Journal of Geriatric Oncology xxx (2020) xxx

Contents lists available at ScienceDirect

Journal of Geriatric Oncology

Comparing the performance of the CARG and the CRASH score for predicting toxicity in older patients with cancer

Imke Ortland [a], Monique Mendel Ott [a], Michael Kowar [b], Christoph Sippel [c], Ulrich Jaehde [a,1], Andreas H. Jacobs [b,*,1], Yon-Dschun Ko [c,1]

[a] Institute of Pharmacy, Clinical Pharmacy, University of Bonn, An der Immenburg 4, 53121 Bonn, Germany
[b] Department of Geriatrics and Neurology, Johanniter Hospital Bonn, Johanniterstr. 1-3, 53113 Bonn, Germany
[c] Department of Oncology and Hematology, Johanniter Hospital Bonn, Johanniterstr. 1-3, 53113 Bonn, Germany

ARTICLE INFO

Article history:
Received 1 September 2019
Received in revised form 31 December 2019
Accepted 31 December 2019
Available online xxxx

Keywords:
CRASH
CARG
Older cancer patients
Toxicity
Onco-geriatrics
Geriatric assessment

ABSTRACT

Objectives: To compare the CARG (Cancer and Aging Research Group) and CRASH (Chemotherapy Risk Assessment Scale for High-Age Patients) score regarding the predictive performance for severe toxicity in older patients with cancer.
Methods: We recruited patients ≥70 years and applied the CARG and CRASH score before the start of systemic cancer treatment. The CARG predicts severe overall toxicity; the CRASH additionally predicts hematologic and nonhematologic toxicity. We captured ≥ grade 3 toxicity according to Common Terminology Criteria for Adverse Events (CTCAE) from medical records. Predictive performance was assessed using logistic regression and the area under the receiver operating characteristic curve (ROC-AUC).
Results: The study cohort comprised 120 patients (50% female, mean age 77.2 years, 57% solid tumors). The median of the CARG (range 0–23) and the combined CRASH (range 0–12) were 9 and 8, respectively. 81% of patients experienced toxicity; 67% showed hematologic toxicity. The predictive performance of the CARG and the combined CRASH was similar for overall toxicity (CARG: Odds ratio per unit increase (OR) 1.266, $P = .015$; ROC-AUC 0.681, $P = .010$; combined CRASH: OR 1.337, $P = .029$; ROC-AUC 0.650, $P = .032$). For hematologic toxicity, the hematologic CRASH was a significant predictor and showed numerically a higher ROC-AUC than the CARG which was not statistically different (CARG: OR 1.048, $P = .462$; ROC-AUC 0.564, $P = .271$; hematologic CRASH: OR 1.602, $P = .007$; ROC-AUC 0.665, $P = .005$).
Conclusion: Both scores exhibited similar predictive performance for toxicity in older patients with cancer.

1. Introduction

Cancer treatment for older patients is associated with a higher toxicity risk than treatment for younger patients [1] and evidence is lacking due to under-representation of older patients in clinical trials [2]. The heterogeneity of older patients' functional reserves makes therapy decision even more complex [3]. For individualizing cancer care in this population, a comprehensive geriatric assessment (CGA) is recommended [4,5]. But being very time-consuming, a CGA is rarely used in clinical routine [6]. Short tools combining geriatric assessment and oncologic parameters were developed for individualized prediction of toxicity during chemotherapy: the CARG (Cancer and Aging Research Group) and the CRASH (Chemotherapy Risk Assessment Scale for High-Age Patients) score [7,8]. Prediction of toxicity may help oncologists as well as patients in the decision-making process regarding cancer therapy [7]. The ASCO guideline for Geriatric Oncology recommends either the CARG or CRASH score for predicting toxicity [5].

The CARG score was developed by Hurria et al. in 2011 and externally validated in 2016 [7,9]. It was designed for a cohort of patients ≥65 years with solid tumors undergoing chemotherapy. The predicted outcome is therapy-associated toxicity of grade 3–5 according to Common Terminology Criteria for Adverse Events (CTCAE). It stratifies patients into three risk categories for overall toxicity: low, mid, and high. CARG offers a quick estimation of the toxicity risk (<5 min [5]).

The CRASH score was developed and internally validated by Extermann et al. in 2012 [8]. It was designed for patients ≥70 years with solid and hematologic tumors, predicting nonhematologic CTCAE grade 3–4 or hematologic grade 4 toxicity during chemotherapy. The CRASH score is divided into three subscores predicting overall toxicity (combined CRASH score), hematologic toxicity (hematologic CRASH

* Corresponding author.
E-mail address: andreas.jacobs@johanniter-kliniken.de (A.H. Jacobs).
[1] Equal contribution.

https://doi.org/10.1016/j.jgo.2019.12.016
1879-4068/

score), and nonhematologic toxicity (nonhematologic CRASH score). It assigns patients to four risk categories: low, mid-low, mid-high, and high. Compared to the CARG, the CRASH is more time-consuming (approximately 20–30 min [5]). Although both scores are recommended in literature [5], it is unclear which score should be favored. The CARG score is a short prediction tool, whereas the CRASH score could possibly give a more detailed prediction regarding different types of toxicity. In a review by Almodovar et al., a panel of six experts judged both scores as feasible tools in NSCLC treatment but considered CARG as the first option in clinical routine due to its ease of use [6]. However, no German study has compared the predictive performance of the CARG and CRASH score so far.

1.1. Aim

This is the first German study comparing the predictive performance of the CARG and CRASH score. Our aim was to compare both scores regarding (I) agreement and (II) predictive performance in order to determine which score is preferable in clinical routine. The secondary aim was to compare the score predictions with physicians' judgements.

2. Methods

2.1. Study Design and Procedures

This was a prospective, single-center observational study. A positive vote of the ethics committee of the Faculty of Medicine of Bonn University was granted and all patients signed an informed consent. The study considered the design and eligibility criteria of the development studies of the CARG and CRASH score where possible. Patients were included if they were ≥70 years, had a diagnosis of a malignancy, had German language skills and were scheduled to start first-line systemic cancer treatment. Exclusion criteria were cognitive impairment (Mini-Mental State Examination <20) or previously started cancer therapy. To enhance applicability in current daily routine, we also included patients being treated with modern agents like targeted therapy. Because our goal was to investigate scores for a broad clinical setting, we additionally considered hematological malignancies which had not been included in the development study of the CARG score and in the CRASH score only regarding some hematologic malignancies like Non-Hodgkin lymphoma. This study was designed as an exploratory trial and we believe that a formal sample size calculation would only be of limited reliability with the available data at this phase. Therefore, we decided not to perform a sample size calculation.

