original reports abstract

A Two-Step Frailty Assessment Strategy in Older Patients With Solid Tumors: A Decision Curve Analysis

Adolfo González Serrano, MD, MSc[1]; Marie Laurent, MD, PhD[1,2]; Thomas Barnay, PhD[3]; Claudia Martínez-Tapia, PhD[1]; Etienne Audureau, MD, PhD[1,4]; Pascaline Boudou-Rouquette, MD[5]; Thomas Aparicio, MD, PhD[6]; Florence Rollot-Trad, MD[7]; Pierre Soubeyran, MD, PhD[8]; Carine Bellera, PhD[9,10]; Philippe Caillet, MD[1,11,12]; Elena Paillaud, MD, PhD[1,11,12]; and Florence Canouï-Poitrine, MD, PhD[1,4]

PURPOSE The intended clinical value of frailty screening is to identify unfit patients needing geriatric assessment (GA) and to prevent unnecessary GA in fit patients. These hypotheses rely on the sensitivity and specificity of screening tests, but they have not been verified.

METHODS We performed a cross-sectional analysis of outpatients age ≥ 70 years with prostate, breast, colorectal, or lung cancer included in the ELCAPA cohort study (ClinicalTrials.gov identifier: NCT02884375) between February 2007 and December 2019. The diagnostic accuracy of the G8 Geriatric Screening Tool (G8) and modified G8 scores for identifying unfit patients was determined on the basis of GA results. We used decision curve analysis to calculate the benefit of frailty screening for detecting unfit patients and avoiding unnecessary GA in fit patients across different threshold probabilities.

RESULTS We included 1,648 patients (median age, 81 years), and 1,428 (87%) were unfit. The sensitivity and specificity were, respectively, 85% (95% CI, 84 to 87) and 59% (95% CI, 57 to 61) for G8, and 86% (95% CI, 84 to 87) and 60% (95% CI, 58 to 63) for the modified G8 score. For decision curve analysis, the net benefit (NB) for identifying unfit patients were 0.72 for G8, 0.72 for the modified G8, and 0.82 for GA at a threshold probability of 0.25. At a threshold probability of 0.33, the NBs were 0.71, 0.72, and 0.80, respectively. At a threshold probability of 0.5, the NBs were 0.68, 0.69, and 0.73, respectively. No screening tool reduced unnecessary GA in fit patients at predefined threshold probabilities.

CONCLUSION Although frailty screening tests showed good diagnostic accuracy, screening showed no clinical benefits over the GA-for-all strategy. NB approaches, in addition to diagnostic accuracy, are necessary to assess the clinical value of tests.

J Clin Oncol 41:826-834. © 2022 by American Society of Clinical Oncology

INTRODUCTION

Geriatric assessment (GA; a comprehensive, multidimensional evaluation of a patient's health status) can identify undetected age-related factors such as frailty.[1] GA in older patients with cancer influences treatment selection, guides geriatric interventions,[2] and predicts mortality[3] and chemotherapy toxicity.[4-6] Hence, American Society of Clinical Oncology,[7] National Comprehensive Cancer Network,[8] and International Society of Geriatric Oncology (SIOG)[1] have recommended GA to optimize decision making in older patients with cancer. However, GA is resource- and time-consuming. Hence, a two-step strategy using frailty screening has been recommended to distinguish between unfit patients (who should undergo GA) and fit patients (who should avoid GA).[9] Although many frailty screening tools are available, in two systematic reviews, the G8 Geriatric Screening Tool (G8) score was one of the most robust because of its high sensitivity and acceptable specificity.[9,10] The G8 score is recommended by the SIOG,[11,12] the European Association of Urology,[13] the European Society of Surgical Oncology, and the European Organisation for Research and Treatment of Cancer.[14,15]

Typically, the evaluation and recommendation to use one screening test over another relies on sensitivity and specificity. However, it is unclear which level of sensitivity and specificity justifies its clinical implementation.[16] In addition, these statistics are of limited clinical value in decision making as they lack any diagnostic meaning because they are conditioned by the future (the confirmed diagnosis) when predicting the past (test result).[17] A physician's decisions should rely on post-test probabilities, that is, the probability of having a disease, given the test results, rather than the contrary.[17] Moreover, sensitivity, specificity, and area under the curve do not provide information about the clinical utility of tests and do not account for the risks or benefits of decisions on the basis of test results.[18]

ASSOCIATED CONTENT

Appendix

Data Supplement

Author affiliations and support information (if applicable) appear at the end of this article.

Accepted on September 9, 2022 and published at ascopubs.org/journal/jco on October 28, 2022: DOI https://doi.org/10.1200/JCO.22.01118

CONTEXT

Key Objective

Typically, the clinical value of screening tools relies on diagnostic accuracy measures such as sensitivity and specificity. However, it is unclear which accuracy level justifies tests' implementation or leads to decisions that would do more harm than good. In older patients with cancer, the recommendation of using frailty screening tools to identify patients who should undergo geriatric assessment (GA) relies on the diagnostic accuracy of screening tools. However, their clinical utility has not been verified.

Knowledge Generated

We used decision curve analysis to evaluate the utility of frailty screening tools for detecting unfit patients or avoiding unnecessary GAs in fit patients. Despite the good diagnostic accuracy of screening tools, we did not observe any benefits of frailty screening over a geriatric-assessment-for-all approach, regardless of the tumor type, clinical stage, or age group.

Relevance *(S.B. Wheeler)*

This analysis examining the utility of frailty screening before GA failed to demonstrate clinical value of prescreening for frailty across multiple tumor types, disease stages, and age groups.*

*Relevance section written by *JCO* Associate Editor Stephanie B. Wheeler, PhD, MPH.

In decision curve analysis (DCA), an intervention's expected risks and benefits are weighted across a range of physician or patient preferences.[18] Moreover, DCA enables simultaneous evaluation of several strategies so that physicians and patients can select the best one.[19] Finally, DCA provides insight into whether making decisions on the basis of test results would do more harm than good.[18]

Therefore, the objective of this study was to evaluate the value (as assessed in a DCA) of frailty screening on the basis of the G8 score or the modified 6-item G8 score (hereafter referred to as the modified G8 score) to identify unfit individuals needing GA and to avoid GA in fit individuals.

METHODS

Design, Setting, and Participants

In this cross-sectional study, we analyzed outpatients with cancer of the multicenter, prospective ELCAPA cohort between February 2007 and December 2019. The ELCAPA study was conducted at 19 geriatric oncology clinics in the Greater Paris region of France. All patients were age ≥ 70 years, had a confirmed cancer diagnosis, and were referred to the geriatric oncology unit for GA before treatment selection. We only included patients for whom complete GA, G8, and modified G8 data sets were available. All patients provided informed consent before inclusion in the study. The study protocol was approved by an independent ethics committee (*CPP Ile-de-France* I, Paris, France; reference: 2019 mai-MS121). The ELCAPA study is registered at ClinicalTrials.gov (identifier: NCT02884375).

Outcome Ascertainment

GA (the reference test). The GA in the ELCAPA study included seven validated scoring tools covering important health domains for older patients with cancer, assessed by an experienced geriatrician using a standardized case report form. The instruments, thresholds, and definition of unfit patients were consistent with development studies for the G8[20] and modified G8 scores.[21] As in the study by Bellera et al,[20] we defined unfit patients as those with at least one abnormal score in the GA domains: activities of daily living (ADL; dependency on at least one item); instrumental activities of daily living (IADL; dependency on at least one item); the Mini-Mental State Examination (≤ 23 out of 30 points); the Mini Geriatric Depression Scale (Mini GDS, ≥ one out of four points); the Mini Nutritional Assessment (MNA, ≤ 23.5 out of 30 points); the Cumulative Illness Rating Scale for Geriatrics (CIRS-G, at least one grade 3 or 4 comorbidity); and the Timed Up and Go test (more than 20 seconds). In patients with prostate cancer (and as recommended by the SIOG), we used only bathing, dressing, toileting, transferring, and feeding items to evaluate ADL. Management of money, medications, transportation, and telephone use were used to evaluate IADL.[22]

Screening tools. The G8 score included the following eight items: food intake and weight loss over the last three months, reduced mobility, neuropsychologic problems, body mass index, administration of more than three daily drugs, self-rated health status, and the patient's age.[20]

The modified G8 score included the following six items: weight loss over the past three months, neuropsychologic problems, administration of more than six daily drugs, self-rated health status, Eastern Cooperative Oncology Group performance status, and the presence of heart failure or coronary disease.[21]

Abnormal G8 and modified G8 scores were defined using the recommended cutoffs (≤ 14 of 17 points and ≥ 6 of 35 points, respectively).[20,21]

Covariates. We included the following baseline covariates to describe the study population and perform sensitivity

analysis: tumor type, clinical stage (metastatic disease or nonmetastatic/unknown), node involvement (positive or negative/unknown), age, and sex.

