Unconventional Views of Frailty

A Comparison of Two Approaches to Measuring Frailty in Elderly People

Kenneth Rockwood,[1,2] Melissa Andrew,[1,2] and Arnold Mitnitski[2]

[1]Division of Geriatric Medicine, [2]Department of Medicine, Dalhousie University and Capital District Health Authority, Halifax, Nova Scotia, Canada.

Background. Many definitions of frailty exist, but few have been directly compared. We compared the relationship between a definition of frailty based on a specific phenotype with one based on an index of deficit accumulation.

Methods. The data come from all 2305 people 70 years old and older who composed the clinical examination cohort of the second wave of the Canadian Study of Health and Aging. We tested convergent validity by correlating the measures with each other and with other health status measures, and analyzed cumulative index distributions in relation to phenotype. To test criterion validity, we evaluated survival (institutionalization and all-cause mortality) by frailty index (FI) score, stratified by the phenotypic definitions as ''robust,'' ''pre-frail,'' and ''frail.''

Results. The measures correlated moderately well with each other ($R = 0.65$) and with measures of function (phenotypic definition $R = 0.66$; FI $R = 0.73$) but less well with cognition (phenotypic definition $R = -0.35$; FI $R = -0.58$). The median FI scores increased from 0.12 for the robust to 0.30 for the pre-frail and 0.44 for the frail. Survival was also lower with increasing frailty, and institutionalization was more common, but within each phenotypic class, there were marked differences in outcomes based on the FI values—e.g., among robust people, the median 5-year survival for those with lower FI values was 85%, compared with 55% for those with higher FI values.

Conclusion. The phenotypic definition of frailty, which offers ready clinical operationalization, discriminates broad levels of risk. The FI requires additional clinical translation, but allows the risk of adverse outcomes to be defined more precisely.

FRAILTY is an important clinical and public health problem. How to define frailty, however, remains controversial (1). Of the many definitions of frailty (2), few have been directly compared with each other (3–5). Given the clinical (6) and public health (7) importance of frailty, we compared two commonly used approaches. The first defines frailty based on a specific phenotype, consisting of five items, any three of which mark a person as recognizably frail (8). The second pays less attention to which items are present in a person who is frail, but rather counts the number of things that people have wrong with them, to propose a frailty index (FI) based on a count of accumulated deficits (9). Each has its advocates—the clinical reproducibility of the former typically seen as a strong point; the mathematical properties, and the ability to make clinical inferences from such properties being seen as strengths of the latter. Still, they have not been compared directly. Here, we were interested in how the measures correlated with each other and with other health measures, and how they predicted adverse outcomes.

METHODS

Setting, Sample, and Measures

We analyzed data from the second clinical examination of the Canadian Study of Health and Aging (CSHA), a cohort study of dementia and other health problems of elderly people (10). In 1990–1991, 9008 community-dwelling elderly people were screened for cognitive impairment using the modified Mini-Mental State Examination (3MS) (11), and those who screened positive, with a sample who screened negative, were invited for a structured clinical examination (12). Five years later, at CSHA-2, the clinical examination cohort was expanded with more persons who had screened negative for cognitive impairment, so that frailty could be better studied (5). The CSHA-2 clinical cohort thus comprises 716 residents of long-term care institutions and 1589 community-dwelling people, of whom 767 had no cognitive impairment, 528 were cognitively impaired but did not meet *Diagnostic and Statistical Manual, Revised 3ʳᵈ Edition* (DSM-III-R) (13) criteria for dementia, and 294 had dementia.

The CSHA-2 data collection protocol was expanded to include physical performance measures, a clinical frailty scale, and a standardized comprehensive assessment. In consequence, there are enough data in the CSHA-2 clinical examination for us to compare the deficit accumulation approach to the phenotypic approach for those people.

