

Frailty Phenotypes and Relations With Surgical Outcomes: A Latent Class Analysis

Furqaan Sadiq, MD, MSCI,* Vanessa L. Kronzer, MD, MSCI,† Troy S. Wildes, MD,* Sherry L. McKinnon, BS,* Anshuman Sharma, MD, MBA,* Daniel L. Helsten, MD,* Lawrence M. Scheier, PhD,‡ Michael S. Avidan, MBBCh,* and Arbi Ben Abdallah, DES, PhD*

BACKGROUND: Frailty is an important concept in the care of older adults although controversy remains regarding its defining features and clinical utility. Both the Fried phenotype and the Rockwood deficit accumulation approaches cast frailty as a “burden” without exploring the relative salience of its cardinal markers and their relevance to the patient. New multifactorial perspectives require a reliable assessment of frailty that can validly predict postoperative health outcomes.

METHODS: In a retrospective study of 2828 unselected surgical patients, we used item response theory to examine the ability of 32 heterogeneous markers capturing limitations in physical, functional, emotional, and social activity domains to indicate severity of frailty as a latent continuum. Eighteen markers efficiently indicated frailty severity and were then subject to latent class analysis to derive discrete phenotypes. Next, we validated the obtained frailty phenotypes against patient-reported 30-day postoperative outcomes using multivariable logistic regression. Models were adjusted for demographics, comorbidity, type and duration of surgery, and cigarette and alcohol consumption.

RESULTS: The 18 markers provided psychometric evidence of a single reliable continuum of frailty severity. Latent class analyses produced 3 distinct subtypes, based on patients’ endorsement probabilities of the frailty indicators: not frail (49.7%), moderately frail (33.5%), and severely frail (16.7%). Unlike the moderate class, severely frail endorsed emotional health problems in addition to physical burdens and functional limitations. Models adjusting for age, sex, type of anesthesia, and intraoperative factors indicated that severely frail (odds ratio, 1.89; 95% confidence interval, 1.42–2.50) and moderately frail patients (odds ratio, 1.31; 95% confidence interval, 1.03–1.67) both had higher odds of experiencing postoperative complications compared to not frail patients. In a 3-way comparison, a higher proportion of severely frail patients (10.7%) reported poorer quality of life after surgery compared to moderately frail (9.2%) and not frail (8.3%) patients ($P < .001$). There was no significant difference among these groups in proportions reporting hospital readmission (5.6%, 5.1%, and 3.8%, respectively; $P = .067$).

CONCLUSIONS: Self-report frailty items can accurately discern 3 distinct phenotypes differing in composition and their relations with surgical outcomes. Systematically assessing a wider set of domains including limitations in functional, emotional, and social activities can inform clinicians on what precipitates loss of physiological reserve and profoundly influences patients’ lives. This information can help guide the current discussion on frailty and add meaningful clinical tools to the surgical practice. (Anesth Analg 2018;127:1017–27)

KEY POINTS

- **Question:** Frailty is a prominent premorbid characteristic of many older adults; however, its conceptualization, measurement, and predictive validity to relevant surgical outcomes remain elusive.
- **Finding:** Frailty can be characterized as a latent continuum denoting severity using 18 heterogeneous markers capturing physical burdens, functional limitations, and diminished emotional and social activities, all of which can be used to characterize 3 distinct phenotypes with specific predictive validity.
- **Meaning:** Frailty assessments should be reliable, parsimonious, and require good discriminative and predictive capability for relevant postoperative outcomes to have clinical utility.

From the *Department of Anesthesiology, Institute of Quality Improvement, Research & Informatics (INQUIRI), Washington University School of Medicine, St Louis, Missouri; †Department of Internal Medicine, Mayo Clinic, Rochester, Minnesota; and ‡LARS Research Institute, Inc, Scottsdale, Arizona. Accepted for publication July 2, 2018.

Funding: Research reported in this publication was supported by the Washington University Institute of Clinical and Translational Sciences grant UL1TR000448, subaward TL1TR000449, from the National Center for Advancing Translational Sciences of the National Institutes of Health. Systematic Assessment and Targeted Improvement of Services Following Yearlong Surgical Outcomes Surveys has been funded by the Washington University Department of Anesthesiology and the Barnes-Jewish Hospital

Copyright © 2018 International Anesthesia Research Society
DOI: 10.1213/ANE.00000000000003695

Foundation (St Louis, MO; Award Reference Number: 7937–77 to M.S.A.). F.S. was supported by a Foundation for Anesthesia Education and Research Medical Student Anesthesia Research Fellowship.

The authors declare no conflicts of interest.

The content of this study is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Health or Barnes-Jewish Hospital Foundation.

Reprints will not be available from the authors.

Address correspondence to Arbi Ben Abdallah, DES, PhD, Department of Anesthesiology, Institute of Quality Improvement, Research & Informatics (INQUIRI), Washington University School of Medicine, 660 S Euclid, Campus Box 8054, St Louis, MO 63110. Address e-mail to aba@anest.wustl.edu.

The concept of frailty has garnered considerable attention in the past few years, becoming a focus of clinical discussion and empirical scrutiny. This is most likely because a substantial body of research documents that frail older people experience poor health outcomes that cannot only be explained by age, functional status, or comorbid diseases. Those who consider frailty a common geriatric “syndrome” broadly define it as a loss of function, decline in physiological reserve, impaired resistance to stressors, and a state of increased susceptibility to disease and illness.¹⁻⁵ Frailty is distinguished from disability or comorbid diseases and aligned more with risk “resulting from age- or disease-associated physiologic accumulation of subthreshold decrements affecting multiple physiologic systems.”⁶ There is now growing evidence documenting that preoperative frailty is a valid indicator of postoperative complications,^{7,8} including delirium,^{9,10} falls,^{11,12} increased length of hospitalization,¹³ discharge to a nursing or assisted living facility,¹⁴ hospital readmissions,¹⁴ and other surgical complications.^{8,15} In addition, frailty among community-dwelling older people is associated with reduced quality of life¹⁶ and is a valid predictor of all-cause mortality.^{8,17,18}

Despite the clinical importance of frailty, there is considerable debate regarding its most prominent features, how it should be measured, and its clinical significance.^{2,3,5} Notably, 2 conceptual models have influenced research examining frailty and its outcomes. Fried et al,¹ using data from the Cardiovascular Health Study, proposed a physical phenotype model based on 5 salient criteria. These include excessive weight loss, exhaustion, slow gait speed, weak handgrip strength, and sedentary behavior defined by low physical activity. The number of criteria present determines a patient’s frailty status, with 0 indicating robust or healthy, 1–2 indicating prefrailty, and ≥ 3 indicating frailty.

Alternatively, Rockwood et al¹⁹ and Rockwood and Mitnitski²⁰ proposed that frailty represents an accumulation of deficits, necessitating markers from multiple domains and a broader collage of symptoms to fully portray its underlying clinical features. In a multifactorial framework, a person’s frailty is based on the sheer number of markers (ie, symptoms, signs, diseases, clinical impressions) indicated out of the total, a proposition that allows for different measures and questionnaire length, without undermining the instruments’ diagnostic utility. When pitted against each other and using the same data source, the deficit approach underestimates mortality less frequently.²¹ Overall, whether one chooses to align with the phenotype or the deficit approach, they both consider frailty as a reflection of deficits, just basing this impression on a smaller subset of muscular weakness and physical burden markers in the former and a larger, if not more holistic, cache of deficits in the latter.