Statistical Analysis

Summary statistics were used to describe participants' demographic and clinical characteristics. The sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated (95% CIs) for each screening tool.

DCA

We used the DCA technique described by Vickers et al.[18] We assumed that reasonable threshold probabilities were 0.25 and 0.33, indicating that missing an unfit patient was three and two times worse than exposing a fit patient to an unnecessary GA (odds of 1:3 and 1:2, respectively). We also assessed a threshold probability of 0.5 (odds 1:1), assuming that missing an unfit patient was the same as exposing a fit patient to an unnecessary GA (this probability corresponds to the underlying hypothesis of evaluating the tests' usefulness with sensitivity and specificity).

We compared the net benefit (NB) of frailty screening with GA-for-all and GA-for-none approaches. The NB in terms of true positives (TP; identified unfit patients) represented the percentage of unfit individuals identified by the test if the test's false-positive rate was zero. NB in terms of true negatives (avoided GAs in fit patients) represented the percentage of fit individuals identified by the test if the test's false-negative (FN) rate was zero. Finally, the tradeoff corresponded to the number needed to test to make the GA worthwhile, given its additional cost.[23]

Sensitivity Analyses

We performed analyses of clinical stage, tumor type, and age group. We performed analyses in which the criterion for abnormal GA was two impaired domains rather than one. We also used an alternative definition of abnormal GA, as described by Martinez-Tapia et al.[24] This definition included recommendations issued by physicians in four clinically relevant domains: nutritional, home, neuropsychologic, and social support. Prescription of recommendations in at least one of the four domains defined an abnormal GA. We compared the geriatric and demographic characteristics of FN and TP patients of the G8 and the modified G8 scores. Finally, we performed pooled analyses of published data for the Vulnerable Elders Survey-13 (VES-13),[25-27] Flemish version of the Triage Risk Screening Tool (fTRST),[28] Cardiovascular Health Study (CHS) instrument,[27] and Senior Adult Oncology Program 2 (SAOP2).[29] See the Data Supplement for a description of these tools (online only). All statistical analyses were performed using the Stata software (version 14.2, Stata Corp, College Station, TX).

RESULTS

Participants

Between February 2007, and December 2019, 2,168 outpatients with breast cancer (n = 1,059 [49%]),

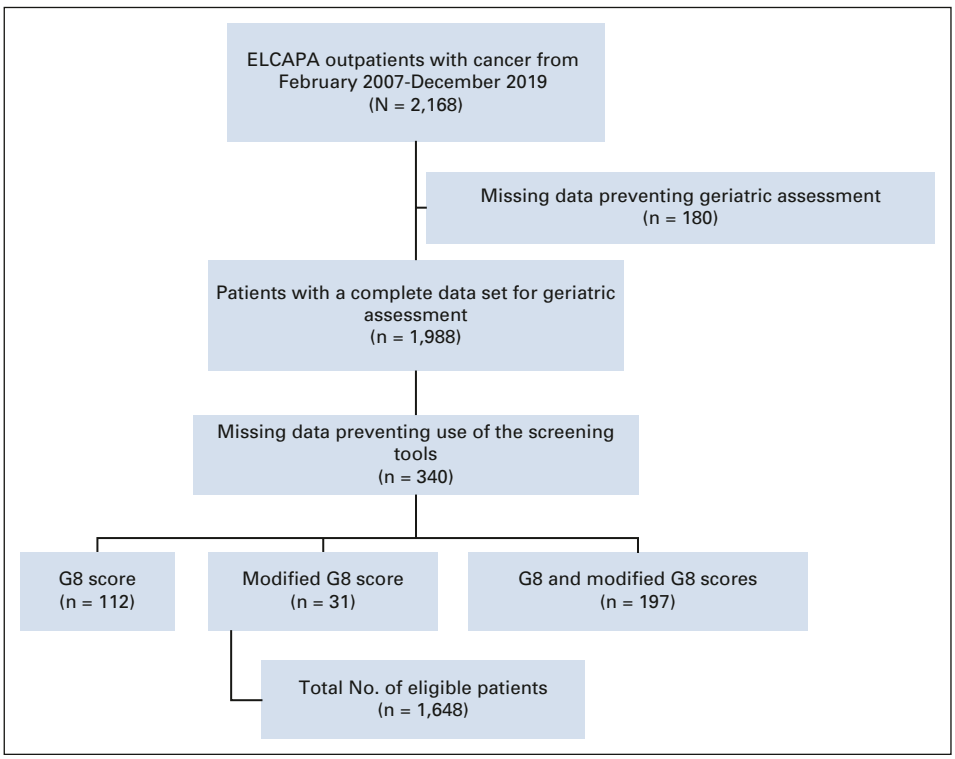

FIG 1. Flow diagram of participants and available data.

colorectal cancer (n = 507 [23%]), prostate cancer (n = 348 [16%]), and lung cancer (n = 254 [12%]) were included in the ELCAPA study. Complete data sets were available for 1,648 patients (76%; Fig 1). The median (interquartile range [IQR]) age was 81 (77-85) years, 1,067 patients (65%) were females, 559 (34%) had metastatic disease, and 1,428 (87%) were unfit. CIRS-G grade 3 or grade 4 comorbidities were the most common impaired domains (58%). For G8, the median (IQR) score was 12 points (10-14), and 1,308 patients (79%) were considered unfit. For the modified G8, the median (IQR) score was 14 points (6-22), and 1,309 patients (79%) were considered unfit. The other characteristics of the study population are summarized in Appendix Table A1 (online only).

Excluded patients (n = 520) were older (median age: 83 years, *v* 81 years for included patients) and were less likely to have CIRS-G grade 3 or grade 4 comorbidities (45% *v* 58%), an impaired Mini GDS (22% *v* 32%), an impaired MNA (27% *v* 47%), an impaired G8 score (67% *v* 79%), and an impaired modified G8 score (63% *v* 79%).

Diagnostic Accuracy of Screening Tools for Identifying Unfit Patients

The G8 score had a sensitivity of 85% (95% CI, 84 to 87), a specificity of 59% (95% CI, 57 to 61), a PPV of 93% (95% CI, 92 to 94), and an NPV of 38% (95% CI, 36 to 41). The modified G8 score had a sensitivity of 86% (95% CI, 84 to 87), a specificity of 60% (95% CI, 58 to 63), a PPV of 93% (92 to 95), and an NPV of 39% (95% CI, 37 to 42).

DCA

For the G8 and modified G8 scores, NB at a threshold probability of 0.25 (odds of 1:3) was 0.72. This result equates to screening 72 patients per 100 with the G8 score, all of whom were found to be unfit. The GA-for-all strategy yielded an NB of 0.82. At a threshold probability of 0.33, the NB for the G8, modified G8 score, and GA-for-all approaches were 0.71, 0.72, and 0.80, respectively. With a threshold probability of 0.50, the NB were 0.68 for G8, 0.69 for modified G8, and 0.73 for the GA-for-all strategy, respectively. The NB data are shown in Figure 2.

We observed negative values when calculating the ability of the screening tools to avoid GAs in fit patients. These negative values represent the number of unfit patients lost if screening was used to decide who should undergo GA. For the G8 score, the numbers of lost patients at probabilities of 0.25, 0.33, and 0.50 were 27, 18, and 5, respectively; for the modified G8 score, they were 26, 17, and 4, respectively (Table 1).

