The Accuracy of Four Frequently Used Frailty Instruments for the Prediction of Adverse Health Outcomes Among Older Adults at Two Dutch Emergency Departments: Findings of the AmsterGEM Study

Carmen S. van Dam, MD*; Marijke C. Trappenburg, MD, PhD; Marieke M. ter Wee, PhD; Emiel O. Hoogendijk, PhD; Henrica C. de Vet, MD, PhD; Yvo M. Smulders, MD, PhD; Prabath W. Nanayakkara, MD, PhD; Majon Muller, MD, PhD; Mike J. Peters, MD, PhD

*Corresponding Author. E-mail: c.vandam@amsterdamumc.nl.

Study objective: Older adults presenting to the emergency department (ED) are at high risk of adverse health outcomes. This study aimed to evaluate the accuracy of 4 frequently used screening instruments for the prediction of adverse health outcomes among older adults in the ED.

Methods: This was a prospective cohort study in patients ≥70 years of age presenting to the ED in 2 hospitals in the Netherlands. Screening instruments included the acutely presenting older patient screening program (APOP) (providing 2 risk scores—functional decline [APOP1] and mortality [APOP2]), the International Resident Assessment Instrument Emergendy Department screener (InterRAI ED), the Identification of Seniors At Risk-Hospitalized Patients (ISAR-HP), and the safety management system (VMS). The primary outcome measure was a composite outcome encompassing functional decline, institutionalization, and mortality at 3 months after ED presentation. Other follow-up time points were 1 and 6 months. Analyses were performed to assess prognostic accuracy.

Results: In total, 889 patients were included. After 3 months, 267 (31%) patients experienced at least 1 adverse outcome. The positive likelihood ratio ranged from 1.67 (VMS) to 3.33 (APOP1), and the negative likelihood ratio ranged from 0.41 (ISAR-HP) to 0.88 (APOP2). Sensitivity ranged from 17% (APOP2) to 74% (ISAR-HP), and specificity ranged from 63% (ISAR-HP) to 94% (APOP2). The area under the curve ranged from 0.62 (APOP2) to 0.72 (APOP1 and ISAR-HP). Calibration was reasonable for APOP1 and VMS. The prognostic accuracy was comparable across all outcomes and at all follow-up time points.

Conclusion: The frailty screening instruments assessed in this study showed poor to moderate prognostic accuracy, which brings into question their usability in the prediction of adverse health outcomes among older adults who present to the ED. [Ann Emerg Med. 2021;■:1-11.]

Please see page XX for the Editor's Capsule Summary of this article.

0196-0644/$-see front matter

https://doi.org/10.1016/j.annemergmed.2021.04.027

INTRODUCTION

Optimizing emergency care for the aging population is a challenge. Older patients are overrepresented in the emergency department (ED), and the number of visits this population makes to the ED is rapidly increasing.[1,2] Older patients stay longer at the ED, are more likely to be hospitalized, and have a higher rate of adverse health outcomes after ED visits compared to younger patients.[1,3-5] Therefore, international guidelines recommend screening for frailty in the ED.[6-9] Frailty is often defined as an aging-related syndrome that encompasses a state of decline in multiple physiological systems accompanied by an increased vulnerability to stressors, leading to an increased risk of adverse health outcomes such as falls, functional decline, hospitalization, institutionalization, and mortality.[3-5,10]

Older patients at risk for adverse health outcomes are the target population for geriatric interventions. In geriatric

Editor's Capsule Summary

What is already known on this topic

Emergency departments have been urged to screen geriatric patients for frailty.

What question this study addressed

How accurate are 4 popular frailty screening instruments?

What this study adds to our knowledge

In this prospective head-to-head comparison in 889 geriatric patients, the 4 instruments displayed prognostic accuracy parameters well below thresholds reliable enough for clinical use.

How this is relevant to clinical practice

These 4 geriatric frailty screening instruments are too inaccurate for reliable clinical use.

emergency medicine, frailty is operationalized by identifying the patients at increased risk of adverse health outcomes.[3,11,12] Since it is not feasible to perform geriatric interventions in all older adults at the ED, frailty screening has been developed to identify the target population that would presumably benefit the most from a full comprehensive geriatric assessment or other geriatric interventions at the ED or during early hospital admission.[1,3,11,13-15] Recently, the assessment of overall effectiveness of frailty screening was listed as the highest priority on the research agenda of geriatric emergency medicine because the most appropriate instrument for the ED is still undetermined.[1,11,14,16]

Many screening instruments were developed for hospitalized patients but are used in the ED—for example, the Identification of Seniors At Risk-Hospitalized Patients (ISAR-HP)[17] and the safety management system (VMS).[18] Recently, researchers developed web-based applications with algorithms to improve the practical use of the screeners in the ED: the acutely presenting older patient[19,20] (APOP) screening program and the International Resident Assessment Instrument (InterRAI) ED screener.[21]

Only a few studies have directly compared different screening instruments in the ED with regard to prediction of adverse health outcomes.[14,22-24] Length of stay during hospital admission, readmission to the ED, and mortality are frequently studied outcome measures, but other clinically relevant outcome measures, such as functional decline and institutionalization, have been investigated less often.[12,25] The follow-up durations of earlier studies were limited and varied between studies, complicating a direct comparison between instruments.[11,25]

The predictive nature of screening instruments for frailty calls for evaluation of predictive accuracy and external validation of the prediction models. However, in contrast to information on diagnostic accuracy, results on predictive accuracy and external validation are scarce.[11] This study aimed to evaluate the prognostic accuracy of 4 frequently used screening instruments for frailty in older adults in 2 Dutch EDs with a follow-up duration of 6 months.

MATERIALS AND METHODS

Study Design

This prospective cohort study—the Amsterdam Geriatric Emergency Medicine study (AmsterGEM)—was conducted in the EDs of 2 Dutch hospitals: a tertiary academic hospital (Amsterdam UMC, location VUmc) and a general community hospital (Amstelland Hospital in Amstelveen). Data were collected from November 2017 until June 2018, daily during office hours and on limited numbers of evenings and weekend days. All participants or their legally authorized representatives provided written informed consent. The study was approved by the medical ethical board of Amsterdam UMC, location VUmc. Local approval was received from the Amstelland Hospital.

Participants

All patients aged ≥70 years attending the ED when the researchers were present were screened for eligibility. Exclusion criteria were high-urgency status (code red according to the Manchester Triage System[26]), a language barrier, and the inability to give informed consent (eg, due to altered mental status in the absence of a caregiver who could provide informed consent by proxy).

