Development of a Frailty Ladder Using Rasch Analysis: If the Shoe Fits

Nancy E. Mayo, PhD[1,2,3], Mylène Aubertin-Leheudre, PhD[4,5], Kedar Mate, PhD[1], Sabrina Figueiredo, PhD[6], Julio Flavio Fiore Jr., PhD[2][7,8], Mohammad Auais, PhD[9], Susan C. Scott, MSc[2], José A. Morais, MD[1,2]

[1]Department of Medicine, McGill University, Montreal, QC; [2]Centre for Outcomes Research and Evaluation, McGill University Health Centre Research Institute, Montreal, QC; [3]School of Physical and Occupational Therapy, McGill University, Montreal, QC; [4]Département des Sciences de l'activité physique, Faculté des Sciences, Université du Québec à Montréal, Montréal, QC; [5]Centre de recherche de l'institut universtaire de Gériatrie de Montréal, Montréal, QC; [6]Health Care Quality Program, The George Washington University, School of Medicine and Health Sciences, Washington, DC; [7]Department of Surgery, McGill University, Montreal, QC; [8]Steinberg-Bernstein Centre for Minimally Invasive Surgery and Innovation, McGill University, Montreal, QC; [9]School of Rehabilitation Therapy, Faculty of Health Sciences, Queen's University, Kingston, ON

ABSTRACT

Background

The current measurement approach to frailty is to create an index of frailty status, rather than measure it. The purpose of this study is to test the extent to which a set of items identified within the frailty concept fit a hierarchical linear model (e.g., Rasch model) and form a true measure reflective of the frailty construct.

Methods

A sample was assembled from three sources: community organization for at-risk seniors (n=141); colorectal surgery group assessed post-surgery (n=47); and hip fracture assessed post-rehabilitation (n=46). The 234 individuals (age 57 to 97) contributed 348 measurements. The frailty construct was defined according to the named domains within commonly used frailty indices, and items drawn to reflect the frailty came from self-report measures. Performance tests were tested for the extent to which they fit the Rasch model.

Results

Of the 68 items, 29 fit the Rasch model: 19 self-report items on physical function and 10 performance tests, including one for cognition; patient reports of pain, fatigue, mood, and health did not fit; nor did body mass index (BMI) nor any item representing participation.

Conclusion

Items that are typically identified as reflecting the frailty concept fit the Rasch model. The Frailty Ladder would be an efficient and statistically robust way of combining results of different tests into one outcome measure. It would also be a way of identifying which outcomes to target in a personalized intervention. The rungs of the ladder, the hierarchy, could be used to guide treatment goals.

Key words: frailty, Rasch analysis, measurement theory

INTRODUCTION

Frailty is a topic of major importance for clinical care and research in aging, geriatric medicine, physical therapy, rehabilitation, and nursing. Interest covers the spectrum from biology, detection, prevention, to treatment.[1-8] But underlying all of these areas is uncertainty about whether frailty is a construct that can be measured or an entity that can only be classified.

There is strong evidence for a frailty syndrome or phenotype[9] that is caused by four converging situations in an older person: systemic illness (which is the accumulation of acute and chronic illness from those on the list for accumulation of deficits), poor genes, poor lifestyle, and poor environment.[1,10-14] When these converge with aging, they create the "perfect storm" that is frailty. As a syndrome or phenotype, it is a classification and the most common way classifying someone as frail or not is if they manifest three of five physical criteria: slowness, weakness, exhaustion, inactivity, and weight loss.[9] This approach renders frailty as a unidimensional, formative construct that is quantified by counting these physical manifestations.

The classification approach is not particularly useful, since being classified as "frail" does not incur special

treatments or services as does being classified as having cancer, or cognitive impairment, or osteoporosis which places people on special treatment and follow-up protocols or makes them eligible for specific medications.

Current definitions of frailty suggest an entity that can be measured. In 2010, Gobbens et al. defined frailty as: "A dynamic state affecting an individual who experiences losses in one or more domains of human functioning (physical, psychological, social), which is caused by the influence of a range of variables and which increases the risk of adverse outcomes."[15] In 2013, Rodríguez-Mañas et al. defined frailty as "a multidimensional syndrome characterized by decreased reserve and diminished resistance to stressors."[16] The most recent definition of frailty has been proposed by the Canadian Frailty Network:[17] "Frailty is a state of increased vulnerability, with reduced physical reserve and loss of function across multiple body systems. This reduces ability to cope with normal or minor stresses, which can cause rapid and dramatic changes in health." Reserve can be measured directly using physiological measures such as muscle mass or tests that have normative data, distance walked in six minutes, for example.[18]

These definitions set up frailty as a reflective construct, something to be measured. Measurement was defined in 1978 by Nunnally[19] as "rules for assigning numbers to objects in such a way as to represent quantities of attributes". Some hundred years earlier Lord Kelvin,[20] responsible for identifying the temperature which defined absolute zero, commented that: "To measure is to know" and "If you cannot measure it, you cannot improve it." Since the 1960s there has been interest in measuring attributes of people on measurement scales that have similar biophysical properties to temperature.[21-23]

While frailty is considered distinct from disability,[16] 85% of the language used to describe frailty[24] is found within the World Health Organization's International Classification of Functioning, Disability and Health (ICF).[25] The link between frailty and disability was illustrated through a recent review of 79 measures used to identify frailty.[26] Of the 25 single-domain frailty instruments, 24 could be mapped to the Body Function, and Activity and Participation components of the ICF. Of the 54 multi-domain frailty instruments, all could be mapped to the ICF with 67% (n = 35) of these instruments linking to at least three components of the ICF and 10% to all five. Van Damme et al., in a 2021 review,[27] identified that the 67 screening measures for frailty covered four domains and all of them included physical function; 73% covered psychological, 52% social, and 78% contextual factors. Almost half (43%) included all four domains.

Another conceptual model that is relevant in the field of frailty is proposed by Guaraldi and Mills on intrinsic capacity.[28] This is a positive construct that leads to a construct that is more actionable in terms of prevention, intervention, and measurement. Intrinsic capacity is something to strive for, whereas frailty is something to avoid. The foundation of intrinsic capacity is function in the broadest sense of this construct and defined by the World Health Organization, of which the ICF is core component.

