Screening for Geriatric Problems in the Emergency Department: Reliability and Validity

JANE MCCUSKER, MD, DRPH, FRANÇOIS BELLAVANCE, PHD, SYLVIE CARDIN, PHD, SYLVAIN TRÉPANIER, MSC, AND THE IDENTIFICATION OF SENIORS AT RISK (ISAR) STEERING COMMITTEE*

Abstract. **Objective:** To determine the test–retest reliability and concurrent criterion validity of a self-report ED screening questionnaire for adverse outcomes in elders. **Methods:** A cohort of 1,885 patients aged ≥65 years were recruited from the EDs of 4 Montreal hospitals. Patients were excluded if they could not be interviewed because of their clinical status or cognitive impairment and no informant was available. The screening questionnaire, administered in the ED, contained 27 items on social, physical, and mental risk factors, medical history, and use of hospital services, medications, and alcohol. A random sample of 404 patients were invited to participate in a clinical assessment 1–3 weeks after the ED visit, that included re-administration of the screening questionnaire, and standardized instruments to assess disability, social resources, depression, alcohol use and abuse, and current medications. **Results:** Study data were collected from 221 patients (54.7%), of whom 193 were included in the test–retest reliability analyses and 213 in the analyses of concurrent validity. The concordance correlation coefficient for test–retest reliability of the risk factor score was 0.78 (95% confidence interval: 0.71, 0.83; $n = 193$). Several screening questions showed moderately good agreement with the appropriate criterion standard, particularly those on visual and hearing impairment, depression, and use of medications. The best subset of 9 screening questions explained approximately half of the variance in the total disability score. **Conclusions:** The screening questionnaire score has good test–retest reliability, but individual screening questions have, at best, modest concurrent validity. The final set of screening questions should be selected based on their predictive validity. **Key words:** aged; screening; functional status; reliability; validity; geriatrics. ACADEMIC EMERGENCY MEDICINE 1998; 5:883–893

GERIATRIC evaluation and management (GEM) programs have been reported from various populations, including inpatients, outpatients, and those in the community.[1] Recent reports suggest that the ED may be a useful site for the identification of older persons with "sentinel health events" that signal risk of functional deterioration.[2,3] Follow-up studies of elders after release from the ED indicate considerable loss of independence,[4,5] substantial rates of hospital admission or death,[5] and underutilization of appropriate community services.[4] A screening and intervention program, to include the identification of those at risk, followed by referrals to appropriate community services, has the potential to prevent some of these adverse outcomes.

Emergency department-based screening programs for elders reported previously reveal high rates of various problems. Gerson and colleagues[6] assessed the feasibility of a brief comprehensive case-finding program in a multicenter study of ED patients aged ≥60 years released to their homes. The participation rate was 75% and multiple problems were detected. In a case-finding and liaison service delivered in the ED by a geriatric nurse clinician, 82% of patients aged ≥65 years were found to have at least 1 "geriatric" problem and 77% reported at least 1 unmet dental or social support need.[7] Although the latter intervention was not effective in reducing mortality or nursing home admissions, members of the intervention group

From the Department of Clinical Epidemiology and Community Studies, St. Mary's Hospital (JM, FB, ST); the Department of Epidemiology and Biostatistics, McGill University (JM, FB); and the Department of Emergency Medicine, Sir Mortimer B. Davis–Jewish General Hospital (SC), Montreal, Quebec, Canada.

Received November 19, 1997; revision received March 27, 1998; accepted April 23, 1998.

Supported by grants from the Ministry of Health and Social Services of the Province of Quebec, the Montreal Regional Board for Health and Social Services, and the St. Mary's Hospital Foundation.

Address for correspondence and reprints: Jane McCusker, MD, DrPH, Department of Clinical Epidemiology and Community Studies, St. Mary's Hospital Center, 3830 Lacombe Avenue, Room 2508, Montreal, Quebec, Canada, H3T 1M5. Fax: 514-734-2652; e-mail: janemc@epid.lan.mcgill.ca

*Project ISAR Steering Committee members included: Marc Afilalo, MD, Jewish General Hospital: Louise Belanger, MSc, Montreal Regional Board for Health and Social Services; Michael Bonnycastle, MD, St. Mary's Hospital Center; Antoinette Colacone, BSc, CCRA, Jewish General Hospital; Elizabeth Healey, MEd, St. Mary's Hospital Center; Pierre Gadoury, MSS, Maisonneuve–Rosemont Hospital; and Maryse Savoie, MSc, Sacré-Coeur Hospital.

had fewer repeat ED visits during the next 3 months than did a nonrandomized control group. Possible explanations of these generally negative results include lack of targeting of those most likely to benefit and lack of coordinated follow-up.[3]

Methods for identifying elders who might benefit from GEM include self-report questionnaires, standardized measures of physical or mental performance, and clinical evaluation. In a pilot study, we found that a brief self-report questionnaire of common geriatric problems predicted repeat ED visits during the next 3 months better than standardized performance tests.[8] Given the difficulty in implementing many performance tests in the ED, as well as insufficient time during the ED visit for a comprehensive geriatric evaluation, we investigated the use of a battery of self-report questions, suitable for completion either by the patient or by an informant, that could be used to target patients for a more complete evaluation conducted at a later time either in the hospital or in the community.

Screening questionnaires have been developed for use with elders in other settings, particularly primary care or community settings.[9-15] These questionnaires may be less reliable and valid when used in the ED, because of acute changes in health and functional status and the stressful circumstances of the ED visit.

