JAMDA xxx (2019) 1.e1–1.e8

JAMDA

journal homepage: www.jamda.com

Original Study

Predictive Accuracy of Frailty Tools for Adverse Outcomes in a Cohort of Adults 80 Years and Older: A Decision Curve Analysis

Eralda Hegendörfer MD, PhD [a,b,*], Bert Vaes MD, PhD [a], Gijs Van Pottelbergh MD, PhD [a], Catharina Matheï MD, PhD [a], Jan Verbakel MD, PhD [a,c], Jean-Marie Degryse MD, PhD [a,b]

[a] Department of Public Health and Primary Care, Katholieke Universiteit Leuven, Belgium
[b] Institute of Health and Society, Université Catholique de Louvain, Belgium
[c] Nuffield Department of Primary Care Health Sciences, University of Oxford, United Kingdom

Keywords:
Adults 80 years or older
frailty
decision curve analysis

A B S T R A C T

Objectives: To compare the predictive performance of 3 frailty identification tools for mortality, hospitalization, and functional decline in adults aged ≥80 years using risk reclassification statistics and decision curve analysis.
Design: Population-based, prospective cohort.
Setting: BELFRAIL study, Belgium.
Participants: 560 community-dwelling adults aged ≥80 years.
Measurements: Frailty by Cardiovascular Health Study (CHS) phenotype, Longitudinal Aging Study Amsterdam (LASA) markers, and Groeningen Frailty Indicator (GFI); mortality until 5.1 ± 0.25 years from baseline and hospitalization until 3.0 ± 0.25 years; and functional status assessed by activities of daily living at baseline and after 1.7 ± 0.21 years.
Results: Frailty prevalence was 7.3% by CHS phenotype, 21.6% by LASA markers, and 22% by GFI. Participants determined to be frail by each tool had a significantly higher risk for all-cause mortality and first hospitalization. For functional decline, only frail by GFI had a higher adjusted odds ratio. Harrell 's C-statistic for mortality and hospitalization and area under receiver operating characteristic curve for functional decline were similar for all tools and <0.70. Reclassification statistics showed improvement only by LASA markers for hospitalization and mortality. In decision curve analysis, all tools had higher net benefit than the 2 default strategies of "treat all" and "treat none" for mortality risk ≥20%, hospitalization risk ≥35%, and functional decline probability ≥10%, but their curves overlapped across all relevant risk thresholds for these outcomes.
Conclusions and Implications: In a cohort of adults aged ≥80 years, 3 frailty tools based on different conceptualizations and assessment sources had comparable but unsatisfactory discrimination for predicting mortality, hospitalization, and functional decline. All showed clinical utility for predicting these outcomes over relevant risk thresholds, but none was significantly superior. Future research on frailty tools should include a focus on the specific group of adults aged ≥80 years, and the predictive accuracy for adverse outcomes of different tools needs a comprehensive assessment that includes decision curve analysis.

As a result of worldwide population ageing and variability of ageing trajectories, the concept of frailty has become increasingly important.[1,2] Frailty identification is recommended because of its association with adverse outcomes such as falls, functional decline, institutionalization, hospitalization, death, and its potential reversibility and prevention.[2,3] Although our understanding of the complexity and heterogeneity of frailty has grown, there is still no widespread consensus on its conceptualization and assessment.[1,3,4] Two main frailty approaches, as unidimensional physical phenotype or multidimensional deficit accumulation index, have been widely used in older adults.[1–3,5,6] Many tools have been developed to identify frailty (up to 67 according to a recent review),[7] yet their validation and accuracy is not adequate.[1,6–11]

The BELFRAIL study was supported by an unconditional grant from Fondation Louvain, Brussels, Belgium (grant number B40320084685).
The authors declare no conflicts of interest.
* Address correspondence to Eralda Hegendörfer, MD, PhD, Academic Center for General Practice, KU Leuven, Kapucijnenvoer 33 blok j-box 7001, 3000 Leuven.
E-mail address: eralda.turkeshi@kuleuven.be (E. Hegendörfer).

https://doi.org/10.1016/j.jamda.2019.08.029
1525-8610/

In the past few years, several studies have compared different frailty tools for their predictive ability of adverse outcomes in older adults in different settings.[8,12–20] One common way of comparing competing risk prediction models is through assessing their discrimination by the concordance (C) statistic either as area under the receiver operating characteristic curve (AUC) for binary outcomes or Harrell's C-statistic for time-dependent outcomes.[21] As the C-statistic does not capture the extent of change in predicted risk between competing models and its clinical utility, risk reclassification statistics and decision curve analysis have been proposed and used as complementary to evaluate the improvement in discrimination and clinical utility of prediction models.[22–25] So far, the comparison of different frailty tools has only been based on odds (ORs) or hazard ratios (HRs) and AUCs.[14,16–19,26,27] Additionally, there is a lack of studies of different frailty tools in community-dwelling older adults aged ≥80 years, the fastest-growing group of older adults, where the accuracy of current frailty tools could improve because of a higher expected prevalence.[10,28]

This study aims to apply and compare the predictive performance of 3 frailty tools for all-cause mortality, first unplanned hospitalization, and functional decline in a community-based prospective cohort of adults aged 80 years and older using risk reclassification statistics and decision curve analysis.

Methods

Study Design and Population

BELFRAIL is a population-based, prospective cohort of adults aged 80 years and older in Belgium. The study protocol has been previously published.[29] Briefly, from November 2008 until September 2009, a total of 567 individuals aged ≥80 years were randomly recruited in 29 general practices, excluding those with severe dementia, in palliative care, or medical emergencies. At baseline, general practitioners recorded sociodemographic and medical history data and performed a standardized physical examination. Two trained clinical research assistants performed at-home standardized comprehensive assessments at baseline and after 1.7 ± 0.21 years. Data on hospitalization were collected until 3.0 ± 0.25 years and on mortality until 5.1 ± 0.25 years from baseline (Figure 1). The study protocol was approved by the Biomedical Ethics Committee of the Medical School of the Universite catholique de Louvain, Belgium, and all participants gave informed consent.

