Research Article

Discrepancy in Frailty Identification: Move Beyond Predictive Validity

Qian-Li Xue, PhD,[1,2,]* Jing Tian, MS,[3,2] Jeremy D. Walston, MD,[1] Paulo H. M. Chaves, MD,[4] Anne B. Newman, MD, MPH,[5] and Karen Bandeen-Roche, PhD[3]

[1]Department of Medicine, School of Medicine, Johns Hopkins University, Baltimore, Maryland. [2]Center on Aging and Health, Johns Hopkins Medical Institutions, Baltimore, Maryland. [3]Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland. [4]Benjamin Leon Jr. Family Center for Geriatric Research and Education, Herbert Wertheim College of Medicine, Florida International University, Miami. [5]Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pennsylvania.

*Address correspondence to: Qian-Li Xue, PhD, Department of Medicine, School of Medicine, Johns Hopkins University, 2024 E. Monument Street, Suite 2-700, Baltimore, MD 21205. E-mail: qxue1@jhu.edu

Abstract

Background: To evaluate the discordance in frailty classification between the frailty index (FI) and the physical frailty phenotype (PFP) and identify factors discriminating those with discordant frailty classification from each other and from those for whom the assessments agree.
Methods: A prospective observational study of older adults aged 65 and older selected from Medicare eligibility lists in four U.S. communities (*n* = 5,362). The PFP was measured by the Cardiovascular Health Study PFP. Participants meeting three or more of the five criteria were deemed frail. The FI was calculated as the proportion of deficits in an a priori selected set of 48 measures, and participants were classified as frail if FI is greater than 0.35.
Results: The prevalence of frailty was 7.0% by the PFP and 8.3% by the FI. Of the 730 deemed frail by either instrument, only 12% were in agreement, whereas 39% were classified as frail by the PFP, but not the FI, and 48% were classified as frail by the FI, but not the PFP. Participants aged 65–72 years or with greater disease burden were most likely to be characterized as being FI-frail, but not PFP-frail. The associations of frailty with age and mortality were stronger when frailty was measured by the PFP rather than the FI.
Conclusions: Despite comparable frailty prevalence between the PFP and the FI, there was substantial discordance in individual-level classification, with highest agreement existing only in the most vulnerable subset. These findings suggest that there are clinically important contexts in which the PFP and the FI cannot be used interchangeably.

Keywords: Cumulative deficits, Construct validation, Geriatric syndrome, Measurement, Vulnerability

Frailty has been theoretically defined as a clinically recognizable state of increased vulnerability. This vulnerability is believed to result from aging-associated decline in reserve and function across multiple physiologic systems such that the ability to cope with everyday or acute stressors is comprised (1). In the absence of a gold standard, the two most commonly cited instruments for frailty assessment in a geriatric population are as follows: the physical frailty phenotype (PFP) and the frailty index (FI) (2). Although a number of epidemiological studies have been conducted to compare the two instruments, the comparisons so far have almost exclusively focused on predictive validity (3–7). However, there has been very little research to ex-

plicate the differences among the older adults the two instruments identify as frail. For example, one study reported that only 30% of those found frail by either method were deemed frail by both methods; 24% were found frail by the PFP, but not the FI, and 46% were found frail by the FI, but not the PFP (8). We believe that there is potential risk to advocating the use of the different frailty assessment tools without a better understanding of the degree of distinction in the identification of vulnerable older adults, as well as the heterogeneity of older adults identified as frail by either instrument. For example, one person is physically frail and the other is mostly cognitively frail, and let us suppose that they are both classified as frail but

by different instruments, that is, the PFP for the former and the FI for the latter. However, their underlying pathophysiology may be very different, and they thus may require different approaches to care. Using existing data from the Cardiovascular Health Study (CHS), this study evaluated the discordance in frailty classification between the FI and the PFP and identified factors discriminating those with discordant frailty classification from each other and from those for whom the assessments agree.

Methods

Study Population

The CHS is a prospective observational study of risk factors for coronary heart disease and stroke in men and women aged 65 years and older. The original cohort consisted of 5,201 men and women (4.6% African American, 57% women) who were randomly selected between June 1989 and May 1990 from Medicare eligibility lists including age-eligible household members in four U.S. communities: Forsyth County, NC; Sacramento County, CA; Washington County, MD; and Allegheny County, PA. A new cohort of 687 African Americans (AA) were subsequently recruited in 1992–1993 at three of the four field centers. Details of the study design are described elsewhere (9,10). Study participants attended clinic visits annually if available through 1999–2000 at which they completed a health questionnaire and a comprehensive physical examination and provided blood specimens. The Institutional Review Board at each participating site approved the study protocol; all study participants signed informed consent. After excluding participants with a history of Parkinson's disease or stroke, the Mini-Mental State Examination score less than 18 in the original cohort or the Modified Mini-Mental State (3MS) Examination score less than 68 in the AA cohort, and those who were taking Sinemet, Aricept, or antidepressant, a combined sample of 5,362 participants (4,793 from the original cohort and 569 from the AA cohort) with available data on the PFP and the FI at study baseline (ie, Year 1989–1990 for the original cohort and Year 1992–1993 for the AA cohort) were used in the current analysis. Because Mini-Mental State Examination was not administered in the AA cohort, equipercentile equating was used to identify the cutoff of 68 for the 3MS that is equivalent to Mini-Mental State Examination of 18.