The variables used to operationalize the phenotypic definition of frailty were similar to those used in the phenotypic definition studies (Table 1) (14). Weight loss was defined as loss of either ≥ 10 pounds or $\geq 5\%$ of body weight in

Table 1. Frailty-Defining Criteria in the Canadian Study of Health and Aging (CSHA), and Comparison of Percentage of People With Each Characteristic in CSHA to the Women's Health and Aging Study (WHAS) and the Cardiovascular Health Study (CHS)

Characteristic	Definition	% of People With This Characteristic in the CSHA	% of People With this Characteristic in	
			WHAS*	CHS*
Weight loss	In last 12 mo, weight has decreased by $\geq$ 10 lb or 5% of body weight	17.5	12.7	7.3
Exhaustion	Feels tired all the time	15.5	14.1	21.3
Low energy expenditure	Unable to walk or needs help to walk	27.0	19.8	24.1
Slowness	TUG > 19 s (based on the random sample of community dwellers brought to clinical examination)	43.2	31.3	38.0
Weakness	Clearly abnormal strength on physical examination	21.8	20.8	26.2
Overall frailty status	Robust	47.7	44.9	33.2
	Pre-frail	35.7	43.8	55.2
	frail	16.5	11.3	11.6

Notes: *Data in these two columns came from Table 1 in Bandeen-Roche et al. *J Gerontol Biol Sci Med Sci.* 2006;61A:262–266.
TUG = Timed Up and Go.

the past year. Exhaustion (poor endurance and energy) was based on self-report of feeling "tired all the time." Low physical activity levels and energy expenditure were operationalized as needing assistance with walking or being unable to walk. Slowness was defined as a time of $\geq$ 19 seconds on the timed up and go (TUG) test (15). Norms for the TUG are not well established in the literature. The 19-second cutoff was chosen on the basis of time distributions in the CSHA, one of the largest population-based studies to include the TUG. The phenotypic definition originally defined slowness as the slowest 20% of the population walking 15 feet (8). As the CSHA-2 clinical sample includes residents of institutions and cognitively impaired older adults, it is not representative of the population at large. A cutoff of 19 seconds approximately identifies the slowest quintile among the random sample of noninstitutionalized individuals brought to clinical examination in the CSHA. Weakness was identified as abnormal strength on physical examination.

The FI was developed using 70 deficits from the clinical examination (5,16). The individual items are available at http://myweb.dal.ca/amitnits/CSHAclinical-variables.jpg. Items included the presence and/or severity of current diseases, ability in activities of daily living (ADL), and physical signs from the clinical and neurological examinations. Each deficit was dichotomized or trichotomized and mapped to the interval 0–1 (i.e., individual items had scores of 0, 0.33, 0.50, 0.67, or 1.0), representing the occurrence and severity of the problem. For each person, a 70-dimensional vector was constructed. For example, a person with seven deficits would have an index score of 7/70 = 0.10. No variable had > 5% missing values; where missing values existed, they were imputed using the relevant mean.

In addition, we compared the two approaches to frailty with other health status measures. Function was summarized using the CSHA function score (5). It is based on the Older American Resources Survey, with 12 instrumental ADL (IADL) and ADL items (17). Functional Reach describes how far forward an individual can move their fully extended arm by bending at the waist; typical performances range from $\geq$ 25 cm to $\leq$ 15 cm (18).

Analysis

We tested convergent validity (validation terminology follows Streiner and Norman) (19) by correlating the measures (Pearson or Spearman, as appropriate) with each other and with other health status measures, and analyzed cumulative index distributions in relation to phenotype. To test criterion validity, we evaluated survival (institutionalization and all-cause mortality) by FI score, stratified by the phenotypic definitions as "robust," "frail," and "pre-frail." Differences in survival were tested using the log-rank test and Cox's F test, as appropriate.

We compared mean FI scores using analysis of variance. We also evaluated the 99% upper limit of the FI distribution, something that appears to vary little by age (20,21). As a second analysis, we repeated this approach for people with no, one, two, three, four, or five of the items that make up the phenotypic definition.