A key observation from the 2016 review by Azzopardi et al.[26] is that virtually all of the 79 measures reviewed quantified frailty as a count of the number of detrimental health indicators such as co-morbidities, medications, abnormal results on blood tests, social situation, and specific functional limitations. The count was used to classify people into frailty categories. From a measurement perspective, the items were used to define the construct, frailty. This formative conceptual model underlies the creation of indices; the value of the index cannot change unless the items change.[29] Targeting the items will change the value on the index, but will do little for the underlying construct.

A reflective conceptual model underlies latent constructs, rather than indices that are formed by the items. In a reflective model, the construct is reflected in the items; many items could reflect the construct, changing the construct will change the items but not the reverse.[29-31] Of these two models, formative models are useful for discrimination, to classify people into groups like frail, not frail, and to predict future health events. To evaluate change, true measures are required. For interventions targeting frailty, the measurement strategy has been to use multiple measures of single indicators of physical, emotional, and cognitive function and report on these separately,[3] with the inherent conceptual and statistical limitations to this approach. Many of these single indicators are those that form the frailty phenotype—strength and gait speed, as key examples.

Many interventions have attempted to change the building blocks of frailty such as strength and gait speed,[3,32,33] and those indicators that were targeted by the intervention did change but the impact on frailty classification was small or nil.

This paper addresses whether frailty could be considered under a reflective conceptual model, where the items related to the phenotypical indicators of frailty are seen as a reflection of frailty rather than forming it. A multitude of items could be chosen to reflect the frailty construct, and the value or quantity of frailty is inferred from observing how people interact with items known to reflect the frailty construct.

One necessary, but not sufficient, requirement of a reflective construct is that it fits a true latent model, such as the Rasch model, named for the Danish mathematician, George Rasch (1929-1980), who provided a method for constructing a measure of by transforming ordinal observations onto an interval scale. The outcome of a Rasch analysis, when the data fit the model, is a unidimensional measure on which items and people are organized hierarchically, by difficulty and ability respectively, on the same measurement scale in natural logarithm linear units or logits. Items that fit a Rasch model would form a measure with a total score that is sufficient to determine that person's value on the underlying construct. The total score is a legitimate representation of the person's "ability", and can be subjected to mathematical transformations such as those needed to estimate change.[34] A feature of a reflective model is that changing the construct will change the items.[30] The evidence for this is plentiful as there are many examples of older persons who become ill or injured and deteriorate and accumulate many of the frailty indicators only to have these

reversed when the source of the illness is identified and treated.[28] While therapy for the consequences of frailty may hasten the recovery of robustness, the fundamental change arises from treating the underlying causes of frailty. If these are not addressed, targeting the functional consequences will have some effect on function but not necessarily reduce frailty. However, that is not to suggest that targeting function will have no impact on frailty, as physical therapies can build reserve.[35,36]

The purpose of this study is to test the extent to which a set of items identified within the frailty concept fit a hierarchical near model and form a latent construct which is reflective of the frailty construct.

METHODS

Cross-sectional and longitudinal data arising from 234 seniors (>55 years) with or at risk for age-related disability or receiving in- or out-patient care at a tertiary care health centre were analyzed for fit to the Rasch model. The sample was assembled from three sources: at risk community based source (n=141);[37] colorectal surgery group assessed post-surgery (n=47);[38] and hip fracture assessed post-discharge from rehabilitation (n=46 measured 1 –to 3 times).[39] All projects had ethical approval from their respective organizations.

The frailty construct was defined according to the named domains within commonly used frailty indices,[26,40] including the domains of mobility, strength, activity, energy, nutritional status, mood, balance, incontinence, cognition, and independence, as examples. This sample of seniors had been assessed within the context of research projects addressing senior's health and the assessments included items reflecting the frailty domains. Table 1 lists the measures that were the source of the frailty items along with their original scoring.

The analysis was carried out within the framework of Rasch measurement theory,[41] an experimental paradigm that tests the extent to which a set of the items form a real measure. The results are reported as recommended by Pallant and Tennant.[42] The ordering of the categories of the rating

TABLE 1.
Measures used for sources of frailty items

Measure	N Items [Original Response options]	Units
Lower Extremity Functional Scale[45]	20 [5]	Difficulty: Extreme, Quite a bit, Moderate, A little, None
RAND-36[44]		
Physical Function Index	10 [3]	Limitation: A lot, A little, Not at all
Self-rated health	1 [5]	Excellent, Very good, Good, Fair, Poor
Pain	1 [6]	Severity: None, Very mild, Mild, Moderate, Severe, Very severe
	1 [5]	Interference: Not at all, A little bit, Moderately, Quite a bit, Extremely
Vitality	4 [6]	Amount of time: All, Most, A good bit, Some, A little, None
Role physical	4 [2]	Interference from physical health: Yes, No
Role emotional	3 [2]	Interference from emotional problems: Yes, No
Social function	1 [5]	Extent of interference: Not at all, Slightly, Moderately, Quite a bit, Extremely
	1 [5]	Time of interference: All, Most, Some, A little, None
Mental Health Index	5 [6]	Amount of time: All, Most, A good bit, Some, A little, None
Montreal Cognitive Assessment (MOCA)[46]	Cognitive impairment	0 to 30
Stair Test[52]		Number of stairs in 20 sec
Sit-to-Stand[53]		Number of sit to stand in 30 sec
Equilibrium[54]	Left, Right leg	Average time (sec) one leg stance
Timed Up and GO (TUG)[55]	Comfortable/Fast	
6 Minute Walk Test (6MWT)[18]		Distance (m) walked 6 min
Forced Expiratory Volume in 1 sec		% predicted
Forced Vital Capacity		% predicted
Total Time in Activity		Total min activity per day
Muscle Mass		% of total body mass
Grip strength[56]	Dominant hand	Best of 3 tries (kg)
Leg strength	Left, Right	Kg
Body Mass Index		Kg/m^2

scales is tested empirically, and if ordering is not met, further experimentation is required before inferences from the ratings are made. Rasch analysis was used to estimate the extent the frailty items fit the Rasch model. If the data fit the model, there is support for a reflective construct but fit is not sufficient; theory must support that construct causes the items, not the reverse.[43] Table 2 lists the iterative steps conducted through Rasch analysis. The partial credit model was used with the RUMM2030 software (https://www.rummlab.com.au/).