Physical Frailty Phenotype

The PFP was measured by the five-criteria CHS PFP: (i) weakness (by grip strength), (ii) slowness (by usual-pace 15-feet walking speed), (iii) low physical activity (by total energy expenditure in kilocalories per week), (iv) weight loss (by self-reported unintentional weight loss of more than 10% in the past year), and (v) exhaustion (by self-report) (1). Participants meeting three or more criteria were classified as frail and those meeting one or two criteria were classified as pre-frail.

Frailty Index

The FI was calculated as the proportion of deficits in an a priori selected set of 48 measures from baseline similar to those used by Kulminski and colleagues (3). These 48 variables are as follows: diseases (pulmonary diseases, heart disease, pulmonary embolus, high blood pressure, diabetes, arthritis, pneumonia, asthma, cancer, nervous/emotional disorder, major ECG abnormality), chronic symptoms (orthostatic hypotension, sleep on more than two pillows to help breathe, awakened by trouble breathing, feeling groggy after waking up in the morning, trouble falling asleep, swelling of feet/ankles, pain in leg on walking, cough, bleeding or bruising easily); symptoms in the past 2 weeks (short of breath, palpitations, dizziness, fatigue, weakness, nausea, indigestion, diarrhea); sensory function (hearing problems, vision problems); physical function (unsafe to stand up without using arms, difficulty with walking, walking 1/2 mile, walking up 10 steps, lifting, reaching out, gripping); psychosocial function (feeling mostly dissatisfied, unhappy, terrible about life); life satisfaction (scored ≥5 on a 10-point Likert scale from 1 [extremely satisfied] to 10 [extremely dissatisfied]); and probably or definitely having no people to talk to when lonely; lack of physical activity engagement in the past 2 weeks (walking for exercise, household chores, mowing lawn, raking lawn, gardening, exercise cycle, dancing, calisthenics exercises; see Supplementary Appendix D for details). Each measure was dichotomized as either presence or absence. Participants were classified as frail if the proportion of deficits (ie, FI) was greater than 0.35, pre-frail if FI greater than 0.2 and less than or equal to 0.35 or non-frail if FI less than or equal to 0.2 (3).

Statistical Analysis

First, we compared the demographic and health characteristics between the groups with concordant versus discordant frailty classification using two-sample *t*-test for continuous variables and chi-square test for categorical variables. Cohen's kappa coefficient was calculated to assess the agreement in frailty classification using the two frailty instruments. Next, within the subset with discordant classification, we implemented the Classification And Regression Tree (CART) analysis to identify characteristics, among 21 baseline covariates in Table 1, that best distinguish those who were deemed frail by the PFP, but not the FI, from those who were frail by the FI, but not the PFP. CART used an automated 10-fold cross-validation to select the optimal tree with the lowest overall misclassification cost (ie, misclassification rates in the learning and testing samples), thus the highest accuracy for prediction (by the area under the receiver operating characteristic curve) (11). Different metrics for measuring node purity including Gini index, entropy, and classification error were used to assess model robustness (12). Finally, we compared the predictive validity of the two frailty instruments for all-cause mortality and incident activities of daily living (ADL) and instrumental activities of daily living (IADL) difficulty. We used the conventional Cox regression for mortality and used discrete-time Cox regression for ADLs and IADLs to account for the fact that incidence of ADL or IADL difficulty could only be observed through self-report at discrete follow-up visits (ie, interval censoring) (13). We conducted the analysis separately for the original and the AA cohorts to accommodate the different length of follow-up and make full use of available data (up to 9 years for the original cohort and up to 6 years for the AA cohort).

As a sensitivity analysis, we reran all models by redefining frailty and pre-frailty by the FI using cutoffs that would yield identical prevalence estimates as those of the PFP. In addition, we reassessed the predictive validity by limiting the maximum follow-up time for the original cohort to 4 years in the Cox models as in the study by Kulminski and colleagues (3).