As we could not replicate two Cardiovascular Health Study (CHS) items ("low physical activity" and "weakness") with performance measures, we evaluated individual CHS items in several ways. First we did additional convergent validation by comparing each item's impact on performance of the Functional Reach, which was not used in either frailty measure. In several other analyses we substituted these two items with other measures. In these analyses, if changing an item gave highly variable results, then the reliability of our conclusions would be suspect. If, in contrast, the reliability of the results did not depend on how the CHS items were operationalized, then the fact that we could not replicate the two CHS items exactly would be of little consequence. Thus we evaluated two other candidate "low physical activity" items being "irregular gait pattern" from the physician's examination and "problems going out alone" and "impaired mobility" from the nursing history. Next, we carried out resampling analyses (22) to evaluate item dependency of the CHS operationalization. To do this, we classified people, as before, such that no items = robust, 1 or 2 = pre-frail, and 3–5 = frail. Instead of using only the five original CHS items, we added to these five another six related items (irregular gait pattern, poor standing posture, poor muscle tone, bradykinesia, impaired mobility, problems
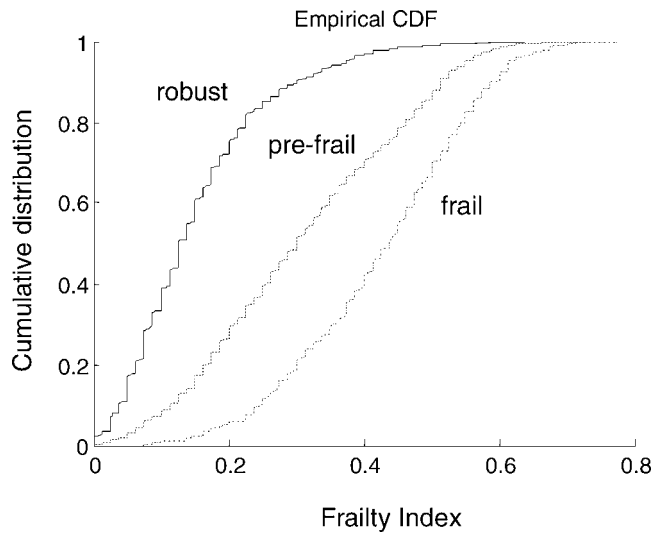
Figure 1. Cumulative distributions of deficit accumulation, classified by the phenotypic definition as robust, pre-frail, and frail. More deficits accumulate with each level of frailty. CDF, cumulative distribution function.

going out alone) and four more (memory changes, sleep changes, feeling sad or depressed, changes in everyday health). To evaluate the performance of this approach, we repeated 100 iterations and calculated means and standard deviations of the median points of the cumulative distributions of the FI values. Next, we again considered 15 variables, but instead of defining frailty according to the number (of five) that were present, we sampled 10 (and correspondingly defined "robust" as when at most 1 was present, "pre-frail" as 2–5, and "frail" as ≥ 6). Finally, we simply repeated these analyses, using randomly selected variables.

RESULTS

The FI and the phenotypic definition are moderately correlated with each other ($R = 0.65$). Each was also correlated with the CSHA function measure (phenotypic definition/function $R = 0.66$; FI/function $R = 0.73$) and with the 3MS (phenotypic definition/3MS $R = -0.35$; FI/3MS $R = -0.58$). The individual items that make up the CHS definition each are associated with impairment on the Functional Reach test. For every centimeter of reach, the odds of weight loss was decreased (odds ratio [OR] = 0.98; 95% confidence interval [CI], 0.97–0.99), and similarly for strength (OR = 0.93, 95% CI, 0.92–0.94), exhaustion (OR = 0.97, 95% CI, 0.96–0.99), slowing (OR = 0.87, 95% CI, 0.86–0.88), and low physical activity (OR = 0.88, 95% CI, 0.86–0.89).

The cumulative distributions of people classified by the phenotypic definition as robust, pre-frail, or frail each have distinct cumulative density distributions (Figure 1). The median values likewise increase across the three classifications (from 0.12 to 0.30 to 0.44, respectively) ($p < .0001$). So too do the 0.99 limits, from 0.50 to 0.62 to 0.70. While many people classified as robust have high FI values, few people with phenotypic frailty have low FI values.

Similar trends are evident in the FI cumulative distributions across the number of items used in the phenotypic
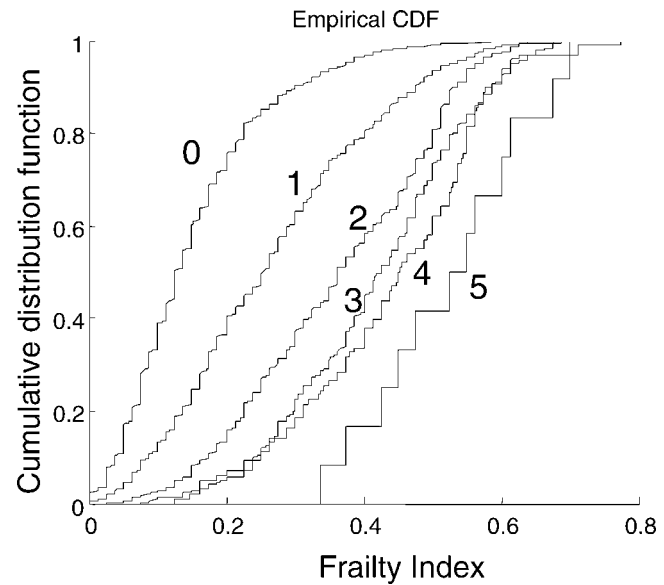
Figure 2. Cumulative distributions of deficit accumulation across the number of items considered in the phenotypic definition of frailty (from no items present to all five items present). The more items included in the frailty definition, the more deficits are accumulated. CDF, cumulative distribution function.