Notwithstanding their applications, both approaches have their own shortcomings. The cumulative deficit approach uses a single additive score and as a result glosses over the salience of different domains. It is quite likely that some markers are more clinically meaningful to the individual. For example, incapacity and its effect on routine activities of daily living (ADL) may have greater clinical importance to the patient because the restrictions on their life are more personally salient, than say unintentional weight loss or low physical activity. Understanding the

relative importance of different domains would help clarify underlying vulnerability and, from a medical standpoint, facilitate rendering differential diagnosis. A summative index also leaves unresolved the “prognostic” value of different items and their relative contribution to the clinical syndrome.²

The Fried phenotype provides an opportunity to examine the role of physical burden with regard to a coherent medical syndrome.²²⁻²⁵ However, even using a reduced set of items, researchers have been unable to obtain consistent phenotypes across different studies.²⁵ This opens the door for different clinical profiles and undermines empirical attempts to clarify a “syndrome.” Moreover, like the frailty index, the predictive accuracy of various markers is never addressed; each of the 5 criteria is weighted equally in determining frailty phenotypes. Sensitivity, specificity, and receiver operating characteristic analyses have not yielded a consensus with respect to the different markers’ ability to differentiate functional outcomes.^{24,26-29} In addition, studies have shown that as few as 3 physical burden items net the same predictive validity as the 5 Fried criteria in predicting incident falls, disability, fractures, and death as far out as 9 years in older women.³⁰ More recently, studies have shown that increasing the number of items with the addition of polypharmacy and self-reported health can enhance predictive validity of the Fried index.³¹ This finding holds for instruments adding measures of cognitive functioning assessing memory and speed of processing.^{32,33}

Additional drawbacks to the Fried phenotype method are that 3 criteria are measured on continuous scales and must be dichotomized using a 20th percentile rule, necessitating a reference population, which can be onerous to clinicians.³⁴ A study of German community-dwelling older people found high correspondence between the traditional lower quintile approach to designate frailty and a population-independent cut-point approach.³⁵ Cut-points were derived from the literature and resulted in higher frailty prevalence rates than those reported using the traditional Fried method. Finally, several studies that derived subtypes matching the Fried phenotypes validated them using functional health markers (ie, ADL), thus conflating associations between frailty and disability-related outcomes.^{19,25,26}

STUDY OBJECTIVES

The present study was designed to resolve 2 pressing concerns, using an integrated set of psychometric and statistical techniques that should help refine our understanding of frailty and improve its predictive accuracy. We first used item response theory (IRT) to ascertain whether a single frailty “liability” can be derived based on markers reflecting physical burden, nutrition/fitness and functional limitations, emotional health, and social activity domains. This portion of the analysis reveals the discriminative properties of markers within domains and their ability to detect frailty severity. We then used latent class analysis (LCA) to derive frailty subtypes using markers identified from the IRT analysis. The LCA can detect whether there are distinct subtypes of frailty that characteristically differ from each other based on a more parsimonious set of markers. Classes were then characterized based on demographic and health covariates and comorbidities that could potentially confound adverse

outcomes. Finally, we used multivariable logistic regression (MLR) to examine relations between class membership and postsurgical complications, readmission, and self-reported change in quality of life.

METHODS

Data Collection and Study Population

The Human Research Protection Office at Washington University School of Medicine approved this study (ID: 201408106) as a substudy of the Systematic Assessment and Targeted Improvement of Services Following Yearlong Surgical Outcomes Surveys (SATISFY-SOS, NCT02032030).³⁶ SATISFY-SOS is an electronic data registry that has collected since 2012 patient-reported outcome surveys from adult patients (age, ≥ 18 years) undergoing general elective surgeries at Barnes-Jewish Hospital in St Louis, MO. Patients were consented to participate in the SATISFY-SOS project during their preoperative clinical evaluation at the Barnes-Jewish Hospital Center for Preoperative Assessment and Planning (CPAP). Specifically, the registry includes patient-reported information from a baseline survey completed preoperatively at the time of consent, and data from follow-up surveys completed by the same patients approximately 30 days and 1 year postoperatively. This article adheres to publication guidelines outlined by the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement for reporting observational studies.

Data available for this study covered 7 months in 2014. Preoperative clinical information came from the hospital's electronic medical record (MetaVision [iMDsoft, Needham, MA]) and is routinely collected during patient's preoperative evaluation, which is conducted at the CPAP. At that time, the total number of surgical patients who were assessed preoperatively at the CPAP and had surgery was 18,735, of which 12,877 consented over the entire 2014 period. The number of actual surgeries performed in the 7-month window selected for this study involved 10,491 patients, of which 7043 (67%) consented to have their data included in the SATISFY-SOS registry and 4042 (57% of those consented) actually completed both baseline and 30-day surveys, as of the beginning of this investigation. We then excluded patients with missing data in any of the selected data fields (demographic variables, frailty items, and patient-reported outcomes), yielding an analysis sample of 2828 patients with complete information.

An extensive literature review produced 52 common frailty measures across different assessment strategies. A total of 32 of these matched up against our medical institution's preoperative anesthesia electronic medical records. Sixteen of these measures are routinely collected as part of preoperative assessment, and an additional 16 were culled from SATISFY-SOS Baseline Survey completed by each consented patient during their first presurgical office visit. The latter set includes 9 items from the baseline Veteran RAND 12 (VR-12; RAND Corp, Santa Monica, CA) plus its aggregate scores as well as each patient's response to 5 additional questions asked in the same survey. All 32 items are either categorical (nominal and ordinal scales) or continuous measures. Although these types of variables do not challenge the robustness of our statistical methods, for simplicity, we dichotomized all the data using

prespecified cut-points relevant to conceptual models of frailty. Supplemental Digital Content, Table 1S, <http://links.lww.com/AA/C524>, shows the final set of measures and comparisons between patients lost to follow-up and the final analytic sample.

Statistical Analysis

Using IRT, we tested a 2-parameter graded response model, which is appropriate for binary and polytomous (categorical) frailty markers.³⁷ The IRT model produces 2 parameters that detail properties of the markers (ie, the item characteristic function). The difficulty (location) parameter indicates the probability of responding to the marker affirmatively at varying levels of frailty (trait) severity. When difficulty parameters are relatively large, they indicate that patients require very high frailty severity scores to endorse the item. The discrimination (slope) parameter reveals how rapidly the probabilities changes with frailty severity. When discrimination (slope) parameters are large, they indicate that the probability of endorsing this item increases with corresponding increases in frailty severity (eg, steeper slopes). Items that discriminate well should have difficulty (location) parameters that generally range from -3 to 3 and discrimination (slope) values ranging from 0 to 3 . The item characteristic curve (ICC) is a function that provides the relationship between the underlying frailty continuum and the estimated probability of scoring a positive value on the individual marker.

The 2-parameter graded response models were tested within each of the 4 domains, including 7 physical burden markers, 7 emotional health markers, 3 markers from the nutritional domain combined with 11 from the fitness and functional limitations domain, and 4 from the social support domain (including employment as a proxy for social activity). Once we obtained a final set of efficient markers, we then used confirmatory factor analysis (CFA) to ascertain the dimensional structure of the resulting frailty markers. The CFA model used weighted least squares means and variance estimation and robust standard errors, which is appropriate for categorical data. We then used LCA to examine frailty phenotypes using the markers that performed well in the IRT analysis. This procedure assigns patients to mutually exclusive classes based on their patterns of endorsing markers from the different domains. The resulting classes are qualitatively discrete so that patients within a class are more "homogeneous" with regard to frailty experiences than patients belonging to a different class. The "best-fitting" LCA model is selected by comparing a base undifferentiated 1-class model with models that increase in complexity and extract additional classes. Specifically, the selected model should provide superior statistical indicators of overall model fit (lower values for Akaike Information Criterion, the Bayesian Information Criterion, and/or the sample size-adjusted Bayesian Information Criterion), a significant Lo-Mendell-Rubin Adjusted Likelihood Ratio Test ($P < .05$) and provide a meaningful interpretation of the resulting classes.³⁸

Based on estimated posterior probabilities, we then assigned patients to their respective classes and examined relations between class membership and 30-day surgical outcomes including postoperative complications (any problem = 1), change in quality of life 30 days after surgery (worse = 1), and hospital readmission (yes = 1) obtained

from the SATISFY-SOS 30-day survey. The latter analysis was conducted using MLR, appropriate for dichotomous outcomes. These analyses controlled for disease comorbidity as well as type and duration of surgery (minutes), surgical risk (cardiac, intermediate, or high), cigarette and alcohol consumption, and demographic factors including age, sex, race (ethnicity), and employment status.