Frailty Identification Tools

Three frailty tools were used: the Cardiovascular Health Study (CHS) frailty phenotype,[30] the Longitudinal Aging Study Amsterdam (LASA) frailty markers,[31] and the Groeningen Frailty Indicator (GFI).[32] They represent the 2 frailty approaches: unidimensional physical phenotype (CHS phenotype) and multidimensional deficit accumulation (LASA markers and GFI). They also cover different assessment sources: self-report (GFI) and performance-based or mixed (CHS phenotype and LASA markers).

The CHS frailty phenotype has 5 components: weight loss, exhaustion, weakness, slowness, and low physical activity[30] (see Supplementary Material). The 9 LASA frailty markers include low body mass index, low peak flow, low Mini Mental State Examination score, urinary incontinence, poor vision or hearing, low mastery/resilience, depression, and low physical activity[31] (see Supplementary Material). For both CHS phenotype and LASA markers, participants with 3 or more components were considered frail.[30,31] The GFI was part of the standardized assessment in the BELFRAIL study.[29] It is a 15-item questionnaire covering self-reported limitations in physical (9 items), cognitive (1 item), psychological (2 items), and social (3 items) domains[33] (see Supplementary Material). Scores range from 0 to 15, and those with scores ≥5 were considered frail. We also investigated "frail by any" if frail by at least 1 of the 3 tools.

Outcomes

Time to all-cause death, first unplanned hospitalization, and functional decline were outcome measurements. The date and cause of hospitalization and death were prospectively reported by the general practitioners. Activities of daily living measured functional limitation. At baseline and follow-up visits, participants described the degree of difficulty with 6 activities: climbing stairs, walking 5 minutes outdoors without resting, getting up and sitting down in a chair, dressing and undressing oneself, using own or public transport, and cutting one's own nails. The response categories ranged from 1 ("No I cannot") to 5 ("Yes, without difficulty"), with a total score of 6 to 30.[29] Functional decline was defined as a decrease of at least 20% from baseline score.

Other Variables

Age, sex, and multimorbidity were used as confounders of the association of frailty with mortality, hospitalization, and functional decline. Multimorbidity was assessed with the unweighted disease count of morbidities reported by the general practitioners at baseline (see Supplementary Material).

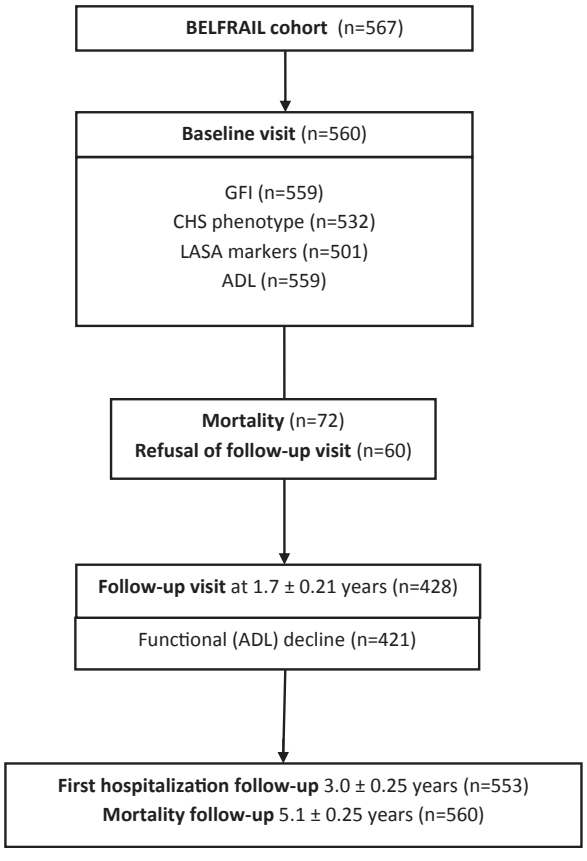

Fig. 1. Flowchart of the data collection in the BELFRAIL cohort study. ADL, activities of daily living.

Table 1
Baseline Characteristics and Outcomes of the Study Population in Total and by Frail Categories According to Each Tool

	Total (n = 560)	CHS Phenotype			GFI			LASA Markers		
		Frail (n = 39)	Robust (n = 493)	*P* Value	Frail (n = 123)	Robust (n = 436)	*P* Value	Frail (n = 108)	Robust (n = 393)	*P* Value
Age, y, mean ± SD	84.7 ± 3.7	86.5 ± 4.8	84.5 ± 3.5	.001*	85.8 ± 4.1	84.4 ± 3.5	<.001*	86.0 ± 4	84.4 ± 3.5	<.001*
Sex, male, n (%)	209 (37.3)	18 (46.2)	183 (37.1)	.263†	24 (19.5)	184 (42.2)	<.001†	30 (27.8)	156 (39.6)	.025†
BMI, mean ± SD	27.4 ± 4.9	27.3 ± 5.4	1.61 ± 0.09	.926*	27.2 ± 5.5	27.5 ± 4.7	.556*	27.1 ± 6.7	27.5 ± 4.3	.481*
ADL score	25 (21, 27)	17 (12, 20)	25 (22, 27)	<.001‡	19 (14, 24)	26 (22, 29)	<.001‡	19.5 (14, 25)	26 (22, 28)	<.001‡
LAPAQ score	61 (18, 98)	8 (0, 28)	70 (29, 102)	<.001‡	18 (0, 58)	71 (31, 104)	<.001‡	16 (16, 56)	76 (40, 106)	<.001‡
MMSE score	28 (25, 29)	26 (22, 28)	28 (26, 29)	.003‡	26 (22, 28)	28 (26, 29)	<.001‡	25.5 (22, 28)	28 (26, 29)	<.001‡
GDS-15 score	2 (1, 4)	6 (3, 8)	2 (1, 4)	<.001‡	5 (3, 7)	2 (1, 3)	<.001‡	5 (2, 7)	2 (1, 3)	<.001‡
Disease count	4 (3, 5)	4 (3, 6)	4 (2, 5)	.276‡	4 (3, 6)	4 (2, 5)	.005‡	5 (3, 7)	4 (2, 5)	<.001‡
Grip strength (kg)	20.9 (15.8, 27.1)	14.8 (10.8, 19.1)	21.4 (16.8, 27.7)	<.001‡	16.5 (13.3, 20.7)	22.5 (17.2, 29.2)	<.001‡	16.8 (13.2, 20.9)	22 (17, 29)	<.001‡
Walking time,§s	10.9 (8.0, 14.7)	19.8 (14.3, 26.6)	10.5 (8.0, 13.7)	<.001‡	13.4 (9.3, 20.9)	10.4 (7.7, 13.5)	<.001‡	14.2 (11.1, 21.4)	10.6 (8.0, 13.7)	<.001‡
Death, n (%)	237 (42.3)	26 (66.7)	192 (38.9)	.001†	72 (58.5)	164 (37.6)	<.001†	69 (63.9)	139 (35.4)	<.001†
First hospitalization, n (%)	284 (50.7)	27 (69.2)	246 (50.5)	.030†	74 (61.7)	210 (48.6)	.013†	69 (65)	108 (46.4)	.001†
ADL decline, n (%)	66 (15.7)	7 (35)	54 (14)	.019†	20 (25)	46 (13.5)	.016†	17 (23.9)	44 (14.1)	.048†