Results

The prevalence of frailty was 7.0% (*n* = 377) by the PFP and 8.3% (*n* = 444) by the FI in the combined sample. The Cohen's kappa coefficient was 0.16 (95% confidence interval = 0.12–0.20). Of the 730 who were classified as frail by either instrument, only 12% (*n* = 91) were in agreement, whereas 39% (*n* = 286) were classified as frailty by the PFP, but not the FI, and 48% (*n* = 353) were classified as

Table 1. Summary of Baseline Characteristics by Concordant Versus Discordant Classification of Frailty Status by the Frailty Index and the Physical Frailty Phenotype in the Combined Cardiovascular Health Study Original and African American Cohorts (*N* = 5,362)

		Agreement		Disagreement			
Characteristic	(1) Overall *n* = 5,362	(2) Not-Frail by Both *n* = 4,632	(3) Frail by Both *n* = 91	(4) Frail by PFP or FI *n* = 639	(5) Frail by PFP, but not by FI *n* = 286	(6) Frail by FI, but not by PFP *n* = 353	*p* Value for (3) vs. (4)[a]
Age (y), mean (*SD*)	72.7 (5.5)	72.4 (5.3)	76.1 (6.4)	74.4 (6.3)	76.8 (6.3)	72.6 (5.7)[‡]	.022
Education (y), mean (*SD*)	13.9 (4.7)	14.0 (4.6)	11.5 (4.6)	12.9 (4.8)	13.1 (4.9)	12.8 (4.8)	.010
Black race, *n* (%)	781 (14.6)	626 (13.5)	23 (25.3)	132 (20.7)	76 (26.6)	56 (16.9)[‡]	.314
Male gender, *n* (%)	2,249 (41.9)	2,013 (43.5)	21 (23.1)	215 (33.7)	109 (38.1)	106 (30.0)*	.044
3MS, mean (*SD*)[b]	89.1 (9.6)	89.6 (9.1)	82.9 (13.5)	86.6 (11.5)	85.4 (12.5)	87.5 (10.7)[‡]	.007
Number of diseases, mean (*SD*)[c]	1.2 (1.0)	1.1 (1.0)	2.5 (1.2)	1.9 (1.1)	1.5 (1.0)	2.2 (1.1)[‡]	<.001
Body mass index, mean (*SD*)	26.7 (4.7)	26.6 (4.5)	27.1 (6.5)	27.3 (5.7)	26.5 (5.6)	28.0 (5.7)[‡]	.711
Marriage, *n* (%)							<.001
Married	3,588 (67.0)	3,183 (68.8)	34 (37.4)	371 (58.2)	54 (53.9)	217 (61.7)	
Widowed	1,291 (24.1)	1,056 (22.8)	39 (42.9)	196 (30.7)	96 (33.6)	100 (28.4)	
Separated/divorced/never married	477 (8.9)	388 (8.4)	18 (19.8)	71 (11.1)	36 (12.6)	35 (9.9)	
Depressive symptoms, *n* (%)	264 (4.9)	141 (3.1)	36 (39.6)	87 (13.6)	28 (9.8)	59 (16.7)*	<.001
Smoking, *n* (%)							.127
Never	2,493 (46.6)	2,148 (46.4)	48 (52.8)	297 (46.5)	146 (51.1)	151 (42.8)	
Former	2,238 (41.8)	1,961 (42.4)	26 (28.6)	251 (39.3)	103 (36.0)	148 (41.9)	
Current	625 (11.7)	517 (11.2)	17 (18.7)	91 (14.2)	37 (12.9)	54 (15.3)	
Income, *n* (%)							<.001
<$16,000	2,065 (41.0)	1,679 (38.7)	64 (75.3)	322 (53.2)	146 (54.3)	176 (52.4)[‡]	
$16,000–$35,000	1,792 (35.6)	1,585 (36.5)	18 (21.2)	189 (31.2)	68 (25.3)	121 (36.0)	
>$35,000	1,175 (23.4)	1,078 (24.8)	3 (3.5)	94 (15.5)	55 (20.4)	39 (11.6)	
Cancer, *n* (%)	781 (14.6)	635 (13.7)	24 (26.4)	122 (19.1)	40 (14.0)	82 (23.2)[‡]	.104
Congestive heart failure, *n* (%)	226 (4.2)	122 (2.6)	29 (31.9)	75 (11.4)	24 (8.4)	51 (14.5)*	<.001
Myocardial infarction, *n* (%)	495 (9.2)	377 (8.1)	19 (20.9)	99 (15.5)	29 (10.1)	70 (19.8)[‡]	.192
Angina, *n* (%)	856 (16.0)	619 (13.4)	42 (46.2)	195 (30.5)	54 (18.9)	141 (39.9)[‡]	.003
Coronary heart disease[d]	1.027 (19.2)	762 (16.5)	47 (51.7)	218 (34.1)	63 (22.0)	155 (43.9)[‡]	.001
Osteoarthritis, *n* (%)[e]	2,707 (51.1)	2,162 (47.2)	80 (88.9)	465 (73.7)	174 (62.4)	291 (82.7)[‡]	.002
Diabetes, *n* (%)	1,564 (29.5)	1,281 (27.9)	37 (43.0)	246 (39.3)	105 (37.4)	141 (40.9)	.508
Difficulty Walking 1/2 mile, *n* (%)	1,046 (19.8)	647 (14.1)	73 (82.0)	326 (52.8)	144 (53.1)	182 (53.6)	<.001
Difficulty climbing 10 steps, *n* (%)	645 (12.5)	348 (7.8)	57 (64.8)	240 (39.2)	87 (32.3)	153 (44.6)*	<.001
ADL difficulty, *n* (%)	48 (0.9)	14 (0.3)	11 (12.2)	23 (3.6)	9 (3.2)	14 (4.0)	<.001
IADL difficulty, *n* (%)	103 (1.9)	39 (0.8)	23 (25.6)	41 (6.5)	22 (7.8)	19 (5.4)	<.001

Notes: 3MS = Modified Mini-Mental State Examination; ADL = activities of daily living; FI = frailty index; IADL = instrumental activities of daily living; PFP = physical frailty phenotype.