Preliminary data management was performed using SAS Version 9.4 (SAS Institute Inc, Cary, NC) and statistical analysis for the IRT and LCA models using the Mplus statistical software package (Muthén & Muthén, Los Angeles, CA).³⁹ Results are presented with exact *P* values and 95% confidence intervals (CIs) for measures of parameter precision. Unadjusted *P* values <.05 were considered statistically significant.

RESULTS

IRT Model

We began with 32 frailty markers reflecting the 4 domains (Supplemental Digital Content, Table 1S, <http://links.lww.com/AA/C524>). The preliminary IRT analysis conducted separately within each domain allowed us to eliminate 14 markers with poor psychometric properties, either having relatively low (or negative) difficulty scores and/or weak discrimination parameters and because the corresponding ICCs visually reinforced poor item performance.

Table 1 shows the IRT results after combining the 18 successful frailty markers from the individual domain analyses into a single model. In the physical burden domain, only poor physical quality of life (based on the VR-12 physical component score <41) had a high discrimination (slope >3.0) index (endorsed by 48% of the sample). Key markers in the fitness and functional domain that discriminated frailty severity included slow physical activities (a proxy for slow gait speed: 34%), limited in moderate ADL (59%), inability to work due to physical health (PH) (12%), and accomplish less than would like due to emotional and/or PH (49%). None of the markers in the emotional health

domain exceeded the benchmark of 3.0; however, poor emotional/mental quality of life (based on the VR-12 mental/psychological component score <41: 13%) and limited social activities due to emotional and/or PH (25%) came near the benchmark of 3.0 (2.43 and 2.36, respectively).

Supplemental Digital Content, Figure 1S, <http://links.lww.com/AA/C524>, shows the ICCs for the 18 items combined, and depicts the characteristic “S” or ogive-shaped curve for each item. The curve is monotonic, reinforcing that items used in the analysis assess frailty across a breadth of range; however, a bulk of the items performs better in distinguishing higher levels of frailty severity. In other words, as patients become increasing frail (higher scores on the x-axis), they should have a higher probability (y-axis) of endorsing the markers used to assess frailty severity. Although not shown, individual ICCs conducted within each domain revealed that 2 of the functional limitation items were more informative at the negative or lower end of the frailty continuum (eg, limited in moderate ADLs and limited climbing stairs). In other words, these items function well to distinguish patients with very low levels of frailty.

One assumption of the IRT model applied is that the underlying latent continuum (trait) is unidimensional.⁴⁰ We tested a CFA model positing a single latent dimension of frailty reflected by the 18 items. Factor loadings were allowed to vary freely, and the data were supplied as tetrachoric correlations. The fit of this unidimensional model was acceptable, comparative fit index = 0.938, Tucker Lewis index = 0.930; root mean square error of approximation = 0.103 (CIs, 0.101–0.106). The unstandardized factor loadings ranged from 0.501 (PH5: muscle weakness) to 1.03 (fitness and functional limitations–item 6 [FFL6]: exhibit slow physical activities). Although some of the benchmark fit indices showed room for improvement (ie, root mean square error of approximation should be <0.08), this would only occur with the addition of post hoc model refinements (ie, correlated errors). This approach is not recommended for IRT

Table 1. Final IRT Model for the 18 Frailty Markers

Item Description	Percent Endorsement of Item (N)	Difficulty (Location)	Standard Error	Discrimination (Slope)	Standard Error
PH1: poor general health status	15.8 (448)	1.55	0.08	1.45	0.07
PH2: worsening physical health	36.2 (1024)	0.65	0.06	1.09	0.05
PH4: physical disability work status	10.6 (301)	2.25	0.08	1.15	0.13
PH5: muscle weakness	16.7 (473)	1.91	0.07	0.99	0.11
PH7: poor physical quality of life	47.9 (1354)	0.07	0.19	3.60	0.03
FFL4: exhibit low energy	17.2 (485)	1.35	0.10	1.70	0.06
FFL6: exhibit slow physical activities	33.6 (950)	0.48	0.23	3.96	0.03
FFL8: worsening disability activities	6.9 (194)	1.74	0.21	2.82	0.06
FFL9: limited in moderate ADLs	58.8 (1663)	-0.25	0.17	3.23	0.03
FFL10: limited ability in kind of work due to PH	11.8 (335)	1.30	0.26	3.64	0.04
FFL11: limited ability to work due to EP	30.2 (855)	0.74	0.08	1.68	0.04
FFL12: accomplish less due to PH/EP	49.0 (1386)	0.04	0.29	4.78	0.03
FFL14: limited climbing flights of stairs	54.6 (1543)	-0.14	0.10	2.01	0.03
MH1: felt downhearted and blue	25.2 (712)	1.07	0.07	1.36	0.05
MH3: felt calm and peaceful	32.5 (920)	0.73	0.07	1.34	0.05
MH4: worsening emotional health	22.7 (641)	1.38	0.07	1.08	0.08
MH7: poor mental quality of life	13.3 (377)	1.36	0.15	2.43	0.05
SASS4: limited social activities	25.1 (709)	0.84	0.12	2.36	0.04
		Mean = 0.95 (SD = 0.72)		Mean = 2.26 (SD = 1.16)	

Abbreviations: ADL, activities of daily living; EP, emotional problem; FFL, fitness and functional limitations; IRT, item response theory; MH, mental health; PH, physical health; SASS, social activities and social support.

Table 2. Model Fit Statistics for (1–10) Latent Class Models of Frailty Phenotypes

Latent Class Count	NFP	Log Likelihood	Akaike (AIC)	Bayesian (BIC)	Adjusted BIC	Entropy	Adj. LRT P Value
1-Class	18	-27,270.83	54,577.66	54,684.71	54,627.52
2-Class	37	-22,008.99	44,091.98	44,312.03	44,194.46	0.928	<.001
3-Class	56	-20,677.84	41,467.68	41,800.73	41,622.80	0.911	.028
4-Class	75	-20,170.14	40,490.27	40,936.32	40,698.02	0.902	.064
5-Class	94	-19,736.48	39,660.97	40,220.02	39,921.35	0.898	.396
6-Class	113	-19,567.40	39,360.79	40,032.84	39,673.80	0.891	.319
7-Class	132	-19,421.54	39,107.07	39,892.12	39,472.71	0.887	.384
8-Class	151	-19,308.10	38,918.20	39,816.25	39,336.47	0.883	.531
9-Class	170	-19,217.60	38,775.21	39,786.25	39,246.10	0.885	.565
10-Class	189	-19,164.91	38,707.82	39,831.86	39,231.34	0.876	.576

Adjusted BIC = sample size-adjusted BIC; entropy = a measure of classification error with higher numbers indicating clear delineation of classes; and Adj. LRT = Lo-Mendell-Rubin adjusted test, testing whether the current (K-class) model is preferred to the ([K-1]-class) model with 1 less class.

Abbreviations: AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion; LRT, likelihood ratio test; NFP, number of free parameters.

