with home living environment, 97.0%	Interview-administered	Home and acute medical ward	Denmark	Danish	NIA
Gobbens and van Assen, 2014	484	80.2 (3.8)	56.8%	- No or primary education, 38.1% - Secondary, 46.4% - Higher, 15.5%	NIA	NIA	- Physical unhealthy, 29.1% - Difficulty in walking, 48% - Poor hearing, 36.6% - Poor vision, 21.4%	- Problems with memory, 9.5% - Feeling nervous or anxious, 30.7% - Feeling down, 40.0%	- Monthly income: ≤900€, 18.7%; 901-2100€, 63.7%; ≥2101€, 17.6% - Living alone, 48.1% - Lack of social relations, 59.0% - Lack of social support, 16.6%	Self-administered	Community–based	The Netherlands	Dutch	92.3%
Gobbens et at., 2014	429	72.6 (5.4, 65-87)	33.3%	- No or primary education, 14.2% - Secondary, 45.0%	NIA	NIA	Multimorbidity, 38.7%	NIA	- Monthly income: <1500€, 16.7%; 1500-3499€, 66.7%;	Self-administered	Web-based survey	The Netherlands	Dutch	98.9%

	Population					Health status/Determinants of frailty				Instrument administration				
Study	**N**	**Age Mean (SD, range) yr**	**Female**	**Education level**	**Health care use**	**Health behaviours**	**Physical**	**Psychological**	**Socioeconomic**	**Mode**	**Setting**	**Country**	**Language**	**Response rate**
				- Higher, 40.8%					≥3500€, 16.6%					
Uchmanowicz et al., 2014	100	68.2 (6.5, 64-71)	58.0%	Secondary education, 51%	NIA	NIA	≥2 chronic diseases, 65.0% -Poor Physical health, 64.0% -Difficulty in walking, 34.0 % -Poor hearing, 51% -Poor vision, 72.0%	- Problems with memory, 22.0% - Feeling nervous or anxious, 77.0%	Living alone, 31.0%	Interview-administered	Primary care facilities	Poland	Polish	100.0%
Andreasen et al., 2015	14	80.6 (69-93)	50.0%	Elementary, 78.6% University / college, 21.4%	NIA	NIA	100% with comorbidity	NIA	NIA	Interview-administered	Community–based	Denmark	Danish	NIA
Coelho et al., 2015	252	79.2 (7.3)	75.8%	Mean (years): 0, 14.3%; 1-4, 63.9%; ≥5, 21.9%	-Contact with healthcare professionals, 71.4% - Hospitalisation last yr, 24.6% -Receiving: personal care, 6.7%; nursing care, 27.8%; and informal care, 19.0%	NIA	- Poor Physical health, 38.9% - Difficulty in walking, 50.0% - Poor hearing, 27.4% - Poor vision, 32.1%	- Cognitive deficit, 52.4% - Depression, 44.8% - Severe anxiety symptoms, 51.6 %	- Monthly household income: ≤500€, 40.9%; ≥501€, 59.1% - Living alone, 52% - Lack of social relations, 59.9% - Lack of social support, 27.0%	Interview-administered	Community-based	Portugal	Portuguese	NIA

	Population					Health status/Determinants of frailty				Instrument administration				
Study	**N**	**Age Mean (SD, range) yr**	**Female**	**Education level**	**Health care use**	**Health behaviours**	**Physical**	**Psychological**	**Socioeconomic**	**Mode**	**Setting**	**Country**	**Language**	**Response rate**
Gobbens et al., 2015	221	84.8 (8.9, 55-101)	63.3%	- No or primary education, 50.2% - Secondary, 41.7% -Higher, 8.1%	-Visits GP last yr: 0, 19.1%; 1-2, 31.4%; 3-4, 27,3%; and ≥5, 22.2% - Hospitalisation last yr, 24.4% -Receiving: personal care, 65.6%; nursing care, 51.8%, and informal care, 59.3%	Healthy lifestyle, 46.1%	- Multimorbidity, 52.1% - Falls, 44.5 %	- Death loved one, 34.8%	- Monthly income ≤600€, 39,4%; 601-900€, 26.1%; 901-1200€, 17.1%; >1200€, 13.9% - Satisfaction residence, 91.4%	Self-administered or interview-administered	Assisting living facilities	The Netherlands	Dutch	60.5%
Mulasso et al., 2015	31	77.0 (8.0)	64.0%	- Primary school (5 years), 52.0% - Secondary, (8 years), 16.0% - Higher, 32.0%	NIA	Healthy lifestyle, 39.0%	Multimorbidity, 45.0%	NIA	- Satisfaction of the home living environment, 100%	NIA	Primary Health Care, and local associations	Italy	Italian	NIA
Roppolo et al., 2015	267	73.4 (6.0, 65-90)	59.9%	- Primary school, 28.8% - Secondary, 43.1% - Higher, 28.1%	Visits to GP, 95.5%	Healthy lifestyle, 44.6%	- Chronic disease, 70.4% - Serious disease, 13.1% - Taking some medication on a regular basis, 88.4%	- Problems with memory, 12.0% - Feeling nervous or anxious, 69.3% - Loss of somebody close, 24.4%	- Living alone, 29.6% - Satisfaction of housing environment, 94.4%	Self-administered	Community-based	Italy	Italian	53.0%

	Population					Health status/Determinants of frailty				Instrument administration				
Study	**N**	**Age Mean (SD, range) yr**	**Female**	**Education level**	**Health care use**	**Health behaviours**	**Physical**	**Psychological**	**Socioeconomic**	**Mode**	**Setting**	**Country**	**Language**	**Response rate**
							- Difficulty in walking, 13.9% - Poor hearing, 11.6% - Poor vision, 29.6%							
Uchmanowicz and Gobbens, 2015	100	Non frail: 62.3 (6.2) Frail 67.9 (10.7)	47.0%	Years of education in frail patients, mean (SD) 11.5 (2.3)	Number of hospitalisations during 1 year in frail patients vs non frail, mean (SD) 1.8 (1.1) vs 1.4 (0.5)	NIA	- Heart Failure, 100% - NYHA class: I, 10.0%; II, 53.0%; III, 35.0%, and IV, 2.0%	- HADS-Depression mean (SD) (Frail vs non frail), 8.8 (4.9) vs 3.4 (2.8) - HADS-Anxiety mean (SD) (Frail vs non frail) 9.5 (4.5) vs 9 (3.3)	- Monthly income: <654€, 97.0%	Interview-administered	Hospital-based	Poland	Polish	NIA
Uchmanowicz et al., 2015	135	69.8 (11.4, 50-92)	39.3%	- Primary education, 17.0% - Vocational education, 37.1% - Secondary, 40.0% - Higher, 5.9%	NIA	NIA	Acute coronary syndrome, 100%	- MMSE, mean (SD) 25.2 (4.0) - MMSE <24 (cognitive impairment), 29.6% - HADS >7 (disturbance), 67.4%	NIA	Self-administered	Hospital-based	Poland	Polish	NIA
Freitag et al., 2016	210	75.31 (5.74)	64.3%	Low, 5.7% Medium, 50.0%	NIA	NIA	- Poor Physical health, 33.3%	- Problems with memory, 5.7%	- Living alone, 41.9% - Lack of social	Self-administered	Urban and rural areas	Germany	German	NIA

	Population					Health status/Determinants of frailty				Instrument administration				
Study	**N**	**Age Mean (SD, range) yr**	**Female**	**Education level**	**Health care use**	**Health behaviours**	**Physical**	**Psychological**	**Socioeconomic**	**Mode**	**Setting**	**Country**	**Language**	**Response rate**
				High, 44.3%			- Difficulty in walking, 41.0% - Poor hearing, 32.4% - Poor vision, 18.6%	- Feeling nervous or anxious, 27.6%	relations, 49.5% - Lack of social support, 10.0%					
Mulasso et al., 2016	267	73.4 (6,0 65-90)	59.9%	- Primary school, 28.8% - Secondary, 43.1% - Higher, 28.1%	Visits to GP, 95.5%	Healthy lifestyle, 44.6%	- Chronic disease, 70.4% - Serious disease, 13.1% - Taking some medication on a regular basis, 88.4% - Difficulty in walking, 13.9% - Poor hearing, 29.6% - Poor vision, 46.4%	- Problems with memory, 12.0% - Feeling nervous or anxious, 69.3% - Feeling down, 61.0% - Loss of somebody close, 24.0%	- Living alone, 29.6% - Lack of social relations, 53.2% - Lack of social support, 22.8% - Satisfaction of housing environment, 94.4%	Self-administered	Primary Health Care, and community associations	Italy	Italian	53.0%
Uchmanowicz et al., 2016	212	70.6 (7.16, 60-90)	29.9%	NIA	NIA	NIA	NIA	- Problems with memory, 19% - Feeling nervous or anxious, 88%	NIA	Interview-administered	Primary Health Care	Poland	Polish	NIA
Chong et al., 2017	210	89.4 (4.6)	69.5%	NIA	NIA	NIA	NIA	- Dementia (Frail vs non frail),	NIA	Interview-administered	Hospital-based	Singapore	English	NIA