definition (Figure 2). Increases are observed in the median values, from 0.12 (when no elements of the definition are present) to 0.54 (when all five are present). The same holds for the 0.99 limit, which increased from 0.50 to 0.70.

The FI is not meant to be dichotomized into frail or robust, but an empirical cut-point for the present purposes is about 0.25 (Figure 3). This value corresponds to the crossing point of the robust and pre-frail groups. Five-year survival was highest for the robust group, and lowest for the frail group (Figure 4A), $p < .05$. Within these strata, however, people with higher FI values (> 0.25) had worse
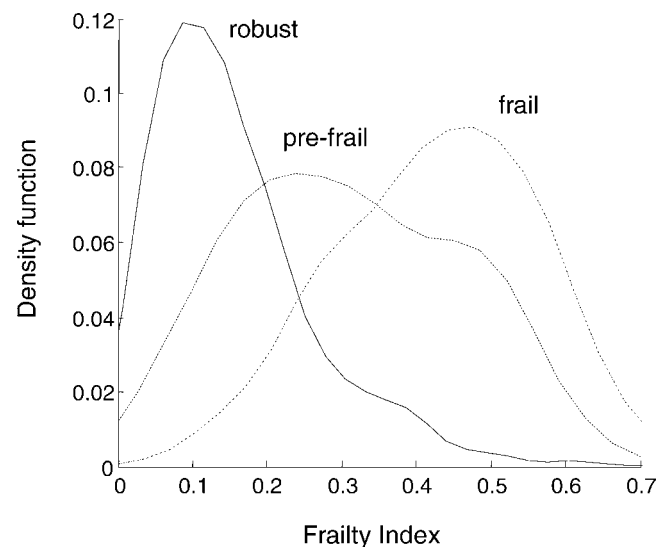
Figure 3. Density distributions of deficits, smoothed by a Gaussian kernel function, for people classified by the phenotypic definition as robust, pre-frail, or frail. The overlap in deficit accumulation between persons who are robust and those who are frail occurs close to the median of the robust, ~0.25.
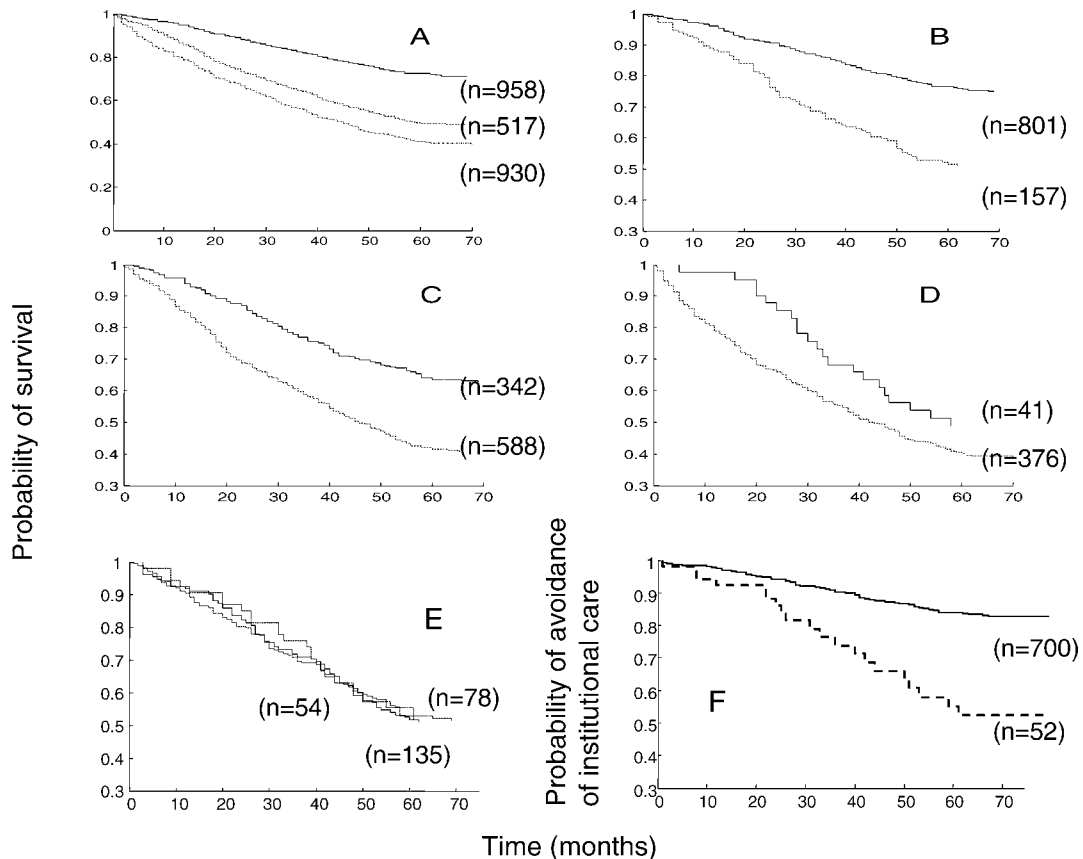
Figure 4. Kaplan–Meier 5-year survival curves for people by the phenotypic definition of frailty. **A,** Robust people (*upper line, solid*) have the highest survival; frail people (*lowest line, dashed*) have the lowest, and the pre-frail are in between. **B,** Five-year survival for people defined as robust, stratified by two levels of the frailty index. This is repeated in **C** for the pre-frail, and in **D** for the frail. Within each phenotypic stratum, people with higher degrees of frailty (≥ 0.25, *dashed lines*) have worse survival than those with less frailty (< 0.25, *solid lines*). **E,** Survival of people with intermediate frailty index values (0.25 ± 0.05) is shown stratified by classification as robust, pre-frail, or frail. **F,** Differences in the proportion of robust people who become institutionalized. *Solid line:* people with frailty index values < 0.25; *dashed line:* people with values ≥ 0.25.