Table 3. Item Endorsement Probabilities in Each of the 3-Class Model

Frailty Indicators	Class 1 Not Frail	Class 2 Moderately Frail	Class 3 Severely Frail
Probability of endorsing the physical health markers			
PH1: poor general health status	0.036	0.200	0.440
PH2: worsening physical health	0.187	0.478	0.650
PH4: physical disability work status	0.022	0.154	0.264
PH5: muscle weakness	0.057	0.252	0.326
PH7: poor physical quality of life	0.056	0.960	0.771
Probability of endorsing the fitness and functional limitation markers			
FFL4: exhibit low energy	0.031	0.207	0.517
FFL6: exhibit slow physical activities	0.006	0.648	0.692
FFL8: worsening disability activities	0.001	0.015	0.379
FFL9: limited in moderate ADLs	0.223	0.978	0.894
FFL10: limited ability in kind of work due to PH	0.000	0.076	0.555
FFL11: limited ability to work due to EP	0.121	0.274	0.899
FFL12: accomplish less due to PH/EP	0.102	0.828	0.968
FFL14: limited climbing flights of stairs	0.249	0.863	0.790
Probability of endorsing the emotional/mental health markers			
MH1: felt downhearted and blue	0.120	0.156	0.836
MH3: felt calm and peaceful	0.164	0.286	0.883
MH4: worsening emotional health	0.126	0.162	0.655
MH7: poor mental quality of life	0.018	0.000	0.744
SASS4: limited social activities	0.035	0.300	0.794
Class membership proportions	49.7%	33.5%	16.7%
Cross-classification probabilities ^a	.959	.968	.933

Abbreviations: ADL, activities of daily living; EP, emotional problem; FFL, fitness and functional limitations; MH, mental health; PH, physical health; SASS, social activities and social support.

^aClassification probabilities for the most likely class membership as defined by the model.

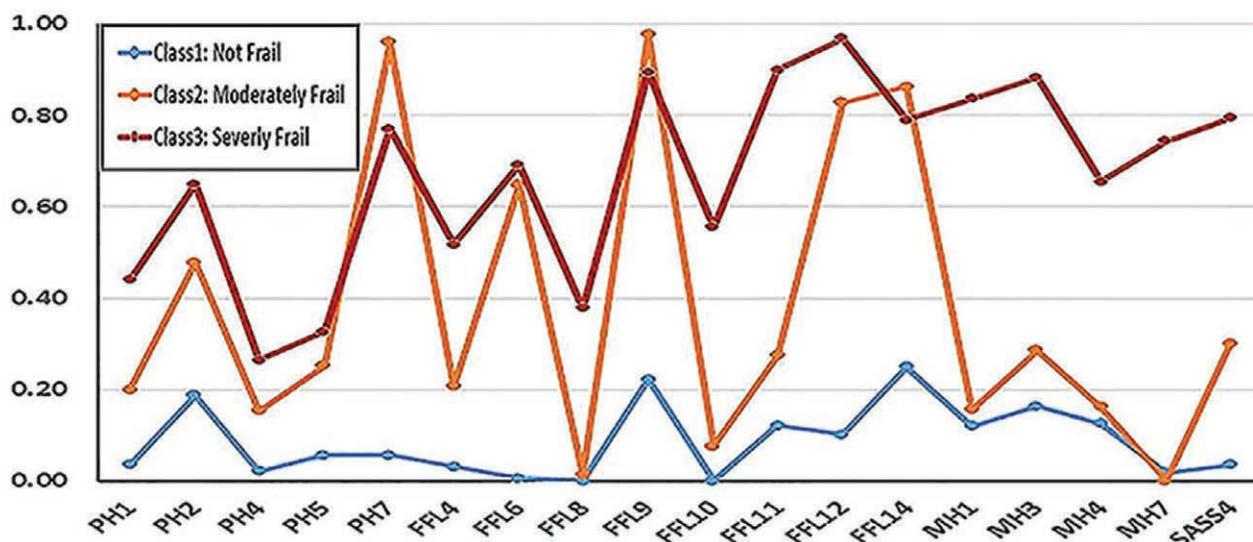
models, as it violates the statistical assumption of local independence (ie, responses are not contingent on one another, even after controlling for the trait).

LCA Model

Table 2 provides fit statistics corresponding to the sequence of LCA models tested with the 18 surviving categorical indicators (dichotomized markers). This included a base 1-class or undifferentiated model (ie, a single undifferentiated population) and up to a 10-class model. Following the rule of parsimony, all model fit indices suggest that the 3-class model provides the best fit and makes the most conceptual sense. With each successive extraction of a class, the Bayesian Information Criterion becomes progressively smaller and shows a negligible change between the 3-class (41,800) and 4-class models (40,936). The corresponding G^2 and L^2/df (similar to an F-statistic), and classification error statistics as well as the adjusted likelihood ratio test P value also support

the 3-class model.⁴¹ Increasing extraction of classes past the 3-class model has a marginal effect on the overall model fit and results in smaller pockets of patients forming unstable classes that are not clinically meaningfully (ie, do not form discrete phenotypes that vary in composition).

Table 3 details the item endorsement probabilities for the 3-class model using the 18 frailty markers. This should be read in conjunction with the Figure, which graphically portrays the 3-class model. A reasonable item endorsement threshold is 60% or higher, so a measure typically must reach this benchmark to be considered an efficient marker of class membership. Class 1 labeled as “Not Frail” (49.7%) consists of patients that endorse a minimum of frailty markers (average item response probability was $\rho = 0.090$). The largest probability, difficulty climbing flights of stairs (0.249), is far below the desired 0.60 threshold. Class 2 (33.5%) consists of “Moderately Frail” patients who endorse 5 markers above the critical threshold (average item response

**Label:****Physical Health Markers**

PH1: Poor general health status
 PH2: Worsening physical health
 PH4: Physical disability work status
 PH5: Muscle weakness
 PH7: Poor physical Quality of Life

Fitness and Functional Limitation Markers

FFL4: Exhibit low energy
 FFL6: Exhibit slow physical activities
 FFL8: Worsening disability activities
 FFL9: Limited in moderate ADL
 FFL10: Limited ability in kind of work due PH
 FFL11: Limited ability to work due to EP
 FFL12: Accomplish less due to PH/EP
 FFL14: Limited climbing flights of stairs

Emotional/Mental Health Markers

MH1: Felt downhearted and blue
 MH3: Felt calm and peaceful
 MH4: Worsening emotional health
 MH7: Poor mental quality of life
 SASS4: Limited social activities

Figure. Endorsement pattern of the 3 latent class phenotypes. ADL indicates activities of daily living; EP, emotional problem; FFL, fitness and functional limitations; MH, mental health; PH, physical health; SASS, social activities and social support.

probability was $\rho = 0.380$). These include poor physical quality of life ($\rho = 0.960$), slow performance of physical activities ($\rho = 0.648$), limited in moderate ADL ($\rho = 0.978$), accomplish less because of physical/emotional problems ($\rho = 0.828$), and limited in climbing flights of stairs ($\rho = 0.863$).

The remaining class, labeled "Severely Frail" (16.7%), consists of patients who endorsed 11 frailty markers across all 3 domains (physical, functional limitations, and emotional health), all exceeding the 0.60 threshold (average item endorsement probability was $\rho = 0.669$). Members of this class endorsed 2 of the physical burden items (worsening PH and poor physical quality of life), but they also endorsed 5 of the functional limitation markers >0.60 (average $\rho = 0.849$). Notably, all 5 of the emotional health markers exceeded 0.60, suggesting that this class experienced a unique level of frailty that combined physical burdens and emotional health problems. Classification based on estimated posterior probabilities was quite sound as evidenced by the very accurate cross-classification probabilities (0.959, 0.968, and 0.933 for classes 1, 2, and 3, respectively).