	Population					Health status/Determinants of frailty				Instrument administration				
Study	**N**	**Age Mean (SD, range) yr**	**Female**	**Education level**	**Health care use**	**Health behaviours**	**Physical**	**Psychological**	**Socioeconomic**	**Mode**	**Setting**	**Country**	**Language**	**Response rate**
								(51.8 vs 16.7%) Delirium (Frail vs non frail), (23.8 vs 4.8%)						
Dong et al., 2017	917	68.6 (6.6, 60-92)	63.8%	Years: ≤6 yrs, 31.6% 7-9 yrs, 28.5% ≥10 yrs, 39.9%	- Hospitalisation last yr, 21.9% - Emergency use, 5.8%	NIA	- Chronic disease, 82.2% - Poor Physical health, 22.2% - ADL disability, 18.0% - IADL disability, 28.6% - Difficulty in walking, 9.3% - Poor hearing, 10.0% - Poor vision, 15.9%	- Cognitive impairment, 8.2% - Problems with memory, 13.2% - Feeling nervous or anxious, 27.9% - Feeling down, 39.0% - Depression, 16.4%	- Living alone, 11.8% - Social relations, 7.7% - Social support, 4.5%	Interview-administered	Community-based	China	Chinese	NIA
Gobbens, 2017	374	79.8 (3,7, 75-95)	30.7%	- Primary education, 8.0% - Vocational/ Secondary, 51.4% -Higher, 40.6%	NIA	Physical inactivity, 22.5%	- Multimorbidity, 43.6% - Difficulty in walking, 35.8% - Poor hearing, 36.1% - Poor vision, 12.6%	- Problems with memory, 4.5% - Feeling nervous or anxious, 24.9% - Bleakness, 43.9% - Loneliness, 57.0%	- Monthly income: <1000€, 2.1%; 1000-1499€, 13.2% - Living alone 36.1% - Insufficient social support, 12.0%	Self-administered	Community-based	The Netherlands	Dutch	NIA

	Population					Health status/Determinants of frailty				Instrument administration				
Study	**N**	**Age Mean (SD, range) yr**	**Female**	**Education level**	**Health care use**	**Health behaviours**	**Physical**	**Psychological**	**Socioeconomic**	**Mode**	**Setting**	**Country**	**Language**	**Response rate**
Gobbens and van Assen, 2017	671	76.6 (4.7, 70-95)	30.6%	- None, 6.6% - Primary, 10.0% - Secondary, 42.5% - Higher, 40.9%	NIA	Physical inactivity, 19.1%	- Multimorbidity, 39.6% - Difficulty in walking, 31.4% - Poor hearing, 30.0% - Poor vision, 10.0%	- Problems with memory, 4.2% - Feeling nervous or anxious, 24.4% - Feeling down, 42.8%	- Monthly income: <2000€, 33.2%; 2000-3500€, 52.3%; >3500€, 14.5% - Living alone, 30.8% - Lack of social relations, 53.1% - Lack of social support, 11.5%	Self-administered	Community-based	The Netherlands	Dutch	NIA
Gonzalez-Colaço Harmand et al., 2017	1278	74 (4.7)	66.4%	- No primary education, 11.8% - Secondary, 23.6% - Higher, 64.4%	NIA	NIA	NIA	MMSE. mean (SD), 27.5 (2.0) (cognitive impairment <24)	NIA	Interview-administered	Home	France	French	NIA
Mulasso et al., 2017	192	73 (6.2)	62.0%	- Primary education, 28.0% - Secondary, 64.0% - Higher, 8.0%	NIA	NIA	- Chronic disease, 68.0% -Chronic diseases, mean (SD) 1.6 (0.9)	MMSE. mean (SD), 27.7 (2.6) (cognitive impairment <24)	NIA	Self-administered and interview-administered	Community-based	Italy	Italian	67.3%
Andreasen et al., 2018	1328	Median (10th and 90th percentiles)77.1 (67.5; 87.7)	50.4%	- Primary education, 83.6%	Length of stay in days, median (10th;90th	NIA	- Comorbidities median (10th;90th percentiles),	Short Portable Mental Status Questionnair	- Living alone, 50.5 % - Satisfied with	Self-administered and interview-	Acute receiving medical units and	Denmark	Danish	47.5%

	Population					Health status/Determinants of frailty				Instrument administration				
Study	**N**	**Age Mean (SD, range) yr**	**Female**	**Education level**	**Health care use**	**Health behaviours**	**Physical**	**Psychological**	**Socioeconomic**	**Mode**	**Setting**	**Country**	**Language**	**Response rate**
				- Secondary, 10.8 % -Higher, 4.7 %	percentiles), 5 (1;15)		1 (0;3) - Charlson comorbidity index score, median (10th;90th percentiles), 1 (0;4)	e, median (10th;90th percentiles), 9 (6;10)	neighbourhood, 94.7%	administered	medical wards			
Chong et al., 2018	210	89.4 (4.6)	69.5%	NIA	NIA	NIA	NIA	- Dementia (Frail vs non frail), (51.8 vs 16.7 %) -Delirium (Frail vs non frail), (23.8 vs 4.8%)	NIA	Interview-administered	Hospital department of geriatric medicine	Singapore	English	98.0%
Renne and Gobbens, 2018	241	76.5 (5.1, 70-90)	48.1%	- No or primary education, 22.4% - Secondary, 59.8% - Higher, 17.8%	NIA	NIA	- Multimorbidity, 30.3% - Poor physical health, 16.6% - Difficulty in walking, 35.7% - Poor hearing, 27.8% - Poor vision, 17.4%	- Problems with memory, 4.6 % - Feeling nervous or anxious, 21.6% - Feeling down, 30.7%	- Living alone, 28.6 % - Lack of social relations, 39.4% - Lack of social support, 11.2%	Self-administered	Primary Health Care	The Netherlands	Dutch	47.5%
Santiago et al., 2018	640	70.5 (8.2)	64.7%	None, 14.8% <5 years, 47.3% >5 years, 37.9%	NIA	NIA	- ADL disability, 9.7% - IADL disability, 47.6%	NIA	NIA	Interview-administered	Primary Health Care	Brazil	Portuguese	66.6%

	Population					Health status/Determinants of frailty				Instrument administration				
Study	**N**	**Age Mean (SD, range) yr**	**Female**	**Education level**	**Health care use**	**Health behaviours**	**Physical**	**Psychological**	**Socioeconomic**	**Mode**	**Setting**	**Country**	**Language**	**Response rate**
Uchmanowicz et al., 2018	330	72.1 (7.9, 60-94)	44.9%	NIA	≤2 Hospitalisation a year, 71.1% ≥3 hospitalisation a year, 62.2%	NIA	- Heart Failure, 100% - NYHA class: I, 5.8%; II, 48.5%; III, 40.3%, and IV, 5.4%	NIA	NIA	Interview-administered	Cardiology ward	Poland	Polish	100.0%
van der Vorst et al., 2018	1027	74.2 (6.1, 65-93)	55.1%	NIA	NIA	- Physically active (30 minutes for day on 5 or more days per week), 37.6% -Non-smoking, 85.6%	NIA	NIA	- Sufficient financial resources, 44.6% - Rural living environment, 54.5% - Being married, 65.5% - Minor ethnicity, 20.7%	Self-administered	Community-based	The Netherlands	Dutch	83.6%
Vrotsou et al., 2018	856	78.1 (4.9)	53.0%	- No or primary education, 81.0% - Secondary, 7.0% - Higher, 12.0%	NIA	Healthy lifestyle, 86.0%	≥2 chronic diseases, 70.0% - Difficulty in walking, 22.0% - Poor hearing, 22.0% - Poor vision, 20.0%	- Problems with memory, 12.0% - Feeling nervous or anxious, 29.0% - Feeling down, 32.0%	- Monthly income: ≤1200€, 62.0%; >1200€, 38% - Living alone, 23.0% - Lack of social relations, 54.0% - Lack of social support, 11.0%	Self-administered	Community-based	Spain	Spanish	95.8%
Hayajneh, 2016	109	67.6 (6.95, 60-88)	38.5%	- None, 28.4%	Hospitalisation last yr, 31.2%	NIA	≥2 chronic diseases, 68.8%	NIA	- Monthly income:	Interview-administered	Community-based	Jordan	Arabic	NIA