RESULTS

From the 234 people in this study, there were 348 valid measurements, as some people contributed more than one

measurement. The mean age (SD) of the at-risk community sample was 71 (5.5) years; the mean ages of the hip fracture and colorectal surgery groups were 73 (6.3), and 77 (9.2) years, respectively. The age ranged from 57 to 97 and, except for the colorectal surgery group with 45% women, approximately 70% were women. A total of 68 items were available for analyses with the 36 items from the RAND-36[44] common to the at-risk community, and the colorectal samples and hip fracture sample having 12 of the 36 items plus all the items from the LEFS. Walking capacity was measured with the 6MWT in two of the data sets and with gait speed in the hip fracture data set. Missing data were generated when items differed across data sources.

All ordinal variables were coded so that the higher value indicated more robustness (less frailty). All continuous items

TABLE 2.
Definition of steps taken to fit the data to the Rasch model

Threshold Order	There should be a logical ordering to the response options such that endorsing a more optimal response option should situate the person at a higher level of the latent trait; more people should endorse a lower response level, and fewer people should endorse a higher response level. If the thresholds are disordered, the response options need to rescored, sometimes reducing the responses to binary. The number of thresholds is equal to the number of response options-1 and reflects the number of "jumps" the person has to make for each item.
Fit to the Rasch Model	The items should line up hierarchically such that those items that need little ability to endorse at the most optimal response level are at the low end and those items requiring more ability to endorse are higher. Overall goodness of model fit is indicated by a non-significant Chi-square test (p>.05) after a Bonferroni adjustment for the number of items. Fit of each item and each person is as important, or even more important, than overall fit. Item and person fit is indicated when fit residual (deviance from pure linearity) values are within ±2.5 and the Chi-square test for fit is non-significant (>0.05). Those items that fail this criterion need to be looked at carefully to ensure their importance in scoring the latent trait. A fit residuals of >+2.5 indicates the item does not fit the latent trait; a fit residual of <-2.5 indicates the item overfits and may be redundant.
Unidimensionality	A requirement of the Rasch model is that a single latent trait is being measured. This is assessed using a principal component analysis (PCA) of the fit residuals. The person-ability estimates derived from all pair-wise comparisons of the two most disparate set of items (those with the highest positive and negative loadings on the first factor) are compared using independent t-tests (equating tests). For a set of items to be considered unidimensional, less than 5% of t values should be outside ± 1.96. When this value is greater than 5%, a binomial test of proportions is used to calculate the 95% confidence interval (CI) around the t-test estimate. Evidence of unidimensionality is still supported if the 5% value falls within the 95%CI.
Response Dependency	Uniqueness of the information provided by the items is a requirement of the Rasch model. Items with pair-wise residual (after controlling for the latent trait) correlations greater than 0.3 could indicate lack of independence of the responses which inflates the reliability. Solutions include creating a super-item which combines the response options across items or choosing the one item that best suits the testing context.
Differential Item Functioning (DIF)	The items should have the same ordering of difficulty across all people being measured defined by personal factors such as gender, age, or education. DIF is an indicator of item bias. Typically, DIF is indicated with a significant F-test from a two-way analysis of variance but the meaningfulness of the difference needs to be considered. Two options are available for items with DIF, deletion or split scoring.
Targeting	An ideally targeted measure should include a set of items that spans the full range of the theoretical latent construct (-4 to +4 logits), and have a mean location of 0 with a standard deviation (SD) of 1. Ideally, the person estimates from this measure should be centered on location 0 with a SD of 1. This corresponds to the standard normal distribution which is centered on 0 with a SD of 1 and ±4 SD represents the theoretical range of the distribution (99% of observations will fall in this range).
Discrimination or Person-Separation	This indicates how well people are differentiated by the spread of the item-difficulty. The person-separation index (PSI) is interpreted like a Cronbach's alpha. The larger the index, the better is the discrimination which facilitates the measurement of change. Values of >0.9 are suitable for measuring within-person change, values >0.7 are suitable for detecting group differences.

were categorized into as many ordinal categories as the distribution would support, usually 10, and rescored until completely ordered thresholds were obtained and with higher values indicating more robustness.

Of the 68 items tested (see Table 1), 29 fit the Rasch model: two of 10 items from the Physical Function Index of the RAND-36,[44] 17 of 20 items from the Lower Extremity Functional Scale (LEFS),[45] and 10 performance tests including one for cognition.[46] For the LEFS, no item retained the five response categories; one item retained four response categories (getting in and out of a car); two items supported three response categories (stairs and standing for one hour); the rest of the fitting items could only support a binary scale. Of the two items retained from the RAND-36, limitation in vigorous activities did not need to be rescored, but the item for moderate activities could only support a binary scale.

Of the 39 items that did not fit, 34 came from the RAND-36, including eight of the items from the Physical Function Index; none of the items from the other seven domains of General Health Perceptions, Pain, Vitality, Mental Health, Role Physical, Role Emotional, or Social role fit. Only three of the 18 LEFS items did not fit, the ones indicating very high functioning (running). The two measures of mobility, fast and comfortable gait speed and Timed-Up-and-Go (TUG), were highly correlated and one super-item was made by combining the two. Leg strength (right and left leg) was redundant with functional measures of strength (Stair Test and Sit-to-Stand). Body mass index (BMI) did not fit.

Grip strength had residual correlations with the sit-to-stand test and with the self-report item from the RAND-36, limitation in doing moderate activities with examples such as moving a table, pushing a vacuum cleaner, bowling, or playing golf, all of which require hand strength; however, the equating tests did not indicate multidimensionality, so both items were retained.

There was differential item functioning (DIF) by sex for lean muscle mass and grip strength as any given value on these tests would indicate more robustness for women than for men. These items were split.