Frail Patient Characteristics

Table 4 shows that compared to patients assigned to the reference not frail class, those in the severely frail class included the largest proportion of female patients, ethnic minorities, unemployed, and a greater proportion of patients whose American Society of Anesthesiologists physical status score

was 3 or higher. A greater proportion of severely frail patients had relatively higher Charlson Comorbidity Index (CCI) scores than those assigned to the not frail and moderately frail classes. Comparison of their baseline health status also indicated that severely frail patients had statistically significant higher proportions of stroke (7.7%), deep vein thrombosis or pulmonary embolism (9.2%), hypertension (57%), and cancer (31%) (not shown in Table 4). Moreover, the CCI score increased across the 3 frailty phenotypes from 15% to 18% to 23% in the not frail, moderately frail, and severely frail classes, respectively. In regard to additional features that could distinguish the classes, the proportion of patients undergoing high-risk surgery increased from 18% to 29.3% and 32.3% in the not frail, moderately frail, and severely frail classes, respectively. The association of frailty with age and general anesthesia was monotonic. The moderately frail class was older on average than both the severely frail and the not frail classes. Also, larger proportions of the patients in the not frail (82.7%) and the severely frail (83.5%) received general anesthesia than those in the moderately frail class (72.4%).

Frailty Phenotypes and Postoperative Outcomes

The overall proportion of patient-reported complications in-hospital and/or 30 days postdischarge was 16.3% (Supplemental Digital Content, Table 2S, <http://links.lww.com/AA/C524>). Among patients reporting in-hospital complications (12.9%), there was a statistically significant

Table 4. Patient Characteristics Across the 3 Latent Class Frailty Phenotypes

Characteristics	Total, N = 2828, n (%)	Not Frail, 1384 (49.7%), %	Moderately Frail 978 (33.5%), %	Severely Frail 466 (16.7%), %	P Value
Female	1687 (59.7%)	57.1	62.3	61.8	.023
Mean age (SD), (y)	58 (14.19)	57 (14.87)	59 (13.83)	56 (12.64)	.005
Non-Caucasian ethnic minorities	314 (11.1%)	10.4	10.3	14.8	.021
Employed	1298 (45.9%)	53.9	41.0	32.4	<.001
BMI (kg/m ²), median (IQR)	30 (25–34)	28 (24–32)	30 (25–35)	30 (26–36)	<.001
ASA class ≥3	1039 (36.7%)	28.1	42.7	49.8	<.001
Charlson Comorbidity Index ≥4	492 (17.4%)	14.8	18.2	23.4	<.001
Surgery type					<.001
Cardiothoracic	199 (7.0%)	7.1	6.5	7.9	
Orthopedic	774 (27.4%)	14.0	44.8	30.7	
Other major surgery ^a	948 (33.5%)	41.3	24.7	28.8	
Other minor surgery	907 (32.1%)	37.6	23.9	32.6	
Surgery risk					<.001
Low risk	958 (33.9%)	43.6	20.5	25.9	
Medium risk	1198 (42.4%)	38.4	50.3	41.9	
High risk	672 (23.8%)	18.0	29.3	32.2	
General anesthesia	2260 (79.9%)	82.7	72.4	83.5	<.001
Surgery duration (min), median (IQR)	102 (59–163)	98 (51–163)	101 (65–157)	113 (69–177)	.001

Abbreviations: ASA, American Society of Anesthesiologists; BMI, body mass index; IQR, interquartile range.

^aOther major surgery: colorectal, hepatobiliary-gastrointestinal, transplant, vascular, otolaryngology, neurosurgery, plastics.

difference ($P < .001$) in the rate of complications, which increased progressively across the 3 phenotypes: not frail (9.4%), moderately frail (13.7%), and severely frail (21.7%). A relatively low percent of patients (4.6%) was readmitted 30 days postdischarge from their surgery. Readmission rates did not differ significantly between the frailty phenotypes ($P < .067$: 3.8%, 5.1%, and 5.6%, respectively, for not frail, moderately frail, and severely frail patients). Quality of life (QoL) reported in 30-day postoperative surveys differed significantly between the 3 frailty phenotypes ($P < .001$: 8.3%, 9.2%, and 10.7%, respectively, for not frail, moderately frail, and severely frail patients reporting “worse” QoL than before surgery).

Table 5 shows the results of the MLR analysis testing whether frailty phenotypes contributed uniquely to postoperative complications, hospital readmission, and quality of life, independent of demographic, baseline comorbidities, intraoperative, and behavioral factors that could contribute to morbidity. Compared to the reference not frail phenotype, moderately frail patients were 31% more likely (odds ratio [OR], 1.31; 95% CI, 1.03–1.68) and severely frail patients were 89% more likely (OR, 1.89; 95% CI, 1.42–2.50) to report postoperative complications ($P < .001$). The log likelihood ratio and Hosmer–Lemeshow goodness of fit tests both showed adequate model fit, with an acceptable C-statistic (receiver operating characteristic = 0.67) indicating adequate model fit.

Although there was a 22% greater likelihood of hospital readmission among the moderately frail (OR, 1.22; 95% CI, 0.80–1.86) and 24% greater likelihood for severely frail patients (OR, 1.24; 95% CI, 0.74–2.06) compared to not frail patients, this difference was not statistically significant ($P = .067$). A single item in the 30-day postoperative survey asks patients, “How would you rate your quality of life now?” Most patients in the study cohort (63%) stated that their quality of life (QoL) was “better” than before surgery, and this trend was representative of all 3 phenotypes (Supplemental Digital Content, Table 2S, <http://links.lww.com/AA/C524>).

Only 9% of patients rated their QoL as “worse” after surgery. While 8.3% of patients assigned to the not frail phenotype reported worse perception of their QoL, 9.2% and 10.7%, respectively, of the patients assigned to the moderate and severe frailty classes reported worse perception of their quality of life ($P < .001$). Only the moderately frail patients were less likely to report worsening QoL compared to the not frail class (OR, 0.57; 95% CI, 0.47–0.69), controlling for the other measures in the model.

DISCUSSION

This retrospective cohort study examined 2 principal issues that revolve around assessing frailty and its implications in surgical settings. First, we addressed whether frailty can be conceptualized as an underlying continuum, reflecting gradations of severity. We felt this to be a pressing issue in the conceptualization of frailty, which has relied up until now on additive “count” approaches, failing to distinguish the salience of markers assessing frailty. The Fried phenotype approach uses 5 physical burdens to make its case regarding a medical syndrome, whereas the deficit accumulation approach uses a wider set of markers ranging from 20 to 70, depending on which index is applied. In either case, the constituent elements of the index or phenotype are all treated equally. This lends no support to figuring out the underlying “cause” of frailty, or which system is failing and needs greatest attention, undermining practitioner’s attempts to render a suitable clinical diagnosis and specifically tailor interventions. Our study shows that it is best to conceptualize frailty as an amalgam of deficits in multiple domains. To make this case, we also used a statistical procedure that produces parameters that are model dependent and independent of sample characteristics within a population. As a result, derivation of a more parsimonious set of frailty markers using the current sample should not hinder researchers from empirically confirming the same markers or their utility in different samples.