Patients were recruited at the inpatient wards of the Johanniter Hospital Bonn. In the routine care of the study center, all patients are discussed in a tumor board for interdisciplinary therapy decisions. However, geriatricians generally don't belong to this tumor board. For older cancer patients, no special standardized procedures like a geriatric assessment exist. The therapy decisions in older patients mainly depend on the oncologists' judgement of the abilities of the patient. An inpatient setting was chosen because we aimed at including rather frail patients where those risk assessment tools may be of higher clinical value due to more complex therapy decisions. In a previous pilot study with 20 patients, our research group verified the feasibility of both scores. Geriatric assessment items of the scores were captured by a study researcher in a patient interview. For the purpose of this study, the questions of the CARG score and the instrumental activities of daily living (IADL) were translated into German by our research group since no German version existed. Required laboratory data and additional patient characteristics were collected from medical records. The CARG score ranged from 0 to 23, the combined CRASH score from 0 to 12, the hematologic CRASH score from 0 to 6, and the nonhematologic CRASH score 0 to 8 [7,8]. Risk categories of the CARG and CRASH score were estimated using the original cut-off values from development studies [7,8]. Before starting cancer therapy, treating physicians were asked to estimate the patient's individual risk (categories: low, medium or high). Physicians were blinded to the score results and hence, were not influenced by the scores either in risk estimation or treatment decisions. Patients were followed until the end of therapy or for a maximum of six cycles. Because the planned treatment itself was also included in the score risk prediction, a follow-up was not pursued any longer if patients completely changed the planned treatment regimen or experienced dose reductions ≥50%.

2.2. Toxicity Assessment

Incidence of severe toxicity during therapy course was obtained from medical records according to Common Terminology Criteria for Adverse Events (CTCAE), version 4.03 [10]. In this study, severe toxicity was defined as CTCAE grade 3 (hospitalization indicated), grade 4 (life-threatening) or grade 5 (treatment-related death). All medical records of the patients were screened for nonhematologic and hematologic toxicity, also considering regular weekly blood controls of hematologic parameters as well as serum electrolytes and kidney and liver function tests. In order to ensure identical procedures in data collection, a standardized form with CTCAE criteria was used, and the screening was always performed by the same researcher. If patients were partly treated in surrounding oncology practices, follow-up data, including laboratory data and medical reports, were also collected there. Considering that the surrounding oncology practices generally belonged to the same regional cancer care network certified by the German cancer society, routine standards were comparable.

2.3. Statistical Analysis

According to which might be appropriate, frequencies were given as numbers and percentages, mean or median. Time to first severe toxicity was assessed. In order to investigate whether both scores predict a similar risk, the correlation and agreement were analyzed. Correlation was estimated by two-sided Spearman's rho (r_s). Chance-corrected agreement was assessed by weighted kappa [11], actual agreement by evaluating CRASH score categories per CARG score categories (exact Chi-square test). For predictive performance, calibration and discrimination were investigated [12]. The proportion of patients with severe toxicity per risk category was analyzed by exact Chi-square test. Risk categories were pooled in two risk categories (low vs high: CARG low/mid vs high; CRASH low/mid-low vs mid-high/high) to ensure a sufficient number of patients per group. Univariate logistic regression was used to assess if scores and physicians' judgments predicted toxicity significantly. To evaluate discrimination, a receiver operating characteristic curve (ROC curve) and the area under the ROC curve (ROC-AUC) were calculated [13]. The ROC-AUC values were compared using the approach of DeLong et al. [14]. Scores were treated as continuous variables for logistic regression and ROC curves in order to allow the use of the maximal information available, physicians' judgements as categorical variables. Analyses were performed using Microsoft Excel© 2007 (Microsoft Corporation, Redmond, WA, USA) and SPSS© 25.0 for Windows (IBM Corporation, Armonk, USA). A p-value of <0.05 was considered statistically significant and 95% confidence intervals were calculated.

3. Results

3.1. Patient Characteristics

Between November 2015 and August 2017, 120 patients were enrolled. Patient characteristics are presented in Table 1 and Supplement 1. The mean age of patients was 77.2 years (standard deviation (SD),

Please cite this article as: I. Ortland, M. Mendel Ott, M. Kowar, et al., Comparing the performance of the CARG and the CRASH score for predicting toxicity in older patients ..., J Geriatr Oncol, https://doi.org/10.1016/j.jgo.2019.12.016

4.5) and gender was equally distributed (female: 60, 50%). The ethnicity of patients was White/Caucasians. The patient cohort demonstrated good performance status (ECOG 0–2: 105, 87.5%) and little comorbidity (Charlson Comorbidity Index 0–2: 108, 90.0%). The most frequent tumor types were hematological (52, 43.3%) and respiratory (29, 24.2%) malignancies. The majority experienced later cancer stages (stage 3 or 4: 87, 72.5%). Most patients were treated with chemotherapy (72, 60.0%) and combinations of chemotherapy with targeted therapy (41, 34.2%). Concurrent radiotherapy was conducted in 37 (30.8%) patients.

3.2. Risk Predictions

Risk predictions of the CARG and the combined CRASH score are illustrated in Fig. 1, results of score items in Supplement 2. The CARG score shows a median of 9 (interquartile range (IQR), 4; range, 4–20). Patients were mainly classified as mid (61, 50.8%) and high (52, 43.3%) risk. Only 7 (5.8%) patients were categorized as low risk. For the combined CRASH score, median was 8 (IQR, 2; range, 2–11). Patients were mostly stratified as mid-high (72, 60.0%); 3 (2.5%) patients as low, 22 (18.3%) as mid-low, and 23 (19.2%) as high risk. For the hematologic CRASH score, median was 4 (IQR, 2; range, 0–6). The majority was classified as mid-high (71, 59.2%); 5 (4.2%) patients as low, 40 (33.3%) as mid-low, and 4 (3.3%) as high. The nonhematologic CRASH score indicates a median of 6 (IQR, 2; range, 0–8). Patients were mostly categorized as mid-high (68,56.7%); 6 (5.0%) as low, 24 (20.0%) as mid-low and 22 (18.3%) as high. For physicians' judgements, 118 estimations were available. Physicians predicted a low risk for 36 (30.5%) patients, mid for 65 (55.1%) patients and high for 17 (14.4%) patients.