survival than those with less frailty (≤ 0.25) ($p < .05$, Figure 4B–D). For example, among people who were phenotypically robust, the median 5-year survival for those with lower FI values was 85%, compared with a 55% median 5-year survival among those with higher FI values. By contrast, survival curves by CHS classification considerably overlap for people with intermediate frailty (Figure 4E). Institutionalization outcomes by CHS classification stratified by FI values show results similar to the mortality analyses (Figure 4F).

Varying exactly which items are used to define frailty phenotypically appeared to have less influence on FI cumulative distributions than did varying the number of items that were used (Figure 5). We first used the original 5 items, and added 10 more, then sampled 5 from these 15 to construct a phenotypic classification of robustness/pre-frailty/frailty (Figure 5A). The results varied little from those of a similar analysis, which used 5 of 15 variables selected at random from the 70 that made up the FI (Figure 5B). By contrast, when we increased the number of variables that we sampled to 10, there was better separation in the cumulative distributions, whether we used the original 5 items and 10 related ones (Figure 5C), or whether we chose 10 from 15

items chosen at random (Figure 5D). There were no significant differences in the distributions (e.g., the mean of the median FI values for people classified as robust was 0.103 ± 0.017 in A, 0.100 ± 0.016 in B, 0.103 ± 0.009 in C, and 0.094 ± 0.01 in D).

DISCUSSION

We compared two approaches to frailty, and showed considerable convergence between the phenotypic definition of frailty (8,14,23,24) and the method of considering frailty in relation to deficit accumulation (9,25–27). Moving from the spectrum of robust to pre-frail to frail (and through a more finely graded approach of counting each of the five deficits in the phenotypic definition) we see an increase in the cumulative distributions of the FI. Adverse outcomes occurred more commonly among people who were frail, however defined.