ADL, activities of daily living; GDS, Geriatric Depression Scale; LAPAQ, LASA Physical Activity Questionnaire; MMSE, Mini-Mental State Examination; SD, standard deviation. Unless otherwise noted, data are presented as median (interquartile range).

*P value based on the Student t test.

†P value based on Pearson χ^2 test.

‡P value based on the Mann-Whitney U test.

§Time to walk 3 m, turn around, and walk back as fast as possible.

Statistical Analysis

Comparisons of baseline and outcome variables between frail and robust participants by each tool were tested with independent Student' t-test (parametric variables), Mann-Whitney U test (nonparametric), and Pearson's chi-squared test (categorical). Agreement between the tools for identifying participants with frailty was measured with Cohen kappa coefficient and displayed graphically with Venn diagram.

Kaplan-Meier curves for all-cause mortality and hospitalization were plotted for each frailty tool using the log-rank test for comparison between frail and robust groups. HRs for mortality and first hospitalization with adjustment for age, sex, and multimorbidity were estimated with Cox proportional hazards regression models. Models were checked for the proportional hazards assumption. ORs for functional decline were estimated with a logistic regression model. The robust group was the reference. Variables were checked for multicollinearity. A 2-tailed probability value $P < .05$ was considered statistically significant.

Harrell's C, AUCs, continuous net reclassification improvement (NRI), integrated discrimination improvement (IDI), and decision curve analysis were used to compare the predictive value and clinical utility of frailty tools. Individual predicted absolute risks (mortality and hospitalization) and probability (functional decline) were calculated for frail by each tool based on regression coefficients.[21] For NRI and IDI, frail by CHS phenotype was used as reference. The continuous NRI is the sum of NRI_{events} and $NRI_{nonevents}$, without a defined risk category. NRI_{events} is the percentage of participants with event (death, hospitalization, or functional decline) who were assigned a higher risk or probability by the alternative tool. $NRI_{nonevents}$ is the percentage of participants without events who were assigned a lower risk or probability by the alternative tool. NRI is considered as sum of improvement in sensitivity (NRI_{event}) and specificity ($NRI_{nonevents}$). IDI is the difference of discrimination slopes (difference of mean predicted risk or probability of participants with and without events) between reference and alternative tools. Relative IDI is IDI over the discrimination slope of the reference tool.[22] It is interpreted as the amount by which the alternative tool increases the separation of mean predicted risk or probability for events and nonevents.[22] Net benefit is the difference between true positives and false positives, weighted by the relative harm of false positives for a chosen risk or probability threshold.[24,25] For decision curve analysis, the net benefit if participants are treated according to risk assigned by frail by each tool is plotted across the range of risks or probabilities for an event and compared with 2 default management strategies if no tool is used: (1) consider all participants as frail and apply an intervention ("treat all") or (2) consider all nonfrail and apply no intervention ("treat none").[24] Frailty by a tool has clinical utility if its net benefit curve is above that of "treat all" or "treat none" for a range of reasonable risk thresholds. The tool with higher net benefit for a certain risk or probability has more clinical utility.[24] Statistical analysis was performed with SPSS

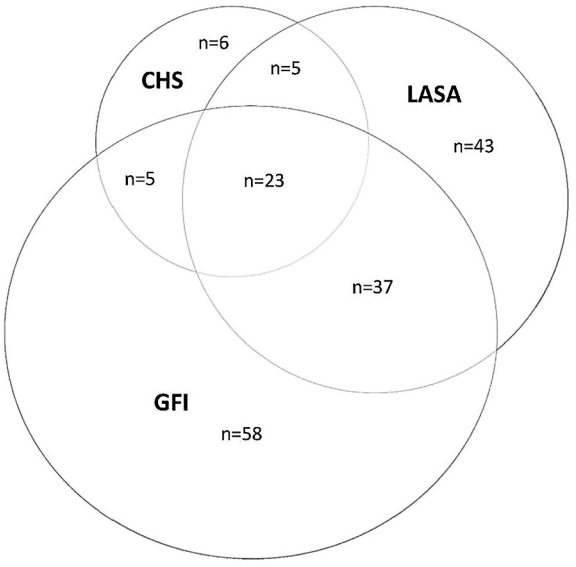

Fig. 2. Venn diagram presenting the extent of overlap between the frail participants identified with the 3 frailty tools (the 177 participants in this diagram are those identified as frail by at least 1 of the tools).