METHODS

Study Design. This study is part of a prospective cohort study that aimed to develop a short screening questionnaire for functional decline from a battery of potential questions administered in the ED to a cohort of patients ≥65 years. We define "screening" to include the detection of those at risk of functional decline and other adverse outcomes, as well as those with current functional disability; the latter is sometimes referred to as "case finding." The final questionnaire, to be reported in a future publication, will be based on the predictive validity of the questions, i.e., their ability to predict adverse outcomes (increased dependence in activities of daily living, death, and institutionalization) during the 6 months following the ED visit.

This study aimed to determine the test–retest reliability (reproducibility) and the concurrent criterion validity of screening questions administered in the ED. We assessed concurrent criterion validity by comparing responses to the screening questions at the ED visit with the results of validated instruments administered at a clinical assessment, conducted shortly after the ED visit. Test–retest reliability was assessed by repeated administration of the screening questions at the clinical assessment.

Written consent to study participation was provided by the patient or, if the patient was disoriented to time or place, by an accompanying family member who also acted as informant for the data collection. The study was approved by the research ethics review boards of the 4 hospitals.

Population and Setting. The study cohort was recruited from patients aged ≥65 years visiting the EDs of 4 acute care, university-affiliated hospitals in Montreal during the daytime weekday shift, over a 3-month period by research assistants with interviewing and/or clinical experience, who were trained and supervised by the project coordinator, a study nurse. One research assistant in each ED interviewed patients in order of arrival. After completing an interview, the research assistant next approached the patient who had arrived most recently (with the shortest duration of stay in the ED). This sampling method aimed to increase the representativeness of all ED visits by preferentially sampling incident cases (those newly arrived in the ED) vs prevalent cases (those still in the ED). Patients were excluded for the following reasons: previous enrollment in the study; transfer to another hospital; nonresidence in Quebec; or language (working knowledge of neither French nor English). Interviews were conducted in the ED either with patients who were medically stable and oriented to time and place, or else with an informant (family member or other accompanying person); therefore, patients were excluded if they were medically unstable or disoriented to either time or place and no informant was available.

Study Protocol. A sample of the study cohort was selected from those who lived within the primary catchment areas of the hospitals. The estimate of the sample size required (n = 200) was based on a desired precision of at least ±10% on the overall estimate of test–retest reliability, with a coefficient of ≥0.50. Systematic sampling was performed using the last digit of the patient number, with a daily sampling ratio varying over the recruitment period from 1 to 3 in 10; this ratio was adjusted as needed to maintain steady enrollment in the sample without overburdening the study nurse. Members of the sample were invited to participate in a clinical assessment in their homes 1–3 weeks after the ED visit. All interviews and assessments were conducted in either French or English.

Clinical assessments were conducted either at the patient's residence or in hospital; if face-to-face assessment was not possible, either a telephone interview or a mail questionnaire was completed. Assessments were conducted by 1 of 2 nurses, who were blinded to the results of the data collected in the ED.

TABLE 1. Schedule of Administration of Study Instruments

Instrument*	ED Visit (T1)	Clinical Assessment (T2)
Screening questionnaire	X	X
SPMSQ	X	X
OARS ADL	X	
SMAF		X
GDS		X
OARS Social Resources		X
Alcohol abuse		X
Medication count		X

*SPMSQ = Short Portable Mental Status Questionnaire; OARS = The Older Americans Resources and Services; ADL = Activities of Daily Living; SMAF = Functional Autonomy Measurement System; GDS = Geriatric Depression Scale.

The interrater reliability of the assessment instruments was measured in a sample of 17 patients from those assessed face-to-face; the sample was one of convenience and was the maximum number that could be jointly assessed within the budgetary and time constraints of the study. These patients were interviewed and assessed by 1 nurse, in the presence of the second nurse who independently completed a questionnaire.

Measurements. Table 1 lists the study instruments administered at the ED visit (T1) and the clinical assessment (T2).

Screening Questionnaire. Potential screening items were developed from a literature review of general factors and specific questions that predicted functional decline. Potential screening questions were circulated among a multidisciplinary group of hospital and community-based health professionals, who were asked to prioritize questions addressing each factor of interest. Because of the hypothesized importance of both functional status *before* the acute ED problem and change in function *since* the onset of this problem, questions on both aspects were included. English and French versions of the questions selected in this consensus process were then pretested in a sample of ED patients to determine ease of comprehension. To verify the accuracy of the translation, the final set of 24 questions (27 items) was back-translated from French to English by a health professional who had not been involved in the development of the questions; 2 formats were used, for administration to the patient or an informant, respectively (Appendix). Ten of these questions, noted in the Appendix, were adapted either verbatim or with very minor changes from a postal questionnaire for elders in the community developed in the province of Quebec.[12]

The screening questions were administered at T1 and again at T2, with the same respondent and method of administration (self-completion or interviewer-administered). At T2, the respondent was asked to answer the questions with regard to the patient's status at the time of the index ED visit. A screening score was computed by adding the number of positive responses to the 25 questions, excluding age and sex. No weighting of items was used because this was not expected to improve the performance of the scale.[16]

Short Portable Mental Status Questionnaire (SPMSQ). The SPMSQ,[17] an observer-rated 10-item questionnaire to evaluate orientation, memory, and concentration, was used to assess cognitive impairment of patients. Using a cutoff of ≥ 3 errors, the SPMSQ has a sensitivity of 84% and a specificity of 89% for the detection of delirium or dementia.[18]

The Older Americans Resources and Services (OARS) ADL Questionnaire. The 14-item OARS activities of daily living (ADL) questionnaire was administered at the ED visit, including questions on basic and instrumental ADL.[19] (This scale would also be used during the 6-month follow-up period to assess functional decline.) Subjects answered these questions with reference to the level of functioning *before* the illness or injury that led to the index ED visit (premorbid level of function). A 5-category impairment scale was computed: none, mild, moderate, severe, and total.[19] The criterion validity of this scale has been reported (Kendall's $\tau = 0.83$), but values for the sensitivity and specificity of different scores are not available.[19]