Data Collection

Data were collected by chart review and interviews with patients and their caregivers in the ED. All of the student researchers who collected the data received detailed information on the study design and the screening instruments and instructions on data collection. The student researchers were extensively trained by a team of geriatric consultants. Moreover, random quality checks were carried out during the data collection. Sociodemographic data and care-related data were obtained at baseline, including living situation, number of prescriptions, and Charlson Comorbidity Index score.[27] Physical status was assessed using the Katz Index of Independence in Activities of Daily Living,[28] with scores

ranging from 0 to 6 (with a score of 0 indicating independence). Cognitive status was assessed using the domain "delirium" from the VMS. In line with the original studies of the instruments, the physical and cognitive status 2 weeks prior to the ED visit was obtained to rule out interference of the acute illness.[17-19,21]

All patients were screened using all 4 screening instruments at baseline. Screening instruments were considered positive based on their original cut-off values. A detailed overview of each instrument can be found in Table E1 (available at http://www.annemergmed.com).

Acutely Presenting Older Patient Screener

The APOP screener is an application based on an algorithm and developed and validated in the Netherlands.[19] The APOP screener shows 2 percentages: one that indicates the risk of functional decline within the next 3 months and one that indicates the risk of mortality, in this paper referred to as APOP1 and APOP2, respectively. The cut-off value indicating the need for further geriatric assessment was recommended as having a risk of $\geq$50% for functional decline or a risk of $\geq$25% for mortality.[19] During this study, the APOP consortium released an optimized version.[20] At that time, we had already assessed patients using the first version; therefore, we decided to continue with this version. The optimized version consists of nearly the same variables and shows comparable predictive properties.[20]

InterRAI ED Screener

The InterRAI ED screener is also an application based on an algorithm, developed in a multicenter multinational cohort study.[21] The InterRAI ED screener stratifies patients into 3 groups: those at low, intermediate, and high risk of adverse health outcomes. Further assessment and interventions are recommended in patients with high risk (score $\geq$5).

Identification of Seniors at Risk-Hospitalized Patients (ISAR-HP)

The ISAR-HP was developed in a Dutch cohort study of hospitalized patients.[17] A score of $\geq$2 is the cut off value for a positive score, indicating an increased risk of adverse health outcomes (frail).

Safety Management System (VMS)

The VMS was also developed in a Dutch cohort study of hospitalized patients.[18] For patients aged 70 to 80 years, a score of $\geq$3 indicates frailty; in patients aged $\geq$80 years, a score of 1 indicates frailty.

Outcome Measurements and Follow-up

Outcome measures were functional decline, institutionalization, and mortality. The primary outcome measure was a composite outcome including all three outcome measures at 3 months after ED presentation. The composite outcome was dichotomous and was considered positive if any adverse health outcome had occurred by the follow-up time point. Follow-up data were collected by student researchers, who were not blinded to baseline data or details of the screening instruments. The student researchers were not involved in the care of these patients. Data on institutionalization and mortality were extracted from the electronic health record and cross-referenced with the general practitioner or caregiver. Follow-up information on functional status was obtained by telephone after 1, 3, and 6 months using a standardized charting form. If the patient was unreachable after 5 attempts by telephone, follow-up data was obtained from the general practitioner. If applicable, the patient was called again at the next follow-up time point.

Functional decline was defined as an increase of 1 or more points in Katz Index of Independence in Activities of Daily Living score compared with baseline. The patient was considered institutionalized if he or she lived at home during baseline but had to stay elsewhere during follow-up (eg, in a nursing home, rehabilitation center, or contemporary health institute). Mortality data was extracted from the electronic health record and cross-referenced with the general practitioner or caregiver. Similarly, the follow-up outcomes ascertained by telephone were cross-referenced with the electronic health record.

Statistical Analysis

Baseline characteristics and descriptive statistics are described as means (standard deviations) or, in case of skewed distribution, medians (interquartile ranges). Prognostic accuracy was evaluated using sensitivity, specificity, and likelihood ratios, calculated using classification tables. Discrimination was quantified by the area under the receiver operating curve (AUC-ROC) and area under the precision-recall curve (AU-PRC) because of an imbalanced distribution of the outcome.[29,30] Calibration was assessed by calibration plots with loess smoothing due to the binary outcome.[31]

There are three subgroup analyses performed: one on hospitalized patients, one on patients that were not institutionalized at baseline (e.g. not living in an institution when they presented at the ED) and one on prolonged length of stay as outcome measure.

All statistical analyses were performed with SPSS version 26 (IBM Corp, Armonk, NY, USA) and R version 3.6 (R Foundation for Statistical Computing, Vienna, Austria).[32] A P value $<.05$ was considered statistically significant.

RESULTS

Patients

In total, 1,601 patients were screened for eligibility, and 712 were excluded (Figure 1). Most patients (n=404) were excluded because no informed consent was given, often due to the absence of a caregiver who could provide consent when the patient was too ill or confused. The exact number of patients with informed consent by proxy was not noted. Furthermore, 134 patients were unapproachable, according to the medical staff at the ED (eg, patients who had just received bad news or patients in extreme pain). An additional 96 were excluded due to their limited length of stay at the ED (these were patients who were admitted to the hospital or transferred to different hospitals before the researcher could approach them). No reason of exclusion was reported for 11 patients. A total of 889 patients were