	Population					Health status/Determinants of frailty				Instrument administration				
Study	**N**	**Age Mean (SD, range) yr**	**Female**	**Education level**	**Health care use**	**Health behaviours**	**Physical**	**Psychological**	**Socioeconomic**	**Mode**	**Setting**	**Country**	**Language**	**Response rate**
Hayajneh, 2019				- Basic or primary, 23.0% - Secondary, 7.3% -Higher, 30.3%			- Hearing impairment, 33.9% - Vision impairment, 53.2%		<523€, 40.4%; 523€-1103€, 33,1%; >1103€, 9.2% - Living alone, 16.5%					
Kendhapedi & Devasenapathy, 2019	408	67.5 (6.62)	56.9%	- No or primary education, 69.6% - Secondary, 20.3% - Higher, 10.1%	NIA	-Smoke/ smokeless, 56.9% -Alcohol current using, 20.3% -Physical activity: some, 16.4%; minimal, 53.7%	Comorbidity, 74.3%	NIA	- Socioeconomic status: Rich, 34.6 %; Middle category, 33.3%; and Poor, 32.1% - Living alone, 20.8%	Interview-administered	Rural area	India	Tamil	59.2%
Op Het Veld et al., 2019a	2420	76.3 (6.6)	60.5%	- No formal or primary education or pre-vocational secondary education, 68.9% - Higher, 31.1%	NIA	NIA	NIA	NIA	Living alone, 39.2%	Self-administered	Community-based	The Netherlands	Dutch	NIA
Op Het Veld et al., 2019b	2420	76.3 (6.6, 65-97)	60.5%	NIA	NIA	NIA	NIA	NIA	Living alone, 39.2%	Self-administered	Community-based	The Netherlands	Dutch	NIA
Santiago et al., 2019	302	70.4 (7.6)	65.9%	Education ≤5 years at school	NIA	- Healthy lifestyle, (frail/non frail)	- Functional dependence in ADL (frail vs non frail),	NIA	Satisfaction of housing environment	Interview-administered	Primary Health Care	Brazil	Portuguese	NIA

	Population					Health status/Determinants of frailty				Instrument administration				
Study	**N**	**Age Mean (SD, range) yr**	**Female**	**Education level**	**Health care use**	**Health behaviours**	**Physical**	**Psychological**	**Socioeconomic**	**Mode**	**Setting**	**Country**	**Language**	**Response rate**
				(frail/ non frail) (79.4 vs 69.1%		(61.1 vs 96.4%) -Smoker in the past/ current smoker (frail/non frail) (49.1 vs 45.1%) -Alcohol, habit in the past/current habit (frail/ non frail) (53.7 vs 50.0 %) - Non regular physical activity (frail /non frail) (82.4 vs 78.4%)	(17.6 vs 0.5%) - Functional dependence in IADL (frail vs non frail), (75.0 vs 36.6%)		(frail vs non frail) (26.9 vs 7.7%)					
Topcu et al., 2019	198	77.7 (5.5, 70-95)	68.7%	- No formal or primary education, 62.2% - Secondary, 12.8 % - Higher, 25.0%	NIA	NIA	NIA	NIA	- Monthly income: <103€, 20.9%; 103-206 €, 48.5%; >206€, 30.6%	Interview-administered	Geriatrics outpatient clinic	Turkey	Turkish	NIA

	Population					Health status/Determinants of frailty				Instrument administration				
Study	**N**	**Age Mean (SD, range) yr**	**Female**	**Education level**	**Health care use**	**Health behaviours**	**Physical**	**Psychological**	**Socioeconomic**	**Mode**	**Setting**	**Country**	**Language**	**Response rate**
Zhang et al., 2019	2167	79.7 (5.6)	60.6%	- No formal or primary education, 27.3% - Secondary, 63.5% - Higher, 9.1%	NIA	-Alcohol, consump of risk, 26.4% - Smoking, 7.3% -Exercise, once a week or less, 28.3%	Multimorbidity, 91.0%	NIA	Living alone, 38.0%	Interview-administered	Community-based	Croatia Greece The Netherlands Spain UK	Croatian Dutch English Greek Spanish	93.2%
Alqahtani et al., 2020	84	72 (4.7)	27.4%	- No formal education, 8.3% - Primary school, 51.2 % - Secondary or more, 40.5%	NIA	Healthy lifestyle, 63.1%	Comorbidity, 75.0%	NIA	- Monthly income: ≤900€, 8.4%; 901-1500€, 19.0%; >1500€,72.6 % -Satisfactory environment, 90.1%	Interview-administered	Outpatient clinic	Saudi Arabia	Arabic	NIA
Giacomini et al., 2020	261	80.76 (7.23)	70.9%	(% no risk of falling/ % with risk of falling) - Illiterate, (5.5/20.7) - 1-4 years (52.4/56.3) - 5-9 years (15.1/11.1) - 10 or more (27.0/11.9)	NIA	NIA	NIA	NIA	Living alone, 23.4%	Interview-administered	Community-based	Brazil	Portuguese	43.4%

	Population					Health status/Determinants of frailty				Instrument administration				
Study	**N**	**Age Mean (SD, range) yr**	**Female**	**Education level**	**Health care use**	**Health behaviours**	**Physical**	**Psychological**	**Socioeconomic**	**Mode**	**Setting**	**Country**	**Language**	**Response rate**
Gobbens and Andreasen, 2020	1328	76.9 (7.5, 65-100)	50.4%	- Primary education, 45.0% - Secondary, 34.4% - Higher, 20.6%	Six months after: 47.8% readmission at hospital	NIA	- Poor physical health, 49.5% - Difficulty in walking, 41.2% - Poor hearing, 27.3% - Poor vision, 26.2%	- Problems with memory, 12.6 % - Feeling down, 48.6% - Feeling nervous or anxious, 41.9%	- Monthly income: ≤ 1070 €, 9.4%; 1071-1470, 33.5%; >1470€, 51.6% - Living alone, 50.6%	Interview-administered	Hospital-based	The Netherlands	Dutch	NIA
Gobbens et al., 2020	180	76.3 (5.1, 70-90)	47.8%	- No formal or primary education, 20.6% - Secondary, 60.0% - Higher, 19.4%	Visits GP last yr, mean (SD) 2.8 (1.1)	NIA	Number of chronic diseases, mean (SD), range 1.1 (1.0), 0-5	NIA	NIA	Self-administered	Community-based	The Netherlands	Dutch	35.5%
Mazoochi et al., 2020	175	68.54 (6.44)	25.7%	- No formal or primary education, 62.3% - Secondary, 13.7% - Higher, 24.0%	NIA	Healthy lifestyle, 69.7%	Multimorbidity, 68.6%	Mental disease, 16.0%	- Monthly income: <196€, 20.6%; 196-390 €, 58.8%; >390 €, 20.6%	Interview-administered	Community-based	Iran	Persian	NIA
Si et al., 2020	305	79.3 (8.4)	57.0%	- No formal or primary education, 57.4% - Secondary, 17.0%	NIA	NIA	-ADL disability, 57.4% -Poor mobility, 53.8% -Malnutrition, 43.9%	- Mild/moderate cognitive impairment, 38.4% - Depression, 23.9%	Inadequate social support, 52.5%	Interview-administered	Nursing homes	China	Chinese	NIA

	Population					Health status/Determinants of frailty				Instrument administration				
Study	**N**	**Age Mean (SD, range) yr**	**Female**	**Education level**	**Health care use**	**Health behaviours**	**Physical**	**Psychological**	**Socioeconomic**	**Mode**	**Setting**	**Country**	**Language**	**Response rate**
				- Higher, 25.6%										
Xie et al., 2020	48	69.5 (6.2)	35.4%	NIA	NIA	- Smoking, 50.0% -Drinking alcohol, 41.7%	Malnutrition 52.1%	- Depression, 77.1% - Insomnia, 37.5%	NIA	Interview-administered	Hospital-based	China	Chinese	NIA
Yang et al., 2020	343	NIA	NIA	NIA	NIA	NIA	NIA	NIA	NIA	Interview-administered	Community-based	China	Chinese	95.3%
Zhang et al., 2020	2250	79.7 (5.7)	60.3%	- No formal or primary education, 27.3% - Secondary, 62.3% Higher level 10.3%	NIA	NIA	-Loss of independence, 25.8% -Limited function, 53.1%	- Poor mental health, 14.4% - Feeling lonely, 46.5%	NIA	Self-administered	Community-based	Croatia Greece The Netherlands Spain UK	Croatian Dutch English Greek Spanish	NIA
Gobbens et al. 2021,	479	80.3 (3.8)	56.8%	- No formal or primary education, 38.1% - Secondary, 46.5% - Higher, 15.4%	NIA	NIA	-Poor physical health, 29.2% -Difficulty in walking, 48.1% -Poor hearing, 36.6% -Poor vision, 21.4%	- Problems with memory, 9.6 % - Feeling nervous or anxious, 31.0% - Feeling down, 40.2%	- Living alone, 47.8% - Lack of social relations, 59.0% - Lack of social support, 16.4%	Self-administered	Community-based	The Netherlands	Dutch	42.0%