Sensitivity Analyses

At probabilities under 0.5, neither of the scores showed NBs over GA, regardless of the tumor type, clinical stage, or age group (Appendix Tables A2 and A3, online only). When using the alternative definition of abnormal GA (according to Martinez Tapia et al[24]), we did not observe NBs for screening over GA. When applying a GA cutoff of two

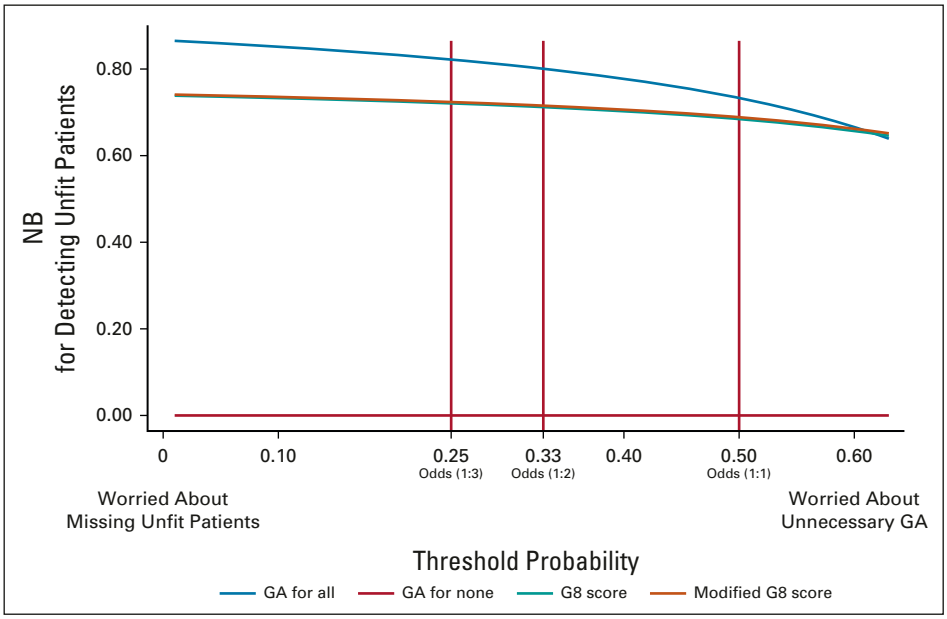

FIG 2. Decision curve for the NB of three different strategies (relative to not doing GA for any patient), with different threshold probabilities. The odds of 1:3 and 1:2 mean that missing an unfit patient is, respectively, three and two times as bad as exposing a fit patient to an unnecessary GA. The odds of 1:1 mean missing an unfit patient is the same as exposing a fit patient to an unnecessary GA. The NB represents the number of patients (per 100) who will be found to be unfit using a given frailty screening tool. G8, G8 Geriatric Screening Tool; GA, geriatric assessment; NB, net benefit.

TABLE 1. Diagnostic Accuracy and Net Benefit for Detecting Unfit Patients and Avoiding Unnecessary GA in Fit Patients for the G8 and the Modified G8 Scores, Using One- and Two-Abnormal-Domain Cutoffs and an Alternative Definition for an Abnormal GA

Result	One-Abnormal-Domain Cutoff		Two-Abnormal-Domain Cutoff		Alternative Abnormal GA Definition	
No. with full data sets (%)	1,648 (100)		1,648 (100)		1,648 (100)	
Prevalence of an abnormal GA, No. (%)	1,428 (87)		1,067 (65)		1,234 (75)	
Result	**G8 Score**	**Modified G8 Score**	**G8 Score**	**Modified G8 Score**	**G8 Score**	**Modified G8 Score**
TP, No.	1,218	1,222	989	985	1,055	1,069
False positives, No.	90	87	319	324	253	240
Sensitivity, % (95% CI)	85 (84 to 87)	86 (84 to 87)	93 (91 to 94)	92 (91 to 94)	85 (84 to 87)	87 (85 to 88)
Specificity, % (95% CI)	59 (57 to 61)	60 (58 to 63)	45 (43 to 48)	44 (42 to 47)	39 (37 to 41)	42 (40 to 44)
PPV, % (95% CI)	93 (92 to 94)	93 (92 to 95)	76 (74 to 78)	75 (73 to 77)	81 (79 to 83)	82 (80 to 84)
NPV, % (95% CI)	38 (36 to 41)	39 (37 to 42)	77 (75 to 79)	76 (74 to 78)	47 (45 to 50)	51 (49 to 54)
Probability threshold of 0.25 (odds 1:3)						
NB of screening	0.72	0.72	0.54	0.53	0.59	0.6
NB of GA	0.82	0.82	0.53	0.53	0.67	0.67
NB difference[a]	–0.10	–0.10	0.01	0.00	–0.08	–0.07
Avoided unnecessary GA (No.)[b,c]	–27	–26	2	1	–23	–20
Tradeoff (No.)	10	10	[d]	[d]	13	15
Probability threshold of 0.33 (odds 1:2)						
NB of screening	0.71	0.72	0.50	0.50	0.56	0.58
NB of GA	0.80	0.80	0.47	0.47	0.62	0.62
NB difference[a]	–0.09	–0.08	0.03	0.03	–0.06	–0.05
Avoided unnecessary GA (No.)[b,c]	–18	–17	6	6	–12	–9
Tradeoff (No.)	11	12	[d]	[d]	17	25
Probability threshold of 0.5 (odds 1:1)						
NB of screening	0.68	0.69	0.41	0.40	0.49	0.5
NB of GA	0.73	0.73	0.29	0.29	0.5	0.49
NB difference[a]	–0.05	–0.04	0.11	0.11	–0.14	–0.01
Avoided unnecessary GA (No.)[b,c]	–5	–4	11	11	–1	0.55
Tradeoff (No.)	21	23	[d]	[d]	92	[d]

NOTE. The odds of 1:3 and 1:2 mean that missing an unfit patient is, respectively, three and two times as bad as exposing a fit patient to an unnecessary GA. The odds of 1:1 mean missing an unfit patient is the same as exposing a fit patient to an unnecessary GA.

Abbreviations: G8, G8 Geriatric Screening Tool; GA, geriatric assessment; NB, net benefit; NPV, negative predictive value; PPV, positive predictive value; TP, true positives.

[a]Negative NB differences mean that screening has no benefit over a GA-for-all approach.

[b]Number per 100 patients.

[c]Negative values represent the number of unfit patients lost if screening is used to decide who should undergo GA.

[d]The tradeoff is not reported as the NB difference favors screening.

impaired domains, both scores showed favorable NBs at threshold probabilities under 0.5 (Table 1).

Compared with FNs, TPs for G8 were more likely to have impaired results in all GA domains, except for CIRS-G grade 3 or grade 4 comorbidities. For the modified G8, TPs were more likely to have impaired results in all GA domains, except for Mini-Mental State Examination and MiniGDS. See Appendix Table A4 (online only).

At threshold probabilities under 0.5, only SAOP2 (cutoff of ≥ 3 impaired domains in GA) showed a benefit for identifying unfit patients and reducing unnecessary GAs in fit patients (Table 2).

DISCUSSION

In the DCA of an outpatient population of older patients with prostate, breast, colorectal, or lung cancer, despite the

TABLE 2. Diagnostic Accuracy and NB for Detecting Unfit Patients and Avoiding Unnecessary GA in Fit Patients, Using the Various Frailty Screening Tools

	VES-13				fTRST	CHS	SAOP2[d]
Result	**Soubeyran[25]**	**Luciani[26]**	**Biganzoli[27]**	**Pooled results**	**Kenis[28]**	**Biganzoli[27]**	**Russo[29]**
No. with full data sets	1,435	419	259	704	937	259	282
Prevalence of an abnormal GA, No. (%)	1,151 (80)	119 (28)	171 (66)	480 (58)	693 (74)	171 (66)	175(62)
Sensitivity, 95% CI	0 to 69	0 to 87	0 to 62	0 to 73	0 to 91	0 to 87	0 to 94
Specificity, 95% CI	0 to 74	0 to 62	0 to 81	0 to 72	0 to 42	0 to 49	0 to 47
Probability threshold of 0.25 (odds 1:3)							
NB of screening	0.53	0.15	0.39	0.38	0.62	0.52	0.52
NB of GA	0.74	0.05	0.55	0.58	0.65	0.55	0.50
NB difference[a]	–0.21	0.11	–0.16	–0.19	–0.03	–0.03	0.02
Avoided unnecessary GA (No.)[b,c]	–64	32	–48	–59	–9	–9	6
Tradeoff (No.)	5	[e]	6	5	33	32	[e]
Probability threshold of 0.33 (odds 1:2)							
NB of screening	0.53	0.11	0.38	0.34	0.60	0.49	0.48
NB of GA	0.70	–0.07	0.49	0.37	0.61	0.49	0.43
NB difference[a]	–0.18	0.18	–0.11	–0.04	–0.01	0.00	0.05
Avoided unnecessary GA (No.)[b,c]	–35	36	–23	–7	–2	0	10
Tradeoff (No.)	6	[e]	9	27	83	400	[e]
Probability threshold of 0.5 (odds 1:1)							
NB of screening	0.50	–0.03	0.34	0.31	0.52	0.40	0.38
NB of GA	0.60	–0.43	0.32	0.36	0.48	0.32	0.24
NB difference[a]	–0.10	0.40	0.02	–0.06	0.04	0.08	0.14
Avoided unnecessary GA (No.)[b,c]	–10	40	2	–6	4	8	14
Tradeoff (No.)	10	[e]	[e]	17	[e]	[e]	[e]

NOTE. The odds of 1:3 and 1:2 mean that missing an unfit patient is, respectively, three and two times as bad as exposing a fit patient to an unnecessary GA. The odds of 1:1 mean missing an unfit patient is the same as exposing a fit patient to an unnecessary GA.