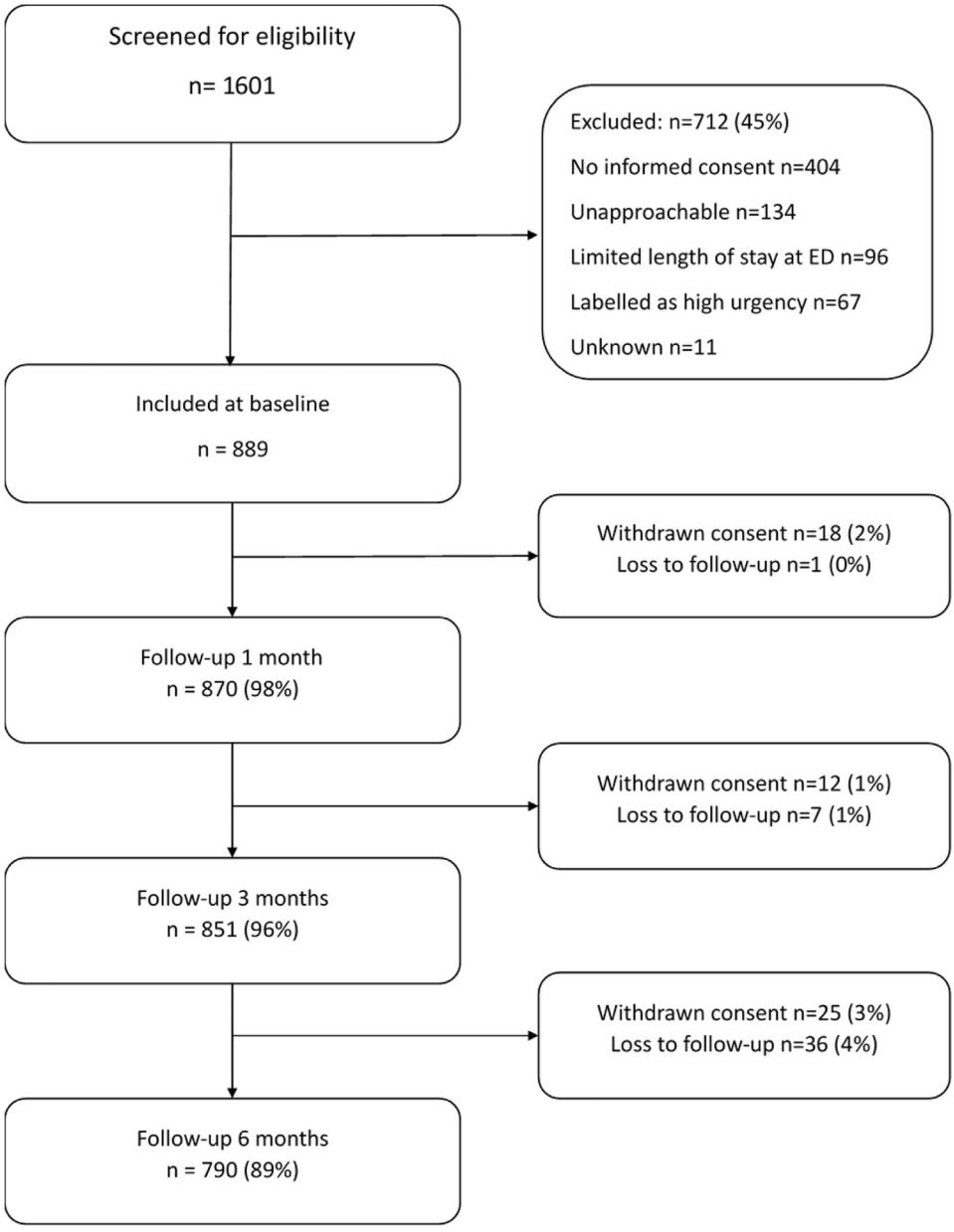

Figure 1. Study population flowchart.

included. Follow-up information was available for 98% of the patients at 1 month, 96% at 3 months, and 89% at 6 months.

Baseline Characteristics

The median age of patients was 78 years (Table 1). Most patients were activities of daily living-independent and lived alone without any home care. The APOP screener indicated that 20% of the patients were at high risk of functional decline and almost 10% were at high risk of mortality. The InterRAI ED screener identified one fourth of the patients as frail with the need for further assessment, and the ISAR-HP and VMS identified half of the patients as frail. Table E2 (available online at http://www.annemergmed.com) shows the distribution of scores for each instrument.

Adverse Health Outcomes

Of all patients, 31% experienced an adverse health outcome within 1 month after ED presentation; this increased slightly to 35% at 6 months (Table 2). Mortality increased at each follow-up time point, from 4% at 1 month to 9% at 3 months and 14% at 6 months. The prevalence of functional decline was highest at 1 month follow-up, and it decreased at the 3- and 6-month time points. Institutionalization remained more or less stable. Figure 2 illustrates the composition of the composite outcomes by showing the overlap of each outcome component.

In the next sections, we describe only the results for the outcomes at 3 months. That is, only the results for the primary outcome measure will be summarized in the text. The results of individual outcomes are presented in the tables. Results at the 1- and 6-month follow-up time points are presented in Tables E3 and E4.

Table 1. Baseline characteristics.

Characteristics	Total N=889
Age in years, median [IQR]	78 [73-83]
Male	467 (48)
Education after age of 14 years	679 (77)
Living situation	
Home without home care	463 (52)
Home with home care*	367 (41)
Other†	59 (7)
Katz Index of Independence in Activities of Daily Living score, median [IQR]	0 [0-1]
Self-reported memory problems	189 (21)
Charlson Comorbidity Index, median [IQR]	5 [4-6]
Number of prescriptions, median [IQR]	5 [3-7]
Included outside office hours	97 (13)
Screening instrument scored positive	
APOP1—risk of functional decline	181 (20)
APOP2—risk of mortality	80 (9)
InterRAI ED screener	205 (23)
ISAR-HP	441 (50)
VMS	387 (44)

IQR, interquartile range.
Numbers are displayed as n (%) unless otherwise specified.
*Including household help.
†Including nursing home, rehabilitation institute, post acute care unit.

Table 2. Adverse health outcomes at 1-, 3-, and 6-month follow-up.

Outcome measure	1 Month n=870	3 Months n=851	6 Months n=790
Composite outcome	269 (31)	267 (31)	280 (35)
Mortality	38 (4)	76 (9)	107 (14)
Functional decline	189 (22)	123 (15)	105 (13)
Institutionalization	98 (11)	112 (13)	107 (14)

Numbers are displayed as n (%). Numbers are cumulative except for functional decline. Functional decline is defined as a decrease of ≥1 point in Katz Index of Independence in Activities of Daily Living score compared to baseline and displayed as a percentage of total population at follow-up.