[a]Two-sample *t*-test assuming unequal variance for continuous variables and chi-square test for categorical variables of differences between groups (3) and (4). [b]3MS was not available in the original cohort; numbers presented here were calculated by combining the 3MS score from the African American cohort with 3MS score converted from the Mini-Mental State Examination score in the original cohort. [c]Number of diseases including cancer, congestive heart failure, myocardial infarction, angina, rheumatoid arthritis, and diabetics. [d]Coronary heart disease is present if myocardial infarction and/or angina is present. [e]Osteoarthritis of hand, shoulder, hip, or knee.

[‡]*p* < .001; *p* < .05 based on two-sample *t*-test for continuous variables and chi-square test for categorical variables between the two discordant groups (ie, Columns 5 and 6).

frailty by the FI, but not the PFP. Compared with those deemed frail by both, those with discordant frailty status were significantly younger, more highly educated, more likely to be male, married, in a higher income category, having lower levels of ADL/IADL difficulty and mobility limitation; they also on average had less burden of disease, fewer depressive symptoms, and higher 3MS score (Table 1).

Within the subset (*n* = 639) with discordant classification of frailty status, compared with those who were classified as frail by the FI (termed "FI-Frail" henceforth), those who were classified as frail by the PFP (termed "PFP-Frail" henceforth) were significantly older, more likely to be male, African American, healthier overall except for a slightly lower mean 3MS score, and had on average higher income and lower BMI. Although the association between subclinical cardiovascular disease and the PFP has been reported in the CHS

(14), it is worth noting that the prevalence of cardiovascular diseases including congestive heart failure, myocardial infarction, and angina was approximately doubled in the FI-Frail group compared with the PFP-Frail group (Table 1). Consistent with those findings, the CART model also showed that age and disease burden were the two strongest predictors of the pattern of discordance (Figure 1). Specifically, people over the age of 72 were more likely to be classified as PFP-Frail, where the percentage of being classified as PFP-Frail was more than doubled in the older group compared with the younger group. Regardless of age, people with greater disease burden were more likely to be classified as FI-Frail. In addition, osteoarthritis was associated with being FI-Frail among the younger-old adults (age ≤ 72 years) with at most one disease. The area under the receiver operating characteristic curve averaged 0.73 for the learning samples and

0.68 for the testing samples; the overall misclassification rate was 0.29 and 0.35 for the learning and testing samples, respectively.

In the CHS original cohort, although frailty was significantly associated with 9-year mortality and incident IADL and ADL difficulty regardless of the choice of the frailty instrument, the associations were uniformly stronger for the PFP compared with the FI for all

Figure 1. Results from the Classification And Regression Tree (CART) analysis showing characteristics that best distinguish those who were deemed frail by the PFP, but not the FI, from those who were frail by the FI, but not the PFP, in the combined Cardiovascular Health Study original and African American cohorts (*n* = 639). PFP = physical frailty phenotype; FI = frailty index.

three outcomes. Specifically, the risk of mortality and incident IADL and ADL difficulty was 2.25, 4.95, and 4.71 times higher in the PFP-Frail group than that of the PFP-not-Frail group after covariate adjustment, whereas the corresponding hazard ratios were 1.38, 2.63, and 3.02 when using the FI (Table 2). The trends remained the same after dropping participants classified as frail by both the PFP and the FI (Table 2).

In our sensitivity analysis, when we used the cutoffs for defining FI-Frail and FI-Pre-Frail that would yield prevalence estimates identical to those of the PFP in the original cohort (ie, 6.6% Frail and 47.6% Pre-frail), the conclusions remained the same (Supplementary Appendix Table B). Shortening the follow-up time to 4 years did not meaningfully change the results of adverse outcome prediction either (Supplementary Appendix Table A).