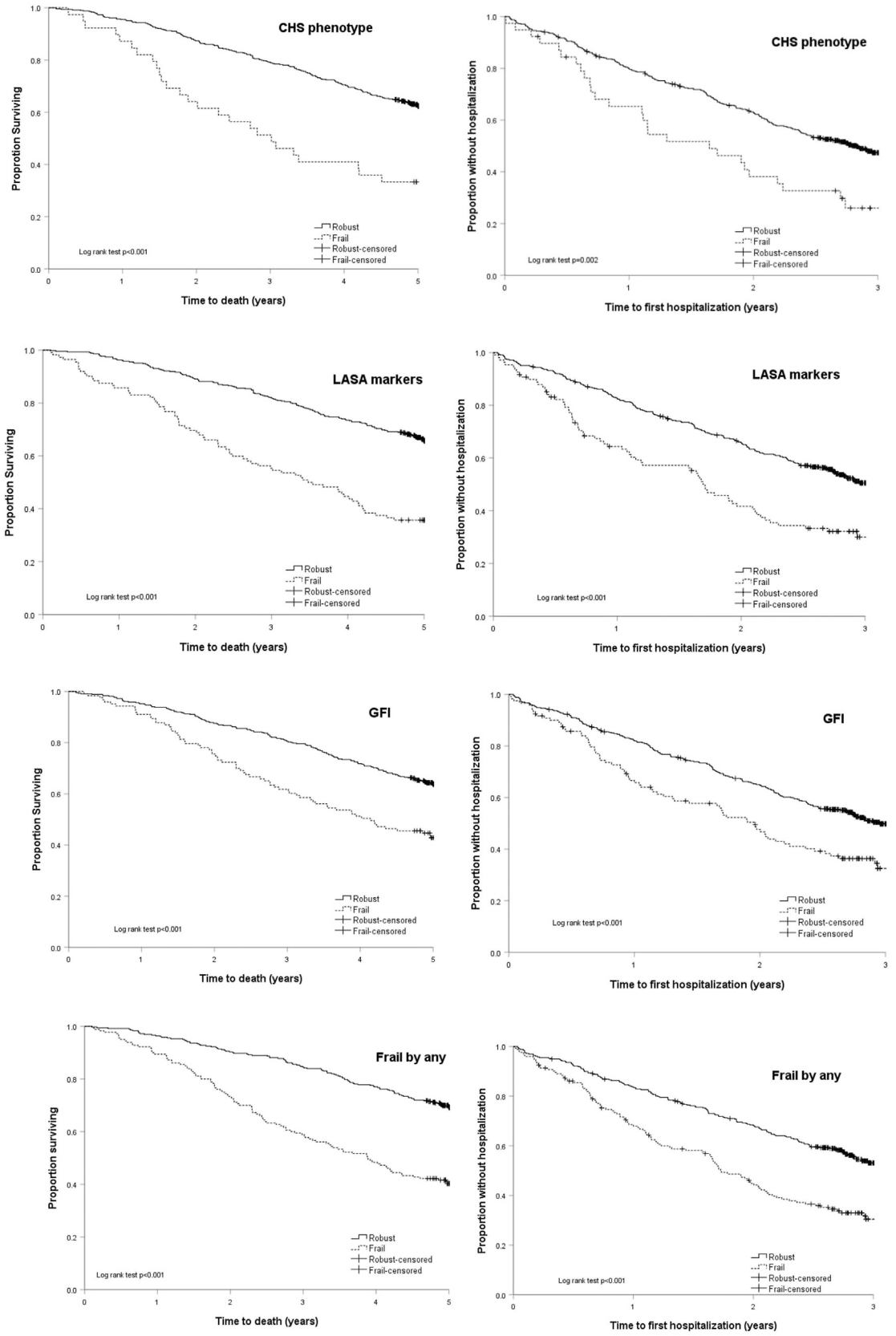

Fig. 3. Kaplan-Meier survival curves for 5-year mortality and 3-year hospitalization for frailty by the different tools. Frail by any: if frail by any of the 3 tools.

Table 2
Association of Frailty by Each of the 3 Tools With Mortality, Hospitalization, and Functional Decline

	Mortality				Hospitalization				Functional Decline			
	Nr	Crude HR (95% CI)	Adjusted HR (95% CI)	Harrell's C (95% CI)	Nr	Crude HR (95% CI)	Adjusted HR (95% CI)	Harrell's C (95% CI)	Nr	Crude OR (95% CI)	Adjusted OR (95% CI)	AUC (95% CI)
CHS	532	2.52 (1.67-3.80)	1.84 (1.20-2.82)	0.65 (0.61-0.69)	523	1.85 (1.24-2.76)	1.63 (1.08-2.45)	0.64 (0.61-0.68)	406	3.31 (1.26-8.67)	2.59 (0.95-7.02)	0.66 (0.58-0.73)
LASA	501	2.40 (1.80-3.21)	1.93 (1.43-2.61)	0.66 (0.62-0.70)	491	1.94 (1.47-2.58)	1.63 (1.21-2.20)	0.65 (0.62-0.69)	384	1.93 (1.02-3.62)	1.69 (0.88-3.28)	0.65 (0.57-0.73)
GFI	559	1.90 (1.44-2.51)	1.62 (1.21-2.16)	0.65 (0.61-0.68)	549	1.61 (1.24-2.10)	1.54 (1.16-2.04)	0.64 (0.61-0.68)	421	2.14 (1.18-3.87)	2.05 (1.10-3.80)	0.67 (0.59-0.74)
Frail by any	541	2.48 (1.91-3.22)	2.09 (1.58-2.77)	0.67 (0.63-0.70)	531	1.95 (1.53-2.49)	1.75 (1.35-2.28)	0.65 (0.62-0.69)	411	2.78 (1.60-4.84)	2.83 (1.57-5.13)	0.69 (0.62-0.76)

CI, confidence interval; Nr, number of participants with all data for the analysis.
Adjusted HR/OR: adjusted for age, sex, multimorbidity; Frail by any: if frail by any of the 3 tools. Reference group: robust group of the population for each tool.