Fit to the model was demonstrated by a chi-square test for goodness of fit associated with a non-significant p value of .283 and a person-separation index of 0.907. Unidimensionality was also supported, although the usual method of testing for unidimensionality (see Table 1) could not be carried out when there is missing data as in this analysis of three merged data sets. Instead, we tested unidimensionality within two groups of variables without missing data and found that fewer than 5% of these equating tests were significant.

Table 3 provides item locations averaged over thresholds, indicating average difficulty of the items. Those items with a logit score less than 0 would be considered "easy" and failing these items would indicate more frailty; items with a logit score of >1 would be considered "difficult" and passing these would indicate more robustness or less frailty.

Figure 1 shows the threshold map of the 29 items. Along the x-axis is the value of the latent frailty construct;

the colored bars represent, for each item, the threshold that has to be passed for a given score (on a logit scale) on the latent construct.

At the most frail end of the scale (-6 logits) would be people who endorse having extreme difficulty getting in and out of a car and who have more than a little difficulty moving in bed (easy items). The next rung of the ladder, at around

TABLE 3.
Mean (over all thresholds) location order of each frailty indicator

Item	Mean Location
Difficulty running on even ground	3.14
Distance walked 6 min (m)	2.65
Forced Expiratory Volume in 1 sec (FEV$_1$) % predicted	2.53
Difficulty walking a mile	2.42
Muscle mass Women	2.24
Grip strength Women	2.17
Limitation vigorous activities	1.98
N sit to stand in 30 sec	1.96
Average one leg stance	1.81
N of stairs in 20 sec	1.77
Muscle mass Men	1.57
Forced Vital Capacity (FVC) % predicted	1.38
TUG combined fast and comfortable	1.14
Cognitive impairment	1.13
Difficulty squatting	1.11
Total min activity per day	0.82
Grip strength Men	0.01
Difficulty standing for 1 hr	-0.32
Difficulty walking 2 blocks	-0.39
Difficulty going up or down 10 stairs	-1.09
Limitation moderate activities	-1.68
Difficulty with hobbies, recreation, sports	-1.85
Difficulty putting on your shoes or socks	-1.99
Difficulty with work or housework	-2.02
Difficulty lifting an object from the floor	-2.03
Difficulty getting into or out of the bath	-2.11
Difficulty performing light household activities	-2.23
Difficulty walking between rooms	-2.74
Difficulty getting into or out of a car	-3.12
Difficulty sitting for 1 hr	-3.45
Difficulty rolling over in bed	-4.82

-4 logits, are people who report any degree of difficulty sitting for one hour. The self-report items are all at the lower end of the scale, and the first of the performance items are at approximately -1 logits, the slowest values on the TUG, and FVC of 30–59% predicted. At the most robust end of the scale

(difficult items) is the longest 6MWT distance, the strongest grip strength for either men or women, and confirmation of having only a little or no difficulty running on even ground.

Figure 2 shows the person-threshold map indicating targeting of the items to the people. The pink bars on the top
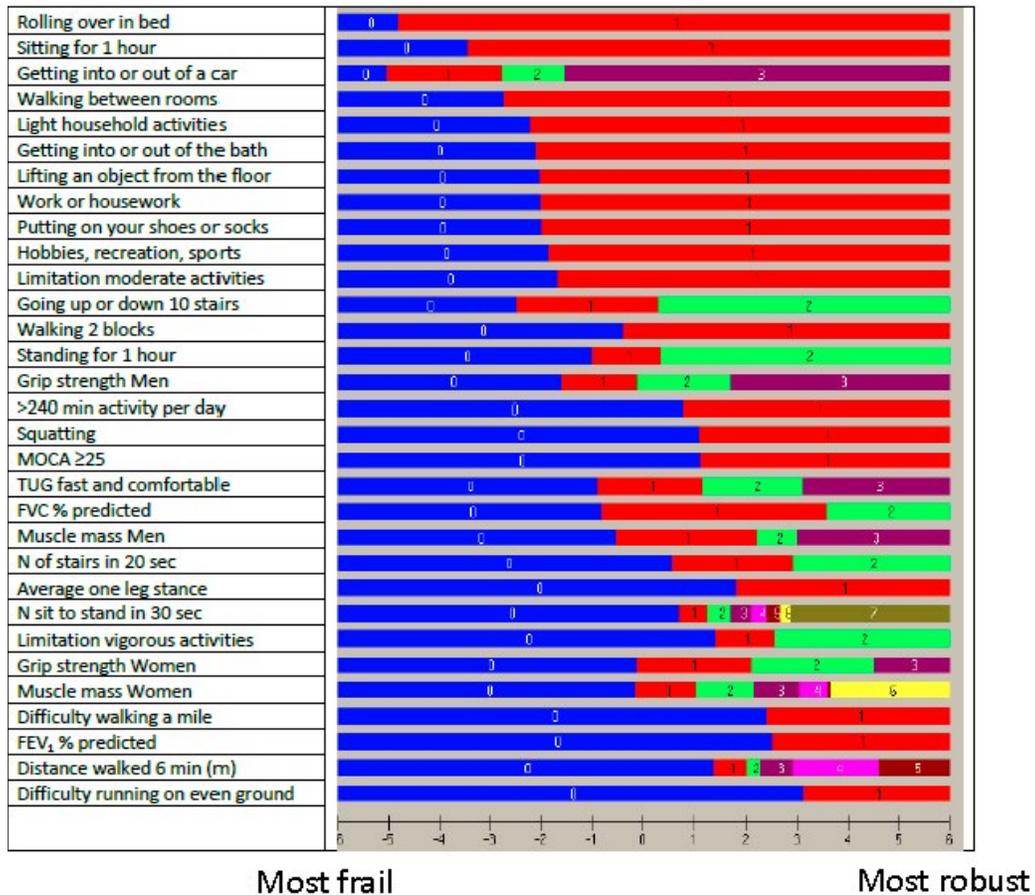
FIGURE 1. Threshold map for Frailty Ladder items
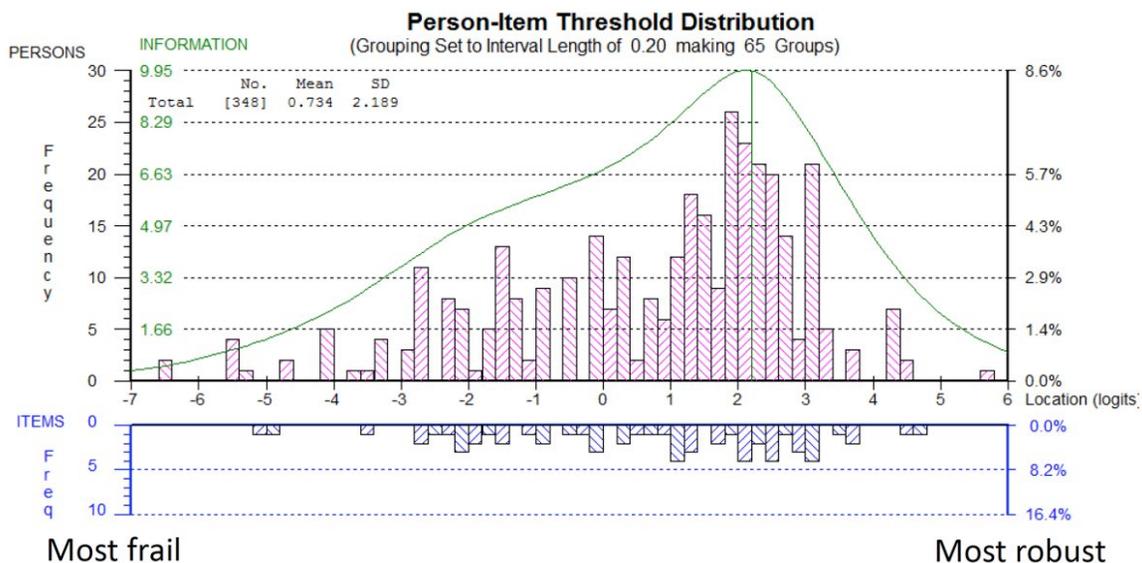
FIGURE 2. Item-threshold map for frailty items showing targeting of the items and persons