Abbreviations: CHS, Cardiovascular Health Study (CHS) instrument; fTRST, Flemish version of the Triage Risk Screening Tool; GA, geriatric assessment; NB, net benefit; SAOP2, Senior Adult Oncology Program 2; VES-13, Vulnerable Elders Survey-13.

[a]Negative NB differences mean that screening has no benefit over a GA-for-all approach.

[b]Number per 100 patients.

[c]Negative values represent the number of unfit patients lost if screening is used to decide who should undergo GA.

[d]The cutoff for defining unfit patients was the existence of ≥ 3 impaired domains in GA.

[e]The tradeoff is not reported as the NB difference favors screening.

good diagnostic accuracy of screening tools, we did not observe any NBs of two-step frailty screening over a one-step, GA-for-all strategy.

In our study, the sensitivity and specificity of G8 and modified G8 (85% and 59%; and 86% and 60%, respectively) coincided with those reported in development studies (85% and 57%[20]; and 89% and 79%,[21] respectively). Only the specificity of the modified G8 was lower in our study, but in contrast to the study by Martinez-Tapia et al,[21] we included only outpatients. Similar to other studies, we observed variations in these measures across tumor types, clinical stages, and age groups.[25,30,31]

High sensitivity is considered the most relevant parameter for identifying unfit patients at risk for adverse outcomes. High specificity is required to limit the number of fit patients who undergo unnecessary GA.[9] However, we showed that these parameters do not necessarily translate into better decisions, thus reinforcing the argument that accuracy measures play a limited role in test implementation in clinical practice.

In our study, the prevalence of unfit patients with at least one impaired GA domain was 87%. Other studies have reported similar results, with a median unfit patient prevalence (IQR) of 88% (87-94).[20,21,25,32,33] When we applied an alternative definition (on the basis of the prescription of geriatric interventions), the prevalence was 75%. We did not observe an NB for screening in these settings relative to the GA. When we applied a cutoff of two impaired GA domains, the prevalence of unfit patients was lower (65%). Several studies have explored this cutoff for defining unfit

patients.[25,26,28] Russo et al[29] used a cutoff of ≥ 3 impaired domains in GA to define an unfit patient when evaluating the SOAP2. In settings using higher cutoffs, we observed NB of screening over GA. However, these results can be explained by the decrease in prevalence and increased NPV produced by increasing the number of impaired domains defining an unfit patient. Although this approach would allow for the identification of the frailest patients, this strategy would rule out GA in many unfit patients who might benefit from targeted interventions.

This study has some limitations. First, we only analyzed patients with complete data sets. However, our study's unfit patient prevalence and diagnostic accuracy results are consistent with those reported in the literature. Second, our population was relatively old (median age: 81 years), so collecting information on younger patients might have been more valuable. However, our sensitivity analyses produced similar results for the various age groups. Third, our data came solely from centers with expertise in geriatric oncology. It would have been interesting to analyze data from primary or secondary care centers, where the prevalence of unfit patients would be lower. However, studies with a lower prevalence of unfit patients have reported similar results.[25,27,28] In addition, the prevalence of unfit adults age ≥ 70 years in the general population is estimated to be 71%.[34] Fourth, the choice of evaluation tools, the number of GA domains to be assessed, and the definition of an unfit patient are still subject to debate. Thus, our results apply to screening tools with the same cutoffs and domains as those studied here. Finally, there is no consensus regarding the threshold probabilities to use. However, we used reasonable threshold probabilities (under 0.5) because those above 0.5 mean that exposing a fit patient to an unnecessary GA is worse than missing an unfit patient.

Our study had some strengths. In particular, we used DCA to directly assess the clinical value of frailty screening using robust, externally validated tools and four of the most frequent types of cancer in older patients.

Our findings suggest that performing GA in all patients (irrespective of frailty screening results) is the most reasonable approach in populations with a high prevalence of unfit individuals. Performing GA on all patients would lead to better decisions and provide the highest benefits if there was a high degree of concern about unfit patients. In addition, performing GA in at least 10 patients would make it worthwhile compared with screening, given its additional costs. However, this tradeoff could be considered high, especially in centers that cannot perform GA. Hence, a time- and resource-saving strategy such as screening is desirable. Nevertheless, direct diagnostic procedures are recommended in high-prevalence settings, as appears to be the case for unfit older patients with cancer.[35]

Although the NBs of GA are higher than those of screening, unfit patients identified by screening would also benefit from GA, which is better than usual care according to two recent randomized controlled trials.[5,6] Additionally, screening can (1) create a framework for the evaluation of older patients with cancer in a busy practice, (2) raise awareness of the importance of GA among physicians, and (3) stimulate discussions about the choice of cancer treatments.[36]

Although accuracy measures do not provide insight into implementing tests in clinical practice or making better decisions, most research papers regarding diagnostic tests or predictive models published in the *Journal of Clinical Oncology*, and other similar journals, report only these performance measures without evaluating clinical utility. In geriatric oncology, for instance, several frailty classifications[37] and predictive tools for chemotherapy toxicity exist[38,39]; some are recommended in practice guidelines,[7,13] whereas others are used in practice. In addition to calibration and discrimination, no study evaluating these tools assessed their clinical usefulness. Moreover, none of the 17 studies included in the most recent systematic review of frailty screening tools[10] evaluated the usefulness of these tools. Therefore, it is hard to know whether using those tests or models is of clinical value or justifies its implementation in clinical practice. As suggested by previous editorials in *Journal of Clinical Oncology*,[40,41] researchers should evaluate the clinical value (eg, using NB approaches) when developing instruments for decision making.

Furthermore, the number, domains, and instruments assessed in a GA differ in their importance for cancer outcomes (eg, toxicity, quality of life, or survival).[42] Therefore, a more comprehensive definition of an abnormal GA is required—perhaps by assigning different weights to various impaired dimensions. In addition, screening tools do not capture all GA domains. Hence, improvements in the identification of impairments in missed domains are desirable. However, when developing decision-making tools, researchers must consider the trade between gaining information and making better decisions, as a good decision tool leads to the best decisions with as little information as possible.

In conclusion, despite the good diagnostic accuracy of screening tools, frailty screening did not provide a benefit over a GA-for-all approach for detecting unfit patients or avoiding GA in fit patients. Optimizing current frailty assessment strategies and evaluating their usefulness with NB approaches is necessary to improve decision making for older patients with cancer in busy practices with limited human resources.

AFFILIATIONS

[1]Inserm, IMRB, Université Paris-Est-Créteil, Créteil, France
[2]Department of Internal Medicine and Geriatrics, Henri Mondor Hospital, AP-HP, Creteil, France
[3]ERUDITE Research Unit, Université Paris-Est-Créteil, Créteil, France
[4]Department of Public Health, Henri Mondor Hospital, AP-HP, Creteil, France
[5]Department of Medical Oncology, Cochin Hospital, AP-HP, Paris, France
[6]Department of Gastroenterology, Saint Louis Hospital, AP-HP, Paris, France
[7]Department of Supportive Care and Geriatric Oncology, Institut Curie, Paris, France
[8]Department of Medical Oncology, Bergonie Institute Comprehensive Cancer Center, Bordeaux, France
[9]Inserm, Bordeaux Population Health Research Center, Epicene Team, UMR 1219, Université de Bordeaux, Bordeaux, France
[10]Inserm CIC1401, Clinical and Epidemiological Research Unit, Bergonié Institute Comprehensive Cancer Center, Bordeaux, France
[11]Department of Geriatrics, Georges Pompidou European Hospital, AP-HP, Paris, France
[12]Paris Cancer Research for Personalized Medicine Institute, Paris, France

CORRESPONDING AUTHOR

Adolfo González Serrano, MD, MSc, Mondor Biomedical Research Institute, Faculty of Health, Université Paris-Est-Créteil, 8 rue du Général Sarrail, 94010 Créteil, France; Twitter: @AGonzalezUro; e-mail: adolfo.gonzalez-serrano@u-pec.fr.