Functional Autonomy Measurement System (SMAF). The main measure of patient functioning used at the clinical assesssment was the SMAF (*le Système de Mesure de l'Autonomie Fonctionelle*), a 29-item clinical instrument for assessing disability in 5 areas: 7 basic ADL items (feeding, washing, dressing, grooming, urinary function, bowel function, and toiletting); 8 instrumental ADL (IADL) items (housekeeping, meal preparation, shopping, laundry, use of telephone, transportation, medication use, and budgeting); 6 mobility items (transfers, walking inside, walking outside, donning prosthesis or orthosis, wheelchair use, and climbing stairs); 3 communication items (hearing, visual, and speech problems); and 5 mental function items (memory, orientation, comprehension, judgment, and behavior).[20] Each item is rated on a 4-point scale: 0 (independent); -1 (requires surveillance or stimulation); -2 (requires help); and -3 (completely dependent). A total score, the sum of the 29 item scores, thus ranges from 0 (completely independent) to -87 (completely disabled); we used a revised scoring system, whereby a score of -0.5 is assigned to each of the ADL, IADL, or mobility items to rate subjects who perform the function independently but with difficulty.[21] The SMAF

TABLE 2. Characteristics of the Study Cohort. and of the Clinical Assessment Sample, by Participation Status

	Study Cohort	Clinical Assessment Sample (n = 404)	
		Participants	Nonparticipants
Total	1,885 (100.0%)	221 (54.7%)	183 (45.3%)
Age			
65–74 years	923 (49.0%)	92 (41.6%)	89 (48.6%)
75–84 years	714 (37.9%)	89 (40.3%)	76 (41.5%)
85+ years	248 (13.2%)	40 (18.1%)	18 (9.8%)
Sex			
Female	1,063 (56.4%)	130 (58.8%)	100 (54.6%)
Male	822 (43.6%)	91 (41.2%)	83 (45.4%)
Hospital			
A	267 (14.2%)	34 (15.4%)	33 (18.0%)
B	537 (28.5%)	57 (25.8%)	53 (29.0%)
C	526 (27.9%)	71 (32.1%)	58 (31.7%)
D	555 (29.4%)	59 (26.7%)	39 (21.3%)
Residence			
Own home	1,589 (84.3%)	175 (79.2%)	155 (84.7%)
With family	59 (3.1%)	6 (2.7%)	6 (3.3%)
Religious community	15 (0.8%)	1 (0.5%)	2 (1.1%)
Foster home or residence	190 (10.1%)	32 (14.5%)	19 (10.4%)
Nursing home	28 (1.5%)	6 (2.7%)	1 (0.6%)
Chronic care hospital	4 (0.2%)	1 (0.5%)	0 (0.0%)
Language			
English	269 (14.3%)	30 (13.6%)	21 (11.5%)
French	1,033 (54.8%)	129 (58.4%)	95 (51.9%)
Other	582 (30.9%)	62 (28.0%)	67 (36.6%)
Missing	1	0	0
Years of education			
0	35 (1.9%)	4 (1.8%)	9 (5.0%)
1–6	436 (23.3%)	47 (21.3%)	38 (20.9%)
7	261 (14.0%)	39 (17.7%)	25 (13.7%)
8–11	536 (28.7%)	60 (27.2%)	51 (28.0%)
12	267 (14.3%)	34 (15.4%)	20 (11.0%)
13+	335 (17.9%)	37 (16.7%)	39 (21.4%)
Missing	15	0	1
Marital status			
Single	157 (8.3%)	21 (9.5%)	15 (8.2%)
Married	955 (50.7%)	106 (48.0%)	103 (56.3%)
Separated/divorced	115 (6.1%)	11 (5.0%)	10 (5.5%)
Widowed	658 (34.9%)	83 (37.6%)	55 (30.1%)
Cognitive impairment			
None/mild	1,298 (68.9%)	159 (72.0%)	113 (61.8%)
Moderate/severe	184 (9.8%)	16 (7.2%)	14 (7.7%)
Not assessed	403 (21.4%)	46 (20.8%)	56 (30.6%)
Informant			
Patient	1,388 (73.6%)	168 (76.0%)	122 (66.7%)
Spouse	136 (7.2%)	18 (8.1%)	18 (9.8%)
Other	361 (19.2%)	35 (15.8%)	43 (23.5%)
Disposition			
Admitted	563 (30.6%)	72 (32.6%)	46 (27.2%)
Released	1,275 (69.4%)	149 (67.4%)	123 (72.8%)
Missing	47	0	14

continued

TABLE 2. (*Continued*)

	Study Cohort	Clinical Assessment Sample (n = 404)	
		Participants	Nonparticipants
OARS/ADL* impairment			
None	661 (35.1%)	68 (30.8%)	64 (35.0%)
Mild	560 (29.7%)	71 (32.1%)	51 (27.9%)
Moderate	370 (19.6%)	42 (19.0%)	37 (20.2%)
Severe	120 (6.4%)	17 (7.7%)	14 (7.7%)
Total	174 (9.2%)	23 (10.4%)	17 (9.3%)
Screening score mean (±SD)	5.90 ± 3.45	5.89 ± 3.48	5.77 ± 3.45

*OARS ADL = The Older Americans Resources and Services Activities of Daily Living.

total disability score has been validated against the nursing time required for care (Pearson's r = 0.88)[20]; values for the sensitivity and specificity of different scores are not available.