ADL, Activities of daily living; GP, General Practitioner; HADS, Hospital Anxiety and Depression Scale; IADL, Instrumental activities of daily living; MMSE, Mini Mental State Examination; NIA, No information available; NYHA, New York Hearth Association classification of heart failure; SD, Standard deviation

Click here to access/download;Table;Supplementary file Table 2.docx

Supplementary file Table 2. Quality of the Tilburg Frailty Indicator development

	Measurement instrument design							Cognitive interview (CI) or other pilot test performed[2]				TOTAL MEASUREMENT INSTRUMENT DEVELOPMENT
	General design requirements					Concept elicitation[1]	Total PROM design	General design requirements	Comprehensibility	Comprehensiveness	Total CI study	
	Clear construct	Clear origin of construct	Clear target population for which the PROM was developed	Clear context of use	PROM developed in sample representing the target population			CI study performed in sample representing the target population				
Gobbens et al., 2010	V	V	V	V	I		I	A	D	D	D	I

V, Very good; A, Adequate; D, Doubtful; I, Inadequate

[1]When the measurement instrument was not developed in a sample representing the target population, the concept elicitation was not further rated

[2]Empty cells indicate that a CI study (or part of it) was not performed

Supplementary file Table 3 Click here to access/download;Table;Supplementary file Table 3.docx

Supplementary file Table 3. Tables on results of studies on measurement properties

Table a. Structural validity, internal consistency, cross-cultural-validity, and reliability

	Country (language) in which the PROM was evaluated	Structural validity			Internal consistency			Cross-cultural validity\ measurement invariance			Reliability		
		n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)
Gobbens et al., 2010	The Netherlands (Dutch)				479	Very good	Physical domain (α=0.70); psychological domain (α=0.63) and social domain (α=0.34) Physical domain (+), psychologic domain (-) and social domain (-)				479	Doubtful	ICC or weighted Kappa not reported (?)
Metzelthin et al., 2010	The Netherlands (Dutch)				532	Inadequate	Internal consistency statistic not calculated on unidimensional scale (?)						
Daniels et al., 2012	The Netherlands (Dutch)												
Gobbens and van Assen, 2012	The Netherlands (Dutch)												
Gobbens, et al., 2012a	The Netherlands (Dutch)												
Gobbens et al., 2012b	The Netherlands (Dutch)				308	Very good	Physical domain (α=0.69); psychological domain (α=0.63) and social domain (α=0.46) Physical domain (-), psychologic domain (-) and social domain (-)						
Gobbens et al., 2012c	The Netherlands (Dutch)												
Santiago et al., 2012	Brazil (Portuguese)												
De Witte et al., 2013	Belgium (Dutch)												
Gobbens et al., 2013	The Netherlands (Dutch)				1031	Very good	Physical domain (α=0.67); psychological domain (α=0.54) and social domain (α=0.51) Physical domain (-), psychologic domain (-) and social domain (-)						

	Country (language) in which the PROM was evaluated	Structural validity			Internal consistency			Cross-cultural validity\ measurement invariance			Reliability		
		n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)
Santiago et al., 2013	Brazil (Portuguese)				118	Very good	Physical domain (α=0.79); psychological domain (α=0.53) and social domain (α=0.38) Physical domain (+), psychologic domain (-) and social domain (-)				101	Adequate	Physical domain: 4 of the 8 items showed kappa index >0.70 Psychological domain: 1 of the 4 items showed kappa index >0.70 Social domain: 2 of the 3 items showed kappa index >0.70. Physical domain (-), psychological domain (-) and social domain (-)
Andreasen et al., 2014	Denmark (Danish)												
Gobbens and van Assen, 2014	The Netherlands (Dutch)				196	Very good	Physical domain (α>0.70); psychological domain (α>0.70) and social domain (α=0.59) Physical domain (+), psychologic domain (+) and social domain (-)						
Gobbens et al., 2014	The Netherlands (Dutch)				429	Very good	Physical domain (α=0.67) (-)						
Uchmanowicz et al., 2014	Poland (Polish)				100	Inadequate	Internal consistency statistic not calculated on unidimensional scale (?)						
Andreasen et al., 2015	Denmark (Danish)												
Coelho et al., 2015	Portugal (Portuguese)				252	Very good	Physical domain (KR-20=0.75); psychological domain (KR-20=0.48) and social domain (KR-20=0.49) Physical domain (+), psychologic domain (-) and social domain (-)				74	Doubtful	Physical domain: 4 of the 8 items showed kappa index >0.70 Psychological domain: 1 of the 4 items showed kappa index >0.70 Social domain: 2 of the 3 items showed kappa index >0.70. Physical domain (-), psychological domain (-), and social domain (-)
Gobbens et al., 2015	The Netherlands (Dutch)												
Mulasso et al., 2015	Italy (Italian)												
Roppolo et al., 2015	Italy (Italian)												
Uchmanowicz and	Poland (Polish)												

	Country (language) in which the PROM was evaluated	Structural validity			Internal consistency			Cross-cultural validity\ measurement invariance			Reliability		
		n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)
Gobbens, 2015													
Uchmanowicz et al., 2015	Poland (Polish)												
Freitag et al., 2016	Germany (German)				210	Very good	Physical domain (α=0.66); psychological domain (α=0.43) and social domain (α=0.36) (physical domain (-), psychological domain (-) and social domain (-))				25	Inadequate	Physical domain (ICC =0.85); psychological domain (ICC=0.75), and social domain (ICC=0.84) (physical domain (+), psychological domain (+) and social domain (+))
Mulasso et al., 2016	Italy (Italian)				267	Very good	Physical domain (α=0.57); psychological domain (α=0.51) and social domain (α=0.36) (physical domain (-), psychological domain (-) and social domain (-))						
Uchmanowicz et al., 2016	Poland (Polish)				212	Very good	Physical domain (α=0.72); psychological domain (α=0.37) and social domain (α=0.59) (physical domain (+), psychological domain (-) and social domain (-))				212	Doubtful	Physical domain: all items showed kappa index >0.90 Psychological domain: all items showed kappa index >0.90 Social domain: all items showed kappa index >0.90 Physical domain (+) Psychological domain (+) Social domain (+)
Chong et al., 2017	Singapore (English)												
Dong et al., 2017	China (Chinese)				917	Very good	Physical domain (α=0.71); psychological domain (α=0.51) and social domain (α=0.25) (physical domain (+), psychological domain (-) and social domain (-))				103	Doubtful	Physical domain (ICC =0.80); psychological domain (ICC=0.65) & social domain (ICC=0.81) (physical domain (+), psychological domain (-) and social domain (+))
Gobbens, 2017	The Netherlands (Dutch)												
Gobbens and van Assen, 2017	The Netherlands (Dutch)												
Gonzalez-Colaço	France (French)												

	Country (language) in which the PROM was evaluated	Structural validity			Internal consistency			Cross-cultural validity\ measurement invariance			Reliability		
		n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)
Harmand et al., 2017													
Mulasso et al., 2017	Italy (Italian)												
Andreasen et al., 2018	Denmark (Danish)												
Chong et al., 2018	Singapore (English)												
Renne and Gobbens, 2018	The Netherlands (Dutch)				241	Very good	Physical domain (α=0.74); psychological domain (α=0.61) and social domain (α=0.51) (physical domain (+), psychological domain (-) and social domain (-))						
Santiago et al., 2018	Brazil (Portuguese)												
Uchmanowicz et al., 2018	Poland (Polish)												
van der Vorst et al., 2018	The Netherlands (Dutch)												
Vrotsou et al., 2018	Spain (Spanish)	829	Very good	Fit indexes were: RMSEA = 0.035 (90% CI: 0.027, 0.042), CFI = 0.961, TLI = 0.953 (+)	829	Very good	Physical domain (KR-20=0.64); psychological domain (KR-20=0.58) and social domain (KR-20=0.22) (physical domain (-), psychological domain (-) and social domain (-))				150	Adequate	The kappa coefficient could not be calculated for item Q12 “no weight loss” for all test replies; it was fair for items Q18, Q21, Q22 and Q25 (k = 0.23-0.34); very good for Q23 (k = 0.98) and moderate for the rest items (k = 0.46-0.57) (-)
Hayajneh, 2016 Hayajneh, 2019	Jordania (Arabic)				109	Very good	Physical domain (KR-20=0.74); psychological domain (KR-20=0.46) and social domain (KR-20=0.39) (physical domain (+), psychological domain (-) and social domain (-))						
Kendhapedi and Devasenapathy, 2019	India (Tamil)												