3.3. Toxicity

One hundred and thirteen patients were available for outcome analysis (further site of treatment unknown, $n = 4$; follow-up data not completely accessible, $n = 3$). The median time between geriatric assessment and first therapy cycle was 1 day (range, 0–53 days) with an interquartile range of 4 days. Almost 90% of patients started the antineoplastic treatment within 1 week after the assessment. Only one patient experienced a 53 days elapse between geriatric assessment and the first therapy cycle. Due to toxicity, 31 (27.4%) patients were forced to completely stop systemic therapy earlier than planned and 11 (10%) changed their treatment to another systemic treatment during therapy course. Ninety-two (81.4%) patients experienced overall toxicity grade ≥ 3; 76 (67.3%) hematologic toxicity grade ≥ 3 and 67 (59.3%) nonhematologic toxicity grade ≥ 3. Most frequent hematologic toxicity grade ≥ 3 was leukopenia (54, 47.8%), most frequent nonhematologic toxicity were infections (37, 32.7%). The proportion of patients with toxicity ≥ grade 3 per score items is provided in Table 2; details regarding toxicity types and CTCAE grades are presented in Supplement 3. Median time to first severe toxicity was 2 weeks. Within the first therapy cycle (or at least within the first 3 weeks), 78 (69.0%) patients showed signs of overall grade ≥ 3 toxicity.

3.4. Comparison of Score Predictions

Proportions of combined CRASH per CARG score predictions, as well as correlation of the CARG with the combined, hematologic, and nonhematologic CRASH score are illustrated in Fig. 2. CARG and combined CRASH score did not suggest a relationship in exact Chi-square test ($P = .394$). The correlation was poor ($r_s = 0.203$, $P = .026$). Chance-corrected agreement was very poor, presenting a weighted kappa of 0.057 (Confidence interval (CI), −0.074-0.188; $P = .394$). These results indicate that both scores predict different risks for patients.

Table 1
Patient characteristics ($n = 120$).

Age		
Mean (SD)	77.2 (4.5)	
Min-max	70–88	
BMI		
Mean (SD)	25.4 (4.2)	
Min-max	15.8–41.4	
Creatinine Clearance (Cockcroft-Gault) [mL/min]		
Mean (SD)	65.6 (21.4)	
Min-max	10–129	
Number of drugs before start of cancer treatment (prescription and non-prescription)		
Median (IQR)	5 (4.75)	
Min-max	0–18	
	n	%
Sex		
female	60	50.0
male	60	50.0
ECOG, classified		
Fully active (0)	40	33.3
Capable of all self-care (1–2)	65	54.2
Limited or no self-care (3–4)	15	12.5
Charlson Comorbidity Index, classified		
No comorbidity (0)	51	42.5
Little comorbidity (1–2)	57	47.5
Moderate comorbidity (3–4)	11	9.2
High comorbidity (≥ 5)	1	0.8
Tumor type		
Respiratory	29	24.2
Digestive/Gastrointestinal	15	12.5
Breast	11	9.2
Unknown Primary	4	3.3
Gynecological	3	2.5
Genitourinary	3	2.5
Hematological	52	43.3
Other	3	2.4
Relapse		
No	104	86.7
Yes	16	13.3
Cancer stage		
I	7	5.8
II	10	8.3
III	29	24.2
IV	58	48.3
Not reported	16	13.3
Treatment type		
Chemotherapy	72	60.0
Targeted/Immunotherapy	7	5.8
Chemotherapy and targeted therapy combination	41	34.2
Number of treatment agents		
Monotherapy	18	15.0
Polytherapy	102	85.0
Treatment intention		
Palliative	62	51.7
Curative	48	40.0
Other	10	8.3
Planned additional therapy		
None	69	57.5
Radiotherapy	31	25.8
Surgery	14	11.7
Radiotherapy and Surgery	6	5.0
Treatment location		
Johanniter hospital	87	77.0
External oncology practices	19	16.8
Johanniter hospital and external oncology practices	7	6.2

3.5. Predictive Performance

For overall toxicity, both scores exhibited a similar predictive performance. The CARG and the combined CRASH score indicated similar ROC-AUC values; 0.681 (CI, 0.551–0.811; $P = .010$) and 0.650 (CI, 0.519–0.782; $P = .032$), respectively. The ROC-AUC values were not statistically different from each other ($P = .726$). ROC curves of both scores are displayed in Fig. 3. In univariate logistic regression, the CARG and the combined CRASH score were both significant predictors of toxicity.

Please cite this article as: I. Ortland, M. Mendel Ott, M. Kowar, et al., Comparing the performance of the CARG and the CRASH score for predicting toxicity in older patients ..., J Geriatr Oncol, https://doi.org/10.1016/j.jgo.2019.12.016