indicate the distribution of the study subjects along the latent construct, and the blue bars represent the location of item-threshold along the same construct. The mean on the subjects without extremes was 0.734 (with extremes: 0.65) which is close to the optimal of 0, with a standard deviation (SD) of 2.189 (with extremes: 2.183), larger than the desired SD of 1.0. The distribution of the people ranged from <-6.0 to >+5; the distribution of the items covered a similar wide range. The items and people are not optimally targeted, as many more people were at the more robust end of the frailty scale and fewer at the frailer end, yielding a skewed distribution.

A simple scoring algorithm, based on the ordinal structure of the data, correlated highly with the logit scale. This provided a more interpretable scoring algorithm with a theoretical range from 0 to 53. This scoring system and the levels of the items for Version 1 of the Frailty Ladder are given in Table 4.

DISCUSSION

The analysis found that the items that are typically identified as reflecting the frailty concept fit with the Rasch model, satisfying one of the necessary criteria for a reflective model. The items that fit the Rasch model covered the majority of domains identified from frailty indices including self-report physical limitations, performance of physical tests, and cognition. These findings suggest that disability is a necessary—but not a sufficient—reflection of frailty,[47] as not all people with disabilities (stroke, Multiple Sclerosis, arthritis, for examples) despite observable disability are frail. Another necessary, but not sufficient, contributor to frailty would be age, but there was no consensus on a particular age cut-off.[16] The observation that the patient-reported outcomes related to fatigue, mood, and health perception did not fit the model support that these are not necessary for frailty. Indeed, there are many older people who report fatigue, or low mood, or low health perception, but who are not "frail". The same argument would apply to participation—it is not necessary for frailty; some quite robust people may not engage in activities and some frail people may still be engaged. Continence was not tested in this model but would not be necessary for frailty as there are many reasons for incontinence beyond frailty. Nutritional status was not tested in this model beyond BMI and so further testing of markers of nutritional status is warranted. The observation that a cognitive item fit with physical function and performance items suggests that the frailty latent is not just purely physical function. An extensive review by Bortone et al.[48] summarized how measures of gait fit with other frailty constructs, including cognition. Here the MoCA[46] fit as a total score, so some unpacking of the cognitive domain is warranted to see which cognitive items fit.

The observation that both self-report and performance-based measures fit together suggests that the source of the information does not define the construct. This was also demonstrated by Theou et al.[49] who observed, using data from some 4,900 people enrolled in The Irish LongituDinal study on Ageing (TILDA), that the best prediction of health

TABLE 4.
The Frailty Ladder Version 1[a]

Linear Score (logit scale)	Item Level
53	6MWT > 600 m
52	Grip strength (W) > 20-23 kg
51	Total muscle mass (M) 70-74% / (W) > 75%
50	FVC$_{pred}$ 75-84%
48.5 (3.0 logits)	Total muscle mass (M) = 60-64% / (W) = 65-69%
48.5	TUGT = 0-6 s
46.5	STS (30 sec) = 32-50 repetitions
46.5	No of stairs (20s) = 6
44.5	6MWT = 526- 600 m
44.5	STS (30 sec) =26-31 repetitions
42	Little to no difficulty running on even ground
42	FEV$_{1pred}$ >90%
42	STS (30 sec) = 23-25 repititions
39.5	Total muscle mass (M) = 55-59%
39.5	6MWT = 481-525 m
37 (2.0 logits)	6MWT = 431-480 m
37	STS (30 sec) = 20-22 repetitions
37	Grip strength (W) > 14-20 kg
35	Single leg stand > 27 s
33.5	STS (30 sec) = 18-19 repetitions
33.5	Grip strength (M) > 20-23 kg
31	STS (30 sec) = 16-17 repititions
31	6MWT = 400-430 m
31	Any limitation in vigorous activities
29.5 (1 logit)	Total muscle mass (W) = 55-59%
29.5	No difficulty in standing for 1 hour
27	MoCA = 25-30
27	> 240 min activities/day
27	TUGT = >6 - 10 s
25	Sit-to-stand (STS 30 sec) = 13-15 repetitions
24	No of stairs (20s) = 22-32
23	No difficulty squatting
21.5 (0.0 logits)	No difficulty in going up and down 10 stairs
21.5	Any difficulty in standing for 1 hour
19.5	Total muscle mass (W) = 50-54%
19.5	Grip strength (M) = 14-20 kg / (W) = 10-14 kg
18	No difficulty in walking 2 blocks