	Country (language) in which the PROM was evaluated	Structural validity			Internal consistency			Cross-cultural validity\ measurement invariance			Reliability		
		n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)
Op Het Veld et al., 2019a	The Netherlands (Dutch)												
Op Het Veld et al., 2019b	The Netherlands (Dutch)												
Santiago et al., 2019	Brazil (Portuguese)												
Topcu et al., 2019	Turkey (Turkish)				198	Inadequate	Internal consistency statistic not calculated on unidimensional scale (?)				16	Doubtful	There was a good agreement between two assessments (ICC=0.99) (+)
Zhang et al., 2019	Five European countries: Spain (Spanish), Greece (Greek), Croatia (Croatian), The Netherlands (Dutch), UK (English)												
Algahtani et al., 2020	Saudi Arabia (Arabic)				84	Very good	Physical domain (KR-20=0.68); psychological domain (KR-20=0.57) and social domain (KR-20=0.42) (physical domain (-), psychological domain (-) and social domain (-))				84	Doubtful	Overall scale ICC=0.86 (+)
Giacomini et al., 2020	Brazil (Portuguese)												
Gobbens and Andreasen, 2020	Denmark (Danish)												
Gobbens et al., 2020	The Netherlands (Dutch)												
Mazoochi et al., 2020	Iran (Persian)												

	Country (language) in which the PROM was evaluated	Structural validity			Internal consistency			Cross-cultural validity\ measurement invariance			Reliability		
		n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)
Si et al., 2020	China (Chinese)												
Xie et al., 2020	China (Chinese)												
Yang et al., 2020	China (Chinese)												
Zhang et al., 2020	Five European countries: Spain (Spanish), Greece (Greek), Croatia (Croatian), The Netherlands (Dutch), UK (English)				2250	Very good	Physical domain (α=0.70); psychological domain (α=0.52) and social domain (α=0.29) (physical domain (+), psychological domain (-) and social domain (-))						
Gobbens et al., 2021	The Netherlands (Dutch)												

CFI, Comparative Fit Index; CI, Confidence interval; KR-20, Kuder-Richardson formula; RMSEA, Root Mean Square Error of Approximation; TLI, Tucker-Lewis Index

Table b. Measurement error, criterion validity, hypothesis testing for construct validity, and responsiveness

	Country (language) in which the PROM was evaluated	Measurement error			Criterion validity			Hypotheses testing for construct validity			Responsiveness		
		n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)
Gobbens et al., 2010	The Netherlands (Dutch)				234	Doubtful	Predictive validity (1 yr) for: *Overall score* - Disability (GARS) (AUC 0.86; 95% CI, 0.81 to 0.92)	479	Very good	Convergent validity with **physical domain**: BMI (r=0.20), TUG test (r=0.36), LAPAQ (r=-0.28), grip strength test (r=-0.27), SFQ, (r=0.53) and four-test balance			

	Country (language) in which the PROM was evaluated	Measurement error			Criterion validity			Hypotheses testing for construct validity			Responsiveness		
		n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)
							- Health care use: hospital admission (AUC 0.61; 95%CI, 0.51 to 0.71); receiving personal care (AUC 0.85; 95% CI 0.78 to 0.92); receiving nursing (AUC 0.77); 95% CI, 0.69 to 0.86; receiving informal care (AUC 0.74; 95% CI, 0.67 to 0.81). *TFI's physical domain* - QoL: WHOQOL-BREF's physical domain (r=-0.71) *TFI's psychological domain* -QoL: WHOQOL-BREF's psychological domain (r=-0.58) *TFI's social domain* - QoL: WHOQOL-BREF's social domain (r=-0.22). Results in line with 6 hypo's (6+) Results not in line with 2 hypo's (2-)			scale (r=0.30); with **psychological domain:** MMSE (r=-0.09), CES-D (r=0.45), HADS-anxiety subscale (r=0.39) and MAS (r=0.40); with **social domain:** loneliness scale (r=0.45) and SSL (r=0.31) Results in line with 8 hypo's (8+). Results not in line with 4 hypo's (4-)			
Metzelthin et al., 2010	The Netherlands (Dutch)							532	Very good	Convergent validity with GFI (r=0.87); SPQ (r=0.42) and GARS (r=0.61) Results in line with 2 hypo's (2+) Result not in line with 1 hypo (1-)			
Daniels et al., 2012	The Netherlands (Dutch)				430	Doubtful	Predictive validity (1 yr) for: disability (GARS) (AUC 0.66; 95% CI, 0.60 to 0.72)); mortality (AUC 0.64; 95% CI, 0.60 to 0.72), and hospital admission (AUC 0.60; 95% CI, 0.52 to 0.67). Results in line with 2 hypo's (2+) Results not in line with 1 hypo's (1-)						
Gobbens and van Assen, 2012	The Netherlands (Dutch)				141	Doubtful	Predictive validity (1 yr and 2 yr) for: disability (GARS) (R^2=0.22 and R^2=0.30); visits GP (R^2=0.04 and R^2=0.07); contacts with HCP (R^2=0.06 and R^2=0.07) Results not in line with 6 hypo's (6-)						
Gobbens et al., 2012a	The Netherlands (Dutch)				266	Doubtful	Predictive validity (1 yr and 2 yr): disability (GARS), (AUC 0.80; 95% CI, 0.75 to 0.86; and AUC 0.81; 95% CI, 0.75 to 0.87); visit GP, (AUC 0.57;	479	Doubtful	Convergent validity with GARS (r=0.56); visits GP (r=0.34); contacts with HCP (r=0.35); WHOQOL-BREF's physical			

	Country (language) in which the PROM was evaluated	Measurement error			Criterion validity			Hypotheses testing for construct validity			Responsiveness		
		n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)
							95% CI, 0.48 to 0.66, and AUC 0.58, 95% CI, 0.47 to 0.70); contacts with HCP, (AUC 0.66, 95% CI, 057 to 0.75, and AUC 0.63; 95% CI, 0.52 to 0.74); hospitalisation, (AUC 0.65; 95% CI, 0.58 to 0.73, and AUC 0.60; 95% CI, 0.51 to 0.69); receiving personal care, (AUC 0.82; 95% CI, 0.76 to 0.89, and AUC 0.81,; 95% CI, 0.74 to 0.88); receiving nursing, (AUC 0.73; 95% CI, 0.65 to 0.81, and AUC 0.71; 95% CI, 0.61 to 0.82); receiving informal care, (AUC 0.73; 95% CI, 0.67 to 0.80, and AUC 0.75; 95% CI, 0.68 to 0.82), and facilities residential care, (AUC 0.81; 95% CI, 0.72 to 0.90, and AUC 0.78; 95% CI, 0.66 to 0.89). Results in line with 12 hypo's (12+) Results not in line with 2 hypo's (2-)			dimension (r= -0.72); WHOQOL-BREF's psychological dimension (r=-0.68); WHOQOL-BREF's social dimension (r=-0.39); WHOQOL-BREF's environmental dimension (r=-0.54): hospitalisation (r=0.08); receiving personal care (r=0.36); receiving nursing (r=0.26); receiving informal care (r=0.31), and facilities residential care (r=0.26). Results in line with 9 hypo's (9+) Results not in line with 3 hypo's (3-)			
Gobbens et al., 2012b	The Netherlands (Dutch)												
Gobbens et al., 2012c	The Netherlands (Dutch)							213	Doubtful	Convergent validity with disability (GARS)(r=0.66); visits GP (r=0.36); contacts with HCP (r=0.43); hospitalisation (r=0.19); receiving personal care (r=0.47); receiving nursing (r=0.34); receiving informal care (r=0.33); use of facilities in residential care (r=0.27); WHOQOL-BREF's physical dimension (r=-0.71); WHOQOL-BREF's psychological dimension (r=-0.69); WHOQOL-BREF's social dimension (r=-0.40), ans WHOQOL-BREF's environmental dimension (r=-0.52) Results in line with 10 hypo's (10+)			