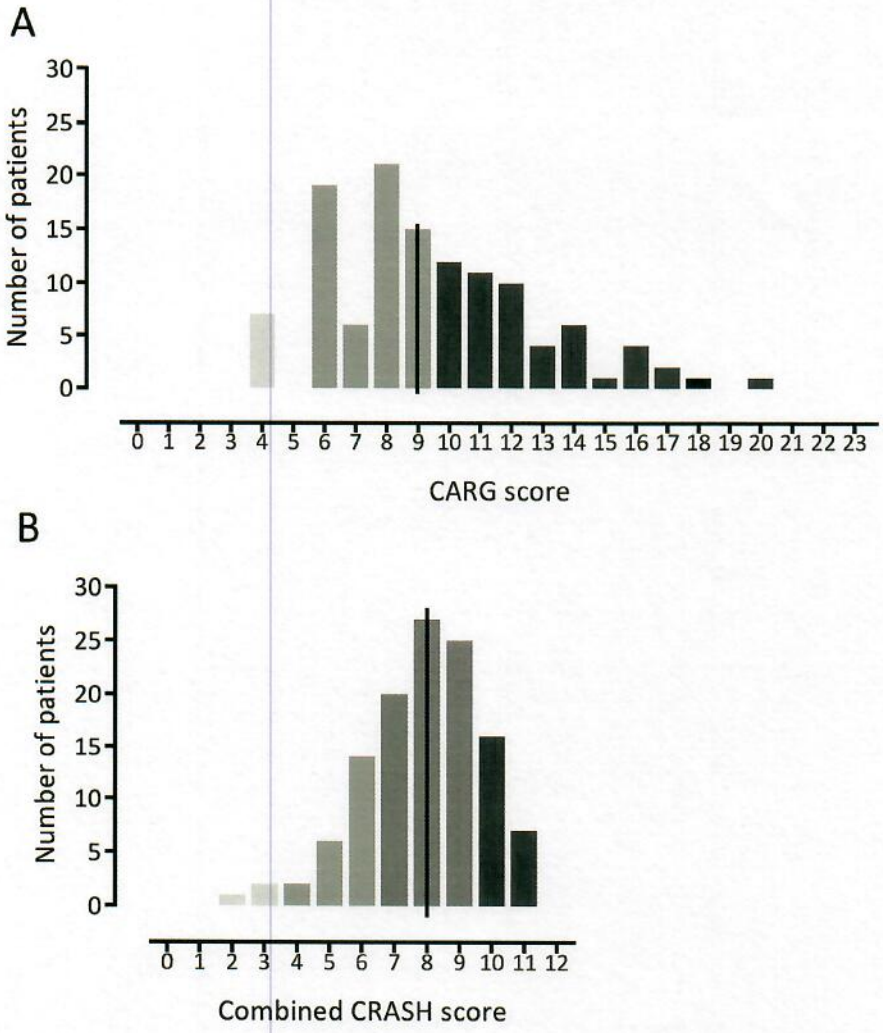

Fig. 1. Results of the CARG (range 0–23 points) (A) and combined CRASH score (range 0–12 points) (B) toxicity predictions. Solid lines show median of score results.

CARG score demonstrated an odds ratio (OR) per unit increase of 1.266 (CI, 1.048–1.530; $P = .015$), combined CRASH score an OR per unit increase of 1.337 (CI, 1.031–1.734; $P = .029$). The proportion of patients with toxicity grade ≥ 3 increased with higher risk category (CARG: $P = .051$; combined CRASH: $P = .382$). The proportion of patients with toxicity per category is presented in Fig. 4.

Regarding hematologic toxicity, the hematologic CRASH score resulted numerically in a higher ROC-AUC than the CARG score; 0.665 (CI, 0.554–0.776; $P = .005$) and 0.564 (CI, 0.445–0.683; $P = .271$), respectively (Fig. 3). However, this difference was not statistically different in the DeLong analysis ($P = .224$). In univariate logistic regression, only the hematologic CRASH score predicted toxicity significantly (hematologic CRASH: OR per unit increase, 1.602; CI, 1.135–2.261; $P = .007$; CARG: OR per unit increase, 1.048; CI, 0.925–1.186; $P = .462$). Using the hematologic CRASH score, the risk increased with a higher risk category ($P = .002$). However, using the CARG score, toxicity only increased slightly ($P = .687$, Fig. 4).

Concerning nonhematologic toxicity, the CARG and nonhematologic CRASH score presented a similar predictive performance. CARG and nonhematologic CRASH score denoted similar ROC-AUC; 0.662 (CI, 0.561–0.763; $P = .003$) and 0.651 (CI, 0.550–0.752; $P = .007$), respectively (Fig. 3). Those were not statistically different from each other ($P = .864$). In univariate logistic regression, CARG and nonhematologic CRASH score were both significant predictors of toxicity. CARG score demonstrated an OR per unit increase of 1.219 (CI, 1.062–1.398; $P = .005$), nonhematologic CRASH score an OR per unit increase of 1.363 (CI, 1.044–1.781; $P = .023$). In both scores, the risk increased with a higher risk category (CARG: $P = .007$; nonhematologic CRASH: $P = .081$; Fig. 4).

3.6. Physicians' Predictive Performance

Physicians' judgements did not indicate an adequate predictive performance for overall toxicity. ROC-AUC was 0.573 (CI, 0.433–0.712; $P = .311$). Physicians did not significantly predict overall toxicity in logistic regression (low vs mid: OR per category increase 1.664, CI 0.585–4.731; $P = .339$; low vs high: OR per category increase 2.240, CI 0.417–12.042, $P = .347$). Toxicity occurred in 25 (75.8%) patients classified as low, 52 (83.9%) as mid and 14 (87.5%) as high category ($P = .576$).

4. Discussion

This is the first German study directly comparing the CARG and CRASH score as a prediction of cancer treatment related toxicity in patients of an advanced age in a clinical routine setting. In our study, the CARG and the CRASH score exhibited a similar predictive performance for severe overall toxicity. For predicting severe hematologic toxicity, the hematologic CRASH score showed numerically a higher ROC-AUC than the CARG score. However, the ROC-AUC of the hematologic

Table 2
CARG and CRASH score items regarding patients with overall CTCAE toxicity grade ≥ 3. In univariate logistic regression, hemoglobin (item of the CARG score) and lactate dehydrogenase (item of the CRASH score) were significantly associated with grade ≥ 3 toxicity.

