

Diagnostic Accuracy of the FRAIL Scale, Groningen Frailty Indicator, Tilburg Frailty Indicator, and PRISMA-7 for Frailty Screening Among Older Adults in Community Settings: A Systematic Review and Network Meta-Analysis

Nhi Thi Vo, MSN,^{1,2} ID Yu-Kang Tu, PhD,^{3,4} ID Kuan-Chia Lin, PhD,⁵ ID Hsiao-Yean Chiu, PhD,^{1,6} ID and Hui-Chuan Huang, PhD¹ ID

¹School of Nursing, College of Nursing, Taipei Medical University, Taipei, Taiwan, China.

²Faculty of Nursing, University of Medicine and Pharmacy, Hue University, Hue City, Vietnam.

³Institute of Epidemiology and Preventive Medicine, National Taiwan University, Taipei, Taiwan, China.

⁴Department of Dentistry, National Taiwan University Hospital, National Taiwan University, Taipei, Taiwan, China.

⁵Institute of Hospital and Health Care Administration, Community Medicine Research Center, Preventive Medicine Research Center, National Yang-Ming University, Taipei, Taiwan, China.

⁶Department of Nursing, Taipei Medical University Hospital, Taipei City, Taiwan, China.

*Address correspondence to: Hui-Chuan Huang, PhD. E-mail: huichuan@tmu.edu.tw

Decision Editor: Patricia C. Heyn, PhD, FGSA, FACRM

Abstract

Background and Objectives: This study aimed to investigate the diagnostic accuracy of four questionnaire-based tools (i.e., the FRAIL scale, Groningen Frailty Indicator [GFI], Tilburg Frailty Indicator [TFI], and PRISMA-7) for screening frailty in older adults.

Research Design and Methods: The 4 databases comprising the Cumulative Index to Nursing and Allied Health Literature, Embase, PubMed, and ProQuest were searched from inception to June 20, 2023. Study quality comprising risks of bias and applicability was assessed via a QUADAS-2 questionnaire. A bivariate network meta-analysis model and Youden's index were performed to identify the optimal tool and cutoff points.

Results: In total, 20 studies comprising 13 for FRAIL, 7 for GFI, 6 for TFI, and 5 for PRISMA-7 were included. Regarding study quality appraisal, all studies had high risks of bias for study quality assessment domains. Values of the pooled sensitivity of the FRAIL scale, GFI, TFI, and PRISMA-7 were 0.58, 0.74, 0.66, and 0.73, respectively. Values of the pooled specificity of the FRAIL scale, GFI, TFI, and PRISMA-7 were 0.92, 0.77, 0.84, and 0.86, respectively. The Youden's index was obtained for the FRAIL scale with a cutoff of 2 points (Youden's index = 0.65), indicating that the FRAIL scale with a cutoff of 2 points was the optimal tool for frailty screening in older adults.

Discussion and Implications: The FRAIL scale comprising 5 self-assessed items is a suitable tool for interview older adults for early frailty detection in community settings; it has the advantages of being short, simple, and easy to respond to.

Keywords: Multiple tests, Multiple thresholds, Sensitivity, Specificity

Frailty is a major health condition associated with aging. The estimated prevalences of pre-frailty and frailty among community-dwelling older adults, which classified participants as frail or pre-frail based on the level of reduced physical functions, were 7.4%–23.0% and 40.2%–48.1%, respectively (Coelho-Junior et al., 2020; Hajek et al., 2022; Kojima et al., 2017). Frailty is characterized by the loss of biological reserve and increased vulnerability to physical and/or psychological stress (Fried et al., 2001). Older adults with frailty may experience a decline in functional abilities; this increases the fall risk which may lead to hospitalization, institutionalization, and death (Chu et al., 2021; Lahousse et al., 2014; Vermeiren et al., 2016). Therefore, early frailty detection is crucial for providing appropriate geriatric care.

Using an optimal assessment method is crucial for detecting early frailty; however, there is no clear consensus on a gold standard for an operational definition of frailty (Abellan van Kan G et al., 2008; Rodríguez-Mañas, 2013). The common theoretical definition of frailty is a clinically recognizable state of increased vulnerability resulting from aging-associated declines in reserve and function across multiple physiologic systems (World Health Organization, 2017). Moreover, evidence on the frailty-related pathophysiology is a single physical frailty or multidimensional construct, but there is little consensus on the underlying domains (Arnadottir et al., 2020). To the present, two methods with biological underpinnings are the frailty phenotype (FP; proposed and validated by Fried and colleagues in the Cardiovascular

Received: March 21 2023; Editorial Decision Date: December 23 2023.

© The Author(s) 2024. Published by Oxford University Press on behalf of The Gerontological Society of America. All rights reserved. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Health Study) and the Frailty Index (FI; proposed and validated by Rockwood and colleagues in the Canadian Study of Health and Aging) which represent common operational definitions of frailty in older populations (Cesari et al., 2014). The FP derived from a phenotype model uses five items (i.e., weight loss, exhaustion, grip strength, physical activity, and gait velocity) measured through self-reported and physical performance examinations to assess the frailty status (Fried et al., 2001). The FI derived from a deficit accumulation model is used to assess frailty syndromes including symptoms, signs, functional impairments, and laboratory abnormalities (Mitnitski et al., 2001). In addition, a comprehensive geriatric assessment (CGA), a multidimensional assessment attributed to age-related health problems, is commonly used to identify and quantify frailty for developing a coordinated plan to improve treatment-related outcomes and quality of life (Lee et al., 2020). The three aforementioned methods which include self-reported items and physical examinations are valid tools for evaluating frailty risk. However, these assessments must be performed by trained healthcare professionals with specialized equipment (e.g., a hand-held dynamometer), sufficient space, and available time (Ambagtsheer et al., 2019; Dent et al., 2016); this may limit the practicality of such assessments in community environments. Therefore, it is imperative to investigate suitable frailty assessment tools for rapid initial screening in community settings and geriatric care centers.

Various screening tools have been developed for frailty detection in older adults (Dent et al., 2016; Panhwar et al., 2019). Among these, four subjective questionnaire-based tools (i.e., the FRAIL scale, Groningen Frailty Indicator [GFI], Tilburg Frailty Indicator [TFI], and PRISMA-7) are widely used for frailty surveys in older adults. The FRAIL scale measures five items: fatigue, resistance, ambulation, illness, and weight loss (Morley et al., 2012). PRISMA-7 measures seven items across four domains: age (>85 years), gender, social support, and daily activities (Raiche et al., 2008). The TFI and GFI measure 15 items across three (i.e., physical, psychological, and social) and four (i.e., physical, cognitive, social, and psychological) domains, respectively (Gobbens et al., 2010b; Peters et al., 2012). These four assessments collect self-reported data. But the advantages of these tools include a quick response (≤ 15 min), no space or equipment requirements (Dent et al., 2016), and acceptable validity and reliability (Dent et al., 2017; Gobbens et al., 2010b; Peters et al., 2012; Saenger et al., 2018; Yaman & Ünal, 2018), and so they can be used for quick surveys of older adults with frailty in community settings. However, the summarized frailty diagnostic accuracy of these four tools compared to different objective frailty assessments (e.g., FP, FI, and CGA) has not been sufficiently examined. Moreover, practical applications of such tools may be restricted by the lack of statistical evaluations and comparisons of their diagnostic accuracy. Therefore, a diagnostic meta-analysis for evaluating and comparing the diagnostic accuracy of these tools in relation to different objective frailty assessments (i.e., FP, FI, and CGA) is crucial.

This meta-analysis aimed to investigate the summarized sensitivity and specificity of the FRAIL scale, GFI, TFI, and PRISMA-7 and compare the diagnostic accuracies of the four tools. Comparison of the diagnostic accuracy of the four tools should enable the identification of the optimal tool for frailty screening of older adults in community settings.

Method

This meta-analysis was conducted in accordance with the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Statement for Diagnostic Test Accuracy Studies (McInnes et al., 2018). The study protocol was registered at PROSPERO (CRD42022328659).

Eligibility Criteria

Eligibility criteria were as follows: (a) older adults aged ≥ 60 years, (b) frailty screening using the FRAIL scale, GFI, TFI or PRISMA-7 against a standard reference (i.e., FP, FI, or CGA), (c) study conducted in community settings including nursing homes, daycare centers, or geriatric outpatient departments, and (d) the study design was an observational study (i.e., cross-sectional or case-control study). Detailed information on the four index scales is shown in [Supplementary Table 1](#).

Literature Search Strategy

According to Medical Subject Headings, the designated keyword combinations of ([“FRAIL Scale” OR “FRAIL Questionnaire”] OR [“PRISMA-7”] OR [“Groningen Frailty Indicator” OR “GFI”] OR [“Tilburg Frailty Indicator” OR “TFI”]) AND (“sensitivity” OR “specificity” OR “reliability” OR “validation” OR “cutoff value”) AND (“frailty” OR “frail”) AND (“older” OR “elderly” OR “older adults” OR “older people”) were used to identify eligible studies. The Cumulative Index to Nursing and Allied Health Literature, Embase, PubMed, and ProQuest databases were searched from inception to June 20, 2023. Reference lists of all relevant papers were searched to identify additional eligible studies. No limitations on publication date or language were applied. Two researchers independently searched for eligible articles and then screened their titles and abstracts. Any disagreement was resolved through discussion. [Supplementary Table 2](#) presents the detailed search strategy.