25.0 (IBM Corp, Armonk, NY), Stata 15.0 (StataCorp, College Station, TX), and SAS University Edition (SAS Institute Inc, Cary, NC).

Results

Baseline Characteristics of the Study Population

Of the 567 participants of the BELFRAIL cohort, 560 had a baseline assessment that included GFI (559 participants) and valid measurements of CHS phenotype components (532) and LASA markers (501) (Figure 1). Frailty prevalence was 7.3% by CHS phenotype, 22% by GFI, and 21.6% by LASA markers. Participants with frailty by each tool had statistically significant worse values for activities of daily living, depression, mental status, physical activity, grip strength, and walking time. They were also older and (except for frail by CHS phenotype) consisted of more women and had higher multimorbidity (Table 1).

Agreement Between Frailty Tools

Of the 177 participants with frailty by any tool, only 23 (13%) were frail by all 3 (Figure 2). The highest concordance was between LASA markers and GFI (48.8% of frail by GFI were also frail by LASA markers and 55.5% vice versa). Within the frail by CHS phenotype, 59% were frail by either LASA or GFI, whereas CHS phenotype identified only 22.8% of frail by GFI and 25.9% of frail by LASA markers. Cohen's kappa coefficients showed moderate agreement between GFI and LASA markers (0.45 95% CI 0.32-0.50) and only fair agreement between CHS phenotype and GFI (0.29, 95% CI 0.20-0.39) or LASA markers (0.35, 95% CI 0.24-0.45).

All-Cause Mortality and Hospitalization

All-cause mortality data were available for all participants, whereas first unplanned hospitalization data were missing for 7 participants. During 3.0 $\pm$ 0.25 years, 284 (50.7%) had at least 1 unplanned hospitalization and at 5.1 $\pm$ 0.25 years, 237 (42.3%) had died (Table 1). For each tool, participants with frailty had significantly higher all-cause mortality and first hospitalization compared with robust ones (Figure 3), even after adjustment (Table 2). Harrell's C for all-cause mortality ranged from 0.65 (95% CI 0.61-0.68) for CHS phenotype or GFI to 0.67 (95% CI 0.63-0.70) for "frail by any." For first hospitalization, the range was 0.64 (95% CI 0.61-0.68) for CHS phenotype or GFI to 0.65 (95% CI 0.62-0.69) for LASA markers or "frail by any" (Table 2).

Based on NRI, compared with CHS phenotype the LASA markers improved risk classification for 75% of alive participants and worsened it for 49.5% of deceased ones, while "frail by any" improved it for 60.9% of alive participants and worsened it for 21% of the deceased (Table 3). For hospitalization, only LASA markers improved risk classification for 80.2% of those without hospitalization and worsened it for 64% of those with hospitalization. Only LASA markers and "frail by any" had a statistically relevant higher IDI for both mortality and hospitalization, increasing the difference of the mean predicted risk of events (death or hospitalization) and nonevents (alive or no hospitalization) with those of the CHS phenotype by 31.6% (LASA markers) and 46.2% ("frail by any") for mortality, and 18.4% (LASA markers) and 18.2% ("frail by any") for first hospitalization (Table 3). In decision curve analysis, frail by each tool had a net benefit superior to "treat all" or "treat none" for mortality risk $\geq$20% and hospitalization risk $\geq$35%. Net benefit curves for each tool overlapped considerably across relevant mortality and hospitalization risk thresholds (Figure 4).

Table 3
Reclassification Improvement Statistics of the Frailty Tools for All-Cause Mortality, Hospitalization, and Functional Decline

	NRI (95% CI)	NRI Events, %	NRI Nonevents, %	IDI (95% CI)	Relative IDI
Mortality					
CHS	Reference tool				
LASA	0.25 (0.11, 0.41)	−49.5	75	0.029 (0.018, 0.040)	0.316
GFI	−0.03 (−0.20, 0.14)	−36.9	33.7	0.005 (−0.002, 0.012)	0.052
Frail by any	0.40 (0.24, 0.56)	−21	60.9	0.041 (0.031, 0.051)	0.462
Hospitalization					
CHS	Reference tool				
LASA	0.16 (0.02, 0.30)	−64	80.2	0.017 (0.011, 0.025)	0.184
GFI	−0.07 (−0.23, 0.09)	−44.3	37.7	−0.0004 (−0.006, 0.005)	−0.005
Frail by any	0.16 (−0.03, 0.33)	−35.9	51.4	0.017 (0.010, 0.024)	0.182
Functional Decline					
CHS	Reference tool				
LASA	−0.008 (−0.29, 0.27)	1.75	−2.58	−0.007 (−0.012, − 0.003)	−6.903
GFI	−0.05 (−0.32, 0.22)	−20.0	14.9	−0.006 (−0.010, −0.001)	−7.450
Frail by any	0.34 (0.08, 0.60)	−23.3	57.5	0.004 (−0.002, 0.010)	−7.745

Frail by any: if frail by any of the 3 tools.

Functional Decline

Data on functional decline at 1.7 ± 0.21 years' follow-up were available for 421 participants, and 66 (15.7%) had declined at least 20% from baseline score. Only frail by GFI and "frail by any" had statistically significant increased adjusted ORs for functional decline (Table 2). All tools had similar AUCs, ranging from 0.65 (95% CI 0.57-0.73) for LASA markers to 0.69 (95% CI 0.62-0.76) for "frail by any" (Table 2).

Based on NRI, only the "frail by any" improved risk classification for 57.5% of those without functional decline and worsened it for 23.3% of those with decline compared with CHS phenotype (Table 3). No tool showed any improvement compared with the CHS phenotype based on IDI. In decision curve analysis, frail by each tool had higher net benefit than "treat all" or "treat none" for ≥10% predicted probability for functional decline. Net benefit curves of all tools overlapped (Figure 4).