The ratings of each area of functioning were based on a combination of observed performance (during the visit), reports of the patient and/or an informant, and judgment of the clinician, usually a nurse. The 2 study nurses were trained in the use of the SMAF by a nurse from the group that developed the instrument, and standardized instructions were developed for this study. The nurses conducted joint patient assessments and standardized their ratings before data collection began for the study.

For each SMAF item, the nurse first assessed current disability and then asked whether there had been any change from the premorbid status. If any change was reported, the premorbid functional level was determined retrospectively. The difference between the current and premorbid total disability scores was computed; a clinically significant change was defined a priori as a difference of ≥5 points, as suggested by the developers of the SMAF (Hébert R, personal communication, 1995).

We computed dichotomous measures of disability (any disability vs none) for several areas of function. For ADL, IADL, mobility, and cognitive impairment, patients were considered to be disabled if they had a score of −1 or less for 1 or more of the items in that area. Patients were considered to have a hearing or visual disability if they had a score of −1 or less for the item in question.

Geriatric Depression Scale (GDS). The GDS, a 30-item questionnaire that has been used extensively in its original English version and also validated in French, was administered to the patient at the clinical assessment.[22,23] Patients were classified as depressed if the GDS score was 11 or more; this cutoff point has a sensitivity of 84% and a specificity of 95% for the detection of clinical depression.[24]

OARS Social Resources Questionnaire. The OARS Social Resources questionnaire, assessing social contacts and resources, was administered to the patient or informant at the clinical assessment. A 5-category impairment scale was computed: none, mild, moderate, severe, and total.[19] The interrater reliability of this scale has been reported,[19] but not its criterion validity.

Alcohol Use/Abuse. Patients were asked about the amount and frequency of alcohol consumed and completed the 4-question CAGE (cut down, annoyed, guilty, eye-opener) questionnaire.[25,26] The CAGE has been validated in older adults; with a cutoff score of ≥1 positive responses, sensitivity is 86–88% and specificity is 78–88%.[27,28] As some other studies have used a cutoff score of ≥2,[6] we used both cutoff scores in our analyses. The French version of the CAGE was that used in a provincial population health survey.[29]

Medications. At the clinical assessment, subjects were asked to produce all their current prescription and nonprescription medications for review by the nurse. These were classified by type of medication and coded into 2 variables: total number of different medications; and any current anxiolytic, sedative, or hypnotic medications. The total number of medications was then grouped into 2 categories, 0–3 and ≥4, to correspond to screening question 19 (Appendix).

Demographic variables also collected at T1 included: language (mother tongue), years of education, marital status, and residence.

Statistical Methods. The concordance correlation coefficient (CCC)[30] was used, with its 95% confidence interval (CI), to measure the test–retest reliability of the screening questionnaire (between T1 and T2) overall and within selected subgroups. A test of homogeneity[31] of the CCCs among the different subgroups was performed. The CCC was also used to evaluate the interrater reliability of

the study nurses for the instruments administered at T2 (SMAF, GDS, CAGE). The test–retest reliability of individual screening questions was measured with the kappa (κ) coefficient and its 95% CI. Changes from T1 to T2 for individual questions were evaluated with McNemar's exact test statistic and corresponding 95% CI. We computed the agreement between selected screening questions from the T1 interview and selected T2 measures, using the sensitivity, specificity, and likelihood ratio (sensitivity/1 − specificity). We used linear regression with the "best subsets" method[32] to identify the best subset of screening questions (T1) for the prediction of the patient's global disability (SMAF total score) at T2. All the statistical analyses were done using SAS for Windows, version 6.12.[33]

RESULTS

A total of 404 subjects were selected for the study sample, of whom 221 (54.7%) received a clinical assessment (93 subjects refused participation at the ED visit and 90 others either declined the assessment after being contacted by the nurse or could not be contacted). Participants did not differ substantially from nonparticipants or from the total study cohort by sex, hospital, residence, first language, education, marital status, cognitive impairment, functional impairment, information source, or disposition (Table 2). The distributions of screening questionnaire scores at the ED visit were also similar in these groups. However, participants were somewhat older than nonparticipants and the study cohort as a whole.

TABLE 3. Positive Response to Screening Questions at T1 (ED Visit), Change to T2 (Clinical Assessment), Kappa Coefficients, and 95% Confidence Intervals (n = 193)