Data Extraction

Two reviewers independently extracted data using a designated data extraction form, including participant characteristics (i.e., age and percentage of women patients), study characteristics (i.e., author name, publication year, country, study design, sample size, cutoff point of the index test, and reference standard with the cutoff point used), and frailty diagnostic accuracy outcomes (i.e., sensitivity, specificity, and original statistics of true positives, false positives, true negatives, and false negatives). In case of insufficient or unclear data, the required data were obtained by contacting the original authors by email. Inconsistencies in data extraction were resolved through discussion.

Methodological Quality Assessment of the Included Studies

The methodological quality of the included studies was independently assessed by two reviewers using the revised Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2; Whiting et al., 2011). QUADAS-2 includes 14 signaling questions divided into four domains (patient selection, index test, reference standard, and flow and timing). Each domain assesses the risk of bias, and the first three domains assess the applicability concerns of review questions. The signaling question responses are *yes*, *no*, and *unclear*. Bias risk is

judged to be low if all signaling questions for a domain are answered with *yes*. The bias risk of a domain is deemed high or unclear if any signaling questions are answered with *no* or *unclear*. Quality assessment results were cross-checked by two reviewers, and any inconsistencies were resolved through discussion.

Data Analysis

The original data of true positives, false positives, true negatives, and false negatives and those calculated according to the reported sensitivity and specificity were used for the data analysis. For studies that adopted the index test against more than one reference standard, data of true positives, false positives, true negatives, and false negatives were divided up for a subsequent unit-of-analysis error (Higgins et al., 2019). Four steps were conducted to complete the data analysis after the extracted data were summarized into a 2×2 table. All analyses were performed using Stata (vers. 15.0; StataCorp LLC, College Station, TX) and WinBUGS 1.4.3. First, a bivariate statistical analysis with a random-effects model was used to estimate the pooled sensitivity and specificity across studies for the four tools. Pairs of sensitivity and specificity measures from each included study were analyzed while considering potential threshold effects and the correlation between sensitivity and specificity (Leeftang, 2014). Significant heterogeneity among included studies was analyzed and identified by $p < .10$ in Cochran's Q tests and an I^2 value of $>25\%$ (Cleophas & Zwinderman, 2009). Second, subgroup analyses for categorical variables (i.e., study setting, country, and type of standard reference) and meta-regression for continuous variables (e.g., age and percentage of female patients), were performed to explain the sources of observed heterogeneity among studies. The subgroup analysis was limited to groups represented by at least two studies to ensure that sufficient data were obtained for analysis. Third, a bivariate network meta-analysis model based on the WinBUGS code was performed to estimate the pooled sensitivity and specificity of the four tools with different cutoff points (Owen et al., 2018). The bivariate network meta-analysis model incorporated constraints on increasing test thresholds and accounted for correlations between multiple test accuracy measures from the same study; this ensured more rigorous absolute sensitivity and specificity measures of diagnostic tests for decision-making (Owen et al., 2018; Rücker, 2018). A Surface Under the Cumulative Ranking Curve (SUCRA) was calculated to assess the diagnostic accuracy of the scales. The SUCRA values range from 0 to 1, and a higher value indicates superior diagnostic accuracy (Salanti et al., 2022). Furthermore, Youden's index was calculated using the formula $J = (\text{sensitivity} + \text{specificity}) - 1$ to determine the diagnostic test and diagnostic threshold with the optimal balance between sensitivity and specificity (Schisterman et al., 2008). Maximization of Youden's index indicates the optimal test and diagnostic threshold for the diagnosis of a disease in a particular patient group (Steinhauser et al., 2016). Finally, Deeks' funnel plot asymmetry test, which is a regression test of the diagnostic log odds ratio against $1/\sqrt{\text{effective sample size}}$, was used to examine publication bias. A p value of $<.10$ represented significant asymmetry for the slope coefficient, indicating significant publication bias (Deeks et al., 2005).

Results

Study Selection

In total, 511 studies were identified. After removing duplicate and irrelevant studies through a title and abstract review, the full texts of 37 eligible studies were examined. Among these studies, five used unsuitable reference standards, three defined both frailty and pre-frailty as frailty cases, three provided insufficient information for data analysis after contacting the original authors, and eight were conducted in a hospital setting (Supplementary Table 3). After identifying two additional studies based on a review of the references of eligible studies, 20 studies were included in the meta-analysis for FRAIL ($n = 13$), GFI ($n = 7$), TFI ($n = 6$), and PRISMA-7 ($n = 5$). Figure 1 presents the detailed study selection process. The reference list of the included studies is shown in Supplementary Table 4.

Study Characteristics

Table 1 shows detailed characteristics of the included studies. Overall, the included sample size was 5,344 and the mean age was 72.1 years in FRAIL studies, and those were 3,480 and 72.0 years in GFI studies, 2,992 and 71.1 years in TFI studies, and 1,681 and 73.3 years in PRISMA-7 studies, respectively. Proportions of female participants in studies using the FRAIL scale, GFI, TFI, and PRISMA-7 were 62.8%, 66.6%, 64.4%, and 57.3%, respectively. Among the 20 studies, 13 studies examined one index test, three studies used two index tests, and four studies used three index tests. Regarding the reference standard, 13 studies used FP, two studies used FI, three studies used CGA, and two studies used both FP and FI. The cutoff point used for frailty definition was ≥ 3 in FP, ≥ 2 in CGA, and ≥ 0.25 or ≥ 0.35 in FI. Regarding the index test, various cutoff points were used in the FRAIL scale (≥ 1 , ≥ 2 , or ≥ 3), GFI (≥ 3 , ≥ 4 , or ≥ 5), TFI (≥ 4 , ≥ 4.5 , ≥ 5 , ≥ 6 , or ≥ 8), and PRISMA-7 (≥ 3). Regarding the country in which the study was conducted, 13 studies were conducted in developing countries (i.e., Brazil, China, Iran, Thailand, Turkey, Vietnam, Indonesia, and Saudi Arabia), and 7 studies were conducted in developed countries (i.e., Australia, Finland, Germany, The Netherlands, and the United Kingdom; International Monetary Fund, 2022).

Study Quality Appraisal

Included studies using the FRAIL scale, GFI, TFI, and PRISMA-7 had low bias risks in terms of applicability concerns, indicating the included studies met review questions. Regarding the bias risk assessment for FRAIL, all studies were rated as having low risk for the criterion of patient selection, indicating that a suitable study design and participant eligibility were used in the studies. Most studies were evaluated as having high risks for the criteria of the index test ($n = 11$) and reference standard ($n = 11$) since the sequence of interpretation between the index test and the reference standard was not clearly described. In the flow and timing domain, 10 studies were at high risk because those studies did not include all patients in the analysis or did not clearly describe the interval between the index test and reference standard (Supplementary Table 5).

In term of GFI, all studies were rated as having low risk for the criterion of patient selection, indicating that a suitable study design and participant eligibility were used in the study. Five studies in the index test domain and six studies in the reference standard domain were classified as high risk due to