Discussion

In a cohort of community-dwelling adults aged ≥80 years, we applied CHS phenotype, LASA markers, and GFI and found different frailty prevalence, with fair to moderate agreement between the tools. All tools had higher net benefit than 2 default strategies of "treat all" and "treat none" for 5-year mortality risk ≥20%, 3-year hospitalization risk ≥35%, and 2-year functional decline probability ≥10%. Yet their net benefit curves overlapped across all relevant risk thresholds.

We could not find studies that had applied and compared different frailty tools in community-dwelling adults aged ≥80 years. In our cohort, frailty prevalence was nearly 3 times higher by GFI and LASA markers compared with CHS phenotype, confirming previous reports of higher prevalence with multidimensional tools.[13,18,28,34] We found lower prevalence than that expected in this age group by previous research (weighted average prevalence of 16% for 80-84 years old and 26% for ≥85 years based on 4 studies of community-dwelling adults ≥65 years old with age-stratified prevalence, 3 of which used the CHS phenotype).[2,28] Differences in frailty prevalence (even when using the same tool) have been previously reported and could reflect differences in study designs, measures of tool components, and ageing trajectories.[35–37] Furthermore, adults aged ≥80 years are survivors of their generation and may show reversed epidemiology.[38,39] The frailty cut-offs of different tools are also based on studies in younger old adults, and different cut-offs may be needed in different age groups or settings of older adults.[2] As expected, the highest but still moderate agreement was between the multidimensional tools, GFI and LASA markers. Although they are both multidimensional, GFI is based only on self-report whereas LASA markers include performance measures

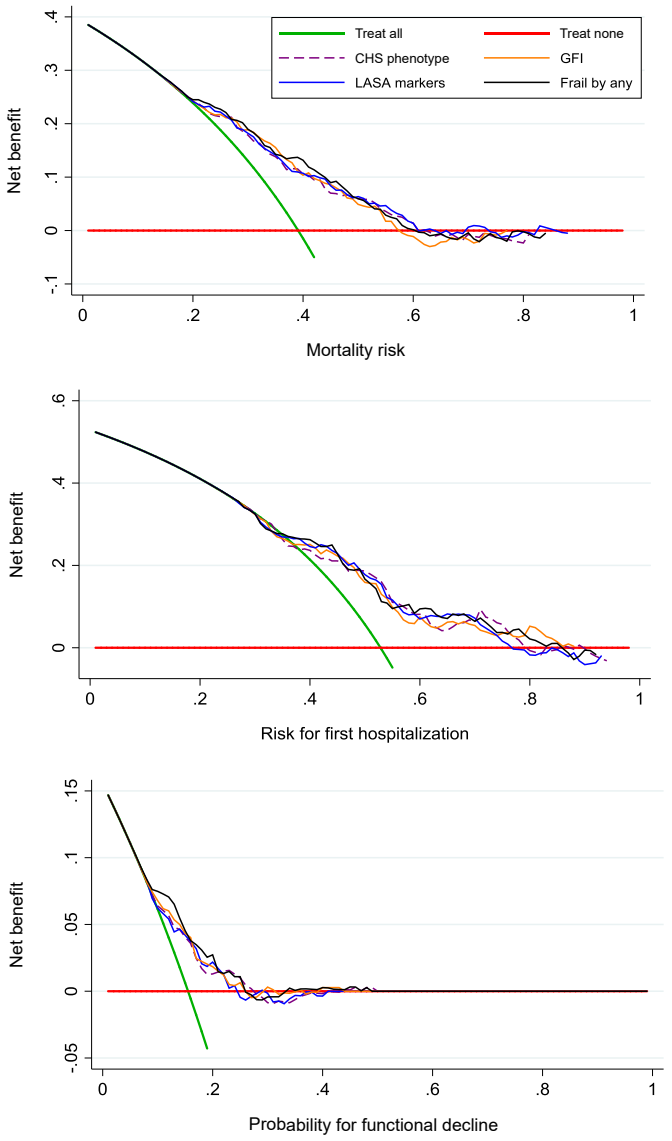

Fig. 4. Decision curve analysis of frailty tools for mortality, hospitalization, and functional decline. Net benefit curves are plotted across risk or probability thresholds for an event (mortality, hospitalization, functional decline) for 6 options: "treat all" as if they are frail, "treat none" considering none is frail, treat according to frailty by CHS phenotype, LASA markers, GFI, or if frail by any tool.

(pulmonary function) and scores of validated scales for mental status, depression, resilience, and physical activity.[31,33]

Previous studies on the predictive performance of different frailty tools have compared mainly ORs or HRs and AUCs.[6,27] In a large comparative study of frailty tools among adults 50 years and older in 11 European countries (mean age 65.3 years), the AUCs for 5-year mortality were similar for GFI [0.70 (0.69-0.72)] and CHS phenotype [0.70 (0.68-0.71)].[13] In a recent comparison study in a Canadian cohort of community-dwelling older adults (mean age 77.7 years) the AUCs for 2-year mortality were 0.69 for CHS phenotype and 0.61 for GFI and for disability 0.68 and 0.66, respectively.[18] We found that HRs and Harrell's C for mortality and hospitalization and ORs and AUCs for functional decline were not significantly different between frailty tools. As in previous studies, Harrell's C and AUC were less than 0.70, thus lacking good discrimination.[18,40] Based on reclassification statistics (NRI and IDI), LASA markers performed better than CHS phenotype for mortality and hospitalization, mainly because of improved specificity. This could be explained by the multidimensional LASA markers and inclusion of pulmonary function that is associated with frailty and adverse outcomes in older adults.[2,41] Decision curve analysis, which assesses the clinical utility of prediction models, showed that if we apply frailty interventions for predicted risks of ≥20% for 5-year mortality, ≥35% for 3-year hospitalization, and ≥10% probability of 2-year functional decline, then all tools would have clinical value, as at these thresholds they are better than "treat all" and "treat none" alternatives. Yet no tool was superior as net benefit curves overlapped across all relevant risk thresholds. These findings in our cohort of adults aged ≥80 years confirm that although the different frailty tools have clinical utility in identifying older adults with higher risk for adverse outcomes, their predictive accuracy is limited and none stands out as the best tool consistently and through a variety of relevant outcomes.