DISCLAIMER

The funders had no role in the study's design, the collection, analysis, and interpretation of data, the writing of the manuscript, or the decision to publish the results.

PRIOR PRESENTATION

Presented orally at the ASCO 2022 annual meeting, Chicago, IL, June 5, 2022.

SUPPORT

Supported by a grant (RINC4) from the French National Cancer Institute (Institut National du Cancer, INCa), Canceropôle Ile-de-France, and Gerontopôle Ile-de-France (Gérond'if). A.G.S. was funded by the EUR-LIVE Graduate School of Research "Life Trajectories and Health Vulnerability." The EUR-LIVE project is funded through the French government's *Investissements d'Avenir*/ANR program (18-EUR-0011).

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Disclosures provided by the authors are available with this article at DOI https://doi.org/10.1200/JCO.22.01118.

DATA SHARING STATEMENT

Restrictions apply to the availability of these data. Data were obtained from the ELCAPA Study Group and are available from the corresponding author with the permission of the ELCAPA Study Group investigators.

AUTHOR CONTRIBUTIONS

Conception and design: Adolfo González Serrano, Thomas Barnay, Florence Canouï-Poitrine
Provision of study materials or patients: Marie Laurent, Pascaline Boudou-Rouquette, Thomas Aparicio, Philippe Caillet, Elena Paillaud, Florence Canouï-Poitrine
Collection and assembly of data: Marie Laurent, Thomas Aparicio, Philippe Caillet, Elena Paillaud, Florence Canouï-Poitrine
Data analysis and interpretation: Adolfo González Serrano, Marie Laurent, Thomas Barnay, Florence Rollot-Trad, Claudia Martínez-Tapia, Etienne Audureau, Boudou-Rouquette, Pierre Soubeyran, Carine Bellera, Philippe Caillet, Elena Paillaud, Florence Canouï-Poitrine
Manuscript writing: All authors
Final approval of manuscript: All authors
Accountable for all aspects of the work: All authors

ACKNOWLEDGMENT

The ELCAPA Study Group consists of geriatricians (Amelie Aregui, Melany Baronn, Mickaël Bringuier, Eric Bouvard, Philippe Caillet, Gaelle Cosqueric, Lola Corsin, Tristan Cudennec, Anne Chahwakilian, Amina Djender, Eric Dupuydupin, Nargess Ebadi, Virginie Fossey-Diaz, Mathilde Gisselbrecht, Charlotte Goldstein, Béatrice Gonzalez, Marie Laurent, Julien Leguen, Madeleine Lefevre, Celine Lazarovici-Nagera, Emmanuelle Lorisson, Josephine Massias, Soraya Mebarki, Galdric Orvoen, Frédéric Pamoukdjian, Anne-Laure Scain, Godelieve Rochette de Lempdes, Florence Rollot-Trad, Gwenaëlle Varnier, Hélène Vincent, Elena Paillaud, Agathe Raynaud-Simon), oncologists (Pascaline Boudou-Rouquette, Etienne Brain, Stéphane Culine, Maxime Frelaut, Djamel Ghebriou, Joseph Gligorov, Stéphane Henault Daniel Lopez-Trabada-Ataz, Olivier Mir, Christophe Tournigand), a digestive oncologist (Thomas Aparicio), a gynecologic oncologist (Cyril Touboul), a radiation oncologist (Jean-Léon Lagrange), nurses (Stephanie Benyahia, Sadia Bonhomme, Alzira Mota, Gwadlys Philocles, Corinne Ouakinine), epidemiologists (Etienne Audureau, Sylvie Bastuji-Garin and Florence Canouï-Poitrine), a medical biologist (Marie-Anne Loriot), a pharmacist (Pierre-André Natella), a biostatistician (Claudia Martinez-Tapia), a clinical research medical doctor (Nicoleta Reinald), clinical research nurses (Sandrine Rello, Melanie Lafage), data managers (Mylène Allain, Clélia Chambraud), and clinical research assistants (Aurélie Baudin, Margot Bobin, Johanna Canovas, Sabrina Chaoui, Lina Iratni, Sonia Garrigou, Sandrine Lacour, Helène Mabungu, Laure Morisset, Besma Saadaoui).

REFERENCES

1. Wildiers H, Heeren P, Puts M, et al: International Society of Geriatric Oncology consensus on geriatric assessment in older patients with cancer. J Clin Oncol 32: 2595-2603, 2014
2. Caillet P, Canoui-Poitrine F, Vouriot J, et al: Comprehensive geriatric assessment in the decision-making process in elderly patients with cancer: ELCAPA study. J Clin Oncol 29:3636-3642, 2011
3. Caillet P, Laurent M, Bastuji-Garin S, et al: Optimal management of elderly cancer patients: Usefulness of the comprehensive geriatric assessment. Clin Interv Aging 9:1645-1660, 2014
4. Corre R, Greillier L, Le Caër H, et al: Use of a comprehensive geriatric assessment for the management of elderly patients with advanced non–small-cell lung cancer: The phase III randomized ESOGIA-GFPC-GECP 08-02 study. J Clin Oncol 34:1476-1483, 2016
5. Li D, Sun CL, Kim H, et al: Geriatric Assessment-Driven Intervention (GAIN) on chemotherapy-related toxic effects in older adults with cancer: A randomized clinical trial. JAMA Oncol 7:e214158, 2021