Within the AA cohort, being PFP-Frail was again more strongly associated with mortality than FI-Frail after covariate adjustment (hazard ratio = 3.53, 95% confidence interval = 1.64, 7.63 vs. hazard ratio = 1.31, 95% confidence interval = 0.60, 2.85). The associations with incident IADL difficulty, however, were stronger for the FI in the AA cohort (Table 3). In the sensitivity analysis using revised cut points for defining FI-Pre-Frailty (>0.185, ≤0.345) and FI-Frailty (>0.345) to yield prevalence estimates comparable to those of the PFP in the AA cohort (ie, 11.4% Frail and 56.1% Pre-frail), the overall results remained the same (Supplementary Appendix Table C).
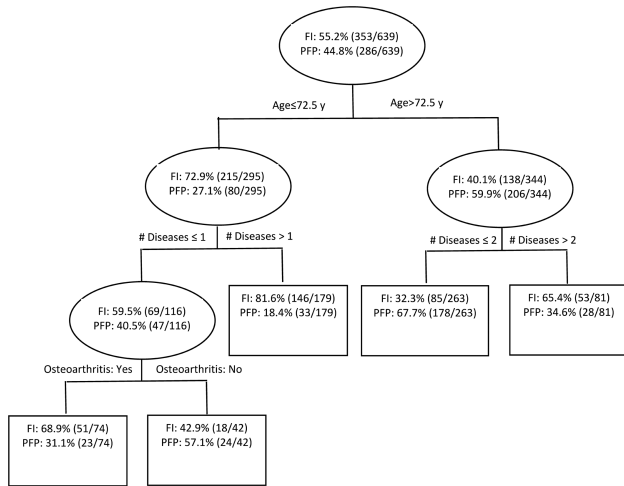
Discussion

A recent frailty consensus document has advocated screening for frailty in older adults using a number of existing frailty instruments

Table 2. Nine-Year Mortality and Incident IADL and ADL Difficulty in the CHS Original Cohort

| Outcome | Sample Size; Incidence Rate (per 100 person-years) | Physical Frailty Phenotype | | | | Frailty Index | | | |
| | | Crude Hazard Ratio (95% CI) *p* Value | | Adjusted Hazard Ratio[a] (95% CI) *p* Value | | Crude Hazard Ratio (95% CI) *p* Value | | Adjusted Hazard Ratio (95% CI) *p* Value | |
		Prefrail	Frail	Prefrail	Frail	Prefrail	Frail	Prefrail	Frail
Mortality	Overall	1.80	3.99	1.46	2.25	1.31	1.99	1.12	1.38
	(*n* = 4,793)	(1.63–1.99)	(3.41–4.65)	(1.32–1.63)	(1.89–2.67)	(1.19–1.45)	(1.70–2.32)	(1.00–1.24)	(1.15–1.66)
	4.01	<.001[b]	<.001	<.001	<.001	<.001	<.001	.043	<.001
	Subset[c]	1.80	3.80	1.45	2.23	1.31	1.69	1.11	1.28
	(*n* = 4,720)	(1.63–2.00)	(3.19–4.51)	(1.31–1.62)	(1.84–2.69)	(1.19–1.45)	(1.42–2.02)	(1.00–1.24)	(1.04–1.56)
	3.93	<.001	<.001	<.001	<.001	<.001	<.001	.048	.017
IADL	Overall	2.46	9.27	1.94	4.95	1.85	3.69	1.55	2.63
	(*n* = 4,579)	(2.10–2.88)	(7.46–11.52)	(1.64–2.28)	(3.88–6.30)	(1.59–2.15)	(2.97–4.58)	(1.32–1.83)	(2.04–3.39)
	2.01	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
	Subset	2.46	8.24	1.94	4.50	1.85	2.93	1.55	2.21
	(*n* = 4,535)	(2.10–2.88)	(6.49–10.45)	(1.64–2.28)	(3.46–5.84)	(1.59–2.15)	(2.30–3.74)	(1.32–1.82)	(1.68–2.91)
	1.95	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
ADL	Overall	2.42	9.33	1.95	4.71	1.89	4.56	1.58	3.02
	(*n* = 4,623)	(2.00–2.93)	(7.26–12.00)	(1.60–2.38)	(3.55–6.26)	(1.58–2.27)	(3.56–5.83)	(1.30–1.92)	(2.25–4.04)
	1.36	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
	Subset	2.42	7.94	1.94	4.30	1.90	3.53	1.57	2.58
	(*n* = 4,571)	(2.00–2.93)	(6.00–10.50)	(1.59–2.36)	(3.16–5.85)	(1.58–2.28)	(2.67–4.67)	(1.29–1.91)	(1.87–3.55)
	1.30	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001

Notes: ADL = activities of daily living; CI = confidence interval; IADL = instrumental activities of daily living.

[a]Adjusted by age, race, gender, Mini-Mental State Examination score, Geriatric Depression Scale score (≥14), and number of diseases (modeled by cubic spline with three knots at the 10th, 20th, and 90th percentile). [b]*p* Value. [c]After excluding participants classified as frail by both the physical frailty phenotype and the frailty index.