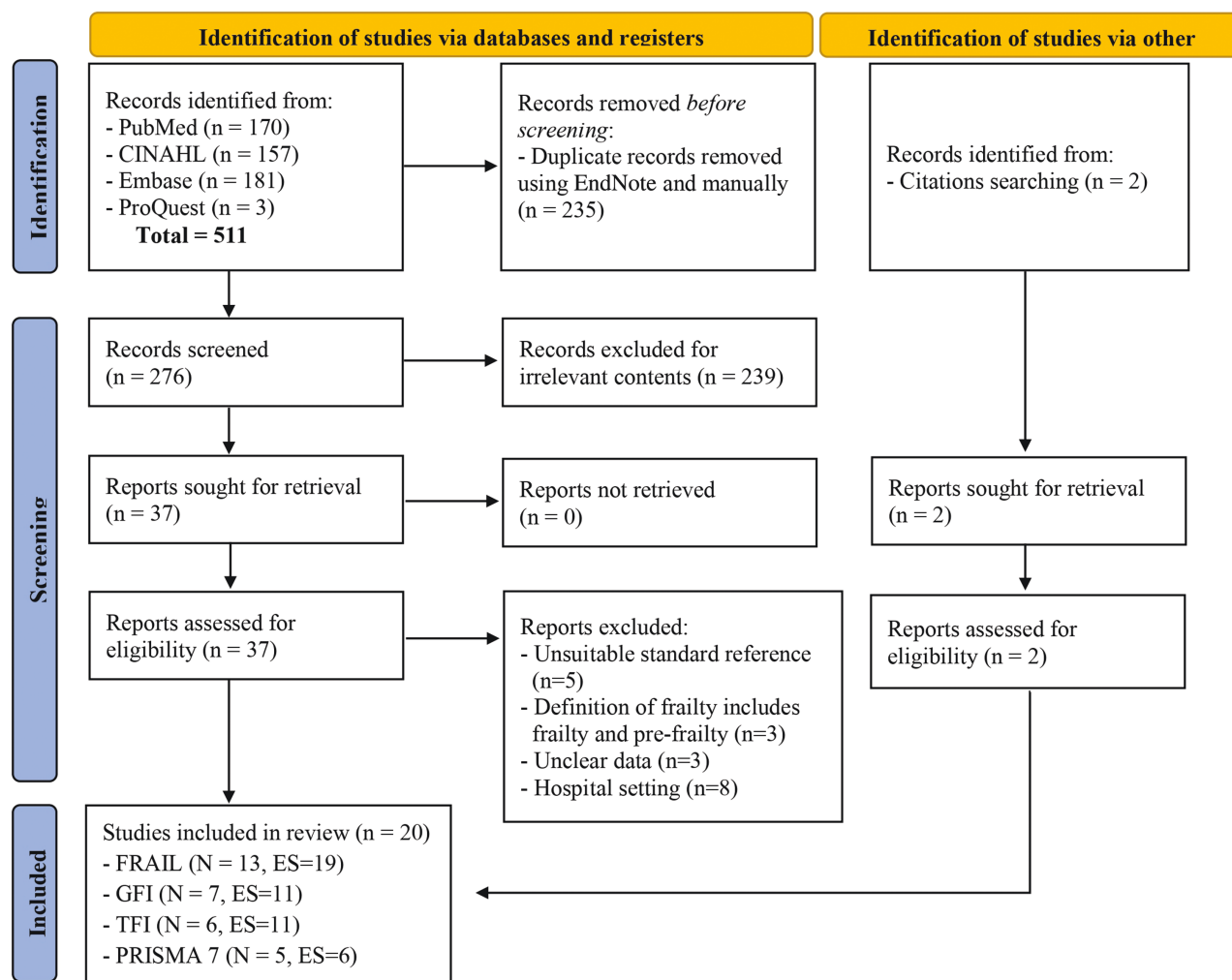


Figure 1. Process of systematic selection of articles using the PRISMA flow chart.

the absence of clear descriptions regarding the sequence of interpretation between the index test and reference standard. In the domain of flow and timing, six studies were categorized as high risk because they did not include all patients in the analysis or clearly delineate the interval between the index test and reference standard (Supplementary Table 6).

Among the six studies evaluating the diagnostic accuracy of the TFI, one study was rated as having a high risk for the criterion of patient selection since frail and nonfrail older adults were included at the beginning of the study. Additionally, six studies were deemed high risk in the criteria of the index test, and similarly, five studies in the reference standard domain, due to a lack of clear description regarding the sequence of interpretation between the index test and reference standard. In the flow and timing domain, all studies were at high risk because the studies did not include all patients in the analysis or did not clearly describe the interval between the index test and reference standard (Supplementary Table 7).

For PRISMA, all studies ($n = 5$) were evaluated as having low risk for the criterion of patient selection. However, three studies and four studies were rated as having a high risk for the criteria of the index test and reference standard, respectively, since the sequence of interpretation between the index test and reference standard was not clearly described. In the flow and timing domain, three studies were at high risk

because those studies did not include all patients in the analysis or did not clearly describe the interval between the index test and reference standard (Supplementary Table 8).

Overall, no study was rated as having low bias risk in all four domains, indicating that the diagnostic accuracy of the four scales could have been affected by the study quality.

Overall Sensitivity and Specificity of the FRAIL Scale, GFI, TFI, and PRISMA-7

Among the 13 studies that evaluated the FRAIL scale, one study had two reference standards and five studies examined two cut-points (i.e., ≥ 2 and ≥ 3); thus 19 data were used for the data analysis. A random-effects model indicated that the pooled sensitivities and specificities of the FRAIL scale were 0.58 and 0.92, respectively (Supplementary Figure 1). Heterogeneity tests showed that the Q statistic and I^2 index indicated substantial heterogeneity in the pooled sensitivity ($Q = 578.35, p < .001, 96.9\%$) and specificity ($Q = 1,652.47, p < .001, 98.3\%$).

In terms of the GFI, seven studies and 11 data were used for the data analysis since two reference standards were adopted in one study and three studies examined two cut-points (i.e., ≥ 3 vs ≥ 4 or ≥ 4 vs ≥ 5). The pooled sensitivities and specificities of the GFI were 0.74 and 0.77, respectively (Supplementary

Table 1. Study Characteristic

Author, year	Study design	Study setting	Cut-point of index test	Reference standard and cut-point	Country	Sample size	Mean age	Women (%)
FRAIL								
Alqahtani and Nasser (2019)	Cross-sectional	Community	≥ 3	FP ≥ 3	Saudi Arabia	47	70.0	34.0
Ambagtsheer et al. (2020)	Cross-sectional	Community	≥ 3	FP ≥ 3	Australia	228	79.0*	54.8
Anh et al. (2022)	Cross-sectional	Geriatric clinic	$\geq 2, \geq 3$	FP ≥ 3	Vietnam	396	72.6	62.1
Aprahamian et al. (2017)	Cross-sectional	Community	≥ 3	FP ≥ 3	Brazil	123	78.6	83.0
Braun et al. (2018)	Cross-sectional	Community	≥ 3	FP ≥ 3 , FI ≥ 0.25	Germany	52	73.0	63.0
Dong et al., 2018	Cross-sectional	Community	$\geq 2, \geq 3$	FP ≥ 3	China	1,188	69.5	69.4
Dwipa et al. (2021)	Cross-sectional	Geriatric clinic	$\geq 2, \geq 3$	FP ≥ 3	Indonesia	101	79.0	43.6
Ibrahim et al. (2022)	Cross-sectional	Community	≥ 3	FP ≥ 3	UK	98	75.0	80.0
Mijnarends et al. (2015)	Cross-sectional	Community	≥ 3	FP ≥ 3	Netherlands	227	74.9	48.5
Salminen et al. (2020)	Cross-sectional	Community	≥ 3	FI ≥ 0.25	Finland	1,188	72.7	57.0
Si et al. (2020)	Cross-sectional	Nursing home	$\geq 2, \geq 3$	CGA ≥ 2	China	305	79.3	57.0
Si et al. (2021)	Cross-sectional	Community	$\geq 1, \geq 3$	CGA ≥ 2	China	1,177	69.3	69.7
Sukkriang and Punsawad (2020)	Cross-sectional	Community	≥ 3	FP ≥ 3	Thailand	214	69.7	50.0
GFI								
Ambagtsheer et al. (2020)	Cross-sectional	Community	≥ 4	FP ≥ 3	Australia	228	79.0*	54.8
Braun et al. (2018)	Cross-sectional	Community	≥ 4	FP ≥ 3 , FI ≥ 0.25	Germany	52	73.0	63.0
Huang et al. (2022)	Cross-sectional	Community	≥ 3	FP ≥ 3	China	350	75.3	68.6
Si et al. (2020)	Cross-sectional	Nursing home	≥ 4	CGA ≥ 2	China	305	79.3	57.0
Si et al. (2021)	Cross-sectional	Community	$\geq 3, \geq 4$	CGA ≥ 2	China	1177	69.3	69.7
Tian et al. (2020)	Cross-sectional	Community	$\geq 3, \geq 4$	FI ≥ 0.25	China	1178	69.5	69.5
Xiang et al. (2020)	Cross-sectional	Nursing home	$\geq 4, \geq 5$	FP ≥ 3	China	190	77.7	56.8
TFI								
Coelho et al. (2015)	Cross-sectional	Community	$\geq 5, \geq 6$	FP ≥ 3	Netherlands	252	79.2	75.8
Dong et al. (2017)	Cross-sectional	Community	$\geq 4, \geq 5$	FP ≥ 3 , FI ≥ 0.35	China	917	68.6	63.8
Mazoochi et al. (2020)	Cross-sectional	Community	≥ 4.5	CGA ≥ 2	Iran	175	68.5	25.7
Ozsoy et al. (2021)	Cross-sectional	Community	≥ 8	FP ≥ 3	Turkey	166	73.0	67.5
Si et al. (2020)	Cross-sectional	Nursing home	≥ 5	CGA ≥ 2	China	305	79.3	57.0
Si et al. (2021)	Cross-sectional	Community	$\geq 4, \geq 5$	CGA ≥ 2	China	1,177	69.3	69.7
PRISMA-7								
Ambagtsheer et al. (2020)	Cross-sectional	Community	≥ 3	FP ≥ 3	Australia	228	79.0*	54.8
Braun et al. (2018)	Cross-sectional	Community	≥ 3	FP ≥ 3 , FI ≥ 0.25	Germany	52	73.0	63.0
Ibrahim et al. (2022)	Cross-sectional	Community	≥ 3	FP ≥ 3	UK	97	75.0	80.0
Salminen et al. (2020)	Cross-sectional	Community	≥ 3	FI ≥ 0.25	Finland	1,090	72.7	57.0
Sukkriang and Punsawad (2020)	Cross-sectional	Community	≥ 3	FP ≥ 3	Thailand	214	69.7	50.0

Notes: CGA = Comprehensive geriatric assessment; FI = Frailty index; FP = Fried's Frailty phenotype; GFI = Groningen *Frailty* Index; PRISMA = Program of Research on Integration of Services for the Maintenance of Autonomy; TFI = Tilburg *Frailty* Indicator. Prospective study shows that all participants were recruited to complete frailty assessments (i.e., FRAIL, PRISMA-7, and FI) at the beginning and then received a 10- and 18-year follow-up to examine the relationship between frailty and mortality risk.