	Country (language) in which the PROM was evaluated	Measurement error			Criterion validity			Hypotheses testing for construct validity			Responsiveness		
		n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)
										Results not in line wit 2 hypo's (2-)			
Santiago et al., 2012	Brazil (Portuguese)												
De Witte et al., 2013	Belgium (Dutch)							178	Very good/Doubtful	Convergent validity with the CFAI (r=0.59) Discriminative validity between aged <75 yr and aged >=75 yr: overall score 3.28 vs 4.36; physical domain 1.20 vs 2.30; psychological domain 1.07 vs 0.96, and social domain 1.00 vs 1.09. Results in line with 4 hypo's (4+) Result not in line with 1 hypo (1-)			
Gobbens et al., 2013	The Netherlands (Dutch)							1031	Very good	Convergent validity with WHOQOL-BREF: TFI's physical domain and WHOQOL-BREF's physical health (r=-0.68); TFI's psychological domain and WHOQOL-BREF's psychological (r=-0.57); TFI's social domain and WHOQOL-BREF's social (r=-0.34). Results in line with 3 hypo's (3+)			
Santiago et al., 2013	Brazil (Portuguese)							118	Doubtful	Convergent validity of the TFI's items with 14 corresponding alternative measures: 8 correlation coefficients were >0.30 and 6 were <0.30. Ranged from r=-0.21 to r=0.96). Results in line with 8 hypo's (8+) Results not in line with 6 hypo's (6-)			
Andreasen et al., 2014	Denmark (Danish)												
Gobbens and van Assen, 2014	The Netherlands (Dutch)				196	Doubtful	Predictive validity (2 and 4 yr) for QoL (WHOQOL-BREF): none of the items of TFI showed correlations>=0.70 with						

	Country (language) in which the PROM was evaluated	Measurement error			Criterion validity			Hypotheses testing for construct validity			Responsiveness		
		n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)
							the different dimensions of WHOQOL-BREF. Result not in line with 1 hypo (1-)						
Gobbens et al., 2014	The Netherlands (Dutch)				355	Doubtful	Predictive validity (2 yr) of physical domain's items for disability (GARS) (R^2=0.02 and R^2=0.00). Result not in line with 1 hypo (1-)						
Uchmanowicz et al., 2014	Poland (Polish)												
Andreasen et al., 2015	Denmark (Danish)												
Coelho et al., 2015	Portugal (Portuguese)							252	Very good	Convergent validity with: TFI's **physical** domain and BMI (r=0.16), TUG (r=0.48), handgrip strength (r=-0.34), COP sway analysis (r ranged from 0.07 to 0.18); TFI's **psychological domain** and MMSE (r=-0.22), GAI (r=0.56), GDS (r=0.58), and TFI's **social domain** with SSSS (r=-0.43). Convergent validity with GFI (AUC 0.89, 95% CI, 0.85 to 0.93), frailty phenotype (AUC 0.75, CI 95%, 0.68 to 0.81), EUROHIS-QOL-8 (r=-0.62), WHOQOL-OLD (r=-0.65), disability (Barthel Index) (AUC 0.72, 95% CI, 0.66 to 0.78), contact with healthcare professional (AUC 0.57, 95% CI 0.49 to 0.65). Results in line with 10 hypo's (10+) Results not in line with 4 hypo's (4-)			
Gobbens et al., 2015	The Netherlands (Dutch)							221	Doubtful	Convergent validity with disability (GARS) (r=0.49), WHOQOL-BREF: physical health (r=-0.67), psychological (r=-0.67),			

	Country (language) in which the PROM was evaluated	Measurement error			Criterion validity			Hypotheses testing for construct validity			Responsiveness		
		n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)
										social relationship (r=-0.42), environmental (r=-0.53)], visits to a GP (r=0.28), contacts with HCP (r=0.22), hospitalisation (r=0.06), receiving personal care (r=0.25), receiving nursing care (r=0.11), receiving informal care (r=-0.08), use of facilities in nursing home/rehabilitation centre (r=0.08) and falls (r=0.29) Results in line with 5 hypo's (5+) Results not in line with 8 hypo's (8-)			
Mulasso et al., 2015	Italy (Italian)												
Roppolo et al., 2015	Italy (Italian)							267	Very good	Convergent validity with CHS index: physical domain (r=0.42); psychological domain (r=0.37), and social domain (r=0.21). Convergent validity with GARS (r=0.48). Results in line with 3 hypo's (3+) Result not in line with 1 hypo (1-)			
Uchmanowicz and Gobbens, 2015	Poland (Polish)							100	Very good/ Doubtful	Convergent validity with SF-36 physical dimension (r=-0.66), SF-36 mental dimension (r=-0.68), HADS-Anxiety subscale (r=0.60), and HADS-depression subscale (r=0.66). Known-groups validity (frailty vs non-frailty people), average value HADS-anxiety subscale (9.5 vs 3.9), HADS-depression subscale (8.8 vs 3.4), SF-36 physical dimensions (32.8 vs 52.4), SF-36 mental dimension (42.9 vs 67.0) Results in line with 8 hypo's (8+)			

	Country (language) in which the PROM was evaluated	Measurement error			Criterion validity			Hypotheses testing for construct validity			Responsiveness		
		n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)
Uchmanowicz et al., 2015	Poland (Polish)							135	Very good/Doubtful	Convergent validity with MMSE (r=-0.60), HADS (r=0.60), ADLs (r=-0.428), and IADLs (r=-0.462). Discriminative validity (results shown figures only) people with frailty had lower values of ADLs, IADLs and MMSE and higher values of the HADS. Results in line with 8 hypo's (8+)			
Freitag et al., 2016	Germany (German)							210	Very good	Convergent validity: **overall score** with QoL (EUROHIS-8) (r=-0.56), physical subscale SF-12 (r=-0.59), psychological subscale SF-12 (r=-0.45); PHQ-9 (r=0.65), GAI (r=0.48), RS-11 (r=-0.43), social support (r=-0.35); **physical domain** with QoL (EUROHIS-8) (r=-0.51), physical subscale SF-12 (r=-0.66); **psychological domain** with psychological subscale SF-12 (r=-0.53), PHQ-9 (r=0.57), GAI (r=0.59), RS-11 (r=-0.407); **social subscale** with social support (t=-0.33) Results in line with 14 hypo's (14+)			
Mulasso et al., 2016	Italy (Italian)							267	Very good	Convergent validity **physical domain** with IPAQ (r=-0.25), BMI (r=0.13), TUG (r=0.40), OLS (r=-0.36), grip strength test (r=-0.34), CES-D (2 items) (r=0.43); **psychological domain**, MMSE (r=-0.10), CES-D (r=0.59), HADS (r=0.56), COPE (r=-0.14); **social domain,** loneliness scale (r=-0.37), LSNS, (r=-0.12). Convergent validity with disability (GARS) (AUC 0.83,			

	Country (language) in which the PROM was evaluated	Measurement error			Criterion validity			Hypotheses testing for construct validity			Responsiveness		
		n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)
										95% CI, 0.75 to 0.92), and falls (AUC 0.61, 95% CI, 0.52 to 0.69) Results in line with 9 hypo's (9+) Results not in line with 5 hypo's (5-)			
Uchmanowicz et al., 2016	Poland (Polish)												
Chong et al., 2017	Singapore (English)				210	Inadequate	Predictive validity for in-hospital mortality (frailty vs non-frailty, OR, 1.06; 95% CI, 0.83-1.34) Not all information for "+" reported(?)						
Dong et al., 2017	China (Chinese)							917	Very good	Convergent validity with alternative measures: **physical domain** with self-rated health status (r=-0.41), walking speed (r=-0.36), TUG test (r=0.41), making telephone calls (r=-0.16), vision measure (r=-0.19), grip strength (r=-0.35), poor endurance (r=0.57); psychological domain with the SPMSQ (r=-0.14), GDS (r=0.67), emotional role (r=-0.53), do at one's own will (r=-0.32); **social domain** with the APGAR scale (r=-0.28), the SSRS (r=-0.30), and living arrangement (r=0.20). Convergent validity with Fried's frailty phenotype (AUC 0.87) and FI (AUC 0.86) for a TFI cut-point of 3. Results in line with 10 hypo's (10+) Results not in line with 5 hypo's (5-)			
Gobbens, 2017	The Netherlands (Dutch)							374	Very good/Doubtful	Convergent validity with SF-12: 5 of the 8 TFI physical dimension's items had correlations >=0.30 with SF-12 physical dimension; 2 of the 4 TFI psychological			