TABLE 4. Continued

Linear Score (logit scale)	Item Level
17	Total muscle mass (M) = 50-54%
16 (-1.0 logits)	TUGT 10.5 - 20.5 s
14.5	FVCpred = 60-74%
14.5	Any difficulty with 10 stairs
12.5	Grip strength (M) = 10-14 kg
12.5	No difficulty in getting into or out of a car
11	Any limitation in moderate activities
10 (-2.0 logits)	Any difficulty in putting on shoes/socks
9	Any difficulty in usual hobbies, recreational or sports
7.5	Any difficulty in getting into or out of bath
7.5	Any difficulty lifting an object
6	Any difficulty in usual work or household activities
5	Any difficulty in performing light household activities
3.5 (-3 logits)	Even a little difficulty in getting into or out of a car.
3.5	Any difficulty in walking between rooms
2 (≈-3.5 logits)	Any difficulty for sitting for 1 hour
1 (-5 logits)	Any difficulty rolling over in bed
0	Quite to moderate difficulty in getting into or out of a car

[a]Here scores are based on the ordinal structure of the data recognizing that many items are at the same logit level (similar shading).

outcomes came when both self-report and performance-based measures were combined.

The other necessary component for a reflective model is evidence of causal relationship from the construct to the items. Figure 3 presents such a model linking frailty causes to frailty consequences which can now be measured with the Frailty Ladder. The causes of frailty are systemic illnesses, unhealthy lifestyle, genes, and the environment.[1,14,50] If these causes were prevented or treated, then frailty would be avoided or reversed, and this reversal would be reflected in the items included in Version 1 of the Frailty Ladder. The reverse is not true: targeting the items will not necessarily change frailty, but physical capacity and performance would likely improve.

Measurement of the effects of frailty interventions is a challenge. A systematic review of 47 studies of exercise interventions targeting frailty[3] found that each study used a multitude of outcome indicators, analyzed separately.

This preliminary work indicates that the Frailty Ladder would be an efficient and statistically robust way of combining the results of different frailty tests into one outcome measure. It would also be a way of identifying what outcomes to target in a personalized intervention targeting frailty. The rungs of the Frailty Ladder, the hierarchy, could be used to guide treatment goals.

The practicality for a clinical setting is that any questionnaire items or tests could be used to identify where a patient is located on the ladder. Not all tests need to be administered. For example, one test that covers a wide range of the linear continuum is the TUG, administered at a comfortable pace and fast as possible. A typical procedure with measures such as this is to test people on the middle item (logit 0) which is grip strength. If the person's grip strength surpasses this threshold, a harder test can be administered. If muscle mass is available, this can also be used to situate the person on the ladder. Questionnaire items are only useful up to a certain point (logit 0). If a person has extreme or quite a bit of difficulty going up or down 10 stairs or more than a little difficulty standing for an hour (items from the LEFS), then the person is at the mid-point of the Frailty Ladder. To situate the person higher up, performance tests are needed. A feature of the Rasch model is that, when the data fit the model, any subset of items can provide the same information as all.

Interestingly, the components of the most widely used index of frailty, Fried's frailty phenotype,[9] (weight loss, exhaustion, slow gait, weak grip, and low level of activity) are represented in the Frailty Ladder. Still, all of the items are at the lowest end of the scale, situated at 7.5 to <27 out of a total score of 53 (-3.5 to 0 on the logit scale). The level of activity that best fit the model of frailty shown here was <240 min of physical activity per week. The range of activity considered "sedentary" is <150 min of moderate or higher intensity activity per week.[51] Thus, Fried's frailty phenotype, while highly sensitive, may be detecting people who may not be able to rebound out of frailty. Detecting people who are descending into frailty and intervening to stop this descent would be a better physical therapeutic and public health approach, and this ladder provides a way of identifying people entering this trajectory.

Strengths and Limitations

There are both strengths and limitations associated with this study. Having three data sets was a strength in that a wide range of abilities was covered, which is a requirement for creating a measure. We particularly selected data sets that had a variety of functional items, including a data set with tests of pulmonary function which are rarely collected in the geriatric context. In this way we made a unique contribution to the field by showing that pulmonary function and cognition fit with the frailty construct. The main limitation of this approach is that no one data source had all the necessary data, but by using Rasch analysis, we were able to create mathematical links between the data sets because of common items and a hierarchy among the items. This was a proof-of-concept study taking advantage of rich data sets with a wide variety
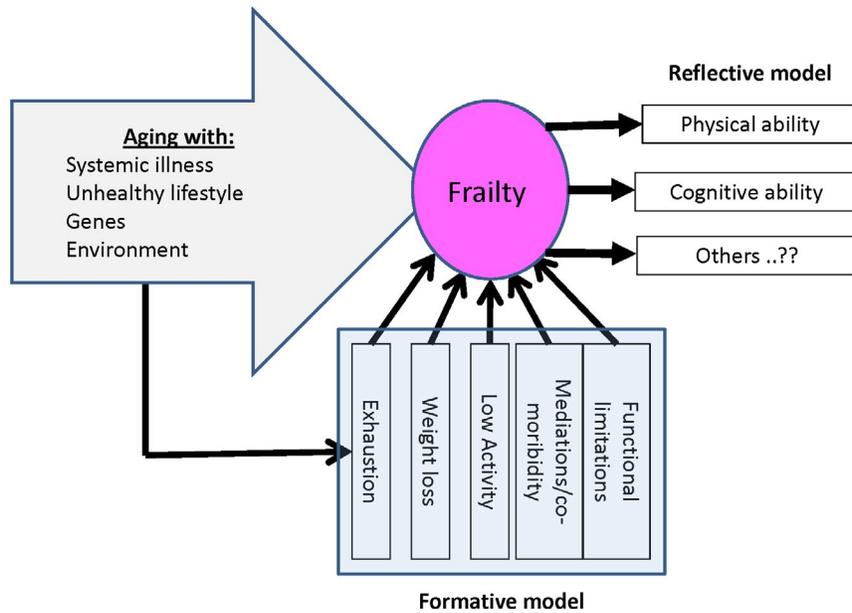
FIGURE 3. Theoretical model of frailty causes and consequences

of measures. Replication of the model in independent data sets is required.