*Convert from median into mean age.

Figure 2). Heterogeneity tests showed that the Q statistic and I^2 index indicated substantial heterogeneity in the pooled sensitivity ($Q = 52.05$, $p < .001$, 80.8%) and specificity ($Q = 105.82$, $p < .001$, 90.6%).

In total, six studies and 11 data comprising one study had two reference standards and three studies used two cut-points (i.e., ≥ 3 vs ≥ 4 or ≥ 4 vs ≥ 5) to examine the diagnostic accuracy of the TFI. Results showed that the pooled sensitivities and specificities of the TFI were 0.66 and 0.84, respectively (Supplementary Figure 3). Heterogeneity tests showed that the Q statistic and I^2 index indicated substantial heterogeneity in the pooled sensitivity ($Q = 62.16$, $p < .001$, 83.9%) and specificity ($Q = 192.71$, $p < .001$, 94.8%).

In terms of PRISMA-7, five studies and six data due to two reference standards from one study were used in data analysis. Results showed that the pooled sensitivities and specificities of the PRISMA-7 were 0.73 and 0.86, respectively (Supplementary Figure 4). Heterogeneity tests showed that the Q statistic and I^2 index indicated substantial heterogeneity in the pooled sensitivity ($Q = 18.52$, $p < .001$, 73.0%) and specificity ($Q = 61.19$, $p < .001$, 91.8%).

Overall, results indicated that the four tools had high summarized specificity but low summarized sensitivity (Table 2). Heterogeneity tests among these studies investigating the four scales were statistically significant (p value in Q test < 0.001 ; $I^2 > 80\%$). Therefore, further subgroup analysis and meta-regression analyses were performed to examine potential factors explaining the heterogeneity among these studies.

Diagnostic Accuracy Ranking

Supplementary Figure 5 shows the network diagram illustrating the relationship between the four scales and the reference standard. Concerning diagnostic accuracy, the SUCRA analysis identified that the FRAIL scale, with a cut-point of 1, ranked highest (82.3%) for sensitivity (Supplementary Figure 6), whereas FRAIL with a cut-point of 3 achieved the top rank (82.3%) for specificity (Supplementary Figure 7). Taking into account the balance between accurately identifying true positives and true negatives in the assessment of frailty in older adults, FRAIL with a cutoff point of 2 emerged as the leading candidate with a SUCRA value of 87.4% (Supplementary Figure 8).

Furthermore, Table 3 shows the diagnostic accuracy ranking of the four tools, considering the respective cutoff points. The findings highlighted that the FRAIL scale, utilizing a cutoff point of 2, exhibited the highest Youden's index at 0.65, accompanied by a sensitivity and specificity of 0.82 each. These results underscore the FRAIL scale, employing a cutoff point of 2, as the optimal tool for frailty screening in older adults.

Subgroup Analysis and Meta-Regression

Results of the subgroup analysis and meta-regression analyses for the four tools are shown in Table 4. For the FRAIL scale, the reference standard and country were significant moderators. Studies conducted in developed countries demonstrated a higher summarized specificity but lower summarized sensitivity compared to those conducted in developing countries (0.84 vs 0.35, $p < .0001$ and 0.66 vs 0.87, $p = .018$). Diagnostic accuracy studies of the FRAIL scale using FI as a reference standard had a lower pooled sensitivity but higher pooled specificity than those using FP as a reference standard (0.15 vs 0.64; $p = .001$ and 0.99 vs 0.90; $p = .001$).

For the GFI, the women percentage was positively correlated with the pooled specificity ($\beta = .023$; $p = .018$). Studies with a sample size of ≥ 300 had a higher pooled specificity than those with a sample size of < 300 (0.72 vs 0.65, $p = .009$).

For the TFI, the female percentage, sample size, country, and reference standard used were significant moderators. The female percentage was negatively correlated with the pooled sensitivity ($\beta = -0.022$; $p = .005$) and pooled specificity ($\beta = -0.027$; $p < .001$). Studies with a sample size of ≥ 300 had a lower pooled sensitivity but higher pooled specificity than those with a sample size of < 300 (0.81 vs 0.93, $p = .009$ and 0.74 vs 0.53, $p = .002$). Studies conducted in developed countries demonstrated a lower summarized specificity than those conducted in developing countries (0.86 vs 0.95, $p = .001$). Studies using CGA as a reference standard had a higher pooled specificity (0.88 vs 0.76; $p = .038$) than those using FP as a reference standard.

For PRISMA-7, the reference standard was correlated with the pooled sensitivity and specificity. Studies using FI as a reference standard had a lower pooled sensitivity (0.82 vs 0.94; $p = .034$) than those using FP as a reference standard.

Table 2. Overall Sensitivity and Specificity Among the Four Subjective Frailty Scales

Test	N	ES	Sensitivity				Specificity			
			Pooled sensitivity	Q test	p Value	I^2	Pooled specificity	Q test	p Value	I^2
FRAIL	13	19	0.58 (0.44–0.71)	578.35	<.001	96.9	0.92 (0.86–0.96)	1652.47	<.001	98.3
GFI	7	11	0.74 (0.68–0.79)	52.05	<.001	80.8	0.77 (0.73–0.81)	105.82	<.001	90.6
TFI	6	11	0.66 (0.58–0.74)	62.16	<.001	83.9	0.84 (0.78–0.89)	192.71	<.001	94.8
PRISMA-7	5	6	0.73 (0.61–0.82)	18.52	<.001	73.0	0.86 (0.79–0.91)	61.19	<.001	91.8

Note: ES = effect size; GFI = Groningen Frailty Index; I^2 = I-square test; N = number of studies; PRISMA = Program of Research on Integration of Services for the Maintenance of Autonomy; Q test = Cochran's Q statistic; TFI = Tilburg Frailty indicator.

Table 3. Ranking of Frailty Screening Scales Based on Youden's Index

Index	Cutoff	Sensitivity (95% CI)	Specificity (95% CI)	Youden index (95% CI)	Ranking
FRAIL	1	0.91 (0.83–0.95)	0.46 (0.29–0.67)	0.37 (0.19–0.57)	11
FRAIL	2	0.82 (0.73–0.90)	0.82 (0.72–0.90)	0.65 (0.52–0.75)	1
FRAIL	3	0.46 (0.34–0.58)	0.94 (0.91–0.97)	0.40 (0.28–0.53)	9
GFI	3	0.87 (0.75–0.93)	0.64 (0.46–0.78)	0.50 (0.32–0.67)	7
GFI	4	0.77 (0.61–0.85)	0.78 (0.65–0.88)	0.54 (0.37–0.69)	5
GFI	5	0.57 (0.33–0.76)	0.90 (0.77–0.96)	0.45 (0.22–0.66)	8
TFI	4	0.82 (0.69–0.90)	0.76 (0.58–0.86)	0.57 (0.35–0.70)	3
TFI	5	0.71 (0.58–0.82)	0.85 (0.72–0.91)	0.56 (0.38–0.68)	4
TFI	6	0.63 (0.44–0.78)	0.89 (0.78–0.94)	0.51 (0.30–0.67)	6
TFI	8	0.44 (0.13–0.69)	0.92 (0.84–0.97)	0.36 (0.04–0.61)	10
PRISMA-7	3	0.80 (0.64–0.90)	0.83 (0.68–0.92)	0.63 (0.40–0.77)	2

Note: GFI = Groningen Frailty Index; PRISMA = Program of Research on Integration of Services for the Maintenance of Autonomy; TFI = Tilburg Frailty Indicator.

However, studies using FI as a reference standard had a higher pooled specificity (0.83 vs 0.59; $p = .024$) than those using FP as a reference standard.

Publication Bias

Deeks' funnel plot asymmetry test indicated no significant publication bias in studies using the FRAIL scale, GFI, TFI, or PRISMA-7 ($p > .10$; [Supplementary Table 9](#)).

Discussion

This meta-analysis confirmed that the FRAIL scale with a cut-off point of 2 is the optimal screening tool for frailty detection in older adults. This is the first diagnostic meta-analysis of numerous studies to use rigorous methodological quality assessments to examine the diagnostic properties of four subjective frailty assessment tools. Thus, the results reliably supported the diagnostic accuracy of these four tools. The use of the FRAIL scale is recommended for rapid frailty detection in older adults in community settings and geriatric care units.