Table 3. Six-Year Mortality and Incident IADL and ADL Difficulty in the African American Cohort

Outcome	Sample Size; Incidence Rate (per 100 person-years)	Physical Frailty Phenotype				Frailty Index			
		Crude Hazard Ratio (95% CI) *p* Value		Adjusted Hazard Ratio[a] (95% CI) *p* Value		Crude Hazard Ratio (95% CI) *p* Value		Adjusted Hazard Ratio (95% CI) *p* Value	
		Prefrail	Frail	Prefrail	Frail	Prefrail	Frail	Prefrail	Frail
Mortality	Overall	2.54	4.47	2.12	3.53	1.40	2.00	1.04	1.31
	(*n* = 569)	(1.39–4.66)	(2.19–9.12)	(1.14–3.94)	(1.64–7.63)	(0.87–2.26)	(1.05–3.79)	(0.60–1.79)	(0.60–2.85)
	2.68	.003[b]	<.001	.018	.001	.164	.038	.902	.493
	Subset[c]	1.80	3.80	1.45	2.23	1.31	1.69	1.11	1.28
	(*n* = 551)	(1.63–2.00)	(3.19–4.51)	(1.31–1.62)	(1.84–2.69)	(1.19–1.45)	(1.42–2.02)	(1.00–1.24)	(1.04–1.56)
	2.54	<.001	<.001	<.001	<.001	<.001	<.001	.048	.017
IADL	Overall	2.03	4.33	1.77	2.84	2.36	5.42	2.05	4.05
	(*n* = 556)	(1.17–3.52)	(2.21–8.48)	(1.01–3.11)	(1.38–5.86)	(1.38–4.03)	(2.89–10.16)	(1.15–3.68)	(1.93–8.50)
	2.90	.011	<.001	.046	.005	.002	<.001	.016	<.001
	Subset	2.03	4.07	1.73	2.72	2.36	5.39	1.94	3.94
	(*n* = 541)	(1.17–3.52)	(1.94–8.53)	(0.98–3.06)	(1.26–5.85)	(1.38–4.03)	(2.73–10.60)	(1.08–3.50)	(1.82–8.50)
	2.76	.011	.002	.057	.011	.002	<.001	.026	<.001
ADL	Overall	1.88	5.40	1.56	3.64	1.90	5.22	1.53	3.39
	(*n* = 555)	(0.95–3.47)	(2.45–12.00)	(0.77–3.17)	(1.57–8.43)	(1.00–3.61)	(2.49–10.94)	(0.76–3.09)	(1.43–8.06)
	1.88	.072	<.001	.214	.003	.050	<.001	.236	.006
	Subset	1.88	3.89	1.67	2.72	1.90	3.76	1.51	2.29
	(*n* = 540)	(0.95–3.75)	(1.56–9.66)	(0.83–3.38)	(1.05–7.09)	(1.00–3.62)	(1.58–8.94)	(0.74–3.09)	(0.84–6.23)
	1.68	.071	.004	.152	.039	.049	.003	.255	.104

Notes: ADL = activities of daily living; CI = confidence interval; IADL = instrumental activities of daily living.

[a]Adjusted by age, race, gender, Mini-Mental State Examination score, Geriatric Depression Scale score (≥14), and number of diseases (modeled by cubic spline with three knots at the 10th, 20th, and 90th percentile). [b]*p* Value. [c]After excluding participants classified as frail by both the physical frailty phenotype and the frailty index.

including the PFP and the FI (15). We believe there is danger to advocating the use of the different frailty screening instruments when there is evidence that they ascertain considerably different groups of individuals as being frail but know little about what distinguishes these groups. Similar to the findings from a previous study (8), although the overall prevalence of frailty was similar between the PFP and the FI, we found substantial discordance in individual-level classification. The greatest agreement tended to occur among the oldest old with a considerable disease burden and disability; in comparison, the observed discordance seemed to exist primarily among younger older adults with higher social economic status and better overall health, suggesting that the concept of frailty may help identify a subclinical subset who might not come under the radar using conventional geriatric risk assessment. Our analysis also revealed that age and disease burden were the strongest predictors of discordance pattern with people at younger age but with greater disease burden more likely be classified as FI-Frail, but not PFP-Frail.

The discordance in frailty classification found by Cigolle and colleagues (8) and confirmed in this study highlights the practical difficulty in choosing a frailty screening tool and interpreting the discrepancies. Although the previous study evaluated demographic and disease characteristics of those identified as frail by each method, they did not seek to identify factors discriminating those with discordant classification from each other and from those for whom the assessments agree. The fact that the discordance preferentially affected the subset of potentially vulnerable older adults who are less likely to be captured by traditional geriatric risk assessment focusing on disability and multimorbidity is important. On the one hand, this

suggests that incorporating frailty assessment into clinical routine could potentially provide prognostic information above and beyond what is already available in a clinical setting to improve clinical care and decision making. On the other hand, the usefulness of such information may vary depending on the frailty instrument of choice and the purpose of frailty assessment (2).

Our finding that people with greater disease burden were more likely to be classified as FI-Frail should not come as a surprise given that the definition of the FI includes diseases and medical conditions and the PFP does not. However, to practitioners who weigh options of different instruments, it is sensible to expect that the different instruments with "frailty" in their name should essentially identify the same people regardless of the particular indicators used in an instrument if they measure the same underlying construct. Contrary to this common perception of measurement equivalence, the observed discrepancy in the classification and distinct instrument-specific profile characteristics provides additional empirical evidence for important distinctions in the constructs for who is frail (16,17). This, in turn, may help clinicians determine the type of frailty and its likely causes; therefore, plan a more targeted intervention to ameliorate functional decline.