6. Mohile SG, Mohamed MR, Xu H, et al: Evaluation of geriatric assessment and management on the toxic effects of cancer treatment (GAP70+): A cluster-randomised study. Lancet 398:1894-1904, 2021
7. Mohile SG, Dale W, Somerfield MR, et al: Practical assessment and management of vulnerabilities in older patients receiving chemotherapy: ASCO guideline for geriatric oncology. J Clin Oncol 36:2326-2347, 2018
8. Dotan E, Walter LC, Browner IS, et al: NCCN Clinical Practice Guidelines in Oncology: Older Adult Oncology. Version 1.2021. https://www.nccn.org/professionals/physician_gls/pdf/senior.pdf
9. Decoster L, Van Puyvelde K, Mohile S, et al: Screening tools for multidimensional health problems warranting a geriatric assessment in older cancer patients: An update on SIOG recommendations. Ann Oncol 26:288-300, 2015
10. Garcia MV, Agar MR, Soo WK, et al: Screening tools for identifying older adults with cancer who may benefit from a geriatric assessment: A systematic review. JAMA Oncol 7:616-627, 2021
11. Pallis AG, Gridelli C, Wedding U, et al: Management of elderly patients with NSCLC; updated expert's opinion paper: EORTC Elderly Task Force, Lung Cancer Group and International Society for Geriatric Oncology. Ann Oncol 25:1270-1283, 2014
12. Biganzoli L, Wildiers H, Oakman C, et al: Management of elderly patients with breast cancer: Updated recommendations of the International Society of Geriatric Oncology (SIOG) and European Society of Breast Cancer Specialists (EUSOMA). Lancet Oncol 13:e148-e160, 2012
13. Mottet N, van den Bergh RCN, Briers E, et al: EAU-EANM-ESTRO-ESUR-SIOG guidelines on prostate cancer—2020 update. Part 1: Screening, diagnosis, and local treatment with curative intent. Eur Urol 79:243-262, 2021
14. Montroni I, Ugolini G, Saur NM, et al: Personalized management of elderly patients with rectal cancer: Expert recommendations of the European Society of Surgical Oncology, European Society of Coloproctology, International Society of Geriatric Oncology, and American College of Surgeons Commission on Cancer. Eur J Surg Oncol 44:1685-1702, 2018
15. Pallis AG, Fortpied C, Wedding U, et al: EORTC elderly task force position paper: Approach to the older cancer patient. Eur J Cancer 46:1502-1513, 2010
16. Vickers AJ, Elkin EB: Decision curve analysis: A novel method for evaluating prediction models. Med Decis Making 26:565-574, 2006
17. Moons KGM, Harrell FE: Sensitivity and specificity should be de-emphasized in diagnostic accuracy studies. Acad Radiol 10:670-672, 2003
18. Vickers AJ, Calster BV, Steyerberg EW: Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. BMJ 352:i6, 2016
19. Fitzgerald M, Saville BR, Lewis RJ: Decision curve analysis. JAMA 313:409-410, 2015
20. Bellera CA, Rainfray M, Mathoulin-Pélissier S, et al: Screening older cancer patients: First evaluation of the G-8 geriatric screening tool. Ann Oncol 23:2166-2172, 2012
21. Martinez-Tapia C, Canoui-Poitrine F, Bastuji-Garin S, et al: Optimizing the G8 screening tool for older patients with cancer: Diagnostic performance and validation of a six-item version. Oncologist 21:188-195, 2016
22. Droz JP, Balducci L, Bolla M, et al: Management of prostate cancer in older men: Recommendations of a working group of the International Society of Geriatric Oncology. BJU Int 106:462-469, 2010
23. Van Calster B, Wynants L, Verbeek JFM, et al: Reporting and interpreting decision curve analysis: A guide for investigators. Eur Urol 74:796-804, 2018
24. Martinez-Tapia C, Laurent M, Paillaud E, et al: Predicting frailty and geriatric interventions in older cancer patients: Performance of two screening tools for seven frailty definitions—ELCAPA cohort. Cancers (Basel) 14:244, 2022
25. Soubeyran P, Bellera C, Goyard J, et al: Screening for vulnerability in older cancer patients: The ONCODAGE prospective multicenter cohort study. PLoS One 9: e115060, 2014
26. Luciani A, Ascione G, Bertuzzi C, et al: Detecting disabilities in older patients with cancer: Comparison between comprehensive geriatric assessment and vulnerable elders survey-13. J Clin Oncol 28:2046-2050, 2010
27. Biganzoli L, Boni L, Becheri D, et al: Evaluation of the cardiovascular health study (CHS) instrument and the Vulnerable Elders Survey-13 (VES-13) in elderly cancer patients. Are we still missing the right screening tool? Ann Oncol 24:494-500, 2013
28. Kenis C, Decoster L, Van Puyvelde K, et al: Performance of two geriatric screening tools in older patients with cancer. J Clin Oncol 32:19-26, 2014
29. Russo C, Giannotti C, Signori A, et al: Predictive values of two frailty screening tools in older patients with solid cancer: A comparison of SAOP2 and G8. Oncotarget 9:35056-35068, 2018
30. Bellera CA, Artaud F, Rainfray M, et al: Modeling individual and relative accuracy of screening tools in geriatric oncology. Ann Oncol 28:1152-1157, 2017
31. Liuu E, Canouï-Poitrine F, Tournigand C, et al: Accuracy of the G-8 geriatric-oncology screening tool for identifying vulnerable elderly patients with cancer according to tumour site: The ELCAPA-02 study. J Geriatr Oncol 5:11-19, 2014
32. Pamoukdjian F, Canoui-Poitrine F, Longelin-Lombard C, et al: Diagnostic performance of gait speed, G8 and G8 modified indices to screen for vulnerability in older cancer patients: The prospective PF-EC cohort study. Oncotarget 8:50393-50402, 2017
33. Velghe A, Petrovic M, De Buyser S, et al: Validation of the G8 screening tool in older patients with aggressive haematological malignancies. Eur J Oncol Nurs 18:645-648, 2014
34. O'Caoimh R, Sezgin D, O'Donovan MR, et al: Prevalence of frailty in 62 countries across the world: A systematic review and meta-analysis of population-level studies. Age Ageing 50:96-104, 2021
35. Wilson JMG, Jungner G: Principles and practice of screening for disease. Public Health Papers 34, 1968. https://apps.who.int/iris/bitstream/handle/10665/37650/WHO_PHP_34.pdf
36. Loh KP, Soto-Perez-de-Celis E, Hsu T, et al: What every oncologist should know about geriatric assessment for older patients with cancer: Young International Society of Geriatric Oncology position paper. JCO Oncol Pract 14:85-94, 2018
37. Ferrat E, Paillaud E, Caillet P, et al: Performance of four frailty classifications in older patients with cancer: Prospective elderly cancer patients cohort study. J Clin Oncol 35:766-777, 2017
38. Hurria A, Togawa K, Mohile SG, et al: Predicting chemotherapy toxicity in older adults with cancer: A prospective multicenter study. J Clin Oncol 29:3457-3465, 2011
39. Hurria A, Mohile S, Gajra A, et al: Validation of a prediction tool for chemotherapy toxicity in older adults with cancer. J Clin Oncol 34:2366-2371, 2016
40. Kerr KF, Brown MD, Zhu K, Janes H: Assessing the clinical impact of risk prediction models with decision curves: Guidance for correct interpretation and appropriate use. J Clin Oncol 34:2534-2540, 2016
41. Vickers AJ: Prediction models: Revolutionary in principle, but do they do more good than harm? J Clin Oncol 29:2951-2952, 2011
42. Canouï-Poitrine F, Martinez-Tapia C, Paillaud E, et al: Geriatric impairments were directly and indirectly associated with mortality in older patients with cancer: A structural equation analysis. J Clin Epidemiol 148:17-26, 2022

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

A Two-Step Frailty Assessment Strategy in Older Patients With Solid Tumors: A Decision Curve Analysis

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/jco/authors/author-center.

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians (Open Payments).

Pascaline Boudou-Rouquette
Honoraria: Ipsen
Consulting or Advisory Role: Takeda
Travel, Accommodations, Expenses: Takeda, PharmaMar

Thomas Aparicio
Honoraria: Servier, AstraZeneca, Pierre Fabre, Amgen
Consulting or Advisory Role: MSD Oncology, Servier, Pierre Fabre, Sirtex Medical
Travel, Accommodations, Expenses: MSD Oncology

Pierre Soubeyran
Honoraria: GlaxoSmithKline
Consulting or Advisory Role: Bristol Myers Squibb, Eisai
Research Funding: Roche
Travel, Accommodations, Expenses: Hospira, Teva, Celgene, AstraZeneca

Carine Bellera
Consulting or Advisory Role: BMS

Philippe Caillet
Honoraria: Pfizer
Consulting or Advisory Role: Pfizer

Elena Paillaud
Honoraria: GlaxoSmithKline, Pfizer, MSD, Novartis, LEO Pharma
Research Funding: Pfizer

No other potential conflicts of interest were reported.

APPENDIX

TABLE A1. Demographic and Clinical Characteristics of the Study Participants

Characteristic	(n = 1,648)
Patients with full data sets, No.	1,648
Age, years, median (IQR)	81 (77-85)
Sex, No. (%)	
Female	1,067 (65)
Male	581 (35)
GA tools	
ADL score, points, median (IQR)	6 (6-6)
ADL score, No. (%)	
Independent	1,634 (99)
Not independent (one or more abnormal items)	11 (1)
Missing data	3 (< 1)
IADL score, points, median (IQR)	7 (4-8)
IADL score, No. (%)	
Independent	728 (44)
Not independent (one or more abnormal items)	873 (53)
Missing data	47 (3)
Cumulative Illness Rating Scale-Geriatric grade 3 or 4 comorbidities, No. (%)	
No	697 (42)
One or more comorbidities	949 (58)
Missing data	2 (< 1)
Mini Mental State Examination, No. (%)	
Normal	1,071 (65)
23 points or less	304 (18)
Missing data	273 (17)
Mini Geriatric Depression Scale, No. (%)	
Normal	1,068 (65)
One or more abnormal items	505 (31)
Missing data	75 (5)
Timed Up and Go test, No. (%)	
20 seconds or less	1,117 (68)
More than 20 seconds	368 (22)
Missing data	163 (10)
Mini Nutritional Assessment, points, median (IQR)	24 (21-26)
Mini Nutritional Assessment, No. (%)	
More than 23.5 points	838 (51)
23.5 points or less	752 (46)
Missing data	58 (4)

(continued in next column)

TABLE A1. Demographic and Clinical Characteristics of the Study Participants (continued)