One of the strengths of this study is the design of BELFRAIL as a population-based prospective cohort of community-dwelling adults aged ≥80 years with data on mortality, first hospitalization, and functional decline. The comprehensive BELFRAIL assessment allowed the application of 3 frailty tools that represent different approaches and assessment sources. We also extended the comparison of their predictive ability beyond the binary AUCs, including time-dependent C-statistic, risk reclassification, and decision curve analysis, that are commonly used for risk prediction models.[42] The measurement modifications for some components of frailty tools are a weakness, yet unavoidable and common in previous research as well.[11,37] We used the CHS phenotype as a reference for calculating NRI and IDI, whereas the comprehensive geriatric assessment is considered as a gold standard assessment for frailty.[2] As expected in a cohort of very old adults, we had loss of data for functional decline because of mortality between the 2 assessments.

Conclusions and Implications

In our cohort of community-dwelling adults aged ≥80 years, 3 frailty tools representing different frailty approaches and assessment sources showed clinical utility for identifying higher risk for functional decline, unplanned hospitalization, and all-cause mortality based on decision curve analysis, but their predictive accuracy was limited and none was robustly superior.

Future research on frailty tools should include a focus on adults aged ≥80 years, who have been underrepresented so far. Although there is no consensus on which frailty tool to use, those based on the multidimensional deficit accumulation approach are preferred and seem to perform better.[2,14,16,43,44] Frailty indexes with a list of health deficits including common laboratory tests have been recently developed and validated and need to be tested in adults aged ≥80 years.[45–47] Because of the multidimensionality and heterogeneity of frailty, it is reasonable to use different tools for different outcomes, settings, and populations of older adults rather than pursue a "one size fits all" approach.[4,11,48,49] Future research on frailty tools should move beyond sensitivity, specificity, and AUCs, and report reclassification statistics and decision curve analysis for different outcomes, for a more comprehensive and clinically relevant evaluation.[24,26,42] While waiting for further research on frailty tools in adults aged ≥80 years, clinical practitioners may choose the easiest-to-use tool in their setting, as identifying frailty by any tool in our study was more beneficial than treating all or none as frail.

Acknowledgments

The authors acknowledge all the general practitioners who recruited and followed up their patients in the BELFRAIL study.

Supplementary Data

Supplementary data related to this article can be found online at https://doi.org/10.1016/j.jamda.2019.08.029.

References

1. Sloane PD, Cesari M. Research on frailty: Continued progress, continued challenges. J Am Med Dir Assoc 2018;19:279–281.
2. Clegg A, Young J, Iliffe S, et al. Frailty in elderly people. Lancet 2013;381:752–762.
3. Morley JE, Vellas B, van Kan GA, et al. Frailty consensus: A call to action. J Am Med Dir Assoc 2013;14:392–397.
4. Looman WM, Fabbricotti IN, Blom JW, et al. The frail older person does not exist: Development of frailty profiles with latent class analysis. BMC Geriatr 2018;18:84.
5. Rockwood K, Howlett SE. Fifteen years of progress in understanding frailty and health in aging. BMC Med 2018;16:220.
6. Bouillon K, Kivimaki M, Hamer M, et al. Measures of frailty in population-based studies: An overview. BMC Geriatr 2013;13:64.
7. Buta BJ, Walston JD, Godino JG, et al. Frailty assessment instruments: Systematic characterization of the uses and contexts of highly-cited instruments. Ageing Res Rev 2016;26:53–61.
8. de Vries NM, Staal JB, van Ravensberg CD, et al. Outcome instruments to measure frailty: A systematic review. Ageing Res Rev 2011;10:104–114.
9. Xue QL, Varadhan R. What is missing in the validation of frailty instruments? J Am Med Dir Assoc 2014;15:141–142.
10. Clegg A, Rogers L, Young J. Diagnostic test accuracy of simple instruments for identifying frailty in community-dwelling older people: A systematic review. Age Ageing 2015;44:148–152.
11. Dent E, Kowal P, Hoogendijk EO. Frailty measurement in research and clinical practice: A review. Eur J Intern Med 2016;31:3–10.
12. Daniels R, van Rossum E, Beurskens A, et al. The predictive validity of three self-report screening instruments for identifying frail older people in the community. BMC Pub Health 2012;12:69.
13. Theou O, Brothers TD, Mitnitski A, Rockwood K. Operationalization of frailty using eight commonly used scales and comparison of their ability to predict all-cause mortality. J Am Geriatr Soc 2013;61:1537–1551.
14. Sutorius FL, Hoogendijk EO, Prins BA, van Hout HP. Comparison of 10 single and stepped methods to identify frail older persons in primary care: Diagnostic and prognostic accuracy. BMC Fam Pract 2016;17:102.
15. Hoogendijk EO, van der Horst HE, Deeg DJ, et al. The identification of frail older adults in primary care: Comparing the accuracy of five simple instruments. Age Ageing 2013;42:262–265.
16. Malmstrom TK, Miller DK, Morley JE. A comparison of four frailty models. J Am Geriatr Soc 2014;62:721–726.
17. Pijpers E, Ferreira I, Stehouwer CD. Nieuwenhuijzen Kruseman AC. The frailty dilemma. Review of the predictive accuracy of major frailty scores. Eur J Intern Med 2012;23:118–123.
18. Bongue B, Buisson A, Dupre C, et al. Predictive performance of four frailty screening tools in community-dwelling elderly. BMC Geriatr 2017;17:262.
19. Lin SM, Aliberti MJR, Fortes-Filho SdQ, et al. Comparison of 3 frailty instruments in a geriatric acute care setting in a low-middle income country. J Am Med Dir Assoc 2018;19:310–314.e3.
20. Turusheva A, Frolova E, Korystina E, et al. Do commonly used frailty models predict mortality, loss of autonomy and mental decline in older adults in northwestern Russia? A prospective cohort study. BMC Geriatr 2016;16:98.
21. Moons KG, Kengne AP, Woodward M, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio) marker. Heart 2012;98:683–690.
22. Pencina MJ, D'Agostino RB Sr, Demler OV. Novel metrics for evaluating improvement in discrimination: Net reclassification and integrated