Screening Question (Number)	T1 (%)	Change (T2 − T1) (%)	95% CI	Kappa	95% CI
Social					
Live alone (Q3)	42.2	−1.6	(−6.5, 3.4)	0.80	(0.71, 0.88)
Special residence (Q4)	19.7	−2.6	(−6.1, 1.0)	0.84	(0.75, 0.94)
Isolation (Q22)	3.6	+3.6	(−1.3, 8.6)	0.05	(−0.13, 0.23)
Lack of support (Q23)	4.2	−0.5	(−4.4, 3.4)	0.24	(−0.06, 0.54)
Inadequate income (Q24)	4.2	+0.5	(−2.3, 3.3)	0.69	(0.44, 0.95)
Physical function (premorbid)					
Activity limitation (Q6)	48.4	+3.6	(−4.4, 11.7)	0.43	(0.30, 0.56)
Need regular help (Q7)	25.9	+6.7	(0.5, 13.0)	0.59	(0.47, 0.71)
Change in function					
Decreased function (Q8)	67.9	−17.4	(−26.9, −7.9)	0.17	(0.04, 0.30)
Increased assistance (Q9)	44.6	−3.3	(−12.0, 5.5)	0.34	(0.20, 0.47)
Medical history					
Heart disease (Q10a)	31.2	+1.6	(−4.6, 7.8)	0.64	(0.52, 0.76)
Diabetes (Q10b)	17.9	+3.2	(−0.6, 6.9)	0.83	(0.73, 0.93)
Cancer (Q10c)	6.5	+2.7	(−1.3, 6.7)	0.59	(0.37, 0.81)
Stroke (Q10d)	4.7	+3.2	(−1.2, 7.5)	0.38	(0.12, 0.64)
Sensory impairment					
Hearing (Q14)	21.8	+4.1	(−1.9, 10.2)	0.57	(0.44, 0.71)
Vision (Q15)	18.7	−3.6	(−8.5, 1.3)	0.65	(0.50, 0.79)
Mental					
Memory (Q16)	11.6	0.0	(−4.4, 4.4)	0.64	(0.47, 0.81)
Depression (Q17)	29.8	+2.1	(−3.7, 8.0)	0.68	(0.56, 0.79)
Bereavement (Q18)	25.5	+5.7	(−1.1, 12.6)	0.50	(0.37, 0.64)
Medications/alcohol					
>3 medications daily (Q19)	45.8	+2.1	(−3.5, 7.7)	0.75	(0.65, 0.84)
Sleeping pills (Q20)	26.1	+4.8	(−0.9, 10.5)	0.67	(0.56, 0.79)
Alcohol daily (Q21)	6.2	0.0	(−3.4, 3.4)	0.64	(0.42, 0.87)
Hospital utilization					
ED visit last month (Q11)	20.7	−8.3	(−14.8, −1.8)	0.33	(0.17, 0.50)
Admission/6 months (Q12)	21.9	+13.0	(6.1, 20.0)	0.46	(0.33, 0.59)
Other					
Poor general health (Q5)	25.4	0.0	(−5.6, 5.6)	0.70	(0.57, 0.82)
2+ falls/6 months (Q13)	11.0	+1.6	(−3.4, 6.6)	0.52	(0.33, 0.71)

TABLE 4. Screening Score Mean at T1 (ED Visit), Change to T2 (Clinical Assessment), and Concordance Correlation Coefficient (CCC), Overall and in Subgroups

Group	*n*	T1 (%)	Change (T2 − T1) (%)	CCC	95% CI	p-value*
Total	193	5.84	+0.25	0.78	(0.71, 0.83)	
Place of clinical assessment						0.14
Home	148	5.24	+0.06	0.77	(0.70, 0.83)	
Other	43	7.72	+0.87	0.64	(0.42, 0.79)	
Missing	2	—	—	—		
Informant						0.85
Patient, not cognitively impaired†	131	5.02	+0.21	0.75	(0.67, 0.82)	
Patient, cognitively impaired†	26	6.22	+0.93	0.73	(0.49, 0.87)	
Other	36	8.56	−0.09	0.70	(0.49, 0.83)	
Time from T1 to T2						0.09
5–11 days	79	5.97	+0.61	0.82	(0.74, 0.88)	
12–20 days	112	5.67	−0.01	0.72	(0.61, 0.80)	
Missing	2	—	—	—		
Disposition at ED visit						0.09
Admitted	61	6.93	+0.77	0.68	(0.52, 0.80)	
Released	132	5.34	+0.01	0.80	(0.73, 0.85)	
ADL impairment (premorbid)						0.52
None–mild	125	4.54	+0.25	0.72	(0.63, 0.80)	
Moderate–total	68	8.24	+0.25	0.67	(0.52, 0.78)	
Increased disability‡						0.14
No	141	5.43	−0.13	0.79	(0.72, 0.85)	
Yes	52	6.96	+1.30	0.68	(0.50, 0.80)	

*P-value of a test of homogeneity of CCCs among the subgroups.
†Cognitive impairment denotes 1 or more errors in questions 1–3 of the SPMSQ at T2.
‡5+ decrease in SMAF score from before ED visit to clinical assessment.

The clinical assessments were conducted within 5–20 days (median 12 days) after the index visit. Of the 221 assessments, 168 were conducted at home, 29 in a hospital, and 15 in a nursing home; 6 screening questionnaires were administered by telephone, and 3 by mail. Interrater reliabilities (CCCs) between the 2 nurse raters were very high (0.97–1.00) for the SMAF current and premorbid total disability, ADL, IADL, mobility, and mental scores, and for the SPMSQ, GDS, and CAGE; they were somewhat lower for the Social Resources scale (0.86) and the SMAF communication subscale (current 0.68; premorbid 0.56).

Test–Retest Reliability. Subjects were included in the analyses of test–retest reliability if the same individual, either patient or informant, completed the screening questionnaire both at T1 and T2. The screening questionnaire was not administered at T2 in 5 cases, and in 21 the informant was different from the T1 informant. Some patients (n = 26) who were not cognitively impaired at T1, and therefore completed the screening questionnaire themselves, were found to be cognitively impaired at T2. These patients were included in the analyses since they answered at T2, despite their impairment. We excluded 2 patients who completed the questionnaire both at T1 and T2 but were not given the SPMSQ at T2. Thus, the reliability sample included 193 subjects.

The frequency of positive responses to the screening questions varied from 67.9% for change in function to <10% for several social factors (isolation, lack of support, inadequate income), history of cancer or stroke, and daily alcohol consumption (Table 3). For most questions, there was little systematic change from T1 to T2 in the percentage reporting the risk factor. However, there was a systematic decrease from T1 to T2 (95% CI excludes 0) in the percentage who reported a decrease in function or an ED visit in the previous month, and an increase in the percentage who reported that they usually needed regular help, or had been admitted to hospital in the previous 6 months. Higher reliabilities ($\kappa \geq 0.80$) were found for the questions on living arrangements (3 and 4), history of diabetes (10b), and memory problems (16). Lower reliabilities ($\kappa < 0.50$) were found for the questions on social isolation and support (22 and 23), premorbid activity limitation (6), acute func-

tional changes (8 and 9), history of stroke (10d), and hospital and ED use (11 and 12).