The overall sensitivity and specificity indicated that the four tools had high specificity (77%–92%) but low sensitivity (58%–74%). A tradeoff exists between the sensitivity and specificity of diagnostic and screening tests. Higher sensitivity corresponds to lower specificity and vice versa. A screening tool with low sensitivity results in more false negatives, and one with high specificity results in fewer false positives ([Treveltham, 2017](#)). Therefore, frail cases identified using these four tools can be confirmed to be frail, but the presence of frailty should not be ruled out for non-frail cases. Because screening tools with low sensitivity may be compensated by frequent screening ([Swift et al., 2020](#)), the sensitivity of using these four subjective tools to detect frailty in older adults can be improved by increasing the survey frequency.

Among the four scales with various cutoff points, results showed that the FRAIL scale with a cutoff point of 2 was the optimal tool for frailty detection with 82% sensitivity and 82% specificity. The FRAIL scale with an original cut-point of 3 showed high specificity (94%) but low sensitivity (46%). A test with 100% sensitivity would be perfect for positive cases, equivalent to a random draw when sensitivity is 50%. A sensitivity of 46% indicated that less than half of frail cases could be detected. A scale with low sensitivity might lead to a high

false negative rate and reduce the precision ([Treveltham, 2017](#)). Thus, applying FRAIL with a cut-point of 2 would increase the precision for frailty detection in older adults. Besides, the decline in physiological reserves is antecedent to physical, psychological, and social frailty ([Gobbens et al., 2010a](#)). One longitudinal study reported that weakness, weight loss, and exhaustion can be used to identify cases with rapid frailty progression ([Xue et al., 2008](#)). Because the FRAIL scale contains items related to weakness, exhaustion, and weight loss, it may provide superior frailty detection compared to the GFI, TFI, and PRISMA-7, which contain items related to physical, psychological, and social frailty. The brief FRAIL scale with a cut-point of 2 can be regularly used to identify suspected early frailty in older adults in the community or geriatric care units; they should then be provided with further examinations and timely management.

A reference standard is used to determine the diagnostic accuracy of an index test ([Santini et al., 2021](#)). In this meta-analysis, subgroup analyses revealed inferior summarized sensitivity in studies using FI as a reference standard compared to the FRAIL scale (15% vs 64%) and PRISMA-7 (82% vs 94%) relative to studies using FP as a reference standard. A possible explanation is that items of the FRAIL scale and PRISMA-7 were dissimilar to those of FI. FI with various items (e.g., 36–40 items) identifies frailty based on accumulating deficits in multidimensional domains such as physical function, multimorbidities, cognition, and psychosocial factors ([Searle et al., 2008](#)). The reduced-item FRAIL scale (five items) and PRISMA-7 (seven items) may insufficiently represent the case definition of FI, leading to decreased summarized sensitivity and increased frailty false negatives. Because FP, FI, and CGA have been recommended as validated comprehensive measurement methods for frailty, item feature consistency should be considered when selecting a reference standard to examine the diagnostic accuracy of an index test. Because a small number of studies applying FI and CGA as reference standards were included in this meta-analysis, more studies should be conducted to examine the diagnostic accuracy of various combinations by pairing different index texts and validated measurement methods.

Results of the present study indicated that country and women percentage were significant moderators of summarized sensitivity and specificity for the FRAIL and TFI scales.

Table 4. Subgroup Analysis and Meta-regression

	ES	β /Sensitivity	<i>p</i> Value	β /Specificity	<i>p</i> Value
FRAIL					
Age	19	−0.062 (−0.218, 0.093)	.434	−0.013 (−0.229, 0.203)	.904
Female percentage	19	−0.018 (−0.063, 0.027)	.430	0.014 (−0.047, 0.075)	.644
Prevalence of frailty	19	−0.800 (−4.104, 2.504)	.635	−0.637 (−5.072, 3.799)	.779
Sample size					
<300	10	0.63	Ref.	0.84	Ref.
≥300	9	0.57	.673	0.91	.495
Reference standard					
FP	13	0.64	Ref.	0.90	Ref.
FI	2	0.15	.001	0.99	.001
CGA	4	0.49	.224	0.84	.293
Country					
Developing	13	0.87	Ref.	0.35	Ref.
Developed	6	0.66	.018	0.84	<.001
Risk of bias					
Index test					
Low risk	5	0.34	Ref.	0.87	Ref.
High risk	14	0.45	.527	0.90	.743
Reference standard					
Low risk	3	0.43	Ref.	0.80	Ref.
High risk	16	0.49	.756	0.88	.603
Flow and timing					
Low risk	6	0.92	Ref.	0.77	Ref.
High risk	13	0.76	.026	0.88	.397
GFI					
Age	11	−0.001 (−0.069, 0.066)	.968	−0.026 (−0.057, 0.004)	.094
Female percentage	11	0.023 (−0.019, 0.066)	.285	0.023 (0.004, 0.041)	.018
Prevalence of frailty	11	0.185 (−1.204, 1.573)	.794	−0.303 (−1.115, 0.510)	.465
Sample size					
<300	5	0.67	Ref.	0.65	Ref.
≥300	6	0.71	.511	0.72	.009
Reference standard					
FP	5	0.79	Ref.	0.74	Ref.
FI	3	0.72	.226	0.79	.140
CGA	3	0.70	.090	0.77	.408
Country					
Developing	8	0.82	Ref.	0.81	Ref.
Developed	3	0.75	.324	0.78	.247
Risk of bias					
Index test					
Low risk	3	0.56	Ref.	0.70	Ref.
High risk	8	0.66	.324	0.74	.247
Flow and timing					
Low risk	2	0.39	Ref.	0.87	Ref.
High risk	9	0.57	.375	0.82	.420
TFI					
Age	11	−0.031 (−0.154, 0.091)	.619	−0.061 (−0.152, 0.031)	.192
Female percentage	11	−0.022 (−0.038, −0.007)	.005	−0.027 (−0.038, −0.016)	<.001
Prevalence of frailty	11	−0.562 (−3.099, 1.976)	.664	0.537 (−1.981, 3.056)	.676
Sample size					
<300	4	0.93	Ref.	0.53	Ref.
≥300	7	0.81	.009	0.74	.002

Table 4. Continued

	ES	β /Sensitivity	<i>p</i> Value	β /Specificity	<i>p</i> Value
Reference standard					
FP	5	0.68	Ref	0.76	Ref
FI	2	0.64	.669	0.79	.261
CGA	4	0.71	.857	0.88	.038
Country					
Developing	9	0.59	Ref.	0.95	Ref
Developed	2	0.67	.643	0.86	.001
PRISMA-7					
Age	6	0.036 (−0.238, 0.310)	.799	−0.111 (−0.326, 0.104)	.312
Female percentage	6	0.059 (−0.008, 0.125)	.086	−0.014 (−0.039, 0.012)	.297
Prevalence of frailty	6	−7.008 (−15.766, 1.751)	.117	4.420 (−2.384, 11.224)	.203
Reference standard					
FP	4	0.94	Ref.	0.59	Ref.
FI	2	0.82	.034	0.83	.024
Risk of bias					
Index test					
Low risk	3	0.83	Ref.	0.65	Ref.
High risk	3	0.75	.341	0.80	.278
Flow and timing					
Low risk	3	0.79	Ref.	0.86	Ref.
High risk	3	0.76	.844	0.86	.973

Note: CGA = Comprehensive geriatric assessment; FI = Frailty index; FP = Fried's Frailty phenotype; GFI = Groningen Frailty Index; PRISMA = Program of Research on Integration of Services for the Maintenance of Autonomy; Ref = Reference; TFI = Tilburg Frailty Indicator.

A spectrum effect derived from inherent variations in a study population, such as setting and patient characteristics (e.g., women), may affect a test's sensitivity and specificity (Mulherin & Miller, 2002). Women may have more frailty caused by greater levels of disability, more comorbidities, and poorer self-rated health, marital status, education level, and social support but they may tolerate this frailty better which reduces the impact on their daily activities (Gordon et al., 2017; Zhang et al., 2018). These variations in demographic characteristics among women may affect the diagnostic accuracy of the TFI scale. In addition, differences in frailty prevalence between different countries may affect the frailty diagnostic accuracy (Harttgen et al., 2013). High disease prevalence tends to be associated with high sensitivity and low specificity for tests under evaluation (Leeftang et al., 2013). Because the included studies did not report the population-level frailty prevalence, the effect of frailty prevalence in different countries on the diagnostic accuracy of the TFI and FRAIL scale could not be fully examined. Because no study examined the diagnostic accuracy of PRISMA-7 in developing countries, more well-designed studies should be conducted to clarify the effect of frailty prevalence in different countries on the diagnostic accuracy of the four tools.