Prognostic Accuracy

As shown in Table 3, the overall sensitivity of all 4 instruments was low; the maximum sensitivity for the composite outcome was 74% (95% confidence interval [CI] 68% to 79%), for the ISAR-HP. The maximum specificity was higher—94% (95% CI 92% to 96%), for APOP2. Figure 3 shows the proportion of predicted adverse health outcomes versus observed adverse health outcomes. The false negative percentages ranged from 9% (ISAR-HP) to 27% (APOP2); the false positive percentages ranged from 4% (APOP2) to 25% (ISAR-HP). This means that with the lowest false negative percentage (ISAR-HP), 27% of all patients with adverse health outcomes were not frail according to the screening instrument, while the lowest false positive percentage (APOP2) still marked 38% of the patients as frail without the occurrence of adverse health outcomes. The correctly predicted percentages (both frail and not frail) ranged from 63% (VMS) to 72% (APOP1 and InterRAI ED screener). APOP1 had the highest positive likelihood ratio (LR) (3.33, 95% CI 2.57 to 4.44). ISAR-HP had the best negative LR (0.41, 95% CI 0.34 to 0.52) (Table 3).

Discrimination was assessed using the AUC. The AUC-ROC of all screening instruments was poor to moderate (Table 3), ranging from 0.62 (95% CI 0.58 to 0.66) (APOP2) to 0.72 (95% CI 0.69 to 0.76) (APOP1 and ISAR-HP). We also constructed precision-recall curves

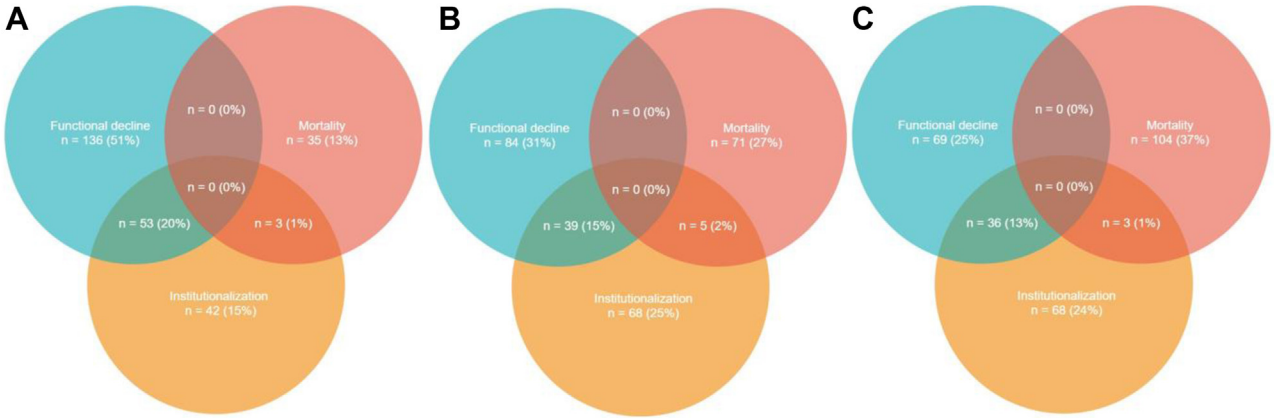

Figure 2. Distribution of adverse outcomes. *A*, adverse outcomes at 1-month follow-up. *B*, adverse outcomes at 3-month follow-up. *C*, adverse outcomes at 6-month follow-up. Functional decline was defined as a decrease of $\geq$1 point in Katz Index of Independence in Activities of Daily Living score compared to baseline and displayed as percentage of total population at follow-up. The composite outcome was considered positive if any adverse health outcome had occurred by the follow-up time point.

Table 3. Prognostic accuracy for adverse health outcomes at 3-month follow-up.

N = 851	Sensitivity	Specificity	LR+	LR-	AU-ROC	AU-PRC
Composite outcome						
APOP1	40 (34-46)	88 (85-91)	3.33 (2.57-4.44)	0.68 (0.62-0.76)	0.72 (0.69-0.76)	0.55
APOP2	17 (13-22)	94 (92-96)	2.83 (1.92-4.56)	0.88 (0.83-0.93)	0.62 (0.58-0.66)	0.45
InterRAI ED Screener	44 (38-50)	86 (83-89)	3.14 (2.45-4.02)	0.65 (0.58-0.73)	0.71 (0.67-0.75)	0.50
ISAR-HP	74 (68-79)	63 (59-67)	2.00 (1.75-2.28)	0.41 (0.34-0.52)	0.72 (0.68-0.76)	0.51
VMS	60 (54-66)	64 (60-68)	1.67 (1.44-1.94)	0.63 (0.53-0.73)	0.70 (0.66-0.73)	0.53
Mortality						
APOP1	37 (26-49)	81 (78-84)	1.95 (1.42-2.74)	0.78 (0.65-0.93)	0.68 (0.62-0.74)	0.27
APOP2	22 (14-33)	92 (90-94)	2.90 (1.78-4.71)	0.85 (0.74-0.95)	0.65 (0.58-0.71)	0.23
InterRAI ED Screener	50 (38-32)	79 (76-82)	2.38 (1.87-3.17)	0.63 (0.50-0.79)	0.71 (0.65-0.76)	0.25
ISAR-HP	76 (65-85)	54 (50-57)	1.65 (1.43-1.92)	0.44 (0.29-0.66)	0.68 (0.62-0.74)	0.25
VMS	59 (47-70)	58 (55-62)	1.40 (1.16-1.75)	0.71 (0.53-0.92)	0.66 (0.60-0.73)	0.24
Functional decline						
APOP1	37 (28-46)	85 (82-88)	2.47 (1.84-3.37)	0.74 (0.65-0.85)	0.67 (0.62-0.72)	0.26
APOP2	13 (8-20)	94 (91-95)	2.17 (1.18-3.56)	0.93 (0.86-1.00)	0.59 (0.53-0.64)	0.23
InterRAI ED Screener	33 (25-42)	82 (79-85)	1.83 (1.39-2.55)	0.82 (0.71-0.92)	0.66 (0.63-0.71)	0.18
ISAR-HP	70 (61-78)	60 (56-64)	1.75 (1.50-2.03)	0.50 (0.38-0.66)	0.68 (0.63-0.73)	0.25
VMS	59 (50-68)	62 (58-66)	1.57 (1.32-1.88)	0.65 (0.52-0.82)	0.64 (0.59-0.70)	0.24
Institutionalization						
APOP1	50 (40-60)	86 (83-89)	3.57 (2.78-4.72)	0.58 (0.48-0.70)	0.72 (0.70-0.77)	0.36
APOP2	18 (11-26)	94 (92-95)	3.00 (1.73-4.64)	0.87 (0.80-0.96)	0.57 (0.51-0.63)	0.21
InterRAI ED Screener	50 (40-60)	84 (91-97)	3.13 (2.41-4.01)	0.60 (0.49-0.72)	0.66 (0.61-0.72)	0.12
ISAR-HP	76 (67-83)	58 (54-62)	1.81 (1.58-2.08)	0.41 (0.30-0.58)	0.69 (0.64-0.74)	0.25
VMS	67 (57-76)	62 (57-66)	1.76 (1.52-2.10)	0.53 (0.40-0.69)	0.68 (0.63-0.74)	0.25

LR+, positive likelihood ratio; LR-, negative likelihood ratio.
Numbers are presented as % (95% CI). Baseline precision-recall curve: mortality 0.17, functional decline 0.17, institutionalization 0.15, composite outcome 0.33.