The test–retest reliability (CCC) of the total risk factor score (excluding age and sex) was 0.78 and did not vary significantly among subgroups defined by: place of clinical assessment, informant, time from T1 to T2, disposition at ED visit, premorbid ADL impairment, and recent change in level of disability (Table 4).

Concurrent Criterion Validity. The most frequent problems detected in the validity subsample were disability in ≥1 IADL (93.2%), the use of ≥4 medications (58.7%), disability in ≥1 ADL (52.9%), use of sedatives or tranquilizers (32.9%), impaired mobility (31.8%), and depression (25.9%). The prevalences of cognitive impairment differed by the method of measurement: 13.0% using the SPMSQ score vs 22.8% using the SMAF.

Selected T1 screening questions were validated against all these clinical criteria except for IADL disability because of very high prevalence (Table 5). For example, 4 screening questions (6–9) were validated against the presence of some ADL impairment: the sensitivities of these items varied between 40.7% and 70.3% and the specificities between 42.6% and 94.8%. In general, the questions were more specific than sensitive. The questions with the highest levels of both sensitivity (≥60%) and specificity (≥90%) were those on visual and hearing impairment (14 and 15) and those on medications (19 and 20). The question on depression (17) had only slightly lower levels of sensitivity and specificity. Most other questions either had lower sensitivity (especially those on isolation, support, and alcohol use) or lower specificity (most of the functional status questions).

The best subset of screening questions (including age as a continuous variable and sex) that predicted SMAF total disability at T2 are shown in Table 6; these 9 variables had an R^2 value of 0.49. Of these, the questions that were most strongly associated with level of disability included: age, need for regular help before the illness (7), and need for increased assistance associated with the illness (9).

TABLE 5. Prevalences of Selected Problems at Clinical Assessment and Sensitivity, Specificity, and Likelihood Ratio of Selected Risk Factor Questions Given in the ED (n = 213)

Problem	Assessed (n)	Prevalence (T2) (%)	Screening Question Number* (T1)	Sensitivity (%)	Specificity (%)	Likelihood Ratio
ADL impairment	204	52.9	6	60.2	65.6	1.75
			7	40.7	94.8	7.82
			8	70.3	42.6	1.22
			9	56.4	75.5	2.31
Mobility impairment	201	31.8	6	67.2	60.6	1.70
			7	50.0	88.3	4.28
			8	79.7	41.8	1.37
			9	73.3	72.9	2.71
Cognitive impairment						
SPMSQ 3+	200	13.0	16	28.0	93.5	4.33
Clinical (SMAF)	206	22.8	16	35.6	96.1	9.24
Hearing	213	20.7	14	63.6	89.9	6.33
Vision	213	11.7	15	80.0	91.0	8.85
Depression: GDS 11+	170	25.9	17	55.8	87.9	4.61
Social resources: severe or total impairment	188	11.2	22	4.8	97.6	1.99
			23	15.0	98.8	12.53
Alcohol:						
CAGE 1+	208	7.2	21	33.3	94.8	6.40
CAGE 2+	208	2.9	21	33.3	93.6	5.18
Medications						
4 or more medications	213	58.7	19	70.2	92.0	8.72
Sedatives/tranquilizers	213	32.9	20	60.0	90.8	6.51

*Numbers of subjects who were assessed for a problem but had missing values for specific screening questions ranged from 0–9.

DISCUSSION

This study is the first, to our knowledge, to assess the reliability and validity of risk factor screening items administered in the ED.

In terms of test–retest reliability, most questions had satisfactory levels of reproducibility. Those concerned with functional status and with change in functional status appeared to be the least reliable; however, the latter reliability estimates should be interpreted with caution, because the test–retest method may not be appropriate when the parameter of interest is unstable. Although respondents were asked to answer these questions at T2 with regard to the patient's health status at T1, it is likely that there were errors of recall, particularly if the patient's health status had changed, either immediately before T1 or between T1 and T2. If in fact these questions were unreliable, they would not be expected to have good validity. However, 2 of these questions, premorbid need for assistance and increased need for assistance, were among the 3 questions most predictive of total disability at T2.

The prevalence rates of the various problems identified in the clinical assessments were generally similar to those reported previously from other North American ED samples. The prevalence of dependence in at least 1 ADL was 52.9% in our sample compared with 67% and 79% in previous studies by Miller et al.[7] and Gerson et al.,[6] respectively. (Note that these studies used different criteria for ADL dependence from ours, and the Gerson study included IADL dependence.) The prevalence of depression was 33% in the Gerson study compared with 25.9% in ours, using similar criteria. The prevalence of cognitive impairment was 22.8% in our sample (using the SMAF) compared with 19% in the Miller study (using other instruments). The prevalence of alcohol abuse based on a CAGE score of ≥2 was 2.9% in our sample and 2.5% in the Gerson study.

Our analyses indicate that a subset of 9 screening questions explains almost half of the variability in disability. Three of the questions are particularly strong predictors of disability: age, premorbid need for assistance, and increased need for assistance. Although age is readily available, patients are rarely asked about their dependency during an ED visit.[34] The use of these brief, intuitively appealing questions could provide substantial information with relatively little effort.