A sample size of ≥ 300 is sufficiently large to evaluate both the sensitivity and specificity of screening and diagnostic tests (Bujang & Adnan, 2016). Based on a rule-of-thumb, our findings revealed that studies with a sample size of ≥ 300 had more stable summarized sensitivity and specificity for the TFI (81% vs 74%) than those with a sample size of < 300 (93% vs 53%). A sufficient sample size may provide more precise estimates with small confidence intervals, ensuring adequate power for evaluating screening or diagnostic tests

(Bachmann et al., 2006; Bujang & Adnan, 2016; Hajian-Tilaki, 2014). Thus, the summarized sensitivity and specificity of the TFI in studies with large sample sizes were more reliable. Because studies with large-sample studies (≥ 300) examining the diagnostic accuracy of the FRAIL scale, GFI, and PRISMA-7 did not significantly differ compared to smaller studies, more studies with large samples (≥ 300) should be conducted to re-examine the diagnostic accuracy.

This meta-analysis used rigorous criteria to include eligible studies, including greater sample sizes, and the quality of the included studies was rigorously assessed, which increased the internal validity. Eligible studies were conducted in various countries, which increased the generalizability of the findings. However, some limitations must be considered when interpreting the findings. First, this review only included studies that recruited individuals aged > 60 years; thus, the generalizability of these findings to other age groups may be limited. Second, older adults with comorbidities (e.g., diabetes, osteoarthritis, or heart disease) could increase the frailty risk which may have affected the diagnostic accuracy of frailty assessment tools. However, the effect of comorbidities on the diagnostic accuracy of the four scales could not be further analyzed due to insufficient data reported in the included studies. Further studies could be considered to clarify this issue. Third, although no language restrictions were used when searching for eligible studies, only studies published in the English language were included in this meta-analysis. A language bias on the resulting estimations should be considered. Fourth, this study only examined the effects of the sample-level frailty prevalence on summarized sensitivity and specificity because the included studies did not provide population-level frailty prevalences. The results

should be interpreted with caution. Fifth, possible moderators related to frailty, such as the nutrition status, comorbidities, and decrease in functional capacity (Fhon et al., 2018; Lorenzo-López et al., 2017), that could explain the heterogeneity among the studies could not be further analyzed because of insufficient data. Finally, only a few studies reported the diagnostic accuracy of the FRAIL scale with a cutoff point of 1 and the TFI with cutoff points of ≥ 6 or ≥ 8 . Future investigations should validate the diagnostic accuracy of these tools with specific cutoff points.

Conclusion

This is the first meta-analysis to examine and compare the summarized sensitivity and specificity of the FRAIL scale, GFI, TFI, and PRISMA-7. Results indicated that these four subjective frailty screening tools had high specificity (77%–92%) but low sensitivity (58%–74%). The FRAIL scale with a cutoff point of 2 and with acceptable sensitivity (82%) and specificity (82%) was the optimal screening tool for detecting frailty in community-dwelling older adults. The FRAIL scale with only five self-reported items has the advantages of accessibility, simplicity, convenience, low time commitment, and less physical discomfort; it is ideal for use in the community or geriatric care units for the rapid identification of frailty among older adults. Regular use of the FRAIL scale can be incorporated into the frailty screening of older adults. Healthcare professionals (e.g., nurses, geriatricians, physical therapists, and geriatric care managers) can use the FRAIL scale to interview older adults for early frailty detection in low-resource settings for physical performance tests and comprehensive frailty assessments.

Supplementary Material

Supplementary data are available at *The Gerontologist* online.

Funding

This study was supported by the Ministry of Science and Technology (MOST 112-2314-B-038-071-MY2).

Conflict of Interest

All authors have no conflicts of interest and source of funding to declare.

Data Availability

The original contributions presented in the study are included in the article and the [Supplementary Table 10](#); further inquiries can be directed to the corresponding author. The study protocol was registered at PROSPERO (CRD42022328659).

References

Abellan van Kan, G., Rolland, Y., Bergman, H., Morley, J. E., Kritchevsky, S. B., & Vellas, B. (2008). The I.A.N.A Task Force on frailty assessment of older people in clinical practice. *Journal of Nutrition, Health & Aging*, 12(1), 29–37. <https://doi.org/10.1007/BF02982161>

- Alqahtani, B. A., & Nasser, T. A. (2019). Assessment of frailty in Saudi community-dwelling older adults: Validation of measurements. *Annals of Saudi Medicine*, 39(3), 197–204. <https://doi.org/10.5144/0256-4947.2019.197>
- Ambagtsheer, R. C., Beilby, J. J., Visvanathan, R., Dent, E., Yu, S., & Braunack-Mayer, A. J. (2019). Should we screen for frailty in primary care settings? A fresh perspective on the frailty evidence base: A narrative review. *Preventive Medicine*, 119, 63–69. <https://doi.org/10.1016/j.ypmed.2018.12.020>
- Ambagtsheer, R. C., Visvanathan, R., Dent, E., Yu, S., Schultz, T. J., & Beilby, J. (2020). Commonly used screening instruments to identify frailty among community-dwelling older people in a general practice (primary care) setting: A study of diagnostic test accuracy. *Journals of Gerontology, Series A: Biological Sciences and Medical Sciences*, 75(6), 1134–1142. <https://doi.org/10.1093/geronag/75.6.1134>
- Anh, D., Nguyen, T., Nguyen, T., & Nguyen, T. V. (2022). The validity of the FRAIL scale in frailty screening among Vietnamese older people. *Aging Medicine and Healthcare*, 13(2), 87–92. <https://doi.org/10.33879/AMH.132.2021.07060>
- Aprahamian, I., Cezar, N. O. C., Izbicki, R., Lin, S. M., Paulo, D. L. V., Fattori, A., Biella, M. M., Jacob Filho, W., & Yassuda, M. S. (2017). Screening for frailty with the FRAIL scale: A comparison with the phenotype criteria. *Journal of the American Medical Directors Association*, 18(7), 592–596. <https://doi.org/10.1016/j.jamda.2017.01.009>
- Arnadottir, S. A., Bruce, J., Lall, R., Withers, E. J., Underwood, M., Shaw, F., Sheridan, R., Hossain, A., & Lamb, S. E.; Pre-FIT Study Group. (2020). The importance of different frailty domains in a population based sample in England. *BMC Geriatrics*, 20(1), 16. <https://doi.org/10.1186/s12877-019-1411-9>
- Bachmann, L. M., Puhon, M. A., ter Riet, G., & Bossuyt, P. M. (2006). Sample sizes of studies on diagnostic accuracy: Literature survey. *British Medical Journal*, 332, 1127–1129. <https://doi.org/10.1136/bmj.38793.637789.2F>
- Braun, T., Grüneberg, C., & Thiel, C. (2018). German translation, cross-cultural adaptation and diagnostic test accuracy of three frailty screening tools: PRISMA-7, FRAIL scale and Groningen Frailty Indicator. *Zeitschrift für Gerontologie und Geriatrie*, 51(3), 282–292. <https://doi.org/10.1007/s00391-017-1295-2>
- Bujang, M. A., & Adnan, T. H. (2016). Requirements for minimum sample size for sensitivity and specificity analysis. *Journal of Clinical and Diagnostic Research*, 10, YE01–YE06. <https://doi.org/10.7860/JCDR/2016/18129.8744>
- Cesari, M., Gambassi, G., van Kan, G. A., & Vellas, B. (2014). The frailty phenotype and the frailty index: Different instruments for different purposes. *Age and Ageing*, 43(1), 10–12. <https://doi.org/10.1093/ageing/aft160>
- Chu, W., Chang, S. F., & Ho, H. Y. (2021). Adverse health effects of frailty: Systematic review and meta-analysis of middle-aged and older adults with implications for evidence-based practice. *Worldviews on Evidence-Based Nursing*, 18, 282–289. <https://doi.org/10.1111/wvn.12508>
- Cleophas, T. J., & Zwinderman, A. H. (2009). Meta-analyses of diagnostic studies. *Clinical Chemistry and Laboratory Medicine*, 47, 1351–1354. <https://doi.org/10.1515/CCLM.2009.317>
- Coelho, T., Santos, R., Paúl, C., Gobbens, R. J., & Fernandes, L. (2015). Portuguese version of the Tilburg Frailty Indicator: Transcultural adaptation and psychometric validation. *Geriatrics & Gerontology International*, 15(8), 951–960. <https://doi.org/10.1111/ggi.12373>
- Coelho-Junior, H. J., Marzetti, E., Picca, A., Calvani, R., Cesari, M., & Uchida, M. C. (2020). Prevalence of prefrailty and frailty in South America: A systematic review of observational studies. *Journal of Frailty & Aging*, 9, 197–213. <https://doi.org/10.14283/jfa.2020.22>
- Deeks, J. J., Macaskill, P., & Irwig, L. (2005). The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *Journal of Clinical Epidemiology*, 58, 882–893. <https://doi.org/10.1016/j.jclinepi.2005.01.016>