The concept of biological age was introduced to explain the heterogeneity in health of people of the same chronological age (18). The FI is considered by some to be a marker of biological aging (19), and a recent study found significant but weak associations between the FI and measures of biological age (20). The fact that those classified as PFP-Frail in the CART model tend to be older is intriguing. Although there is important distinction between biological age and chronological age, frailty as an aging-related phenomenon

is expected to be correlated with chronological age, which is confirmed by the steeper relationship between age and prevalence of PFP-Frailty relative to FI-Frailty shown in Figure 2. A separate logistic model for frailty classification within the discordant subgroup confirmed that the odds of being PFP-Frail was 1.72 (95% confidence interval = 1.42, 2.09; $p < .001$) times of the odds of being FI-Frail for every 5-year increase in age after adjusting for covariates in Table 1. This, coupled with our finding of PFP-Frailty having a stronger association with mortality than FI-Frailty after adjusting for chronological age, as discussed in the next paragraph, raises the interesting question of whether or not the PFP is a better marker of biological aging than the FI.

Given the dominance of discussion on predictive validity when comparing frailty instruments (16), we analyzed the associations of PFP-Frailty versus FI-Frailty with mortality and incident IADL and ADL difficulty. We found PFP-Frailty to be a stronger and independent predictor of mortality in both the CHS original and AA cohorts. The associations with incident ADL and IADL difficulty were also stronger for the PFP in the original cohort, suggesting that the PFP may do a better job at predicting the loss of function, which is what older adults are most concerned about. However, we caution against generalization of this finding because the associations became less consistent in the AA cohort. This discrepancy may be explained by the between-cohort differences in sample characteristics at baseline, including lower education, income, and Mini-Mental State Examination, and higher prevalence of congestive heart failure, osteoarthritis, diabetes, and mobility limitation in the AA cohort, all of which are known to be associated with the development of physical disability (18), as well as the measurement focus of FI on comorbidities and disability.

It is worth pointing out that our finding of the mortality association was in contradiction with the finding of Kulminski and colleagues using public access data from the CHS original cohort showing a greater power of the FI for discriminating the risk of death compared with the PFP (3). The differences for our study versus the Kulminski's study include longer follow-up (9 years vs. 4 years) and lower baseline frailty prevalence by the FI in our study (8.0% vs. 19.8%; which could be the result of using different cutoffs

for some of the component criteria of the FI). As indicated by our sensitivity analysis, reducing the study follow-up to 4 years produced similar findings. We endeavored to obtain data or programming code used in the previous study, but were unsuccessful, as a result, we were only able to recreate variables mimicking those reported in the Kulminski's study without certainty, which may also have contributed to the discrepancy.

Given the exploding interest in frailty research, accurate diagnosis of frailty is a fundamental first step in developing intervention strategies targeting frailty itself or for ameliorating its damaging health consequences. Although the frail people identified by the PFP exhibited different characteristics from those identified by the FI, one could argue that if the difference were not exceedingly large, it may still be that comparable populations are being identified—just with measurement error. Conversely, if the differences are remarkable as indicated by our findings, the underlying populations may be quite different. The fact that different frailty instruments may identify different subtypes of "frail" older adults not only limits our ability to cross-validate scientific findings generated using different instruments, but also impedes progress on etiologic discovery, the key to developing frailty prevention and treatment strategies. By focusing on the factors explaining the discordant frailty classification between the two most commonly used frailty instruments, the results from this study may help advance the discussion about the meaning of the word "frailty," which has varied widely from syndromic manifestations of underlying physiologic vulnerability captured by the PFP to the broad and multifactorial deficits including impairment, disability, and comorbidity measured by the FI. Although there is continuing debate on the dissociability of frailty from diseases and disability regarding their unique versus overlapping pathways, it is time to move beyond predictive validity to examine consistency of frailty diagnosis and its implications, as well as differential effectiveness of the different frailty instruments in accomplishing the aim(s) of frailty ascertainment.

Supplementary Material

Supplementary data is available at *The Journals of Gerontology, Series A: Biological Sciences and Medical Sciences* online.

Funding

Figure 2. Prevalence of being PFP-Frail (thick solid line), FI-Frail (thick dashed line), PFP-Frail/Pre-Frail (thin solid line), and FI-Frail/Pre-Frail (thin dashed line) by age in the combined Cardiovascular Health Study original and African American cohorts (*N* = 5,362). The distance between the solid lines (or between dashed lines) represents the prevalence of being PFP-Pre-Frail (or FI-Pre-Frail). PFP = physical frailty phenotype; FI = frailty index.