Table 5. Multivariable Logistic Regressions Predicting Postoperative Outcomes: Complications, Readmission, and Quality of Life 30 Days After Surgery

Predictors	Postoperative Complications (Overall) 460 (16.3)		Hospital Readmission 129 (4.6%)		Quality of Life ^a 244 (9.0%)	
	OR	95% CI	OR	95% CI	OR	95% CI
Age (y)	1.00	0.99–1.01	1.01	0.99–1.02	1.00	0.99–1.01
Female	0.99	0.80–1.23	0.97	0.67–1.42	1.05	0.89–1.24
Ethnic minority	1.17	0.85–1.61	0.64	0.32–1.29	0.75 ^b	0.57–0.99
Currently employed	0.82	0.65–1.03	1.02	0.67–1.54	0.97	0.81–1.16
Charlson comorbidity ≥ 4	1.54 ^c	1.12–1.94	2.01 ^c	1.36–2.98	1.25 ^b	1.03–1.52
Frailty ^d						
Moderately frail	1.31	1.03–1.68	1.22	0.80–1.86	0.57 ^e	0.47–0.69
Severely frail	1.89 ^c	1.42–2.50	1.24 ^c	0.74–2.06	0.71	0.56–0.90
Risk of surgery ^e						
Medium risk	1.15	0.89–1.50	1.73	1.07–2.79	0.74	0.61–0.89 ^b
High risk	1.72 ^c	1.28–2.30	1.57	0.91–2.70	0.78	0.62–0.99
General anesthesia	1.70 ^c	1.24–2.34	3.30 ^c	1.63–6.68	1.71 ^c	1.37–2.14
Length of surgery	1.00 ^c	1.00–1.01	1.00	1.00–1.01	1.00	1.00–1.01
Current tobacco use	1.05	0.85–1.30	1.04	0.72–1.51	1.18	0.98–1.41
Current alcohol use	0.96	0.77–1.21	1.08	0.72–1.62	0.92	0.78–1.09
Intercept	
No. observations						
Read	2828		2828		2828	
Used	2793		2784		2681	
LLR						
$\chi^2(11)$	131.607		53.816		98.371	
Probability > χ^2	<.0001		<.0001		<.0001	
C-statistic (ROC)	0.669		0.690		0.610	
Hosmer and Lemeshow						
$\chi^2(8)$	9.051		9.478		6.534	
Probability > χ^2	0.338		0.304		0.588	

Abbreviations: CI, confidence interval; LLR, log likelihood ratio; OR, odds ratio; QoL, quality of life; ROC, receiver operating characteristic; SATISFY-SOS, Systematic Assessment and Targeted Improvement of Services Following Yearlong Surgical Outcomes Surveys.

^aQoL came from a single question in SATISFY-SOS 30-d survey (“How would you rate your quality of life now”) with response formats expressed on a 3-point scale (better, the same, or worse). For this logistic regression, QoL was dichotomized (same/worse = 1).

^bP < .05.

^cP < .01.

^dThe reference category is the “not frail” class.

^eThe reference category is “low risk surgery.”

The present findings also extend previous work with the Fried phenotypes by showing that classes are not based on who has “more deficits” but rather it is the unique composition of deficits that distinguishes class membership. In the case of the severely frail patients, they endorsed a wide range of markers in the physical burden, functional limitations, emotional and social domains, reinforcing that frailty affects people in multiple ways, not just hindering their physical capabilities. Of paramount importance is that, although the IRT and LCA statistical procedures address different pressing questions, they both focus exclusively on the pattern of responses, in the case of the former as a latent continuum indicating severity and in the case of the latter yielding homogeneous subtypes. Together, these techniques offer a much richer picture of frailty and at the same time provide deeper insight into the “experience” of frailty from the patients’ perspective. Indeed, when taken together, the different approaches contribute to the current discussion pertaining to definitions, conceptualization, and utility of the frailty concept.^{2,4,42,43}

Surprisingly, several markers from the original pool of 32 candidates were eliminated through the IRT analysis although they are quite popular components of many frailty instruments. These included body mass index, slow walking speed, limited mobility and CCI, history

of falls, incontinence, dizziness, the American Society of Anesthesiologists III or IV physical status, low oxygen saturation, measures of cognitive functioning (ie, memory, concentration, and orientation), a psychiatric assessment of dementia, and all the nutrition and weight loss items (eg, nutrition, body mass index, or weight loss). These markers had relatively low discrimination and/or difficulty levels within the current sample.

The items that were selected by the IRT model included a single physical burden item, a proxy for the Fried slow gait item, functional limitations assessing exhaustion, low physical activity, worsening disability, limitation in ADL, limitations in work from physical and emotional health problems, accomplishing less, and restrictions on climbing stairs, and only one of 4 social support measures assessing whether patients felt their physical or emotional health problems interfered with their routine social activities (ie, visiting friends and joining other social activities). The 4 emotional health markers selected by the IRT model included feeling downhearted and blue, calm/peaceful, emotional health status change, and overall quality of mental health.

The final set of 18 frailty indicators that survived the IRT procedure included 5 physical burden markers, 8 functional limitation markers, 4 mental/emotional health markers, and a single social activities marker. Overall, the compilation

of this final pool of markers indicated that patients were unable to do what they were accustomed to as part of their normal routine activities and this was attributed to a multitude of limitations.

It is important to note that like Fried et al,¹ we too obtained 3 phenotypes; however, the composition of the phenotypes differed greatly. Fried et al¹ felt that having more burdens accentuated differences in the phenotype classes, however, we found that classes differed based on their composition in addition to the sheer number of markers defining each class. Looman et al⁴⁴ also found very heterogeneous classes in their analyses of data obtained from older people residing in the Netherlands. A careful examination of their class structure indicates very similar classes to our findings albeit with a minor distinction regarding the role of psychological frailty. In our study, the distinction of being assigned to the moderately frail class was based primarily on endorsing physical deficits, whereas assignment to the severely frail class entailed endorsing not only physical burdens but also fitness and functional limitations and a wide range of emotional problems that collectively hindered patients' activity levels. These "typological" distinctions may evade detection during a clinical examination because they require sifting through a large number of responses and mixing and matching to come up with a clear picture of the patient's day-to-day living and their various impairments. They are, however, important indicators of the most severely frail patient.

Validation of the phenotypes in terms of 3 relevant surgical outcomes also shows how class membership differed in terms of real-world experiences. These experiences included returning to the hospital, worsening quality of life after surgery, and complications from surgery. Despite the fact that the logistic models produced modest concordance statistics (ie, C-statistics varying between 0.6 and 0.7), indicating that additional covariates are perhaps needed to gain model precision, overall, this information should be quite useful to clinicians whose objectives include relating perioperative information to surgical outcomes as part of the decision process determining the suitability of patients for surgery. Importantly, the increased risk associated with severely frail patients cannot be attributed to potential demographic (ie, age) or disease comorbidity, and intraoperative anesthetic care, and detrimental lifestyle factors (ie, smoking), which were statistically controlled. Nevertheless, we cannot exclude confounding from unmeasured variables. Interestingly, in contrast to findings from other studies,⁴⁵⁻⁴⁷ we were not able to firmly predict readmission from frailty class membership.

We also found that moderately frail patients had a lower odds of reporting poor quality of life after surgery compared to the not frail, whereas we should expect greater frailty severity to be associated with poorer functional outcomes and limitations. There are many reasons for this surprising finding, some of which may tap into expectations of relief from pain and other "cognitive or motivational" factors that can differentiate the frailty phenotypes but remain unmeasured. Moderately frail individuals may resign themselves to the plethora of limitations they experience and not get their hopes up for a better quality of life even with the surgical intervention. As a clinical phenotype, they do not have

as many problems or functional limitations as the severely frail and thus, may not feel that their quality of life has been adversely affected at that point in time. They also comprise the largest percentage of individuals who reported their quality of life was better, and only a very small percentage (<10%) reported their condition as "worse" after surgery, thus limiting the possibility of optimally detecting why they felt improved quality of life after surgery.