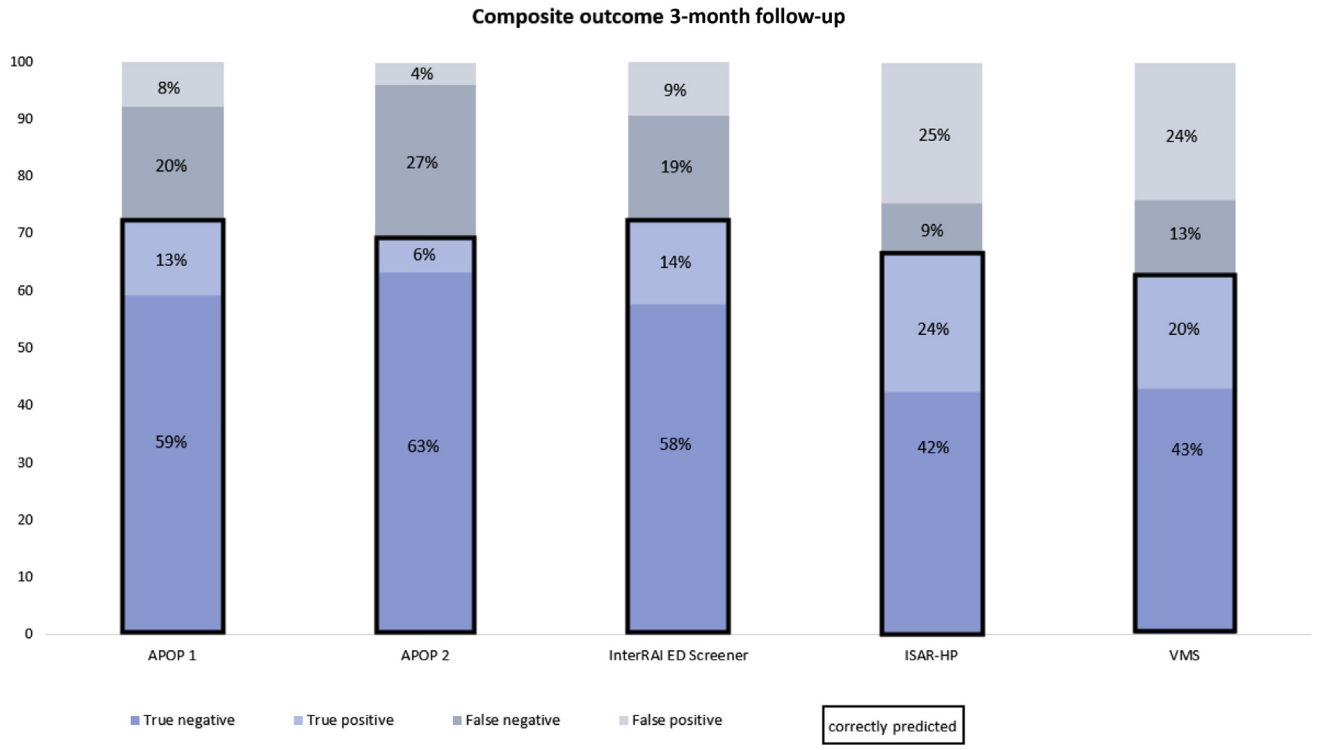

Figure 3. The proportion of predicted adverse health outcomes versus observed outcomes. True negative reflects the percentage of total population with a negative screening result and no adverse health outcomes. True positive reflects the percentage of total population with a positive screening result and one or more adverse health outcomes. False negative reflects the percentage of total population with a negative screening result but one or more adverse health outcomes. False positive reflects the percentage of total population with a positive screening result and one or more adverse health outcomes. Correctly predicted reflects the percentage of total population with screening result in concordance with experienced health outcomes.

(Table 3, Figure E1, [available online at http://www.annemergmed.com]) because of imbalanced classification—the majority of the patients experienced no adverse health outcomes. The AU-PRC ranged from 0.45 to 0.55 (with a baseline of 0.33 indicating random discrimination for the composite outcome and 1 reflecting a perfect discrimination).

Calibration plots showed a poor calibration for APOP2, the InterRAI ED screener, and ISAR-HP (Figure 4). Calibration was reasonable for APOP1 and VMS, but with a wide confidence interval. This result is also illustrated in Figure E2 (available online at http://www.annemergmed.com), showing an increasing prevalence of adverse health outcomes with higher scores on the screening instrument for APOP1 and VMS—in contrast to APOP2, the InterRAI ED screener, and ISAR-HP, which showed a discordant prevalence of adverse health outcomes.

Subgroup Analyses

Subgroup analyses were performed: without nursing home patients, only hospitalized patients, and with a prolonged length of stay as outcome measure (Tables E5 to E8). These different groups also showed poor to moderate results with regard to prognostic accuracy.

LIMITATIONS

Our study has limitations. First, we selected only 4 instruments, including ISAR-HP and VMS, which were developed for hospitalized patients. Yet, in the Netherlands, it is compulsory to use these instruments at hospital admission, which makes it efficient to use the same instruments in the ED. Second, we only used the cut-off values of the original studies, to reflect their use in daily practice. However, we believe that different cut-off values would not change our conclusions; the ROC, PRC, and calibration plots were not based on cut-off values but on the gradual scores of the instruments, and these tests showed comparable poor to moderate results. Third, recall bias on functional decline might have occurred, since one fifth of all patients reported memory problems. However, we believe that recall bias was limited due to the clear definition of the Katz Index of Independence in Activities of Daily Living