	Study patients (n = 113)			
	Patients with grade ≥ 3 toxicity		Odds ratio (95% CI)	p-value
	n	%		
CARG				
Age, classified [years]				
<72	7	87.5	–	
≥72	85	81.0	0.607 (0.071–5.218)	0.649
Cancer type				
Other	76	80.0	–	
GI/GU tumor	16	88.9	2.000 (0.423–9.457)	0.382
Dose				
Standard	78	80.4	–	
Reduced	14	87.5	1.705 (0.357–8.148)	0.504
Number of treatment agents				
Monotherapy	12	80.0	–	
Polytherapy	80	81.6	1.111 (0.284–4.349)	0.880
Hemoglobin [g/dL]				
≥10 (female), ≥11 (male)	55	75.3	–	
<10 (female), <11 (male)	37	92.5	4.036 (1.110–14.683)	0.034
Creatinine clearance, Jeliffe [mL/min]				
≥34 mL/min	81	79.4	–	
<34 mL/min	11	100.0	nd	
Hearing abilities				
Good/excellent	54	76.1	–	
Fair/worse	38	90.5	2.991 (0.932–9.594)	0.065
Falls in past six months				
0	73	80.2	–	
≥1	19	86.4	1.562 (0.416–5.860)	0.509
Medication intake				
No assistance	89	80.9	–	
Requires assistance	3	100.0	nd	
Limited in walking one block				
Not limited at all	49	81.7	–	
Limited	43	81.1	0.965 (0.374–2.494)	0.942
Decreased social activity				
A little/ none of the time	60	76.9	–	
Some/most/all of the time	32	91.4	3.200 (0.876–11.687)	0.078
CRASH				
Diastolic blood pressure [mmHg]				
≤72	38	79.2	–	
>72	54	83.1	1.292 (0.499–3.346)	0.598
IADL				
26–29	73	80.2	–	
10–25	19	86.4	1.562 (0.416–5.860)	0.509
LDH (U/L)				
>167	89	84.0	–	
≤167	3	42.9	6.980 (1.432–34.035)	0.016
ECOG performance score				
0	28	73.7	–	
1–2	50	83.3	1.786 (0.663–4.811)	0.252
3–4	14	93.3	5.000 (0.580–43.071)	0.143
MMSE				
30	14	70.0	–	
<30	78	83.9	2.229 (0.738–6.725)	0.155
MNA				
28–30	13	86.7	–	
<28	79	80.6	0.640 (0.133–3.077)	0.577
Therapy toxicity (MAX2)				
0 (0.00–0.44)	18	75.0	–	
1 (0.45–0.57)	37	78.7	1.233 (0.387–3.928)	0.723
2 (<0.57)	37	88.1	2.467 (0.663–9.176)	0.178

nd, not determinable; CI, confidence interval; GI, gastrointestinal; GU, genitourinary; IADL, Instrumental Activities of Daily Living; LDH, lactate dehydrogenase; ECOG, Eastern Cooperative Oncology Group Performance Status; MMSE, Mini-Mental State Examination; MNA, Mini Nutritional Assessment; MAX2, chemotherapy toxicity index

CRASH score was not statistically different from the ROC-AUC of the CARG score. This might have been caused by the small sample size. Since the clinical judgement indicated a rather low predictive performance, the CARG and CRASH score might be reasonable tools for supporting clinical judgement.

In general, the closer a ROC-AUC is to 1, the better the discrimination [15]. ROC of the CARG and the combined CRASH score were relatively far from 1, showing a ROC-AUC of 0.681 and 0.650, respectively. However, these ROC-AUC results were close to those reached in the validation study of the CARG (0.65) and combined CRASH score (0.64) [8,9]. Usually, tools tend to perform better in their derivation dataset than during external validation. Since the CARG score was originally not developed for hematological malignancies and immune or targeted therapies, this might have attenuated the predictive performance as well. However, we aimed at investigating if the CARG score can be used in a broader patient population including non-chemotherapy antineoplastic regimens and hematological malignancies. For hematologic toxicity, the hematologic CRASH showed numerically a higher ROC-AUC than the CARG score (AUC-ROC 0.665 vs 0.564) although this was not statistically different. This was congruent with our expectations especially as hematologic CRASH score was developed for this type of toxicity. We are only aware of one other study which compared the CARG and the CRASH score. Zhang et al. compared the scores in 106 patients ≥70 years with solid carcinoma and found an AUC-ROC of 0.77 for the CRASH score and 0.76 for the CARG score. In this study, the CARG and CRASH score were positively correlated with each other (Spearman's correlation coefficient = 0.689) [16]. The predictive value of the CARG score has been assessed in different studies. In a recent study by Moth et al., the CARG score did not show a predictive value for patients with solid tumors (AUC-ROC 0.52; OR per unit increase 1.04, $P = .54$; low 58%, mid 47%, high 58%; $P = .4$) [17]. Contrary to our study, Moth et al. did not include hematologic malignancies. For patients with prostate cancer, Alibhai et al. could not demonstrate a predictive value for the CARG score either (AUC-ROC 0.54; OR per unit increase 1.09, $P = .58$). Toxicity increased with the CARG score category but not significantly (low 0%, mid 17%, high 27%; $P = .65$) [18]. However, the study of Alibhai et al. was limited by a relatively small sample size of 46 patients. In patients with lung cancer, Nie et al. observed that toxicity incidence increased significantly with a higher CARG risk category (low 37,5%, mid 37,5%, high 25%; $P < .001$) [19].

Physicians' judgement versus CARG score predictions was also investigated [17,18]. Similar to our results, these studies did not observe adequate predictive performance of physicians' judgements. However, both studies did not find adequate predictive value of the CARG score either [17,18]. However, contrary to these results, in our patient cohort, the CARG and the CRASH score demonstrated satisfactory predictive performance. Our findings underline the value of both scores in supporting physicians in the decision-making process. Instead of using an exact percentage as Moth et al. [17], we asked physicians to classify risks as low, mid or high because – not being trained for this - we did not expect physicians to be able to estimate risks in such a detailed manner. However, physicians could have had different perceptions about the meaning of "low, mid or high".

Only poor agreement between CARG and CRASH score predictions was found. Possible reasons for this are the varying inclusion of therapy toxicity in the score predictions (for example the CARG score considers mono−/polytherapy regimens whereas the CRASH score uses the MAX2 score) as well as the inclusion of different aspects of the geriatric assessment (for instance the CARG score considers social aspects whereas the CRASH score includes nutritional aspects). Although only poor agreement was found, the predictive performance was similar which might be explained by the relatively low prediction performance. Since the prediction performance was not very high, this leaves space for incorrect predictions in each score. If one score does not predict correctly, the other score may do so. We assume that the scores predict

Please cite this article as: I. Ortland, M. Mendel Ott, M. Kowar, et al., Comparing the performance of the CARG and the CRASH score for predicting toxicity in older patients ..., J Geriatr Oncol, https://doi.org/10.1016/j.jgo.2019.12.016