	Country (language) in which the PROM was evaluated	Measurement error			Criterion validity			Hypotheses testing for construct validity			Responsiveness		
		n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)
										dimension's items had correlations >=30 with SF-12 mental dimension; 0 of the 3 TFI social dimension's items had correlations >=0.30 with any SF-12 dimension. Discriminative validity between frailty and non-frailty: DF-12 physical dimension (48.4 vs 78.4), and mental dimension (60.8 vs 82.9). Results in line with 9 hypo's (9+) Results not in line with 8 hypo's (8-)			
Gobbens and van Assen, 2017	The Netherlands (Dutch)							671	Very good	Convergent validity with the six facets of WHOQoL-OLD: **physical domain**, 4 of the 6 correlations were >=-0.30; **psychological domain,** 4 of the 6 correlations were >=-0.30, and social domain, 1 of the 6 correlations were >=-0.30. Results in line with 9 hypo's (9+) Results not in line with 9 hypo's (9-)			
Gonzalez-Colaço Harmand et al., 2017	France (French)				1278	Doubtful	Predictive validity (12 yr) for mortality (AUC, 0.59; 95%CI, 0.55 to 0.64); disability (Katz index) (AUC, 0.66; 95% CI, 0.61 to 0.72); institutionalisation (AUC, 0.65; 95% CI, 0.59 to 0.71), hospitalisation (AUC, 0.60; 95% CI, 0.54 to 0.65), and falls (AUC, 0.65; 95% CI, 0.60 to 0.70) Results in line with 3 hypo's (3+) Results not in line with 2 hypo's (2-)	1278	Very good	Convergent validity with Fried's phenotype (Kappa index=0.25), Rockwood index (Kappa index=0.33) Results not in line with 2 hypo's (2-)			
Mulasso et al., 2017	Italy (Italian)				192	Inadequate	Predictive validity (1 yr) for falls: overall (OR 1.3, 95% CI, 1.1-1.6). TFI physical dimension, OR 1.5 (95% CI, 1.2-1.9); TFI psychological dimension,	192	Very good	Discriminative validity between fallers and non-fallers (mean score): overall TFI (6.3 vs 3.8); TFI physical domain (2.6 vs 1.7);			

	Country (language) in which the PROM was evaluated	Measurement error			Criterion validity			Hypotheses testing for construct validity			Responsiveness		
		n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)
							1.5 (95% CI, 1.1-2.4), and TFI social dimension 1.4 (95% CI, 0.9-2.2) Not all information for "+" reported (?)			psychological domain (1.9 vs 1.3); social domain (1.1 vs 1.0). Results in line with 3 hypo's (3+) Result not in line with 1 hypo (1-)			
Andreasen et al., 2018	Denmark (Danish)				1328	Doubtful	Predictive validity (6 mos.) for unplanned readmission or death (AUC 0.67; CI 95% 0.64 to 0.70) Result in line with 1 hypo (1+)						
Chong et al., 2018	Singapore (English)				206	Doubtful	Predictive validity (6 and 12 mos.) for all-cause mortality (AUC 0.68; 95% CI, 0.59 to 0.77 and 0.69; 95% CI, 0.61 to 0.77) Results in line with 2 hypo's (2+)						
Renne and Gobbens, 2018	The Netherlands (Dutch)							241	Very good/Doubtful	Convergent validity with WHOQOL-OLD: TFI overall score (r=-0.68); TFI physical dimension (r=-0.59); TFI psychological dimension (r=-0.54), and TFI social dimension (r=-0.46). Discriminative validity with WHOQOL-OLD: frailty people scored lower than non-frailty people in 6 dimensions of WHOQOL-OLD. Results in line with 10 hypo's (10+)			
Santiago et al., 2018	Brazil (Portuguese)				640	Inadequate	Predictive validity (1 yr) for falls OR 1.9; 95% CI, 1.1 to 3.4), hospitalisation (OR 1.8; 95% CI, 1.1 to 3.0), ADL (OR						

	Country (language) in which the PROM was evaluated	Measurement error			Criterion validity			Hypotheses testing for construct validity			Responsiveness		
		n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)
							2.5; 95% CI, 1.1 to 5.3), IADL (OR 1.6; 95% CI, 1.1 to 2.5) and death (OR 2.7; CI 95% 1.0 to 7.3). Not all information for "+" reported (?)						
Uchmanowicz et al., 2018	Poland (Polish)				330	Very good	Predictive validity (1 yr) for hospitalisations: TFI overall score (r=0.14); physical dimension (r=0.09); psychological dimension (r=0.04), and social dimension (r=0.19). Results not in line with 4 hypo's (4-)						
Van der Vorst et al., 2018	The Netherlands (Dutch)				859	Inadequate	Predictive validity (2 yr) for ADL dependency (GARS) (OR=2.1; 95% CI, 1.5 to 3.0). Not all information for "+" reported (?)						
Vrotsou et al., 2018	Spain (Spanish)	150	Adequate	LoA were calculated but MIC was not defined (?)				856	Very good/ Doubtful	Convergent validity with Fried's scale (r=0.49), self-assessed health (r=0.44), TUG (r=0.39), SPPB (r=-0.38), EQ VAS (r=-0.32), gait speed (r=-0.29), Charlson index (r=0.16), Charlson age-adjusted index (r=0.18), and Lawton scale(r=0.13) Discriminative validity between frailty and non-frailty groups according SPPB (4.1 vs 2.4), GFST (4.7 vs 2.8), and between age groups 70-75 yr, (3.5); 76-85 yr, (3.2), and >85 yr, (3.6). Results in line wit 7 hypo's (7+) Results not in line with 5 hypo's (5-)			
Hayajneh, 2016 Hayajneh, 2019	Jordania (Arabic)							109	Very good (convergent validity) Doubtful (discri	Convergent validity with GDS (r=0.52), SF-36 Physical domain (r=-0.36), and SF-36 Social domain (r=-0.52). Discriminative validity (mean overall score) between older aged 60-70 yr and >=71 yr (6.5 vs 7.9); people who had comorbidities and			

	Country (language) in which the PROM was evaluated	Measurement error			Criterion validity			Hypotheses testing for construct validity			Responsiveness		
		n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)
									minative validity)	those who did not have (5.6 vs 7.6). Results in line with 4 hypo's (4+) Result not in line with 1 hypo (1-)			
Kendhapedi and Devasenapathy, 2019	India (Tamil)							406	Very good	Convergent validity with SFES (OR 5.1, 95% CI, 3.3 to 8.1), and falls (OR 1.86; 95% CI, 1.10 to 3.14) Result in line with 2 hypo (2+)			
Op Het Veld et al., 2019a	The Netherlands (Dutch)				2420	Very good	Predictive validity (2 yr) for IADL and ADL dependency (GARS) (Sen., 72,7% & Spe., 45.7%), mortality (Sen., 80.6% & Spe. 36.5%), and hospitalisation (Sen. 70.5% & Spe. 44.1%) Not all information for "+" reported (?)						
Op Het Veld et al., 2019b	The Netherlands (Dutch)				2420	Very good	Predictive validity (2 yr) for mortality (AUC 0.62, 95% CI, 0.58 to 0.66), hospitalisation (AUC 0.61, 95% CI, 0.58 to 0.63), and increase in IADL and ADL (GARS) (AUC 0.64, 95% CI, 0.61 to 0.66) Results not in line with 3 hypo's (3-)						
Santiago et al., 2019	Brazil (Portuguese)							302	Very good	Convergent validity with CHS index: TFI overall score (r=0.68); TFI physical dimension (r=0.74); TFI psychological dimension (r=0.32), and TFI social dimension (r=0.06). Results in line with 3 hypo's (3+) Result not in line with 1 hypo (1-)			
Topcu et al., 2019	Turkey (Turkish)												
Zhang et al., 2019	Five European countries: Spain (Spanish), Greece (Greek), Croatia (Croatian), The Netherlands							2167	Very good	Discriminative validity between frailty and non-frailty people in SF-12 physical component (36.6 vs 46.1), in SF.12 mental component (46.1 vs 55.4) Results in line with 2 hypo's (2+)			