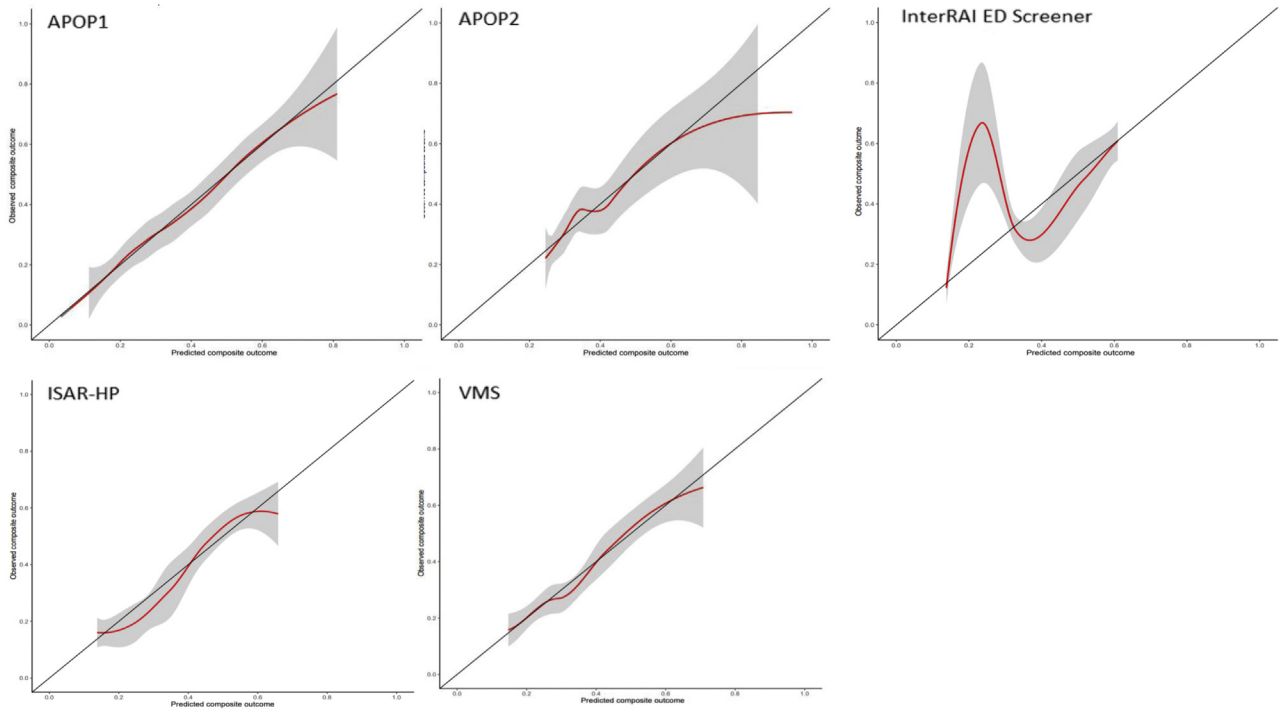

Figure 4. Calibration plot for the composite outcome at 3 months. This figure shows the prediction of the screening instrument on the x-axis and the composite outcome of the y-axis. Loess smoothing is used to estimate the observed probabilities of the outcome in relation to the predicted probabilities reflecting in the red line with the grey area as the confidence interval.33 The black line is the 45° line of perfect prediction. APOP = Acutely Presenting Older Patient, APOP1 = APOP screener positive for functional decline (risk within 3 months $\geq$ 50%) APOP2 = APOP screener positive for mortality (risk within 3 months $\geq$ 25%), composite outcome: mortality, functional decline, and institutionalization, ISAR-HP = Identification of Seniors at Risk – Hospitalized Patients, VMS = Safety Management System

questions which were used to assess functional status. Additionally, telephonic follow-up is commonly used with this population, and the prevalence of adverse outcomes in this study is consistent with the literature. Finally, our primary outcome was a composite outcome, so that our results could be compared with those of the original studies and other validation studies.[18-20,23,38,43] A pitfall of using composite outcomes is that their results might be driven by components of lesser importance. As our Venn diagram shows, all outcomes contributed more or less equally to the composite outcome at the 3- and 6-month follow-up time points. Because mortality, functional decline, and institutionalization might not be equally important to patients and clinicians, we also reported on the outcomes individually.

Further research is needed on the effectiveness and feasibility of a total screening program (including interventions as a comprehensive geriatric assessment and advance care planning) by frailty screening selected group.[11,16] We suggest further research on the prognostic accuracy and implementation of the clinical opinion (eg, the Clinical Frailty Scale[44] as a standardized clinical opinion) to evaluate its applicability. Also, a 2-step approach for use of screening instruments can be investigated (eg, one with a high negative predictive value and, subsequently, one with a high positive predictive value). This might result in a better identification of older patients at risk.

DISCUSSION

This study confirmed that a substantial proportion of older adults in the ED are at risk for adverse health outcomes after their ED visits. This observation supports the concept of frailty screening in the ED to identify those at risk as a target population for geriatric interventions. However, our observations illustrate that frequently used screening instruments have poor to moderate prognostic accuracy for adverse health outcomes up to 6 months after ED presentation.

None of the investigated instruments stands out or can be recommended. The results were comparable across all outcomes and at all follow-up time points. The correctly predicted percentage (both frail and not

frail) seemed reasonable but depended strongly on the prevalence of the outcome. The AU-ROC and AU-PRC—which account for prevalence—were poor to moderate. This was also illustrated by other clinically useful parameters: the lowest false negative percentage reflected that 27% of the patients with adverse health outcomes were not frail according to the screening instrument, while the lowest false positive percentage still marked 38% of the patients as frail without the occurrence of adverse health outcomes. Calibration seemed moderate for APOP1 and VMS but showed wide confidence intervals.

The prevalence of frailty and adverse health outcomes observed in our cohort is consistent with rates seen in previous literature.[1,17,18-23,33-35] Compared with the original and other validation studies of the screening instruments, our study showed worse prognostic accuracies.[17-21,33,36-38] Possible explanations for the worse results are differences in outcome measurements, population characteristics, and health care systems.[11,33] For example, the ISAR-HP has not been evaluated for institutionalization, and the InterRAI ED screener has not been evaluated for mortality.[17,21,33,38] Differences in population characteristics were age, living situation, and admission status. The original study of the ISAR-HP included patients aged 65 years and older, and that of the InterRAI ED screener included patients aged 75 years and older.[17,21] In our study, we included all patients presenting to the ED, including those sent home after their ED visits, while other studies only included admitted patients.[17,18,36] However, additional analyses in admitted patients showed similar results. Nursing home patients were also included in our study, but these patients were already institutionalized at baseline and could not experience the adverse health outcome of institutionalization during the follow-up period. Again, a subgroup analysis of predictive accuracy with the exclusion of nursing home patients did not show different results.