Our data must be interpreted with caution. As with others who have replicated the work (28), we did not have each of the variables operationalized exactly as proposed by Fried and colleagues (8), although even within the phenotypic definition reports there are subtle differences (14). This is
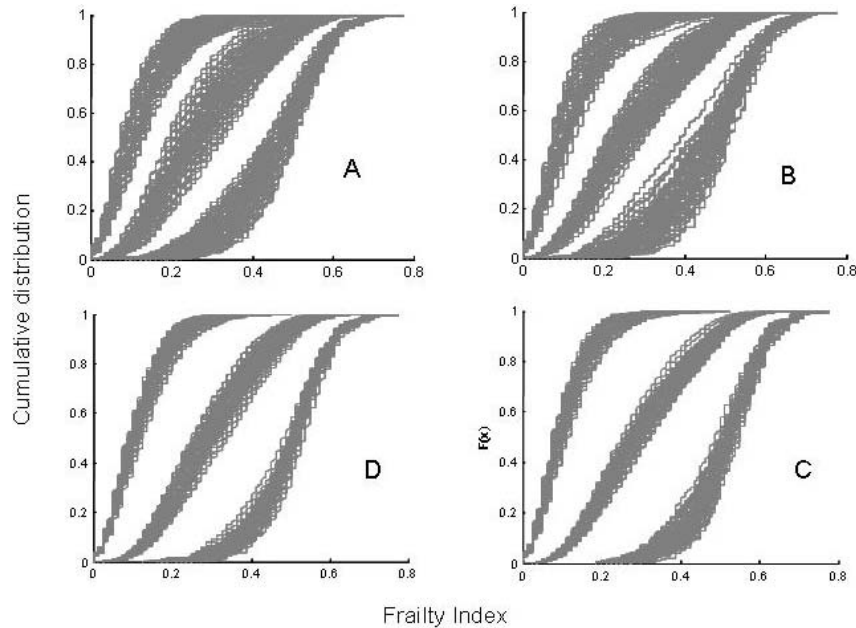
Figure 5. Cumulative distributions of frailty index scores in relation to the items that are used to operationally classify people as robust (*left*), pre-frail (*middle*), or frail (*right*). **A,** The 5 items that make up the Cardiovascular Health Study (CHS) definition are included in a set of 15 items related to defining frailty in relation to weight loss, weakness, slowness, tiredness, or low physical activity. Lines represent estimates of the frailty index distribution for varying operational definitions used to classify frailty, based on random sampling of 5 variables from the set of the 15 related items, i.e., the 5 that make up the definition, plus 10 related items. **B,** Robust/pre-frail/frail status is defined according to the presence of any five items, selected randomly from the list of variables (http://myweb.dal.ca/amitnits/CSHAclinical-variables.jpg), but excluding any of the 15 previously used items. Note that the distributions of the frailty index scores are similar in each panel. **C,** Repeat of the random sampling done in the variables considered in **A,** but now 10 items at a time (cf. 5 in **A**) are used to classify people as robust, pre-frail, or frail. Similarly, **D** uses the same items as in **B,** but increases the number of items used in each iteration to 10. In general, the distributions of the frailty index scores are similar no matter which specific items are used to classify people as robust, pre-frail, or frail. The precision of the estimates (reflected in the standard deviations of the median values, or here in the width of the distribution of the lines) decreases as more item are considered.

less of a problem with the FI approach, which need not use the same items, or even the same number of items, to estimate the proportions that represent the index's values. Indeed, random selection of variables yield comparable estimates (22), although to evaluate changes in individuals over time, it remains necessary to compare like with like (29,30). Still, the prevalence estimates for individual items between our work and earlier work are very close (Table 1). In general, our estimates are slightly higher, reflecting that our sample is older (mean age 81.6 years, range 69–109 years) and included more people selected for cognitive impairment compared with CHS/Women's Health and Aging Study (WHAS) estimates for people aged 70–79 years (14). In addition, the analyses show that varying the two measures that we could not exactly replicate for the CHS definition (i.e., low physical activity and weakness) is unlikely to have a large effect on our results. In particular, Figure 5 suggests that it is not plausible that more precise operationalization of low physical activity and of weakness would have an effect as big as simply considering more measures in a frailty phenotype definition. Even so, only a head-to-head comparison of the two approaches, each operationalized according to accepted conventions, can clarify their relative contributions.

Despite the convergence of the two approaches in these analyses, differences remain. Perhaps the most important is conceptual. The FI approach does not assume that the elements (or groups of elements) that make up frailty are statistically independent. In consequence, we are less per-suaded of the need to begin with a clinical syndrome of distinct elements. That decision affects both our operational program and how we understand some of the phenotypic definition work. For example, a recent latent class analysis has suggested that three clusters of frailty from the phe-notypic definition are identifiable (14). The presence of a dose-response in the FI by accumulation of the items that make up the phenotypic definition would suggest, however, that finer grades are possible still. In our view, given that the three syndromes came from a consideration of five elements, their robustness needs further testing not just by cross-validation but by revisiting the latent class analysis to con-sider more elements.