In general, the concurrent validity of most of the screening questions was not sufficiently high for these questions to be used alone. The questions on social isolation and lack of support (22 and 23) were unreliable and detected only a small percentage of those with severely impaired social resources, using OARS criteria. Hébert and colleagues[12] found that elders living with others were more likely to decline in their functioning than those who lived alone; this result is likely explained by the greater need of contact and support among those elders living with others. The screening question on depression is almost identical to one previously recommended for screening internal medicine patients[35] that appeared to have concurrent validity similar to the GDS in one small study.[36]

TABLE 6. Best Subset of Screening Questions (T1) for the Prediction of Disability at T2*

Question (Number)†	Coefficient	SE	p-value
Age	0.03	0.01	<0.001
Live alone (Q3)	−0.24	0.11	0.02
Poor general health (Q5)	0.34	0.13	0.009
Premorbid need for regular help (Q7)	0.55	0.14	<0.001
Increased assistance (Q9)	0.51	0.11	<0.001
Diabetes (Q10b)	0.27	0.13	0.04
Hearing (Q14)	0.22	0.13	0.09
Memory (Q16)	0.37	0.17	0.03
Isolation (Q22)	−0.57	0.28	0.04

R^2 = 0.49 Adjusted R^2 = 0.46 for the above "best" subset model
R^2 = 0.54 Adjusted R^2 = 0.46 for the full model with 27 items

*Multiple linear regression model; disability was measured as the log of 1 minus SMAF total score; n = 182 subjects with answers on all questions.
†Actual age was used in the model as a continuous variable; a score of 1 was given for "yes" and 0 for "no" to questions 3, 7, 9, 10b, and 16, and vice-versa for questions 5, 14, and 22.

LIMITATIONS AND FUTURE QUESTIONS

The patient groups excluded from this study may limit its generalizability. Patients were recruited only during weekdays; those visiting the EDs at different times may be more seriously ill,[8] but, among those eligible for screening, there is no reason to suppose that the screening questions would differ in their reliability or validity. Patients were excluded from the study cohort if they could not be interviewed because of their clinical status or cognitive impairment and no informant was available. Although screening in the ED using self-report questions is obviously not possible for these patients, they may very well be at increased risk of adverse outcomes, and other methods of management may need to be considered. Clinical assessments could be completed in just over half of those eligible; however, there were only small differences between participants and nonparticipants based on their characteristics at the ED visit, and the distributions of screening scores in these groups were almost identical to those of the entire study cohort.

The potential for recall bias in several of the measures should be noted, particularly during repeat administration of the screening questions at T2, and in the assessment of change in the SMAF total disability score.

This study was part of a larger effort to develop a self-report ED screen to identify elders at risk of adverse outcomes. The final choice of a short screening questionnaire will therefore be based primarily on its predictive validity. We hope, therefore, to make recommendations for a screening questionnaire when the analyses from follow-up of the cohort have been completed.

CONCLUSIONS

The results of this study indicate that screening elders in the ED for risk factors for functional decline and other adverse outcomes is feasible, the screening questionnaire score has good test–retest reliability, but individual screening questions have, at best, modest concurrent validity. The final set of screening questions should be selected based on their predictive validity.