Characteristic	(n = 1,648)
GA, No. (%)	
Normal	220 (13)
At least 1 abnormal domain	1,428 (87)
Impaired domains, No. (%)	
0	210 (13)
1	371 (23)
2	350 (21)
3 or more	717 (43)
G8 score items	
Neuropsychologic problems, No. (%)	
Severe dementia or depression	123 (7)
Mild dementia	518 (31)
No neuropsychologic problems	1,007 (61)
More than three prescription drugs per day, No. (%)	
Yes	1,174 (71)
No	474 (29)
Health status compared with other people of the same age, No. (%)	
Not as good	185 (11)
As good	487 (30)
Better	619 (38)
Does not know	357 (22)
Decrease in food intake, No. (%)	
Severe	144 (9)
Moderate	441 (27)
No decrease	1,063 (65)
Weight loss, No. (%)	
More than 3 kg	342 (21)
Not known	143 (9)
Between 1 and 3 kg	284 (17)
No weight loss	879 (53)
Mobility, No. (%)	
Bed- or chair-bound	101 (6)
Able to get out of bed/chair but does not go out	345 (21)
Goes out	1,202 (73)
Body mass index, No. (%)	
Less than 19 kg/m^2	103 (6)
Between 19 and 21 kg/m^2	160 (10)
Not between 21-23 kg/m^2	274 (17)

(continued on following page)

TABLE A1. Demographic and Clinical Characteristics of the Study Participants (continued)

Characteristic	(n = 1,648)
More than 23 kg/m^2	1,111 (67)
Age, years, No. (%)	
< 80	636 (39)
80-85	620 (38)
> 86	392 (24)
G8 score, points, median (IQR)	12 (10-14)
G8 score, No. (%)	
More than 14 points	340 (21)
14 points or less	1,308 (79)
Modified G8 score items	
Neuropsychologic problem, No. (%)	
No neuropsychologic problem	1,007 (61)
Dementia or depression	641 (39)
More than six prescription drugs per day, No. (%)	
No	732 (44)
Yes	916 (56)
Health status compared with other people of the same age, No. (%)	
Does not know/as good/better	1,463 (89)
Not as good	185 (11)
Eastern Cooperative Oncology Group performance status, No. (%)	
0	358 (22)
1	636 (39)
2-4	654 (40)
Heart failure or coronary disease, No. (%)	
No	942 (57)
Yes	706 (43)
Weight loss, No. (%)	
No weight loss	879 (53)
Between 1 and 3 kg	284 (17)
More than 3 kg or unknown	485 (29)
Modified G8 score, points, median (IQR)	14 (6-22)
Modified G8 score, No. (%)	
< 6 points	339 (21)
6 points or more	1,309 (79)
Tumor-related characteristics	
Tumor type, No. (%)	
Prostate	254 (15)
Breast	849 (52)
Colorectal	362 (22)
Lung	183 (11)
Node involvement, No. (%)	
N0	804 (49)
N1	562 (34)
Nx	282 (17)
Clinical stage, No. (%)	
Nonmetastatic	920 (56)
Metastatic	559 (34)
Mx	169 (10)

(continued in next column)

TABLE A1. Demographic and Clinical Characteristics of the Study Participants (continued)

Abbreviations: ADL, activities of daily living; G8, G8 Geriatric Screening Tool; GA, geriatric assessment; IADL, instrumental activities of daily living; IQR, interquartile range; M, metastasis; N, node.

TABLE A2. Diagnostic Accuracy and NB for Detecting Unfit Patients and Avoiding Unnecessary GA in Fit Patients for the G8 and the Modified G8 Scores

Result	Prostate		Breast		Colorectal		Lung	
No. with full data sets (%)	254 (15)		849 (52)		362 (22)		183 (11)	
Prevalence of an abnormal GA, No. (%)	213 (84)		684 (81)		352 (97)		179 (98)	
Result	**G8 Score**	**Modified G8 Score**	**G8 Score**	**Modified G8 Score**	**G8 Score**	**Modified G8 Score**	**G8 Score**	**Modified G8 Score**
TP, No.	161	188	586	555	312	312	159	167
False positives, No.	12	12	68	67	6	6	4	2
Sensitivity (95% CI)	76 (70 to 81)	88 (84 to 92)	86 (83 to 88)	81 (79 to 84)	89 (85 to 92)	89 (85 to 92)	89 (84 to 93)	93 (90 to 97)
Specificity (95% CI)	71 (65 to 76)	71 (65 to 76)	59 (55 to 62)	59 (56 to 63)	40 (35 to 45)	40 (35 to 45)	0 (0 to 0)	50 (43 to 57)
PPV (95% CI)	93 (90 to 96)	94 (91 to 97)	90 (88 to 92)	89 (87 to 91)	98 (97 to 100)	98 (97 to 100)	98 (95 to 100)	99 (97 to 100)
NPV (95% CI)	36 (30 to 42)	54 (48 to 60)	50 (46 to 53)	43 (40 to 47)	9 (6 to 12)	9 (6 to 12)	0 (0 to 0)	14 (9 to 19)
Probability threshold of 0.25 (odds 1:3)								
NB of screening	0.62	0.72	0.66	0.63	0.86	0.86	0.86	0.91
NB of GA	0.79	0.79	0.74	0.74	0.96	0.96	0.97	0.97
NB difference[a]	–0.17	–0.06	–0.08	–0.11	–0.11	–0.11	–0.11	–0.06
Avoided unnecessary GA (No.)[b,c]	–51	–18	–24	–35	–32	–32	–33	–19
Tradeoff (No.)	6	16	13	9	9	9	9	16
Probability threshold of 0.33 (odds 1:2)								
NB of screening	0.61	0.72	0.65	0.61	0.85	0.85	0.86	0.91
NB of GA	0.76	0.76	0.71	0.71	0.96	0.96	0.97	0.97
NB difference[a]	–0.15	–0.04	–0.06	–0.09	–0.10	–0.10	–0.11	–0.06
Avoided unnecessary GA (No.)[b,c]	–30	–8	–12	–19	–21	–21	–22	–12
Tradeoff (No.)	7	24	17	11	10	10	9	17
Probability threshold of 0.5 (odds 1:1)								
NB of screening	0.59	0.69	0.61	0.57	0.85	0.85	0.85	0.90
NB of GA	0.68	0.68	0.61	0.61	0.94	0.94	0.96	0.96
NB difference[a]	–0.09	0.02	0.00	–0.04	–0.10	–0.10	–0.11	–0.05
Avoided unnecessary GA (No.)[b,c]	–9	2	0	–4	–10	–10	–11	–5
Tradeoff (No.)	11	[d]	849	27	10	10	9	18

NOTE. The data are presented by tumor type. The odds of 1:3 and 1:2 mean that missing an unfit patient is, respectively, three and two times as bad as exposing a fit patient to an unnecessary GA. The odds of 1:1 mean missing an unfit patient is the same as exposing a fit patient to an unnecessary GA.

Abbreviations: G8, G8 Geriatric Screening Tool; GA, geriatric assessment; NB, net benefit; NPV, negative predictive value; PPV, positive predictive value; TP, true positives.

[a]Negative NB differences mean that screening has no benefit over a GA-for-all approach.

[b]Number per 100 patients.

[c]Negative values represent the number of unfit patients lost if screening is used to decide who should undergo GA.

[d]The tradeoff is not reported as the NB difference favors screening.