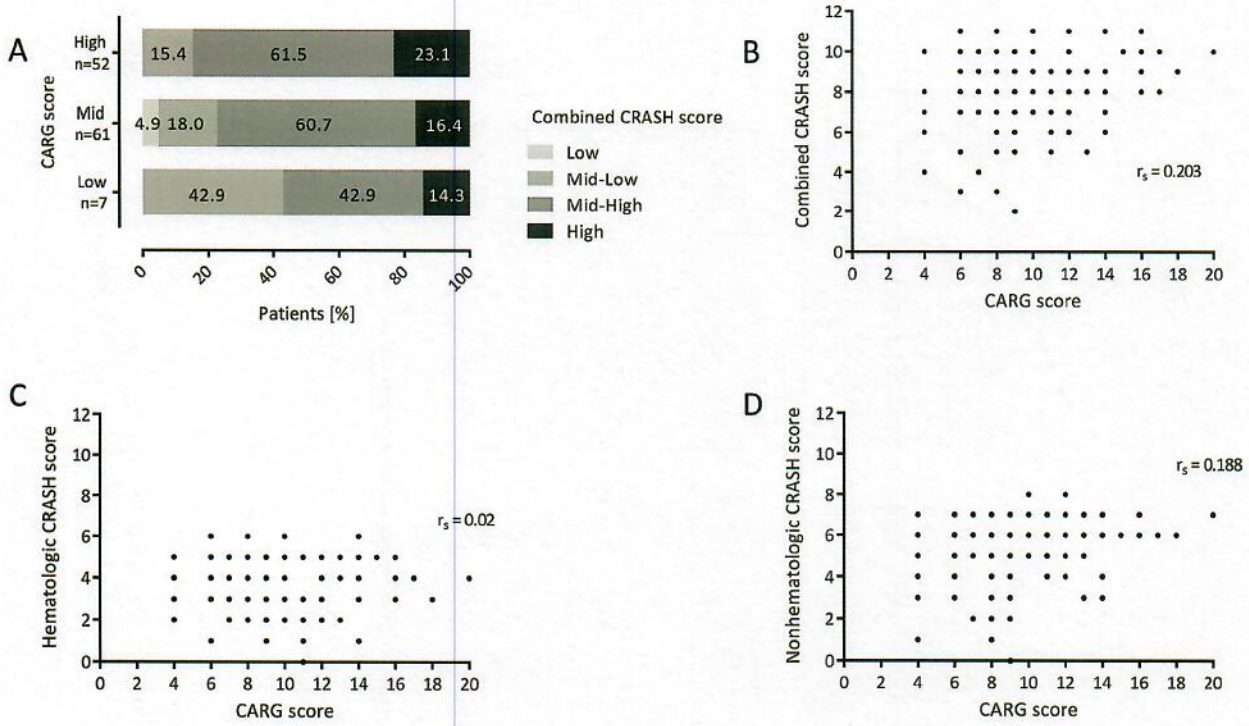

Fig. 2. Agreement (A) and correlation (B, C, D) of risk categories by the CARG and the combined CRASH score: (A) Increasing CARG risk categories show only a slight increase of percentage of higher combined CRASH risk categories within a CARG risk category. (B, C, D) Scatterplot demonstrates only low correlation of the CARG and combined, hematologic, and non-hematologic CRASH score results; r_s, Spearman's Rho.

complementary risks and that each score might predict accurately for different patients. Further studies should be conducted in order to investigate which score predicts adequately for which type of patient.

Our study population consisted of mainly fit patients, not representing a typical cohort of geriatric patients. This was also the case in the development studies of the CARG and the CRASH score (CARG: Karnofsky Performance Status >70: 80%; CRASH: ECOG 0–2 98% [7,8]). This might be due to the selection bias of non-frail patients in studies. As our study design was based on the development study of the CARG score as well as the CRASH score, our study featured different eligibility criteria than previous studies. The differences in patient cohort (hematologic malignances, older patients, starting therapy in

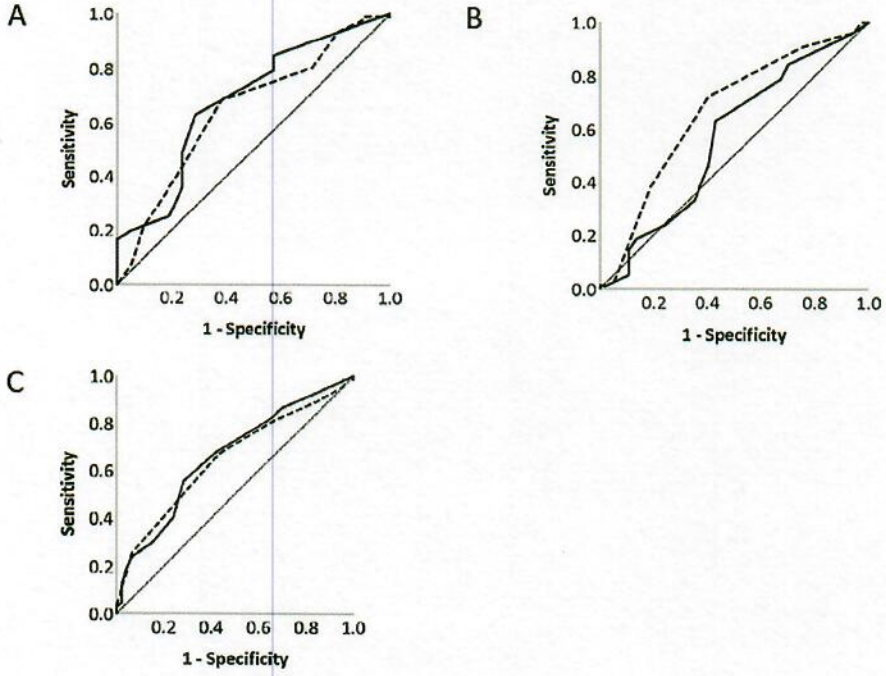

Fig. 3. ROC curves of the CARG and CRASH score for overall (A), hematologic (B) and nonhematologic (C) toxicity predictions. Solid line: CARG score, Dashed line: combined (A), hematologic (B) or nonhematologic (C) CRASH score.