Although we do not see the need to begin with the clinical syndrome, we still aim to end up there. In consequence, we have cross-validated the deficit accumulation approach by counting the items in a standardized Comprehensive Geriatric Assessment (31,32). Still, clinical use of the FI remains to be fully demonstrated, which is why, even given some evidence for cross-validation of the clinical phenotype (14,28), we note that not every attempt at cross-validation has had the same success (33) and that questions about the final formulation of the phenotype, and whether it is one or many, remain (24).

The phenotypic definition allows for mechanisms to be explored, by testing for shared elements in the pathophysi-ology of each item (24). In contrast, some element of

tautology seems unavoidable—for example, there are sufficient elements of parkinsonism in the definition to make explanations based on basal ganglia disease likely mechanistic candidates. Mathematical explorations, recently replicated independently (34,35) in the FI, also have mechanistic implications. In particular, the identification of limits to frailty (21), the identification of decreasing relative heterogeneity of fitness with age (36), and the elaboration of a stochastic model of transitions between degrees of frailty (29,30) each illustrate how studying the behavior of the system—compared with its component parts—can yield insights into mechanisms (27). In consequence, there is much yet to be learned from studying the deficits as a group, and not isolating them into clusters. In short, there remains ample reason to continue to endorse the view that it is too early to settle on one definition of frailty (37,38). Instead, researchers should make clear what they mean when they use the word, and should continue to explore how each definition contributes to our overall understanding of the variable vulnerability of people of the same chronological age.

CORRESPONDENCE

Address correspondence to Kenneth Rockwood, MD, Centre for Health Care of the Elderly, 1421-5955 Veterans' Memorial Lane, Halifax, Nova Scotia, Canada, B3H 2E1. E-mail: kenneth.rockwood@dal.ca