TABLE A3. Diagnostic Accuracy and NB for Detecting Unfit Patients and Avoiding Unnecessary GA in Fit Patients for the G8 and the Modified G8 Scores

	Clinical Stage		Age Groups, years		
Result	**M0 and Mx**	**Metastatic**	**< 80**	**80-90**	**> 90**
No. with full data sets (%)	1,089 (66)	559 (34)	732 (44)	819 (50)	97 (6)
Prevalence of an abnormal GA, No. (%)	900 (83)	528 (94)	599 (82)	735 (90)	94 (97)

Result	**G8 Score**	**Modified G8 Score**	**G8 Score**	**Modified G8 Score**	**G8 Score**	**Modified G8 Score**	**G8 Score**	**Modified G8 Score**	**G8 Score**	**Modified G8 Score**
TP, No.	763	753	455	469	445	486	681	649	92	87
False positives, No.	73	74	17	13	32	43	55	41	3	3
Sensitivity (95% CI)	85 (83 to 87)	84 (81 to 86)	86 (83 to 89)	89 (86 to 91)	74 (71 to 77)	81 (78 to 84)	93 (91 to 94)	88 (86 to 91)	98 (95 to 100)	93 (87 to 98)
Specificity (95% CI)	61 (58 to 64)	61 (58 to 63)	45 (41 to 49)	58 (54 to 62)	76 (73 to 79)	68 (64 to 71)	35 (31 to 38)	51 (48 to 55)	0 (0 to 0)	0 (0 to 0)
PPV (95% CI)	91 (90 to 93)	91 (89 to 93)	96 (95 to 98)	97 (96 to 99)	93 (91 to 95)	92 (90 to 94)	93 (91 to 94)	94 (92 to 96)	97 (93 to 100)	97 (93 to 100)
NPV (95% CI)	46 (43 to 49)	44 (41 to 47)	16 (13 to 19)	24 (20 to 27)	40 (36 to 43)	44 (41 to 48)	35 (32 to 38)	33 (30 to 37)	0 (0 to 0)	0 (0 to 0)
Probability threshold of 0.25 (odds 1:3)										
NB of screening	0.68	0.67	0.80	0.83	0.59	0.64	0.81	0.78	0.94	0.89
NB of GA	0.77	0.77	0.93	0.93	0.76	0.76	0.86	0.86	0.96	0.96
NB difference[a]	–0.09	–0.10	–0.12	–0.09	–0.16	–0.11	–0.05	–0.09	–0.02	–0.07
Avoided unnecessary GA (No.)[b,c]	–27	–30	–36	–29	–50	–34	–15	–24	–6	–22
Tradeoff (No.)	11	10	8	11	6	8	20	13	50	14
Probability threshold of 0.33 (odds 1:2)										
NB of screening	0.67	0.66	0.80	0.83	0.59	0.63	0.80	0.76	0.93	0.88
NB of GA	0.74	0.74	0.92	0.92	0.73	0.73	0.85	0.85	0.95	0.95
NB difference[a]	–0.07	–0.08	–0.12	–0.09	–0.14	–0.09	–0.05	–0.09	–0.02	–0.07
Avoided unnecessary GA (No.)[b,c]	–14	–16	–24	–18	–28	–19	–10	–18	–4	–14
Tradeoff (No.)	14	12	9	11	7	11	21	12	48	14
Probability threshold of 0.5 (odds 1:1)										
NB of screening	0.63	0.62	0.78	0.82	0.56	0.61	0.77	0.74	0.92	0.87
NB of GA	0.65	0.65	0.89	0.89	0.64	0.64	0.79	0.79	0.94	0.94
NB difference[a]	–0.02	–0.03	–0.11	–0.07	–0.07	–0.03	–0.02	–0.05	–0.02	–0.07
Avoided unnecessary GA (No.)[b,c]	–2	–3	–11	–7	–7	–3	–2	–5	–2	–7
Tradeoff (No.)	52	34	10	14	14	32	44	19	49	14

NOTE. The data are presented by clinical stage and age group. The odds of 1:3 and 1:2 mean that missing an unfit patient is, respectively, three and two times as bad as exposing a fit patient to an unnecessary GA. The odds of 1:1 mean missing an unfit patient is the same as exposing a fit patient to an unnecessary GA.

Abbreviations: G8, G8 Geriatric Screening Tool; GA, geriatric assessment; M0, nonmetastatic disease; M1, metastatic disease; Mx, no assessment of metastatic disease; NB, net benefit; NPV, negative predictive value; PPV, positive predictive value; TP, true positives.

[a]Negative NB differences mean that screening has no benefit over a GA-for-all approach.

[b]Number per 100 patients.

[c]Negative values represent the number of unfit patients lost if screening is used to decide who should undergo GA.

TABLE A4. Demographic and Clinical Characteristics of TP and False-Negative Patients for the G8 and the Modified G8 Scores

Characteristic	G8 Score				Modified G8			
	TP	FN	OR[a] (95% CI)	*P*[b]	TP	FN	OR[a] (95% CI)	*P*[b]
No. with full data set	1,218	210			1,222	206		
Age, years, No. (%)								
< 80	445 (37)	154 (73)	1	< .001	486 (40)	113 (55)	1	< .001
81-90	681 (56)	54 (26)	4.36 (3.13 to 6.08)		649 (53)	86 (42)	1.75 (1.29 to 2.38)	
> 90	92 (8)	2 (1)	15.92 (3.88 to 65.39)		87 (7)	7 (3)	2.89 (1.30 to 6.41)	
Sex of participants, No. (%)								
Male	435 (36)	96 (46)	1	.006	476 (39)	55 (27)	1	< .001
Female	783 (64)	114 (54)	1.52 (1.13 to 2.04)		746 (61)	151 (73)	0.57 (0.41 to 0.79)	
ADL scores, No. (%)								
Independent	1,075(88)	207(99)	1	< .001	1,080 (89)	202 (98)	1	< .001
Not independent (one or more abnormal items)	140(12)	3(1)	8.99 (2.84 to 28.47)		139 (11)	4 (2)	6.50 (2.38 to 17.76)	
IADL scores, No. (%)								
Independent	397 (33)	123 (63)	1	< .001	382 (32)	138 (72)	1	< .001
Not independent (one or more abnormal items)	790 (67)	73 (37)	3.35 (2.45 to 4.59)		808 (68)	55 (28)	5.31 (3.79 to 7.42)	
Cumulative Illness Rating Scale-Geriatric grade 3 or 4 comorbidities, No. (%)								
No	397 (33)	80 (38)	1	.11	373 (31)	104 (51)	1	< .001
One or more comorbidities	820 (67)	129 (62)	1.28 (0.95 to 1.74)		849 (69)	100 (49)	2.37 (1.75 to 3.20)	
Mini Mental State Examinations, No. (%)								
Normal	700 (72)	151 (82)	1	.007	719 (73)	132 (78)	1	.20
23 points or less	270 (28)	34 (18)	1.71 (1.15 to 2.55)		266 (27)	38 (22)	1.29 (0.87 to 1.89)	
Mini Geriatric Depression Scales, No. (%)								
Normal	693 (61)	155 (75)	1	< .001	714 (62)	134 (67)	1	.20
One or more abnormal items	452 (39)	53 (25)	1.91 (1.37 to 2.66)		438 (38)	67 (33)	1.23 (0.89 to 1.68)	
Mini Nutritional Assessments, No. (%)								
More than 23.5 points	418 (36)	200 (99)	1	< .001	444 (38)	174 (86)	1	< .001
23.5 or less	749 (64)	3 (1)	119.46 (37.96 to 375.88)		724 (62)	28 (14)	10.13 (6.68 to 15.36)	

(continued on following page)

TABLE A4. Demographic and Clinical Characteristics of TP and False-Negative Patients for the G8 and the Modified G8 Scores (continued)

Characteristic	G8 Score				Modified G8			
	TP	FN	OR[a] (95% CI)	*P*[b]	TP	FN	OR[a] (95% CI)	*P*[b]
Tumor types, No. (%)								
Prostate	161 (13)	52 (25)	1	< .001	188 (15)	25 (12)	1	< .001
Breast	586 (48)	98 (47)	1.93 (1.32 to 2.82)		555 (45)	129 (63)	0.57 (0.36 to 0.91)	
Colorectal	312 (26)	40 (19)	2.52 (1.60 to 3.97)		312 (26)	40 (19)	1.04 (0.61 to 1.76)	
Lung	159 (13)	20 (10)	2.57 (1.47 to 4.50)		167 (14)	12 (6)	1.85 (0.90 to 3.80)	
Nodal involvements, No. (%)								
No	557 (56)	111 (60)	1	.31	554 (56)	114 (63)	1	.089
Yes	433 (44)	73 (40)	1.18 (0.86 to 1.63)		438 (44)	68 (37)	1.33 (0.96 to 1.84)	
Clinical stages, No. (%)								
Nonmetastatic	623 (58)	125 (63)	1	.16	616 (57)	132 (69)	1	.001
Metastatic	455 (42)	73 (37)	1.25 (0.91 to 1.71)		469 (43)	59 (31)	1.70 (1.23 to 2.37)	

Abbreviations: ADL, activities of daily living; FN, false negatives; G8, G8 Geriatric Screening Tool; IADL, instrumental activities of daily living; OR, odds ratio; TP, true positives.

[a]OR for the probability of being a TP.

[b]Chi-squared test.