	Country (language) in which the PROM was evaluated	Measurement error			Criterion validity			Hypotheses testing for construct validity			Responsiveness		
		n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)
	(Dutch), UK (English)												
Alqahtani et al., 2020	Saudi Arabia (Arabic)							84	Very good	Convergent validity with SPPB (r=-0.38); TUG test (r=0.38); gait speed (r=-0.34), grip strength (r=-0.32, PHQ-9 (r=0.45), and WHOQOL-BREF (r=-0.56); TFI physical domain with SPPB (r=-0.39), TUG (r=0.38), gait speed (r=-0.37), and grip strength (r=-0.30); TFI psychological domain with PHQ-9 (r=0.51). Results in line with 11 hypo's (11+)			
Giacomini et al., 2020	Brazil (Portuguese)							261	Very good/ Doubtful	Convergent validity with fall risk score (OR 6.7, 95% CI 3.9 to 11.6) Discriminative validity between frailty and non-frailty people in % risk of falls (66.7% vs 33.3%) Results in line with 2 hypo's (2+)			
Gobbens and Andreasen, 2020	Denmark (Danish)				1328	Inadequate	Predictive validity (6 mos.) for readmission and mortality. Not all information for "+" reported (?)						
Gobbens et al., 2020	The Netherlands (Dutch)				180	Doubtful	Predictive validity (1 yr) for disability (GARS) (physical domain r=0.49, psychological domain r=0.36 and social domain r=0.26), visits GP (physical domain r=0.39, psychological domain r=0.26 and social domain r=0.18), contacts with HCP (physical domain r=0.39, psychological domain r=0.15 and social domain r=0.20), hospitalisation (physical domain r=0.10, psychological domain r=0.12 and social domain r=0.02), receiving personal care (physical domain r=0.19, psychological domain r=0.11 and social domain r=0.15), receiving nursing (physical domain r=0.33, psychological domain						

	Country (language) in which the PROM was evaluated	Measurement error			Criterion validity			Hypotheses testing for construct validity			Responsiveness		
		n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)
							r=0.06 and social domain r=0.22), and falls (physical domain r=0.25, psychological domain r=0.11 and social domain r=0.18). Results not in line with 21 hypo's (21-)						
Mazoochi et al., 2020	Iran (Persian)				175	Very good	Concurrent validity with the CGA+Expert panel (AUC 0.92) Result in line with 1 hypo (1+)						
Si et al., 2020	China (Chinese)				305	Very good	Concurrent validity with the CGA (AUC 0.80, 95% CI, 0.75 to 0.85) Result in line with 1 hypo (1+)						
Xie et al., 2020	China (Chinese)							48	Very good	Convergent validity with Fried's phenotype(r=0.91) Result in line with 1 hypo (1+)			
Yang et al., 2020	China (Chinese)							343	Very good	Convergent validity with Fried's phenotype (AUC 0.89) cut off 5.5, falls (AUC 0.69, 95%0.75), and hospitalisations (AUC 0.67, 95% CI 0.62 to 0.73). Result in line with 3 hypo (3+)			
Zhang et al., 2020	Five European countries: Spain (Spanish), Greece (Greek), Croatia (Croatian), The Netherlands (Dutch), UK (English)							2250	Very good	Convergent validity: TFI's physical domain with SF-12 physical component (r=-0.56), GARS (r=0.57), GARS-ADL (r=0.56); TFI's psychological domain with SF-12 mental component (r=-0.55), mental well-being (r=-0.65); TFI's social domain with loneliness score (r=0.58) Results in line with 6 hypo's (6+)			
Gobbens et al., 2021	The Netherlands (Dutch)				479	Very good	Predictive validity (7 yr) for mortality: overall score (AUC 0.70, 95% CI, 0.65 to 0.76), physical domain (AUC 0.70, 95% CI, 0.65 to 0.75), psychological domain (AUC 0.65, 95% CI, 0.60 to 0.71), and social domain (AUC 0.64, 95% CI 0.59 to 0.69). Results in line with 3 hypo's (3+) Result not in line with 1 hypo (1-)						

AUC, Area under the curve; ADL, activities of daily living; APGAR, Adaptability, Partnership, Growth, Affection and Resolve; BMI, Body Mass Index; CES-D, Center for Epidemiologic Studies Depression Scale; CGA, Comprehensive Geriatric Assessment; CHS, Cardiovascular Health Study; CI, Confidence Interval; COP, center of pressure; COPE, Coping Orientation to Problems Experienced; FI, Frailty Index; GAI, Geriatric Anxiety Inventory; GARS, Groningen Activity Restriction Scale; GFI, Groningen Frailty Indicator; GDS, Geriatric Depression Scale; GFI, Groningen Frailty Indicator; GFST, Gérontopôle Frailty Screening Tool; GP, general practitioner; HADS, Hospital Anxiety and Depression Scale; HCP, Healthcare professionals; IADL, instrumental activities of daily living; IPAQ, International Physical Activity Questionnaire, LoA, Limits of Agreement;; LAPAQ, LASA Physical Activity Questionnaire; LSNS, Lubben Social Network Scale; MAS, Mastery Scale; MIC, minimal important change; MMSE, Mini-Mental State Examination; OLS, One Leg Standing; OR, Odds Ratio; QoL, Quality of life; PHQ, Patient Health Questionnaire; RS-11, Resilience Scale; SFES, Short Falls Efficacy Scale; SFQ, Shortened Fatigue Questionnaire; SF, Short Form Health Survey; SPQ, Sherbrooke Postal Questionnaire; SPMSQ, Short Portable Mental Status Questionnaire; SPPB, Short Physical Performance Battery; SSL, Social Support List; SSRS, Social Support Rating Scale; SSSS, Social Support Satisfaction Scale; TFI, Tilburg Frailty Indicator; TUG, Timed Up and Go Test; VAS, Visual Analogue Scale; WHOQOL, World Health Organization Quality of Life Questionnaire

Title

The Tilburg Frailty Indicator: a psychometric systematic review

Authors

Juan-José Zamora-Sánchez[1,2] MSc, Ana-María Urpí-Fernández[1] PhD, Meritxell Sastre-Rus[3] PhD, Iris Lumillo-Gutiérrez[2,4] PhD, Vicente Gea-Caballero[5,6,7] PhD, Lina Jodar-Fernández[8] MSc, Iván Julián-Rochina[9,10] PhD, Edurne Zabaleta-del-Olmo[1,11,12] PhD

Affiliations

[1]Gerència Territorial de Barcelona, Institut Català de la Salut, Barcelona, Spain

[2]School of Nursing, Universitat de Barcelona, Barcelona, Spain

[3]Escola Universitària d'Infermeria Gimbernat, Universitat Autònoma de Barcelona, Sant Cugat del Vallès, Barcelona, Spain

[4]Chronic Disease Management Team, Baix Llobregat Centre Primary Care Service, Costa de Ponent Primary Care Directorate, Gerència Territorial Metropolitana Sud, Institut Català de la Salut, Cornellà de Llobregat, Spain

[5]Deanery, Faculty of Health Sciences, Valencian International University, Valencia, Spain

[6]Nursing School "La Fe", University of Valencia, Valencia, Spain

[7]Instituto de Investigación Sanitaria del Hospital Universitario La Paz (IdiPAZ), Madrid, Spain

[8]Montbaig Primary Care Centre, Delta Primary Care Service, Costa de Ponent Primary Care Directorate, Gerència Territorial Metropolitana Sud, Institut Català de la Salut, Viladecans, Spain

[9]Nursing Department, Universitat de València, Valencia, Spain

[10]Frailty and Cognitive Impairment Group (FROG), Universitat de València, Valencia, Spain

[11]Nursing Department, Faculty of Nursing, Universitat de Girona, Girona, Spain

[12]Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain

Corresponding author

Edurne Zabaleta-del-Olmo

Gran Via Corts Catalanes, 587 àtic

08007 Barcelona (Spain)

Phone: +34 934 824 105

emails: ezabaleta@idiapjgol.org; nieves.zabaleta@udg.edu

Additional emails and orcid ids

Juan-José Zamora-Sánchez email: jjzamora.bcn.ics@gencat.cat ORCID iD: https://orcid.org/0000-0002-3487-7789

Ana-María Urpí-Fernández email: amurpics@gmail.com ORCID iD: https://orcid.org/0000-0002-2446-1077

Meritxell Sastre-Rus email: meritxell.sastre@eug.es ORCID iD: https://orcid.org/0000-0002-7043-0706

Iris Lumillo-Gutiérrez email: irislumillo.bcn.ics@gencat.cat ORCID iD: https://orcid.org/0000-0002-4787-3939

Vicente Gea-Caballero email: vagea@universidadviu.com ORCID iD: https://orcid.org/0000-0001-8607-3195

Lina Jodar-Fernández email: ljodar@ambitcp.catsalut.net ORCID iD: https://orcid.org/0000-0003-4316-6724

Iván Julián-Rochina email: ivan.julian@uv.es ORCID iD: https://orcid.org/0000-0002-1782-0019

Edurne Zabaleta-del-Olmo ORCID iD: https://orcid.org/0000-0002-5072-8548