- Dent, E., Kowal, P., & Hoogendijk, E. O. (2016). Frailty measurement in research and clinical practice: A review. *European Journal of Internal Medicine*, 31, 3–10. <https://doi.org/10.1016/j.ejim.2016.03.007>
- Dent, E., Lien, C., Lim, W. S., Wong, W. C., Wong, C. H., Ng, T. P., Woo, J., Dong, B., de la Vega, S., Hua Poi, P. J., Kamaruzzaman, S. B. B., Won, C., Chen, L. K., Rockwood, K., Arai, H., Rodriguez-Mañas, L., Cao, L., Cesari, M., Chan, P., ... Flicker, L. (2017). The Asia-Pacific clinical practice guidelines for the management of frailty. *Journal of the American Medical Directors Association*, 18(7), 564–575. <https://doi.org/10.1016/j.jamda.2017.04.018>
- 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. *Archives of Gerontology and Geriatrics*, 73, 21–28. <https://doi.org/10.1016/j.archger.2017.07.001>
- Dong, L., Qiao, X., Tian, X., Liu, N., Jin, Y., Si, H., & Wang, C. (2018). Cross-cultural adaptation and validation of the FRAIL scale in Chinese community-dwelling older adults. *Journal of the American Medical Directors Association*, 19(1), 12–17. <https://doi.org/10.1016/j.jamda.2017.06.011>
- Dwipa, L., Apandi, M., Utomo, P. P., Hasmirani, M., Rakhimullah, A. B., & Yulianto, F. A. (2021). Adaptation and validation of the Indonesian version of the FRAIL scale and the SARC-F in older adults. *Asian Journal of Gerontology and Geriatrics*, 16(1), 40–47. <https://doi.org/10.12809/ajgg-2020-436-0a>
- Fhon, J. R. S., Rodrigues, R. A. P., Santos, J. L. F., Diniz, M. A., Santos, E. B. D., Almeida, V. C., & Giacomini, S. B. L. (2018). Factors associated with frailty in older adults: A longitudinal study. *Revista de Saude Publica*, 52, 74. <https://doi.org/10.11606/S1518-8787.2018052000497>
- Fried, L. P., Tangen, C. M., Walston, J., Newman, A. B., Hirsch, C., Gottdiener, J., Seeman, T., Tracy, R., Kop, W. J., Burke, G., & McBurnie, M. A.; Cardiovascular Health Study Collaborative Research Group (2001). Frailty in older adults: Evidence for a phenotype. *Journals of Gerontology, Series A: Biological Sciences and Medical Sciences*, 56, M146–M156. <https://doi.org/10.1093/gerona/56.3.m146>
- Gobbens, R. J., Luijckx, K. G., Wijnen-Sponselee, M. T., & Schols, J. M. (2010a). Towards an integral conceptual model of frailty. *Journal of Nutrition, Health & Aging*, 14, 175–181. <https://doi.org/10.1007/s12603-010-0045-6>
- Gobbens, R. J., van Assen, M. A., Luijckx, K. G., Wijnen-Sponselee, M. T., & Schols, J. M. (2010b). The Tilburg frailty indicator: Psychometric properties. *Journal of the American Medical Directors Association*, 11, 344–355. <https://doi.org/10.1016/j.jamda.2009.11.003>
- Gordon, E. H., Peel, N. M., Samanta, M., Theou, O., Howlett, S. E., & Hubbard, R. E. (2017). Sex differences in frailty: A systematic review and meta-analysis. *Experimental Gerontology*, 89, 30–40. <https://doi.org/10.1016/j.exger.2016.12.021>
- Hajek, A., Kretzler, B., & König, H. H. (2022). Prevalence of prefrailty and frailty among older adults in Germany: A systematic review, meta-analysis and meta-regression. *Frontiers in Medicine*, 9, 870714. <https://doi.org/10.3389/fmed.2022.870714>
- Hajian-Tilaki, K. (2014). Sample size estimation in diagnostic test studies of biomedical informatics. *Journal of Biomedical Informatics*, 48, 193–204. <https://doi.org/10.1016/j.jbi.2014.02.013>
- Harttgen, K., Kowal, P., Strulik, H., Chatterji, S., & Vollmer, S. (2013). Patterns of frailty in older adults: Comparing results from higher and lower income countries using the Survey of Health, Ageing and Retirement in Europe (SHARE) and the Study on Global AGEing and Adult Health (SAGE). *PLoS One*, 8, e75847. <https://doi.org/10.1371/journal.pone.0075847>
- Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2019). *Cochrane handbook for systematic reviews of interventions: Cochrane book series*. John Wiley & Sons.
- Huang, E. Y. Z., Cheung, J., Liu, J. Y. W., Kwan, R. Y. C., & Lam, S. C. (2022). Groningen Frailty Indicator-Chinese (GFI-C) for pre-frailty and frailty assessment among older people living in communities: Psychometric properties and diagnostic accuracy. *BMC Geriatrics*, 22(1), 788. <https://doi.org/10.1186/s12877-022-03437-1>
- Ibrahim, K., Mullee, M. A., Cox, N., Russell, C., Baxter, M., Tilley, S., Yao, G. L., Zhu, S., & Roberts, H. C. (2022). The feasibility and acceptability of assessing and managing sarcopenia and frailty among older people with upper limb fracture. *Age and Ageing*, 51(1), afab252. <https://doi.org/10.1093/ageing/afab252>
- International Monetary Fund. (2022). *World economic outlook*. IMF Publishing. <https://www.imf.org/en/Publications/WEO/weo-database/2022/October>
- Kojima, G., Iliffe, S., Taniguchi, Y., Shimada, H., Rakugi, H., & Walters, K. (2017). Prevalence of frailty in Japan: A systematic review and meta-analysis. *Journal of Epidemiology*, 27, 347–353. <https://doi.org/10.1016/j.je.2016.09.008>
- Lahousse, L., Maes, B., Ziere, G., Loth, D. W., Verlinden, V. J. A., Zillikens, M. C., Uitterlinden, A. G., Rivadeneira, F., Tiemeier, H., Franco, O. H., Ikram, M. A., Hofman, A., Brusselle, G. G., & Stricker, B. H. (2014). Adverse outcomes of frailty in the elderly: The Rotterdam Study. *European Journal of Epidemiology*, 29, 419–427. <https://doi.org/10.1007/s10654-014-9924-1>
- Lee, H., Lee, E., & Jang, I. Y. (2020). Frailty and comprehensive geriatric assessment. *Journal of Korean Medical Science*, 35, e16. <https://doi.org/10.3346/jkms.2020.35.e16>
- Leeftang, M. M. (2014). Systematic reviews and meta-analyses of diagnostic test accuracy. *Clinical Microbiology and Infection*, 20, 105–113. <https://doi.org/10.1111/1469-0691.12474>
- Leeftang, M. M., Rutjes, A. W., Reitsma, J. B., Hooft, L., & Bossuyt, P. M. (2013). Variation of a test's sensitivity and specificity with disease prevalence. *Canadian Medical Association Journal*, 185, E537–E544. <https://doi.org/10.1503/cmaj.121286>
- Lorenzo-López, L., Maseda, A., de Labra, C., Regueiro-Folgueira, L., Rodríguez-Villamil, J. L., & Millán-Calenti, J. C. (2017). Nutritional determinants of frailty in older adults: A systematic review. *BMC Geriatrics*, 17, 108. <https://doi.org/10.1186/s12877-017-0496-2>
- Mazoochi, F., Gobbens, R. J. J., Lotfi, M. S., & Fadayevatan, R. (2020). Diagnostic accuracy of the Tilburg Frailty Indicator (TFI) for early frailty detection in elderly people in Iran. *Archives of Gerontology and Geriatrics*, 91, 104187. <https://doi.org/10.1016/j.archger.2020.104187>
- McInnes, M. D. F., Moher, D., Thombs, B. D., McGrath, T. A., Bossuyt, P. M., Clifford, T., Cohen, J. F., Deeks, J. J., Gatsonis, C., Hooft, L., Hunt, H. A., Hyde, C. J., Korevaar, D. A., Leeflang, M. M. G., Macaskill, P., Reitsma, J. B., Rodin, R., Rutjes, A. W. S., Salameh, J. -P., ... Willis, B. H.; the PRISMA-DTA Group. (2018). Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: The PRISMA-DTA statement. *Journal of the American Medical Association*, 319, 388–396. <https://doi.org/10.1001/jama.2017.19163>
- Mijnarends, D. M., Schols, J. M., Meijers, J. M., Tan, F. E., Verlaan, S., Luiking, Y. C., Morley, J. E., & Halfens, R. J. (2015). Instruments to assess sarcopenia and physical frailty in older people living in a community (care) setting: Similarities and discrepancies. *Journal of the American Medical Directors Association*, 16(4), 301–308. <https://doi.org/10.1016/j.jamda.2014.11.011>
- Mitnitski, A. B., Mogilner, A. J., & Rockwood, K. (2001). Accumulation of deficits as a proxy measure of aging. *Scientific World Journal*, 1, 323–336. <https://doi.org/10.1100/tsw.2001.58>
- Morley, J. E., Malmstrom, T. K., & Miller, D. K. (2012). A simple frailty questionnaire (FRAIL) predicts outcomes in middle aged African Americans. *Journal of Nutrition, Health & Aging*, 16, 601–608. <https://doi.org/10.1007/s12603-012-0084-2>
- Mulherin, S. A., & Miller, W. C. (2002). Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation. *Annals of Internal Medicine*, 137, 598–602. <https://doi.org/10.7326/0003-4819-137-7-200210010-00011>
- Owen, R. K., Cooper, N. J., Quinn, T. J., Lees, R., & Sutton, A. J. (2018). Network meta-analysis of diagnostic test accuracy studies identifies