References

1. Stuck AE, Siu AL, Wieland DG, Adams J, Rubenstein LZ. Comprehensive geriatric assessments: a meta-analysis of controlled trials. Lancet. 1993; 342:1032–6.
2. Sanders AB. Emergency department utilization by elder persons [commentary]. Acad Emerg Med. 1996; 3:658–9.
3. Rubenstein LZ. The emergency department: a useful site for CGA? J Am Geriatr Soc. 1996; 44:601–2.
4. Currie C, Lawson P, Robertson C, Jones A. Elderly patients discharged from accident and emergency departments—their dependency and support. Arch Emerg Med. 1984; 1:205–13.
5. Rosenfeld T, Fahey P, Price M, Leeder S. The fate of elderly patients discharged from the accident and emergency department of a general teaching hospital. Community Health Stud. 1990; 14:365–72.
6. Gerson LW, Rousseau EW, Hogan TM, Bernstein E, Kalbfleisch N. Multicenter study of case finding in elderly emergency department patients. Acad Emerg Med. 1995; 2: 729–34.
7. Miller DK, Lewis LM, Nork MJ, Morley JE. Controlled trial of a geriatric case-finding and liaison service in an emergency department. J Am Geriatr Soc. 1996; 44:513–20.
8. McCusker J, Healey E, Bellavance F, Connolly B. Predictors of repeat emergency department visits in the elderly. Acad Emerg Med. 1997; 4:581–8.
9. Bowns I, Challis D, Tong MS. Case finding in elderly people: validation of a postal questionnaire. Br J Gen Pract. 1991; 41: 100–4.
10. Freer CB. Consultation-based screening of the elderly in general practice: a pilot study. J R Coll Gen Pract. 1987; 37: 455–6.
11. Frederiks CM, Margreet J, TeWierik MJ, Visser A, Sturmans F. The functional status and utilization of care of elderly people living at home. J Community Health. 1990; 15:307–71.
12. Hébert R, Bravo G, Korner-Bitensky N, Voyer L. Predictive validity of a postal questionnaire for screening community-dwelling elderly individuals at risk of functional decline. Age Ageing. 1996; 25:159–67.
13. Kerse N, Clark F. Case finding in the elderly: a postal questionnaire. N Z Med J. 1994; 107:33–6.
14. Pannill FC. A patient-completed screening instrument for functional disability in the elderly. Am J Med. 1991; 90: 320–7.
15. Boult C, Dowd B, McCaffrey D, Boult L, Hernandez R, Krulewitch H. Screening elders for risk of hospital admission. J Am Geriatr Soc. 1993; 41:811–7.
16. Streiner DL, Norman GR. From items scales. In: Health Measurements Scales: A Practical Guide to Their Development and Use. New York, NY: Oxford University Press, 1994, pp 67–70.
17. Pfeiffer E. A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. J Am Geriatr Soc. 1975; 23:433–41.
18. Smyer MA, Hofland BF, Jonas EA. Validity study of the Short Portable Mental Status Questionnaire for the Elderly. J Am Geriatr Soc. 1979; 27:263–9.
19. Fillenbaum GG. Multidimensional Functional Assessment of Older Adults: The Duke Older Americans Resources and Services Procedures. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988, pp 34, 143–5.
20. Hébert R, Carrier R, Bilodeau A. The functional autonomy measurements system (SMAF): description and validation of an instrument for the measurement of handicaps. Age Aging. 1988; 17:293–302.
21. Desrosiers J, Bravo G, Hébert R, Dubuc N. Reliability of the revised functional autonomy measurement system (SMAF) for epidemiological research. Age Ageing. 1995; 24:402–6.
22. Vézina J, Laprise R, Gourgue M, Parenteau P, Fréchette M. A comparison of the diagnostic performance of the Beck Depression Inventory and the Geriatric Depression Scale using ROC analyses. Can J Aging. (in press), 1998.
23. Bourque P, Blanchard L, Vézina J. Étude psychométrique de l'échelle de dépression gériatrique. (Psychometric study of the Geriatric Depression Scale.) Can J Aging. 1990; 9:348–55.
24. Yesavage JA, Brink TL, Rose TL, et al. Development and validation of a geriatric depression screening scale: a preliminary report. J Psychiatr Res. 1983; 17(1):37–49.
25. Ewing JA. Detecting alcoholism: the CAGE questionnaire. JAMA. 1984; 252:1905–7.
26. King M. At risk drinking among general practice attenders: validation of the CAGE questionnaire. Psychol Med. 1986; 16:213–7.
27. Buchsbaum DG, Buchanan RG, Welsh J, Centor RM, Schnoll SH. Screening for drinking disorders in the elderly using the CAGE questionnaire. J Am Geriatr Soc. 1992; 40: 662–5.
28. Jones TV, Lindsey BA, Yount P, Soltys R, Farani-Enayat B. Alcoholism screening questionnaires: are they valid in elderly medical outpatients? J Gen Intern Med. 1993; 8:674–8.
29. Quebec Health and Social Survey, 1992–1993. Gouvernement du Québec, Conseil québécois de la recherche sociale.
30. Lin LI-K. A concordance correlation coefficient to evaluate reproducibility. Biometrics. 1989; 45:255–68.
31. Cooper H, Hedges L. The Handbook of Research Synthesis. New York, NY: Russell Sage Foundation, 1994, pp 266–7.
32. Neter J, Wasserman W, Kutner MH. Applied Linear Statistical Models. Third edition. Homewood, IL: Irwin, 1990, pp 443–52.
33. SAS Institute Inc. SAS/STAT Software: Changes and Enhancements through Release 6.12. Cary, NC: SAS Institute Inc., 1997, pp 219–46, 1003–28.
34. Hedges JR, Singal BM, Rousseau EW, et al. Geriatric patient emergency visits. Part II: perceptions of visits by geriatric and younger patients. Ann Emerg Med. 1992; 21:808–13.
35. Lachs MS, Feinstein AR, Cooney LM, et al. A simple procedure for general screening for functional disability in elderly patients. Ann Intern Med. 1990; 112:699–706.
36. Mahoney J, Drinka TJK, Abler R, et al. Screening for depression: single question versus GDS. J Am Geriatr Soc. 1994; 42:1006–8.

(The appendix follows)

APPENDIX

Risk Factor Questionnaire

1. What is your age? *(65–74 years, 75–84, 84 or over)*
2. Your sex? *(male, female)*
3. Do you live alone? *(yes, no)**
4. Do you live in a foster home, nursing home, or residence for elder persons? *(yes, no)*
5. In comparison with other people your own age, do you think you are in good health? *(yes, no)**
6. *Before* the illness or injury that brought you to the emergency department, did you have any health problems that required you to limit your activities? *(yes, no)**
7. *Before* the illness or injury that brought you to the emergency department, did you need someone to help you on a regular basis? *(yes, no)**
8. *Since* the illness or injury that brought you to the emergency department, has there been a decrease in your ability to get about? *(yes, no)*
9. *Since* the illness or injury that brought you to the emergency department, have you needed more help than usual to take care of yourself? *(yes, no)*
10. During the past year, have you had:
    a) Heart disease *(yes, no)*
    b) Diabetes *(yes, no)*
    c) Cancer *(yes, no)*
    d) Stroke *(yes, no)*
11. Have you visited a hospital emergency department during the past month (not counting this visit)? *(yes, no)*
12. Have you been hospitalized for one or more nights during the past 6 months (excluding a stay in the emergency department)? *(yes, no)*
13. Did you trip or fall 2 or more times during the past 6 months? *(yes, no)*
14. In general, do you hear well? *(yes, no)**
15. In general, do you see well? *(yes, no)**
16. In general, do you have serious problems with your memory? *(yes, no)**
17. In general, do you feel sad and depressed? *(yes, no)**
18. During the past year, have you been affected by the death of a person close to you, or another serious event? *(yes, no)*
19. Do you take more than 3 different medications every day? *(yes, no)**
20. Do you take pills to help you sleep? *(yes, no)*
21. Do you drink alcohol (wine, beer, etc.) every day, not counting drinking with meals? *(yes, no)*
22. Do you usually have daily contact with other people? *(yes, no)*
23. In case of need, can you count on someone close to you? *(yes, no)**
24. Do you usually have enough income to meet your daily needs? *(yes, no)*

A risk factor score ranging from 0 to 25 was computed from the responses to all items except age and sex (1 point was given for answers of "yes" to items 3–4, 6–13, and 16–21, and for answers of "no" to items 5, 14, 15, and 22–24).

---

*Questions adapted from postal questionnaire of Hébert et al.[12]