Please cite this article as: I. Ortland, M. Mendel Ott, M. Kowar, et al., Comparing the performance of the CARG and the CRASH score for predicting toxicity in older patients ..., J Geriatr Oncol, https://doi.org/10.1016/j.jgo.2019.12.016

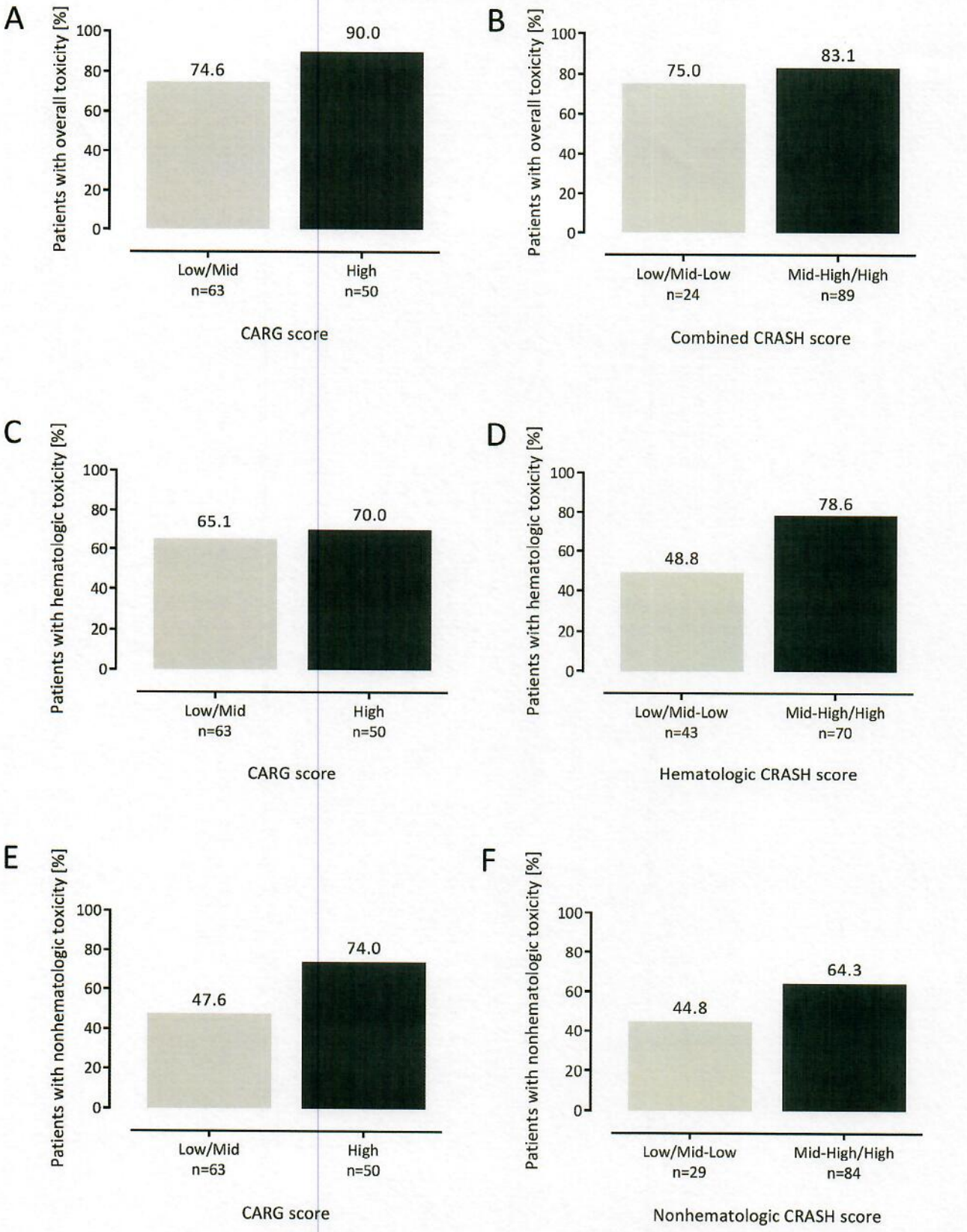

Fig. 4. Percentage of patients with overall (A-B), hematologic (C-D) or nonhematologic (*E*-F) toxicity per risk category of the CARG or CRASH score. CARG and CRASH are grouped into two risk categories. For overall and nonhematologic toxicity, CARG and CRASH categories indicate similar differences in percentage of patients with toxicity. For hematologic toxicity, the hematologic CRASH demonstrates higher differences between risk categories than the CARG score.

Please cite this article as: I. Ortland, M. Mendel Ott, M. Kowar, et al., Comparing the performance of the CARG and the CRASH score for predicting toxicity in older patients ..., J Geriatr Oncol, https://doi.org/10.1016/j.jgo.2019.12.016

an inpatient setting) might explain why overall toxicity incidence is higher in our study (81%) than in the CARG score development study (53%) [7]. Toxicity incidence in the CRASH score development study (64%) was lower as well, however in CRASH score development, only grade 4 hematologic toxicity was considered [8]. Those high toxicity rates in this study may generally limit the predictive ability of the tools due to the lack of non-events. Furthermore, some results of the score items were unevenly distributed between the patients, leading to low patient numbers in one category of the score items and consequently to wide confidence intervals in the logistic regression per score item. This might limit the generalizability of results. Moreover, few patients received targeted or immunotherapy alone.

4.1. Limitations and Strengths

The strength of this study is being the first German study to compare the CARG and the CRASH score in a clinical routine setting. Moreover, this study fills a gap in knowledge, being the first CARG score study which also includes patients with hematologic malignancies and new therapy forms. This enhances transportability into current routine care. For the CRASH score, to our knowledge, this is the first German study investigating it in a patient cohort different from the development study.

A limitation of our study is the retrospective collection of toxicity from medical records. Moreover, the moderate sample size might influence generalizability of results. Furthermore, including patient data from external oncology practices might bear potential for bias due to different standard procedures or documentation. However, including patients partly continuing treatment in an outpatient setting has the benefit of avoiding selection bias and ensures a patient cohort close to clinical routine. Also, identical procedures in data collection were applied to ensure consistent data quality. The CARG score was not validated for the German language and was therefore used orally. For the sake of using a self-administered CARG questionnaire in the future, it would be an interesting field for further research to linguistically validate those questions.

4.2. Implication for Practice

Risk of therapy-related toxicity is an important factor for therapy decision in older patients with cancer [20]. In our patient cohort, the scores suggest an adequate predictive performance whereas the predictive performance of the clinical judgement is rather low. Therefore, our results imply that the CARG and the CRASH score might both be useful for supporting clinical decision-making. As both scores present similar predictive performance, in general, none could be recommended above the other. However, the CARG score tends to take less time and hence might be preferable in a busy routine. The CRASH score, on the contrary, might be useful if a more detailed estimation of hematologic toxicity is required. Considering that both scores predicted different risks with similar overall prediction performance, using both simultaneously might provide complementary information for clinicians. Further prospective studies with a larger sample size are needed to corroborate those results. Moreover, since targeted therapy and immunotherapy play an increasing role in cancer therapy, future studies should explore toxicity prediction of these therapies in more detail.