In line with our observations, previous studies investigating the prognostic accuracy of other frailty screening instruments (eg, Identification of Seniors At Risk or Triage Risk Screening Tool) and other adverse health outcomes showed poor to moderate accuracy.[11,14,22-24,33,39] Calibration plots and precision-recall curves have not been frequently reported in studies on these instruments.

Frailty might to be too extensive to determine with standardized and quick instruments.[11,33] The calibration results do indicate that a higher score has a higher probability of an adverse outcome. Additionally, the individual components of the screening instruments (eg, age, functional status, living situation, falls) are known to be associated with adverse health outcomes. However, despite this association, accurate prediction for individual patients seems almost impossible. Therefore, the goal of frailty screening in the ED can shift, from predicting adverse health outcomes to creating awareness about this vulnerable population among health care professionals. Implementation of routine frailty screening can force physicians to evaluate the functional status and treatment goals of older adults already in an early phase at the ED. Frailty screening can warrant awareness about the care needs of this vulnerable population and trigger beneficial interventions, such as a comprehensive geriatric assessment.[5,12,15,40-42] The choice of a certain instrument should be based on its practical abilities that fit the local aim and that support purposeful care for older adults in the ED. Frailty screening results of specific instruments should be thoughtfully interpreted in daily practice and should not lead clinical decisions because of their poor to moderate prognostic accuracy on an individual patient level.[42]

Our study is one of the few large cohort studies to directly compare multiple screening instruments with a follow-up duration of up to 6 months and to report extensively on overall predictive performance, discrimination, and calibration (instead of only diagnostic characteristics). We also included a wide range of patients (from nursing home patients to community-dwelling patients and patients who presented for all specialties at the ED), making this a generalizable population.

In conclusion, this study confirmed the increased risk of adverse health outcomes in older adults presenting to the ED, which seems to justify the use of frailty screening instruments to identify a target population for geriatric interventions. However, the frailty screening instruments assessed in this study showed poor to moderate accuracy, which brings into question their usability in the prediction of adverse health outcomes among older adults presenting to the ED.

The authors would like to thank all participants and the student researchers who contributed to the data collection. We would also like to thank the EDs of Amsterdam UMC, location VUmc and the Amstelland Hospital for their collaboration. We thank HJ de Vries for his contribution to the statistical analyses.

Supervising editor: Steven M. Green, MD. Specific detailed information about possible conflict of interest for individual editors is available at https://www.annemergmed.com/editors.

Author affiliations: From the Department of Internal Medicine and Geriatrics (van Dam, Trappenburg, Muller, Peters), the Department of Epidemiology and Data Science (ter Wee, Hoogendijk, de Vet),

the Department of Internal Medicine and Vascular Medicine (Smulders, Peters), and the Department of Internal Medicine and Acute Medicine (Nanayakkara), Amsterdam UMC, location VUmc, Amsterdam, the Netherlands.

Author contributions: CSD conceived and designed the analysis, collected the data, contributed to the data and analysis tools, performed the analysis, and wrote the paper. MCT conceived and designed the analysis, collected the data, contributed to the data and analysis tools, and wrote the paper. MMW conceived and designed the analysis, contributed to data and analysis tools, and wrote the paper. EOH conceived and designed the analysis, contributed to data and analysis tools, and wrote the paper. YMS conceived and designed the analysis, contributed to data and analysis tools, and wrote the paper. PWN conceived and designed the analysis, contributed to data and analysis tools, and wrote the paper. MM conceived and designed the analysis, contributed to data and analysis tools, and wrote the paper. MJP conceived and designed the analysis, collected the data, contributed to the data and analysis tools, and wrote the paper.

All authors attest to meeting the four ICMJE.org authorship criteria: (1) Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND (2) Drafting the work or revising it critically for important intellectual content; AND (3) Final approval of the version to be published; AND (4) Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding and support: By *Annals* policy, all authors are required to disclose any and all commercial, financial, and other relationships in any way related to the subject of this article as per ICMJE conflict of interest guidelines (see www.icmje.org). The authors have stated that no such relationships exist.

Publication dates: Received for publication April 15, 2021. Revisions received April 15, 2021. Accepted for publication April 19, 2021.