Limitations

There are several important limitations associated with this study. The surgical patients originated from a single center reflecting the demographic composition of this particular hospital and the region it primarily serves. In addition, most of the surgeries were elective, suggesting the need to validate the current findings with different and more heterogeneous cohorts of surgical patients to confirm that frailty subtypes exist independent of environmental conditions, cultural issues, and practice specialties.

In several instances, we used self-reported proxies as "markers" to canvass the different frailty domains. With the Fried approach, this is not so problematic given that only 5 physical burden items are needed to replicate the phenotypes. However, the inclusion of other domains assessing functional limitations, social activities, emotional, and cognitive performance measures requires new instrumentation to be introduced into the discussion. This could potentially lead to extraction of uniquely different subtypes, as we found with the inclusion of 3 domains not considered by Fried et al.¹ Additional registry-based studies with reported instrumentation representing as many domains as possible are needed to test the replicability of these findings using reasonable proxies.

Missing data may also present another limitation, albeit 70% of the original registry database had complete data. Analyses to detect sample bias did not indicate significant differences between the subgroups with complete versus missing data. Still, subtle if not systematic differences may exist between these subgroups although they remain unmeasured. We also dichotomized ordinal and continuous indicators used in the models. There is strong statistical evidence that with regression or other variable-centered approaches, this would undermine our ability to detect underlying statistical relations.⁴⁸ However, LCA is a person-centered technique that incorporates statistical information for each frailty marker based on patient endorsement (ie, yes/no). In this case, binary information is appropriate to indicate a patient giving a positive response (ie, "yes, this happens to me"). It is the presence versus absence of the symptom rather than its magnitude or level of intensity that matters in this type of analysis. Accordingly, there should be minimal loss of information from this data management procedure to fashion patterns of frailty markers and posit an underlying vulnerability.

CONCLUSIONS

This study distinguished subtypes of frailty based on common markers readily used in clinical practice. Distinguishing phenotypes based on symptomatology paints a more refined clinical picture of frailty and may help to clarify features of the underlying syndrome.^{2,49,50} We extended current

frailty conceptual models that primarily emphasize physical burden markers by including compromised functioning in social, cognitive, and emotional health domains as well as physical functional limitations that may influence a patient's will to live, thriving, and resilience.⁵⁰⁻⁵³ We also were able to identify a group of characteristically older frail patients that was most vulnerable to postoperative complications, suggesting that efficient risk assessments to detect frailty preoperatively in a systematic, objective fashion could be integrated into clinical practice.

Clinicians benefit from the progressive refinement of risk assessments that have important practice ramifications. A prerequisite to support this goal requires establishing an instrument's psychometric properties. The present study does this by clarifying frailty as a reliable and valid, albeit, multifaceted syndrome, including physical burden deficits, functional limitations, and problems in living that underlie emotional health and social support. This more holistic view of frailty should have ramifications for multiple settings, including research, clinical, surgical, and policy initiatives. ■■

ACKNOWLEDGMENTS

The authors thank TL1 fellowship directors Bradley A. Evanoff, MD, MPH (Professor, Internal Medicine, Director, Institute of Clinical and Translational Sciences, Washington University School of Medicine, St Louis, Missouri), Jay F. Piccirillo, MD, FACS (Professor, Otolaryngology, Head and Neck Surgery, Director, Clinical Outcomes Research Office, Washington University School of Medicine, St Louis, Missouri), Susan Stark, PhD, OTR/L, FAOTA (Assistant Professor of Occupational Therapy, Neurology and Social Work, Codirector of TL1 Predoctoral Training program, Washington University School of Medicine, St Louis, Missouri), and Jeffrey F. Peipert, MD, PhD (Robert J. Terry Professor of Obstetrics and Gynecology, Vice Chair of Clinical Research, Washington University School of Medicine, St Louis, Missouri) who served as scientific advisors and critically reviewed the original study proposal; Reem Mustafa, MD, PhD (Assistant Professor, Departments of Internal Medicine and Biomedical and Health Informatics, University of Missouri-Kansas City [UMKC], Kansas City, Missouri) for academic support; Will Godfrey, MA (Database Manager [former], Department of Anesthesiology, Institute of Quality Improvement, Research & Informatics, Washington University School of Medicine, St Louis, Missouri) for technical and data support; and all preoperative clinic staff for their contribution to Systematic Assessment and Targeted Improvement of Services Following Yearlong Surgical Outcomes Surveys.

DISCLOSURES

Name: Furqaan Sadiq, MD, MSCI.

Contribution: This author helped conceptualize the study design, including data acquisition and preparation as well as the statistical data modeling and interpretation; and write, revise, and edit the manuscript.

Name: Vanessa L. Kronzer, MD, MSCI.

Contribution: This author helped prepare and critically review the manuscript.

Name: Troy S. Wildes, MD.

Contribution: This author helped prepare and critically review the manuscript.

Name: Sherry L. McKinnon, BS.

Contribution: This author helped prepare and critically review the manuscript.

Name: Anshuman Sharma, MD, MBA.

Contribution: This author helped conceptualize the study design, including statistical data modeling and interpretation; and write, revise, and edit the manuscript.

Name: Daniel L. Helsten, MD.

Contribution: This author helped prepare and critically review the manuscript.

Name: Lawrence M. Scheier, PhD.

Contribution: This author helped conceptualize the study design, including statistical data modeling and interpretation; and write, revise, and edit the manuscript.

Name: Michael S. Avidan, MBBCh.

Contribution: This author helped conceptualize the study design, including statistical data modeling and interpretation; and write, revise, and edit the manuscript.

Name: Arbi Ben Abdallah, DES, PhD.

Contribution: This author helped conceptualize the study design, including data acquisition and preparation as well as the statistical data modeling and interpretation; and write, revise, and edit the manuscript.

This manuscript was handled by: Robert Whittington, MD.

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.
2. Bergman H, Ferrucci L, Guralnik J, et al. Frailty: an emerging research and clinical paradigm—issues and controversies. *J Gerontol A Biol Sci Med Sci.* 2007;62:731-737.
3. Hogan DB, MacKnight C, Bergman H; Steering Committee, Canadian Initiative on Frailty and Aging. Models, definitions, and criteria of frailty. *Aging Clin Exp Res.* 2003;15:1-29.
4. Rockwood K. What would make a definition of frailty successful? *Age Ageing.* 2005;34:432-434.
5. Abellan van Kan G, Rolland YM, Morley JE, Vellas B. Frailty: toward a clinical definition. *J Am Med Dir Assoc.* 2008;9:71-72.
6. Fried LP, Ferrucci L, Darer J, Williamson JD, Anderson G. Untangling the concepts of disability, frailty, and comorbidity: implications for improved targeting and care. *J Gerontol A Biol Sci Med Sci.* 2004;59:255-263.
7. Dasgupta M, Rolfson DB, Stolee P, Borrie MJ, Speechley M. Frailty is associated with postoperative complications in older adults with medical problems. *Arch Gerontol Geriatr.* 2009;48:78-83.
8. Lin HS, Watts JN, Peel NM, Hubbard RE. Frailty and post-operative outcomes in older surgical patients: a systematic review. *BMC Geriatr.* 2016;16:157.
9. Pol RA, van Leeuwen BL, Visser L, et al. Standardised frailty indicator as predictor for postoperative delirium after vascular surgery: a prospective cohort study. *Eur J Vasc Endovasc Surg.* 2011;42:824-830.
10. Brown CH 4th, Max L, LaFlam A, et al. The association between preoperative frailty and postoperative delirium after cardiac surgery. *Anesth Analg.* 2016;123:430-435.
11. Schultz M, Rosted E, Sanders S. Frailty is associated with a history with more falls in elderly hospitalised patients. *Dan Med J.* 2015;62:A5058.
12. Kojima G. Frailty as a predictor of future falls among community-dwelling older people: a systematic review and meta-analysis. *J Am Med Dir Assoc.* 2015;16:1027-1033.
13. Boyd CM, Xue QL, Simpson CF, Guralnik JM, Fried LP. Frailty, hospitalization, and progression of disability in a cohort of disabled older women. *Am J Med.* 2005;118:1225-1231.
14. Makary MA, Segev DL, Pronovost PJ, et al. Frailty as a predictor of surgical outcomes in older patients. *J Am Coll Surg.* 2010;210:901-908.
15. McIsaac DI, Bryson GL, van Walraven C. Association of frailty and 1-year postoperative mortality following major elective noncardiac surgery: a population-based cohort study. *JAMA Surg.* 2016;151:538-545.
16. Kojima G, Iliffe S, Jivraj S, Walters K. Association between frailty and quality of life among community-dwelling older