Author Contributions

All authors contributed to the study concept and design, manuscript editing, reviewing and approving the final manuscript. Data acquisition: IO, MMO, MK, CS, AHJ, YDK. Statistical analysis: IO. Data analysis and interpretation: IO, UJ, AHJ, YDK. Manuscript preparation: IO.

Declaration of Competing Interest

IO, MK, CS have nothing to disclose. UJ, AHJ, YDK report grants from the Herbert-Worch foundation during the conduct of the study. MMO reports as financial activities outside the submitted work being currently employed at Grünenthal GmbH, however this was not the case during her main contribution to this study.

Acknowledgement

The study was funded by the Herbert-Worch foundation. The funder had no role in study design, data collection and analysis, preparation of the manuscript, or decision to publish. We kindly thank the staff of the Johanniter Hospital Bonn and the oncology practices of the Rhein-Sieg area (ZAHO Bonn Loestraße; ZAHO Siegburg (Dr. S. Fronhoffs); Praxiskooperation Bonn-Euskirchen-Rheinbach-Wesseling; Praxisnetzwerk Onkologie Rheinsieg; Dr. H. P. Jungbluth, Neuwied; Praxis für Innere Medizin, Hämatologie, Onkologie und hausärztliche Versorgung Bonn-Bad Godesberg (Dr. P. F. Schwindt); and Gemeinschaftspraxis für Hämatologie und Onkologie Rhein-Ahr) for their kind support. Especially, we express our gratitude to all patients who participated in our study despite their difficult situation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jgo.2019.12.016.

References

[1] Muss HB, Berry DA, Cirrincione C, Budman DR, Henderson IC, Citron ML, et al. Toxicity of older and younger patients treated with adjuvant chemotherapy for node-positive breast cancer: the Cancer and Leukemia Group B experience. J Clin Oncol 2007;25:3699–704.
[2] Scher KS, Hurria A. Under-representation of older adults in cancer registration trials: known problem, little progress. J Clin Oncol 2012;30:2036–8.
[3] Soto-Perez-de-Celis E, Li D, Yuan Y, Ming Lau Y, Hurria A. Functional versus chronological age: geriatric assessments to guide decision making in older patients with cancer. Lancet Oncol 2018;19:e305–16.
[4] Extermann M, Aapro M, Bernabei R, Cohen HJ, Droz J-P, Lichtman S, et al. Use of comprehensive geriatric assessment in older cancer patients: Recommendations from the task force on CGA of the International Society of Geriatric Oncology (SIOG). Crit Rev Oncol Hematol 2005;55:241–52.
[5] Mohile SG, Dale W, Somerfield MR, Schonberg MA, Boyd CM, Burhenn PS, et al. Practical assessment and management of vulnerabilities in older patients receiving chemotherapy: Asco guideline for geriatric oncology. J Clin Oncol 2018;36:2326–47.
[6] Almodovar T, Teixeira E, Barroso A, Soares M, Queiroge HJ, Cavaco-Silva J, et al. Elderly patients with advanced NSCLC: the value of geriatric evaluation and the feasibility of CGA alternatives in predicting chemotherapy toxicity. Pulmonology 2019;25:40–50.
[7] Hurria A, Togawa K, Mohile SG, Owusu C, Klepin HD, Gross CP, et al. Predicting chemotherapy toxicity in older adults with cancer: a prospective multicenter study. J Clin Oncol 2011;29:3457–65.
[8] Extermann M, Boler I, Reich RR, Lyman GH, Brown RH, DeFelice J, et al. Predicting the risk of chemotherapy toxicity in older patients: the chemotherapy risk assessment scale for high-age patients (CRASH) score. Cancer 2012;118:3377–86.
[9] Hurria A, Mohile S, Gajra A, Klepin H, Muss H, Chapman A, et al. Validation of a prediction tool for chemotherapy toxicity in older adults with cancer. J Clin Oncol 2016;34:2366–71.
[10] U.S. Department of Health and Human Services Health. Common Terminology Criteria for Adverse Events Version 4.0 (v4.03). https://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03/CTCAE_4.03_2010-06-14_QuickReference_5x7.pdf. Accessed date: 14 July 2019.
[11] Cohen J. Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. Psychol Bull 1968;70:213–20.
[12] Royston P, Moons KGM, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. BMJ 2009;338:b604.
[13] Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29–36.
[14] DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a non-parametric approach. Biometrics 1988;44:837–45.
[15] Akobeng AK. Understanding diagnostic tests 3: receiver operating characteristic curves. Acta Paediatr 2007;96:644–7.
[16] Zhang J, Liao X, Feng J, Yin T, Liang Y. Prospective comparison of the value of CRASH and CARG toxicity scores in predicting chemotherapy toxicity in geriatric oncology. Oncol Lett 2019;18:4947–55.

Please cite this article as: I. Ortland, M. Mendel Ott, M. Kowar, et al., Comparing the performance of the CARG and the CRASH score for predicting toxicity in older patients ..., J Geriatr Oncol, https://doi.org/10.1016/j.jgo.2019.12.016

[17] Moth EB, Kiely BE, Stefanic N, Naganathan V, Martin A, Grimison P, et al. Predicting chemotherapy toxicity in older adults: comparing the predictive value of the CARG toxicity score with oncologists' estimates of toxicity based on clinical judgement. J Geriatr Oncol 2019;10:202–9.
[18] Alibhai SMH, Aziz S, Manokumar T, Timilshina N, Breunis H. A comparison of the CARG tool, the VES-13, and oncologist judgment in predicting grade 3 + toxicities in men undergoing chemotherapy for metastatic prostate cancer. J Geriatr Oncol 2017;8:31–6.
[19] Nie X, Liu D, Li Q, Bai C. Predicting chemotherapy toxicity in older adults with lung cancer. J Geriatr Oncol 2013;4:334–9.
[20] Puts MTE, Tapscott B, Fitch M, Howell D, Monette J, Wan-Chow-Wah D, et al. A systematic review of factors influencing older adults' decision to accept or decline cancer treatment. Cancer Treat Rev 2015;41:197–215.

Please cite this article as: I. Ortland, M. Mendel Ott, M. Kowar, et al., Comparing the performance of the CARG and the CRASH score for predicting toxicity in older patients ..., J Geriatr Oncol, https://doi.org/10.1016/j.jgo.2019.12.016