REFERENCES

1. Aminzadeh F, Dalziel WB. Older adults in the emergency department: a systematic review of patterns of use, adverse outcomes, and effectiveness of interventions. *Ann Emerg Med*. 2002;39:238-247.
2. Samaras N, Chevalley T, Samaras D, et al. Older patients in the emergency department: a review. *Ann Emerg Med*. 2010;56:261-269.
3. Hoogendijk EO, Afilalo J, Ensrud KE, et al. Frailty: implications for clinical practice and public health. *Lancet*. 2019;394:1365-1375.
4. Bonin-Guillaume S, Durand AC, Yahi F, et al. Predictive factors for early unplanned rehospitalization of older adults after an ED visit: role of the caregiver burden. *Aging Clin Exp Res*. 2015;27:883-891.
5. Goldstein J, Hubbard RE, Moorhouse P, et al. The validation of a care partner-derived frailty index based upon comprehensive geriatric assessment (CP-FI-CGA) in emergency medical services and geriatric ambulatory care. *Age Ageing*. 2015;44:327-330.
6. Dent E, Martin FC, Bergman H, et al. Management of frailty: opportunities, challenges, and future directions. *Lancet*. 2019;394:1376-1386.
7. Multimorbidity: clinical assessment and management. NICE guideline [NG56]. National Institute for Health and Care Excellence. https://www.nice.org.uk/guidance/ng56. Accessed April 15, 2021.
8. Goldstein J, McVey J, Ackroyd-Stolarz S. The role of emergency medical services in geriatrics: bridging the gap between primary and acute care. *CJEM*. 2016;18:54-61.
9. American College of Emergency Physicians, American Geriatrics Society, Emergency Nurses Association. et al. Geriatric emergency department guidelines. *Ann Emerg Med*. 2014;63:e7-e25.
10. Soong JTY, Poots AJ, Bell D. Finding consensus on frailty assessment in acute care through Delphi method. *BMJ Open*. 2016;6:e012904.
11. Carpenter CR, Mooijaart SP. Geriatric screeners 2.0: time for a paradigm shift in emergency department vulnerability research. *J Am Geriatr Soc*. 2020;68:1402-1405.
12. Theou O, Squires E, Mallery K, et al. What do we know about frailty in the acute care setting? A scoping review. *BMC Geriatr*. 2018;18:139.
13. VMS. Praktijkgids Kwetsbare Ouderen. VMS Veiligheidsprogramma. 2009. https://www.vmszorg.nl/wp-content/uploads/2017/11/web_2009.0104_praktijkgids_kwetsbare_ouderen.pdf. Accessed April 15, 2021.
14. Carpenter CR, Shelton E, Fowler S, et al. Risk factors and screening instruments to predict adverse outcomes for undifferentiated older emergency department patients: a systematic review and meta-analysis. *Acad Emerg Med*. 2015;22:1-21.
15. Ellis G, Gardner M, Tsiachristas A, et al. Comprehensive geriatric assessment for older adults admitted to hospital. *Cochrane Database Syst Rev*. 2017;9:CD006211.
16. Mooijaart SP, Nickel CH, Conroy SP, et al. A European Research Agenda for Geriatric Emergency Medicine: a modified Delphi study. Eur Geriatr Med. *Published online November 21, 2020*. https://doi.org/10.1007/s41999-020-00426-8.
17. Hoogerduijn JG, Buurman BM, Korevaar JC, et al. The prediction of functional decline in older hospitalised patients. *Age Ageing*. 2012;41:381-387.
18. Heim N, van Fenema EM, Weverling-Rijnsburger AWE, et al. Optimal screening for increased risk for adverse outcomes in hospitalised older adults. *Age Ageing*. 2015;44:239-244.
19. de Gelder J, Lucke JA, de Groot B, et al. Predicting adverse health outcomes in older emergency department patients: the APOP study. *Neth J Med*. 2016;74:342-352.
20. de Gelder J, Lucke JA, Blomaard LC, et al. Optimization of the APOP screener to predict functional decline or mortality in older emergency department patients: cross-validation in four prospective cohorts. *Exp Gerontol*. 2018;110:253-259.
21. Costa AP, Hirdes JP, Heckman GA, et al. Geriatric syndromes predict postdischarge outcomes among older emergency department patients: findings from the interRAI Multinational Emergency Department Study. *Acad Emerg Med*. 2014;21:422-433.
22. Wou F, Gladman JRF, Bradshaw L, et al. The predictive properties of frailty-rating scales in the acute medical unit. *Age Ageing*. 2013;42:776-781.
23. Lewis ET, Dent E, Alkhouri H, et al. Which frailty scale for patients admitted via Emergency Department? A cohort study. *Arch Gerontol Geriatr*. 2019;80:104-114.
24. Pijpers E, Ferreira I, Stehouwer CDA, et al. The frailty dilemma. Review of the predictive accuracy of major frailty scores. *Eur J Intern Med*. 2012;23:118-123.
25. Jørgensen R, Brabrand M. Screening of the frail patient in the emergency department: a systematic review. *Eur J Intern Med*. 2017;45:71-73.
26. Mackway-Jones K, Marsden J, Windle J. *Emergency Triage: Manchester Triage Group*. 3rd Edition. London: BMJ Books; 2014.
27. Charlson ME, Pompei P, Ales KL, et al. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40:373-383.
28. Katz S, Ford AB, Moskowitz RW, et al. Studies of illness in the aged. The Index of ADL: a standardized measure of biological and psychosocial function. *JAMA*. 1963;185:914-919.

29. Hosmer DW, Hosmer T, Le Cessie S, et al. A comparison of goodness-of-fit tests for the logistic regression model. *Stat Med*. 1997;16:965-980.
30. Bewick V, Cheek L, Ball J. Statistics review 14: logistic regression. *Crit Care*. 2005;9:112-118.
31. Austin PC, Steyerberg EW. Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. *Stat Med*. 2014;33:517-535.
32. DescTools. *Tools for descriptive statistics. R package version 0.99.37*. Signorell A, et al. 2020.
33. Heeren P, Devriendt E, Wellens NIH, et al. Old and new geriatric screening tools in a Belgian emergency department: a diagnostic accuracy study. *J Am Geriatr Soc*. 2020;68:1454-1461.
34. Taylor A, Broadbent M, Wallis M, et al. The predictive validity of the interRAI ED screener for predicting re-presentation within 28 days for older adults at a regional hospital emergency department. *Australas Emerg Care*. 2019;22:149-155.
35. Calf AH, Lubbers S, Van Den Berg AA, et al. Clinical impression for identification of vulnerable older patients in the emergency department. *Eur J Emerg Med*. 2020;27:137-141.
36. Oud FMM, de Rooij SEJA, Schuurman T, et al. Predictive value of the VMS theme "Frail elderly": delirium, falling and mortality in elderly hospital patients. *Ned Tijdschr Geneeskd*. 2015;159:A8491.
37. Schuijt HJ, Oud FMM, Bruns EJR, et al. Does the Dutch Safety Management Program predict adverse outcomes for older patients in the emergency department? *Neth J Med*. 2020;78:244-250.
38. de Gelder J, Haenen E, Lucke JA, et al. Optimising the ISAR-HP to screen efficiently for functional decline in older patients. *Neth J Med*. 2017;75:379-385.
39. Michalski-Monnerat C, Carron PN, Nguyen S, et al. Assessing older patients' vulnerability in the emergency department: a study of InterRAI ED screener accuracy. *J Am Geriatr Soc*. 2020;68:2914-2920.
40. Ellis G, Marshall T, Ritchie C. Comprehensive geriatric assessment in the emergency department. *Clin Interv Aging*. 2014;9:2033-2043.
41. Elliott A, Hull L, Conroy SP. Frailty identification in the emergency department-a systematic review focussing on feasibility. *Age Ageing*. 2017;46:509-513.
42. Hollinghurst J, Housley G, Watkins A, et al. A comparison of two national frailty scoring systems. *Age Ageing*. Published online November 30, 2020. https://doi.org/10.1093/ageing/afaa252
43. Meldon SW, Mion LC, Palmer RM, et al. A brief risk-stratification tool to predict repeat emergency department visits and hospitalizations in older patients discharged from the emergency department. *Acad Emerg Med*. 2003;10:224-232.
44. Rockwood K, Song X, MacKnight C, et al. A global clinical measure of fitness and frailty in elderly people. *CMAJ*. 2005;173:489-495.