discrimination improvement for normal variables and nested models. Stat Med 2012;31:101–113.
23. Van Calster B, Vickers AJ, Pencina MJ, et al. Evaluation of markers and risk prediction models: Overview of relationships between NRI and decision-analytic measures. Med Decis Making 2013;33:490–501.
24. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. Br Med J 2016;352:i6.
25. Van Calster B, Wynants L, Verbeek JFM, et al. Reporting and interpreting decision curve analysis: A guide for investigators. Eur Urol 2018;74:796–804.
26. Rockwood K, Theou O, Mitnitski A. What are frailty instruments for? Age Ageing 2015;44:545–547.
27. Vermeiren S, Vella-Azzopardi R, Beckwée D, et al. Frailty and the prediction of negative health outcomes: A meta-analysis. J Am Med Dir Assoc 2016;17:1163.e1–1163.e17.
28. Collard RM, Boter H, Schoevers RA, Oude Voshaar RC. Prevalence of frailty in community-dwelling older persons: A systematic review. J Am Geriatr Soc 2012;60:1487–1492.
29. Vaes B, Pasquet A, Wallemacq P, et al. The BELFRAIL (BFC80+) study: A population-based prospective cohort study of the very elderly in Belgium. BMC Geriatr 2010;10:39.
30. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: Evidence for a phenotype. J Gerontol A Biol Sci Med Sci 2001;56:M146–M156.
31. Puts MT, Lips P, Deeg DJ. Sex differences in the risk of frailty for mortality independent of disability and chronic diseases. J Am Geriatr Soc 2005;53:40–47.
32. Steverink N, Slaets J, Schuurmans H, Van Lis M. Measuring frailty: Developing and testing the GFI (Groningen frailty indicator). Gerontologist 2001;41:236–237.
33. Schuurmans H, Steverink N, Lindenberg S, et al. Old or frail: What tells us more? J Gerontol A Biol Sci Med Sci 2004;59:M962–M965.
34. Shamliyan T, Talley KMC, Ramakrishnan R, Kane RL. Association of frailty with survival: A systematic literature review. Ageing Res Rev 2013;12:719–736.
35. Wong CH, Weiss D, Sourial N, et al. Frailty and its association with disability and comorbidity in a community-dwelling sample of seniors in Montreal: A cross-sectional study. Aging Clin Exp Res 2010;22:54–62.
36. Chen CY, Wu SC, Chen LJ, Lue BH. The prevalence of subjective frailty and factors associated with frailty in Taiwan. Arch Gerontol Geriatr 2010;50:S43–S47.
37. Theou O, Cann L, Blodgett J, et al. Modifications to the frailty phenotype criteria: Systematic review of the current literature and investigation of 262 frailty phenotypes in the Survey of Health, Ageing, and Retirement in Europe. Ageing Res Rev 2015;21:78–94.
38. Rapp MA, Gerstorf D, Helmchen H, Smith J. Depression predicts mortality in the young old, but not in the oldest old: Results from the Berlin Aging Study. Am J Geriatr Psychiatry 2008;16:844–852.
39. de Ruijter W, Westendorp RG, Assendelft WJ, et al. Use of Framingham risk score and new biomarkers to predict cardiovascular mortality in older people: Population based observational cohort study. Br Med J 2009;338:a3083.
40. Hosmer DW, Lemeshow S. Applied Logistic Regression. 2nd ed. New York, NY: John Wiley & Sons; 2000.
41. Hegendorfer E, Vaes B, Andreeva E, et al. Predictive value of different expressions of forced expiratory volume in 1 second (FEV1) for adverse outcomes in a cohort of adults aged 80 and older. J Am Med Dir Assoc 2017;18:123–130.
42. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): Explanation and elaboration. Ann Intern Med 2015;162:W1–W73.
43. Kulminski AM, Ukraintseva SV, Kulminskaya IV, et al. Cumulative deficits better characterize susceptibility to death in elderly people than phenotypic frailty: Lessons from the Cardiovascular Health Study. J Am Geriatr Soc 2008;56:898–903.
44. Gonzalez-Colaco Harmand M, Meillon C, Bergua V, et al. Comparing the predictive value of three definitions of frailty: Results from the Three-City study. Arch Gerontol Geriatr 2017;72:153–163.
45. Hoogendijk EO, Theou O, Rockwood K, et al. Development and validation of a frailty index in the Longitudinal Aging Study Amsterdam. Aging Clin Exp Res 2017;29:927–933.
46. Blodgett JM, Theou O, Howlett SE, Rockwood K. A frailty index from common clinical and laboratory tests predicts increased risk of death across the life course. Geroscience 2017;39:447–455.
47. Clegg A, Bates C, Young J, et al. Development and validation of an electronic frailty index using routine primary care electronic health record data. Age Ageing 2018;47:319.
48. Cigolle CT, Ofstedal MB, Tian Z, Blaum CS. Comparing models of frailty: The Health and Retirement Study. J Am Geriatr Soc 2009;57:830–839.
49. Sternberg SA, Schwartz AW, Karunananthan S, et al. The identification of frailty: A systematic literature review. J Am Geriatr Soc 2011;59:2129–2138.