REFERENCES

1. Fisher AL. Just what defines frailty? *J Am Geriatr Soc.* 2005;53:2229–2230.
2. 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(3 Suppl):1–29.
3. Chin A, Paw MJ, Dekker JM, Feskens EJ, Schouten EG, Kromhout D. How to select a frail elderly population? A comparison of three working definitions. *J Clin Epidemiol.* 1999;52:1015–1021.
4. Canadian Study of Health and Aging Working Group. Disability and frailty among elderly Canadians: a comparison of six surveys. *Int Psychogeriatr.* 2001;13(Suppl 1):159–167.
5. Rockwood K, Song X, MacKnight C, et al. A global clinical measure of fitness and frailty in elderly people. *CMAJ.* 2005;173:489–495.
6. 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.
7. Woo J, Goggins W, Sham A, Ho SC. Public health significance of the frailty index. *Disabil Rehabil.* 2006;28:515–521.
8. Fried LP, Tangen CM, Walston J, et al. Cardiovascular Health Study Collaborative Research Group. Frailty in older adults: evidence for a phenotype. *J Gerontol Med Sci.* 2001;56A:M146–M156.
9. Mitnitski A, Mogilner A, Rockwood K. Accumulation of deficits as a proxy measure of aging. *Scientific WorldJournal.* 2001;1:323–336.
10. Canadian Study of Health and Aging: study methods and prevalence of dementia. CMAJ. 1994;150:899–913.
11. Teng EL, Chui HC. The Modified Mini-Mental State (3MS) examination. *J Clin Psychiatry.* 1987;48:314–318.
12. Graham JE, Rockwood K, Beattie BL, McDowell I, Eastwood R, Gauthier S. Standardization of the diagnosis of dementia in the Canadian Study of Health and Aging. *Neuroepidemiology.* 1996;15:246–256.
13. American Psychiatric Association. *Diagnostic and Statistical Manual, Revised 3rd Edition.* Washington, DC: APA Press; 1987.
14. Bandeen-Roche K, Xue Q-L, Ferrucci L, et al. Phenotype of frailty: characterization in the Women's Health and Aging Study. *J Gerontol A Biol Sci Med Sci.* 2006;61A:262–266.
15. Podsiadlo D, Richardson S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc.* 1991;39:142–148.
16. Mitnitski A, Song X, Skoog I, et al. Relative fitness and frailty of elderly men and women in developed countries, in relation to mortality. *J Am Geriatr Soc.* 2005;53:2184–2189.
17. George LK, Fillenbaum GG. OARS methodology. A decade of experience in geriatric assessment. *J Am Geriatr Soc.* 1985;33:607–615.
18. Duncan PW, Weiner DK, Chandler J, Studenski S. Functional reach: a new clinical measure of balance. *J Gerontol.* 1990;45:M192–M197.
19. Streiner D, Norman G. *Health Measurement Scales: A Guide to Their Development and Use.* 3rd Ed. Oxford, U.K.: Oxford University Press; 2003.
20. Rockwood K, Mogilner AJ, Mitnitski A. Changes with age in the distribution of a frailty index. *Mech Ageing Dev.* 2004;125:517–519.
21. Rockwood K, Mitnitski A. Limit to deficit accumulation in elderly people *Mech Ageing Dev.* 2006;127:494–496.
22. Rockwood K, Mitnitski A, Song X, Steen B, Skoog I. Long-term risks of death and institutionalization of elderly people in relation to deficit accumulation at age 70. *J Am Geriatr Soc.* 2006;54:975–979.
23. Ferrucci L, Guralnik JM, Studenski S, Fried LP, Cutler GB Jr, Walston JD; Interventions on Frailty Working Group. Designing randomized, controlled trials aimed at preventing or delaying functional decline and disability in frail, older persons: a consensus report. *J Am Geriatr Soc.* 2004;52:625–634.
24. Walston J, Hadley EC, Ferrucci L, et al. Research agenda for frailty in older adults: toward a better understanding of physiology and etiology: summary from the American Geriatrics Society/National Institute on Aging Research Conference on Frailty in Older Adults. *J Am Geriatr Soc.* 2006;54:991–1001.
25. Rockwood K, Mitnitski AB, MacKnight C. Some mathematical models of frailty and their clinical implications. *Rev Clin Gerontol.* 2003;12:109–117.
26. Mitnitski AB, Song X, Rockwood K. The estimation of relative fitness and frailty in community-dwelling older adults using self-report data. *J Gerontol Med Sci.* 2004;59A:M627–M632.
27. Rockwood K, Mitnitski AB. Review Article. Frailty in relation to the accumulation of deficits. *J Gerontol A Biol Sci Med Sci.* 2007;62A:722–727.
28. Gill TM, Gahbauer A, Allore HG, Han L, Transitions between frailty states among community-living older persons. *Arch Int Med.* 2006;166:418–423.
29. Mitnitski AB, Bao L, Rockwood K. Going from bad to worse: a stochastic model of transitions in deficit accumulation, in relation to mortality. *Mech Ageing Dev.* 2006;127:490–493.
30. Mitnitski A, Bao L, Skoog I, Rockwood K. A cross-national study of transitions in deficit counts in two birth cohorts: implications for modeling ageing. *Exp Gerontol.* 2007;42:241–246.
31. Jones DM, Song X, Rockwood K. Operationalizing a frailty index from standardized comprehensive geriatric assessment. *J Am Geriatr Soc.* 2004;52:1929–1933.
32. Jones D, Song X, Mitnitski A, Rockwood K. Evaluation of a frailty index based on a comprehensive geriatric assessment in a population based study of elderly Canadians. *Aging Clin Exp Res.* 2005;17:465–471.
33. Blaum CS, Xue QL, Michelon E, Semba RD, Fried LP. The association between obesity and the frailty syndrome in older women: the Women's Health and Aging Studies. *J Am Geriatr Soc.* 2005;53:927–934.
34. Goggins WB, Woo J, Sham A, Ho SC. Frailty index as a measure of biological age in a Chinese population. *J Gerontol A Biol Sci Med Sci.* 2005;60A:1046–1051.
35. Kulminski A, Yashin A, Ukraintseva S, et al. Accumulation of health disorders as a systemic measure of aging: findings from the NLTCS data. *Mech Ageing Dev.* 2006;127:840–848.
36. Mitnitski A, Rockwood K. Decrease in the relative heterogeneity of health with age: a cross-national comparison. *Mech Ageing Dev.* 2006;127:70–72.
37. Rockwood K. What would make a definition of frailty successful? *Age Ageing.* 2005;34:432–434.
38. Rockwood K. Frailty and its definition: a worthy challenge. *J Am Geriatr Soc.* 2005;53:1069–1070.