- and ranks the optimal diagnostic tests and thresholds for health care policy and decision-making. *Journal of Clinical Epidemiology*, 99, 64–74. <https://doi.org/10.1016/j.jclinepi.2018.03.005>
- Ozsoy, G., Ates Bulut, E., Gurbinar, B., Ilcin, N., & Isik, A. T. (2021). Determination of an optimal frailty cutoff score of Tilburg frailty indicator and frailty associated factors in community-dwelling Turkish older adults. *Annals of Geriatric Medicine and Research*, 25(4), 294–300. <https://doi.org/10.4235/agmr.21.0086>
- Panhwar, Y. N., Naghdy, F., Naghdy, G., Stirling, D., & Potter, J. (2019). Assessment of frailty: A survey of quantitative and clinical methods. *BMC Biomedical Engineering*, 1, 7. <https://doi.org/10.1186/s42490-019-0007-y>
- Peters, L. L., Boter, H., Buskens, E., & Slaets, J. P. (2012). Measurement properties of the Groningen Frailty Indicator in home-dwelling and institutionalized elderly people. *Journal of the American Medical Directors Association*, 13, 546–551. <https://doi.org/10.1016/j.jamda.2012.04.007>
- Raiche, M., Hebert, R., & Dubois, M. F. (2008). PRISMA-7: A case-finding tool to identify older adults with moderate to severe disabilities. *Archives of Gerontology and Geriatrics*, 47, 9–18. <https://doi.org/10.1016/j.archger.2007.06.004>
- Rodríguez-Mañas, L., Féart, C., Mann, G., Viña, J., Chatterji, S., Chodsko-Zajko, W., Gonzalez-Colaço Harmand, M., Bergman, H., Carcaillon, L., Nicholson, C., Scuteri, A., Sinclair, A., Pelaez, M., van der Cammen, T., Beland, F., Bickenbach, J., Delamarche, P., Ferrucci, L., Fried, L. P., Gutiérrez-Robledo, L. M.; FOD-CC group (Appendix 1). (2013). Searching for an operational definition of frailty: A Delphi method based consensus statement: The frailty operative definition-consensus conference project. *Journals of Gerontology, Series A: Biological Sciences and Medical Sciences*, 68(1), 62–67. <https://doi.org/10.1093/gerona/gls119>
- Rücker, G. (2018). Network meta-analysis of diagnostic test accuracy studies. In: Biondi-Zoccai, G. (Eds.), *Diagnostic meta-analysis*. Springer. https://doi.org/10.1007/978-3-319-78966-8_13
- Saenger, A., Caldas, L. F., Raiche, C. P., M., da, & Motta, L. B. (2018). Identifying the loss of functional independence of older people residing in the community: Validation of the PRISMA-7 instrument in Brazil. *Archives of Gerontology and Geriatrics*, 74, 62–67. <https://doi.org/10.1016/j.archger.2017.09.008>
- Salanti, G., Nikolakopoulou, A., Efthimiou, O., Mavridis, D., Egger, M., & White, I. R. (2022). Introducing the treatment hierarchy question in network meta-analysis. *American Journal of Epidemiology*, 191(5), 930–938. <https://doi.org/10.1093/aje/kwab278>
- Salminen, M., Viljanen, A., Eloranta, S., Viikari, P., Wuorela, M., Vahlberg, T., Isoaho, R., Kivelä, S. L., Korhonen, P., Irjala, K., Löppönen, M., & Viikari, L. (2020). Frailty and mortality: An 18-year follow-up study among Finnish community-dwelling older people. *Aging Clinical and Experimental Research*, 32(10), 2013–2019. <https://doi.org/10.1007/s40520-019-01383-4>
- Santini, A., Man, A., & Voidăzan, S. (2021). Accuracy of diagnostic tests. *Journal of Critical Care Medicine [Universitatea de Medicina si Farmacie din Targu-Mures]*, 7, 241–248. <https://doi.org/10.2478/jccm-2021-0022>
- Schisterman, E. F., Faraggi, D., Reiser, B., & Hu, J. (2008). Youden index and the optimal threshold for markers with mass at zero. *Statistics in Medicine*, 27, 297–315. <https://doi.org/10.1002/sim.2993>
- Searle, S. D., Mitnitski, A., Gahbauer, E. A., Gill, T. M., & Rockwood, K. (2008). A standard procedure for creating a frailty index. *BMC Geriatrics*, 8, 24. <https://doi.org/10.1186/1471-2318-8-24>
- 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. *The Journal of Nutrition, Health & Aging*, 24(2), 188–193. <https://doi.org/10.1007/s12603-019-1301-z>
- Si, H., Jin, Y., Qiao, X., Tian, X., Liu, X., & Wang, C. (2021). Comparison of 6 frailty screening tools in diagnostic properties among Chinese community-dwelling older people. *Geriatric Nursing*, 42(1), 276–282. <https://doi.org/10.1016/j.gerinurse.2020.08.017>
- Steinhauser, S., Schumacher, M., & Rücker, G. (2016). Modelling multiple thresholds in meta-analysis of diagnostic test accuracy studies. *BMC Medical Research Methodology*, 16, 97. <https://doi.org/10.1186/s12874-016-0196-1>
- Sukkriang, N., & Punsawad, C. (2020). Comparison of geriatric assessment tools for frailty among community elderly. *Heliyon*, 6(9), e04797. <https://doi.org/10.1016/j.heliyon.2020.e04797>
- Swift, A., Heale, R., & Twycross, A. (2020). What are sensitivity and specificity? *Evidence-Based Nursing*, 23, 2–4. <https://doi.org/10.1136/ebnurs-2019-103225>
- Tian, X., Qiao, X., Dong, L., Liu, N., Si, H., Jin, Y., Liu, X., & Wang, C. (2020). Cross-cultural adaptation and psychometric properties of the Groningen Frailty Indicator (GFI) among Chinese community-dwelling older adults. *Geriatric Nursing*, 41(3), 236–241. <https://doi.org/10.1016/j.gerinurse.2019.10.002>
- Trevethan, R. (2017). Sensitivity, specificity, and predictive values: Foundations, pliabilitys, and pitfalls in research and practice. *Frontiers in Public Health*, 5, 307. <https://doi.org/10.3389/fpubh.2017.00307>
- Vermeiren, S., Vella-Azzopardi, R., Beckwée, D., Habbig, A. -K., Scafoglieri, A., Jansen, B., Bautmans, I., Bautmans, I., Verté, D., Beyer, I., Petrovic, M., De Donder, L., Kardol, T., Rossi, G., Clarys, P., Scafoglieri, A., Catrysse, E., de Hert, P., & Jansen, B. (2016). Frailty and the prediction of negative health outcomes: A meta-analysis. *Journal of the American Medical Directors Association*, 17(1163), 1163.e1–1163.e17. <https://doi.org/10.1016/j.jamda.2016.09.010>
- Whiting, P. F., Rutjes, A. W. S., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., Leeflang, M. M. G., Sterne, J. A. C., & Bossuyt, P. M. M.; QUADAS-2 Group. (2011). QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine*, 155, 529–536. <https://doi.org/10.7326/0003-4819-155-8-201110180-00009>
- World Health Organization. (2017). *WHO clinical consortium on healthy ageing. Report of consortium meeting 1–2 December 2016 in Geneva, Switzerland (WHO/FWC/ALC/17.2)*. Licence: CC BY-NC-SA 3.0 IGO. <https://www.who.int/publications/i/item/WHO-FWC-ALC-17.2>
- Xiang, W., Cheng, Y., Li, Z., Han, J., & Li, K. (2020). Cross-cultural adaptation and validation of the Groningen Frailty Indicator in Chinese nursing home residents. *Aging Clinical and Experimental Research*, 32(6), 1035–1042. <https://doi.org/10.1007/s40520-019-01178-7>
- Xue, Q. L., Bandeen-Roche, K., Varadhan, R., Zhou, J., & Fried, L. P. (2008). Initial manifestations of frailty criteria and the development of frailty phenotype in the Women's Health and Aging Study II. *Journals of Gerontology, Series A: Biological Sciences and Medical Sciences*, 63, 984–990. <https://doi.org/10.1093/gerona/63.9.984>
- Yaman, H., & Ünal, Z. (2018). The validation of the PRISMA-7 questionnaire in community-dwelling elderly people living in Antalya, Turkey. *Electronic Physician*, 10(9), 7266–7272. <https://doi.org/10.19082/7266>
- Zhang, Q., Guo, H., Gu, H., & Zhao, X. (2018). Gender-associated factors for frailty and their impact on hospitalization and mortality among community-dwelling older adults: A cross-sectional population-based study. *PeerJ*, 6, e4326. <https://doi.org/10.7717/peerj.4326>