Author's Contributions

Q.-L.X. participated in concept design, interpretation, securing funding, drafting, critical revision, and approval of the article. J.T. participated in data analysis, interpretation, critical revision, and approval of the article. P.H.M.C. participated in critical revision and approval of the article. A.B.N. participated in data collection, critical revision, and approval of the
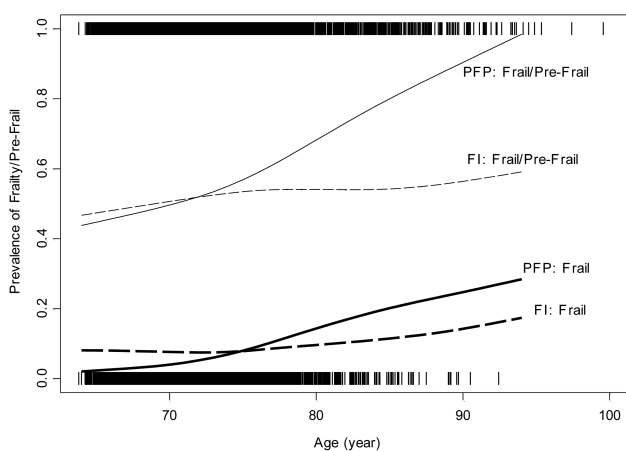
article. K.B.-R. participated in concept design, interpretation, securing funding, critical revision, and approval of the article.

Conflict of Interest

None reported.

References

1. Fried LP, Tangen CM, Walston J, et al.; Cardiovascular Health Study Collaborative Research Group. Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci*. 2001;56:M146–M156. doi:10.1093/gerona/56.3.M146

2. 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. doi:10.1016/j.arr.2015.12.003

3. Kulminski AM, Ukraintseva SV, Kulminskaya IV, Arbeev KG, Land K, Yashin AI. 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. doi:10.1111/j.1532-5415.2008.01656.x

4. 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. doi:10.1111/jgs.12420

5. Mitnitski A, Fallah N, Rockwood MR, Rockwood K. Transitions in cognitive status in relation to frailty in older adults: a comparison of three frailty measures. *J Nutr Health Aging*. 2011;15:863–867. doi:10.1007/s12603-011-0066-9

6. Rockwood K, Andrew M, Mitnitski A. A comparison of two approaches to measuring frailty in elderly people. *J Gerontol A Biol Sci Med Sci*. 2007;62:738–743. doi:10.1093/gerona/62.7.738

7. Woo J, Leung J, Morley JE. Comparison of frailty indicators based on clinical phenotype and the multiple deficit approach in predicting mortality and physical limitation. *J Am Geriatr Soc*. 2012;60:1478–1486. doi:10.1111/j.1532-5415.2012.04074.x

8. 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. doi:10.1111/j.1532-5415.2009.02225.x

9. Fried LP, Borhani NO, Enright P, et al. the cardiovascular health study: design and rationale. *Ann Epidemiol*. 1991;1:263–276.

10. Tell GS, Fried LP, Hermanson B, Manolio TA, Newman AB, Borhani NO. Recruitment of adults 65 years and older as participants in the Cardiovascular Health Study. *Ann Epidemiol*. 1993;3:358–366. doi:10.1016/1047-2797(93)90062-9

11. Steinberg D, Golovnya M. *CART 6.0 User's Manual*. San Diego, CA: Salford Systems; 2006.

12. Breiman L, Friedman J, Stone C, et al. *Classification and Regression Trees*. Monterey, CA: Wadsworth; 1984.

13. Prentice RL, Gloeckler LA. Regression-analysis of grouped survival data with application to breast-cancer data. *Biometrics*. 1978;34:57–67. doi:10.2307/2529588

14. Newman AB, Gottdiener JS, Mcburnie MA, et al.; Cardiovascular Health Study Research Group. Associations of subclinical cardiovascular disease with frailty. *J Gerontol A Biol Sci Med Sci*. 2001;56:M158–M166. doi:10.1093/gerona/56.3.M158

15. Morley JE, Vellas B, van Kan GA, et al. Frailty consensus: a call to action. *J Am Med Dir Assoc*. 2013;14:392–397. doi:10.1016/j.jamda.2013.03.022

16. Xue QL, Varadhan R. What is missing in the validation of frailty instruments? *J Am Med Dir Assoc*. 2014;15:141–142. doi:10.1016/j.jamda.2013.11.007

17. Walston JD, Bandeen-Roche K. Frailty: a tale of two concepts. *BMC Med*. 2015;13:185. doi:10.1186/s12916-015-0420-6

18. Klemera P, Doubal S. A new approach to the concept and computation of biological age. *Mech Ageing Dev*. 2006;127:240–248. doi:10.1016/j.mad.2005.10.004

19. Mitnitski AB, Graham JE, Mogilner AJ, Rockwood K. Frailty, fitness and late-life mortality in relation to chronological and biological age. *BMC Geriatr*. 2002;2:1.

20. Mitnitski A, Howlett SE, Rockwood K. Heterogeneity of human aging and its assessment. *J Gerontol A Biol Sci Med Sci*. 2017;72:877–884. doi:10.1093/gerona/glw089