- people: a systematic review and meta-analysis. *J Epidemiol Community Health*. 2016;70:716–721.
17. Kulmala J, Nykänen I, Hartikainen S. Frailty as a predictor of all-cause mortality in older men and women. *Geriatr Gerontol Int*. 2014;14:899–905.
 18. Lee DH, Buth KJ, Martin BJ, Yip AM, Hirsch GM. Frail patients are at increased risk for mortality and prolonged institutional care after cardiac surgery. *Circulation*. 2010;121:973–978.
 19. Rockwood K, Song X, MacKnight C, et al. A global clinical measure of fitness and frailty in elderly people. *CMAJ*. 2005;173:489–495.
 20. Rockwood K, Mitnitski A. Frailty in relation to the accumulation of deficits. *J Gerontol A Biol Sci Med Sci*. 2007;62:722–727.
 21. 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.
 22. Xue QL. The frailty syndrome: definition and natural history. *Clin Geriatr Med*. 2011;27:1–15.
 23. Bandeen-Roche K, Xue QL, Ferrucci L, et al. Phenotype of frailty: characterization in the women's health and aging studies. *J Gerontol A Biol Sci Med Sci*. 2006;61:262–266.
 24. Xue QL, Bandeen-Roche K, Varadhan R, Zhou J, Fried LP. Initial manifestations of frailty criteria and the development of frailty phenotype in the Women's Health and Aging Study II. *J Gerontol A Biol Sci Med Sci*. 2008;63:984–990.
 25. Op het Veld LP, van Rossum E, Kempen GJ, de Vet HC, Hajema K, Beurskens AJ. Fried phenotype of frailty: cross-sectional comparison of three frailty stages on various health domains. *BMC Geriatr*. 2015;15:77.
 26. Forti P, Rietti E, Pisacane N, Olivelli V, Maltoni B, Ravaglia G. A comparison of frailty indexes for prediction of adverse health outcomes in an elderly cohort. *Arch Gerontol Geriatr*. 2012;54:16–20.
 27. Lee L, Patel T, Costa A, et al. Screening for frailty in primary care: accuracy of gait speed and hand-grip strength. *Can Fam Physician*. 2017;63:e51–e57.
 28. Kiely DK, Cupples LA, Lipsitz LA. Validation and comparison of two frailty indexes: the MOBILIZE Boston Study. *J Am Geriatr Soc*. 2009;57:1532–1539.
 29. Gale CR, Martyn CN, Cooper C, Sayer AA. Grip strength, body composition, and mortality. *Int J Epidemiol*. 2007;36:228–235.
 30. Ensrud KE, Ewing SK, Taylor BC, et al; Study of Osteoporotic Fractures Research Group. Frailty and risk of falls, fracture, and mortality in older women: the study of osteoporotic fractures. *J Gerontol A Biol Sci Med Sci*. 2007;62:744–751.
 31. 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.
 32. Malmstrom TK, Morley JE. Frailty and cognition: linking two common syndromes in older persons. *J Nutr Health Aging*. 2013;17:723–725.
 33. Rosado-Artalejo C, Carnicero JA, Losa-Reyna J, et al. Global performance of executive function is predictor of risk of frailty and disability in older adults. *J Nutr Health Aging*. 2017;21:980–987.
 34. Romero-Ortuno R, Soraghan C. A Frailty Instrument for primary care for those aged 75 years or more: findings from the Survey of Health, Ageing and Retirement in Europe, a longitudinal population-based cohort study (SHARE-FI75+). *BMJ Open*. 2014;4:e006645.
 35. Saum KU, Müller H, Stegmaier C, Hauer K, Raum E, Brenner H. Development and evaluation of a modification of the Fried frailty criteria using population-independent cutpoints. *J Am Geriatr Soc*. 2012;60:2110–2115.
 36. Helsten DL, Ben Abdallah A, Avidan MS, et al. Methodologic considerations for collecting patient-reported outcomes from unselected surgical patients. *Anesthesiology*. 2016;125:495–504.
 37. Hambleton RK, Swaminathan H, & Rogers JH. *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage; 1991.
 38. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Modeling*. 1999;6:1–55.
 39. Muthén LK, Muthén BO. *Mplus: Statistical Analysis With Latent Variables: User's Guide*. Muthén & Muthén: Los Angeles, CA; 1998–2012.
 40. Embretson SE. Multidimensional measurement from dynamic tests: abstract reasoning under stress. *Multivariate Behav Res*. 2000;35:505–542.
 41. Lubke GH, Muthén B. Investigating population heterogeneity with factor mixture models. *Psychol Methods*. 2005;10:21–39.
 42. Hubbard RE, Story DA. Patient frailty: the elephant in the operating room. *Anaesthesia*. 2014;69(suppl 1):26–34.
 43. Partridge JS, Harari D, Dhesi JK. Frailty in the older surgical patient: a review. *Age Ageing*. 2012;41:142–147.
 44. Looman WM, Fabbriotti IN, Blom JW, et al; TOPICS-MDS Research Consortium. The frail older person does not exist: development of frailty profiles with latent class analysis. *BMC Geriatr*. 2018;18:84.
 45. Robinson TN, Wu DS, Pointer L, Dunn CL, Cleveland JC Jr, Moss M. Simple frailty score predicts postoperative complications across surgical specialties. *Am J Surg*. 2013;206:544–550.
 46. Anpalahan M, Gibson SJ. Geriatric syndromes as predictors of adverse outcomes of hospitalization. *Intern Med J*. 2008;38:16–23.
 47. Hewitt J, Moug SJ, Middleton M, Chakrabarti M, Stechman MJ, McCarthy K; Older Persons Surgical Outcomes Collaboration. Prevalence of frailty and its association with mortality in general surgery. *Am J Surg*. 2015;209:254–259.
 48. MacCallum RC, Zhang S, Preacher KJ, Rucker DD. On the practice of dichotomization of quantitative variables. *Psychol Methods*. 2002;7:19–40.
 49. de Vries NM, Staal JB, van Ravensberg CD, Hobbelen JS, Olde Rikkert MG, Nijhuis-van der Sanden MW. Outcome instruments to measure frailty: a systematic review. *Ageing Res Rev*. 2011;10:104–114.
 50. Gill TM, Gahbauer EA, Allore HG, Han L. Transitions between frailty states among community-living older persons. *Arch Intern Med*. 2006;166:418–423.
 51. Pialoux T, Goyard J, Lesourd B. Screening tools for frailty in primary health care: a systematic review. *Geriatr Gerontol Int*. 2012;12:189–197.
 52. Bergland A, Kirkevold M. Thriving—a useful theoretical perspective to capture the experience of well-being among frail elderly in nursing homes? *J Adv Nurs*. 2001;36:426–432.
 53. Stewart DE, Yuen T. A systematic review of resilience in the physically ill. *Psychosomatics*. 2011;52:199–209.