CONCLUSIONS

Decades of research on frailty have consolidated our understanding of frailty, but measurement has not kept up with modern psychometric developments. This study bridges the measurement gap by enhancing our capacity to develop and test effective interventions to prevent and reverse frailty. Based on the results of this study, we propose that frailty is: "A dynamic health state experienced by an older person who, in the face of systemic illness or health threat, fails to maintain physiological and functional benchmarks expected for age and sex."

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST DISCLOSURES

We have read and understood the Canadian Geriatrics Journal's policy on conflicts of interest disclosure and declare there are none.

FUNDING

REFERENCES

1. Vina J, Tarazona-Santabalbina FJ, Perez-Ros P, Martinez-Arnau FM, Borras C, Olaso-Gonzalez G, et al. Biology of frailty: Modulation of ageing genes and its importance to prevent age-associated loss of function. Mol Aspects Med. 2016;50:88–108.

2. Cesari M, Costa N, Hoogendijk EO, Vellas B, Canevelli M, Perez-Zepeda MU. How the Frailty Index may support the allocation of health care resources: an example from the INCUR Study. J Am Med Dir Assoc. 2016;17(5):448–50.

3. Theou O, Stathokostas L, Roland KP, Jakobi JM, Patterson C, Vandervoort AA, et al. The effectiveness of exercise interventions for the management of frailty: a systematic review. J Aging Res. 2011;2011:569194.

4. Theou O, Jones GR, Jakobi JM, Mitnitski A, Vandervoort AA. A comparison of the relationship of 14 performance-based measures with frailty in older women. Appl Physiol Nutr Metab. 2011;36(6):928–38.

5. Cesari M, Prince M, Thiyagarajan JA, De Carvalho IA, Bernabei R, Chan P, et al. Frailty: An Emerging Public Health Priority. J Am Med Dir Assoc. 2016;17(3):188–92.

6. Cesari M, Nobili A, Vitale G. Frailty and sarcopenia: from theory to clinical implementation and public health relevance. Eur J Intern Med. 2016;35:1–9.

7. Vina J, Salvador-Pascual A, Tarazona-Santabalbina FJ, Rodriguez-Manas L, Gomez-Cabrera MC. Exercise training as a drug to treat age associated frailty. Free Radic Biol Med. 2016; 98:159–64.

8. Marzetti E, Calvani R, Tosato M, Cesari M, Di BM, Cherubini A, et al. Physical activity and exercise as countermeasures to physical frailty and sarcopenia. Aging Clin Exp Res. 2017;29(1):35–42.

9. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, et al. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci. 2001;56(3):M146–M57.

10. Andrew MK, Mitnitski AB, Rockwood K. Social vulnerability, frailty and mortality in elderly people. PLoS One. 2008;3(5):e2232.

11. Brinkman S, Voortman T, Kiefte-de Jong JC, van Rooij FJA, Ikram MA, Rivadeneira F, et al. The association between lifestyle and overall health, using the frailty index. Arch Gerontol Geriatr. 2018;76:85–91.

12. Livshits G, Ni Lochlainn M, Malkin I, Bowyer R, Verdi S, Steves CJ, et al. Shared genetic influence on frailty and chronic widespread pain: a study from TwinsUK. Age Ageing. 2018;47(1):119–25.

13. Lang PO, Michel JP, Zekry D. Frailty syndrome: a transitional state in a dynamic process. Gerontology. 2009;55(5):539–49.

14. Levasseur M, Genereux M, Bruneau JF, Vanasse A, Chabot E, Beaulac C, et al. Importance of proximity to resources, social support, transportation and neighborhood security for mobility and social participation in older adults: results from a scoping study. BMC Public Health. 2015;15(1):503.

15. Gobbens RJ, Luijkx KG, Wijnen-Sponselee MT, Schols JM. In search of an integral conceptual definition of frailty: opinions of experts. J Am Med Dir Assoc. 2010;11(5):338–43.

16. Rodriguez-Manas L, Feart C, Mann G, Vina J, Chatterji S, Chodzko-Zajko W, et al. Searching for an operational definition of frailty: a Delphi method based consensus statement: the frailty operative definition-consensus conference project. J Gerontol A Biol Sci Med Sci. 2013;68(1):62–67.

17. Canadian Frailty Network. What is frailty? Available from: http://www.cfn-nce.ca/frailty-in-canada/

18. Enright PL, McBurnie MA, Bittner V, Tracy RP, McNamara R, Arnold A, et al. The 6-min walk test - A quick measure of functional status in elderly adults. Chest. 2003;123(2):387–98.

19. Nunnally JC. Psychometric theory, 2nd edition. New York: Mc Graw-Hill; 1978.

20. Saxon D. In praise of Lord Kelvin [feature]. Physics World. Published Dec. 17, 2007. Available from: https://physicsworld.com/a/in-praise-of-lord-kelvin

21. Rasch G. Probabilistic models for some intelligence and attainment tests. Chicago: University of Chicago Press; 1980.

22. Masters GN. The key to objective measurement in the psychosocial sciences. Australian Council for Educational Research; 2001.

23. Conrad KJ, Smith EV. International Conference on objective measurement: applications of Rasch analysis in health care. Med Care. 2004;42(1).

24. Hogan DB, MacKnight C, Bergman H. Models, definitions, and criteria of frailty. Aging Clin Exp Res. 2003;15(3 Suppl):1–29.

25. World Health Organization. International classification of functioning, disability and health. Geneva: WHO; 2001.

26. Azzopardi RV, Vermeiren S, Gorus E, Habbig AK, Petrovic M, Van Den Noortgate N, et al. Linking frailty instruments to the International Classification of Functioning, Disability, and Health: a systematic review. J Am Med Dir Assoc. 2016;17(11):1066.

27. Van Damme JK, Lemmon K, Oremus M, Neiterman E, Stolee P. Understanding frailty screening: a domain mapping exercise. Can Geriatr J. 2021;24(2):154–61.

28. Guaraldi G, Milic J. The interplay between frailty and intrinsic capacity in aging and HIV infection. AIDS Res Hum Retroviruses. 2019;35(11-12):1013–22.

29. Stenner AJ, Burdick DS, Stone MH. Formative and reflective models: can a Rasch analysis tell the difference? Rasch Measure Transact. 2008;22(1):1152–53.

30. Stenner AJ. Point-biserial fit indices. Rasch Measure Transact. 1995;9(1):416.

31. Peterson CH, Gischlar KL, Peterson NA. Item construction using reflective, formative, or Rasch measurement models: implications for group work. J Special Group Work. 2017;42(1):17–32.

32. Suikkanen S, Soukkio P, Aartolahti E, Kääriä S, Kautiainen H, Hupli MT, et al. Effect of 12-month supervised, home-based physical exercise on functioning among persons with signs of frailty: a randomized controlled trial. Arch Phys Med Rehabil. 2021;102(12):2283–90.

33. Daryanti Saragih I, Yang YP, Saragih IS, Batubara SO, Lin CJ. Effects of resistance bands exercise for frail older adults: a systematic review and meta-analysis of randomised controlled studies. J Clin Nurs. 2022;31(1-2):43–61.

34. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. Health Technol Assess. 2009;13(12):1–200.

35. Mayo NE, Moriello C, Scott SC, Dawes D, Auais M, Chasen M. Pedometer-facilitated walking intervention shows promising effectiveness for reducing cancer fatigue: a pilot randomized trial. Clin Rehabil. 2014;28(12):1198–209.

36. Carli F, Charlebois P, Stein B, Feldman L, Zavorsky G, Kim DJ, et al. Randomized clinical trial of prehabilitation in colorectal surgery. Br J Surg. 2010;97(8):1187–97.

37. Barbat-Artigas S, Pion CH, Leduc-Gaudet JP, Rolland Y, Aubertin-Leheudre M. Exploring the role of muscle mass, obesity, and age in the relationship between muscle quality and physical function. J Am Med Dir Assoc. 2014;15(4):303.

38. Fiore JF, Jr., Castelino T, Pecorelli N, Niculiseanu P, Balvardi S, Hershorn O, et al. Ensuring early mobilization within an enhanced recovery program for colorectal surgery: a randomized controlled trial. Ann Surg. 2017;266(2):223–31.

39. Auais M, Morin SN, Finch L, Ahmed S, Mayo N. Toward a meaningful definition of recovery after hip fracture: comparing two definitions for community-dwelling older adults. Arch Phys Med Rehabil. 2018;99(6):1108–15.

40. 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(9):1537–51.

41. Andrich D. Rasch models for ordered response categories. Enc Stat Behav Sci. 2005;4:1698–707.

42. Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). Br J Clin Psychol. 2007;46(Pt 1):1–18.

43. de Vet HC, Terwee CB, Mokkink LKD. Measurement in medicine: a practical guide. Cambridge, UK: Cambridge University Press; 2011.

44. Hays RD, Morales LS. The RAND-36 measure of health-related quality of life. Ann Med. 2001;33(5):350–57.

45. Binkley JM, Stratford PW, Lott SA, Riddle DL, North American Orthopaedic Rehabilitation Research Network. The Lower Extremity Functional Scale (LEFS): scale development, measurement properties, and clinical application. Phys Ther. 1999;79(4):371–83.

46. Nasreddine ZS, Phillips NA, Bedirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc. 2005;53(4):695–99.

47. Mayo NE, Bronstein D, Scott SC, Finch LE, Miller S. Necessary and sufficient causes of participation post-stroke: practical and philosophical perspectives. Qual Life Res. 2014;23(1):39–47.

48. Bortone I, Sardone R, Lampignano L, Castellana F, Zupo R, Lozupone M, et al. How gait influences frailty models and health-related outcomes in clinical-based and population-based studies: a systematic review. J Cachexia Sarcopenia Muscle. 2021;12(2):274–97.

49. Theou O, O'Connell MD, King-Kallimanis BL, O'Halloran AM, Rockwood K, Kenny RA. Measuring frailty using self-report and test-based health measures. Age Ageing. 2015;44(3):471–77.

50. Lang PO, Michel JP, Zekry D. Frailty syndrome: a transitional state in a dynamic process. Gerontology. 2009;55(5):539–49.

51. Owen N, Healy GN, Matthews CE, Dunstan DW. Too much sitting: the population health science of sedentary behavior. Exerc Sport Sci Rev. 2010;38(3):105–13.

52. Nightingale EJ, Pourkazemi F, Hiller CE. Systematic review of timed stair tests. J Rehabil Res Dev. 2014;51(3):335-50.

53. Bohannon RW. Reference values for the five-repetition sit-to-stand test: a descriptive meta-analysis of data from elders. Percept Motor Skills. 2006;103(1):215–22.

54. Bohannon RW, Larkin PA, Cook AC, Gear J, Singer J. Decrease in timed balance test scores with aging. Phys Ther. 1984;64(7):1067–70.

55. Bohannon RW. Reference values for the timed up and go test: a descriptive meta-analysis. J Geriatr Phys Ther. 2006;29(2):64–68.

56. Bohannon RW. Test-retest reliability of hand-held dynamometry during a single session of strength assessment. Phys Ther. 1986;66(2):206–09.

Correspondence to: Nancy E Mayo, bsc(pt), msc, phd, Department of Medicine School of Physical and Occupational Therapy, McGill University, Division of Clinical Epidemiology Division of Geriatrics, McGill University Health Center (MUHC), Center for Outcomes Research and Evaluation (CORE), MUHC-Research Institute, 5252 de Maisonneuve, Office 2B:43, Montreal, QC, H4A 3S5
Email: nancymayo@mcgill.ca