CLINICAL SCIENCE OF NUTRITION

Original Article

DOI: 10.5152/ClinSciNutr.2022.222051

Using Proxy-Reported SARC-F on Behalf of Self-Assessment in Older Adults: Examining its Reliability and Agreement with Patient Responses

Serdar Özkök[1], Meryem Merve Ören[2], Çağlar Özer Aydın[1], Hümeyra Özalp[1], Cihan Kılıç[1], Yaşagül Koç[3], Hafize Doğan[4], Mehmet Akif Karan[1], Gülistan Bahat[1]

[1]Division of Geriatrics, Department of Internal Medicine, Istanbul University, Faculty of Medicine, İstanbul, Turkey
[2]Department of Public Health, Istanbul University, Faculty of Medicine, İstanbul, Turkey
[3]Department of Nursing, Istanbul Kayisdagi Nursing Home, İstanbul, Turkey
[4]Department of General Practice, Istanbul Kayisdagi Nursing Home, İstanbul, Turkey

Cite this article as: Özkök S, Ören MM, Aydın ÇÖ, et al. Using proxy-reported SARC-F on behalf of self-assessment in older adults: examining its reliability and agreement with patient responses. *Clin Sci Nutr.* 2023;4(3):88-97.

ABSTRACT

Objective: SARC-F is a self-reported questionnaire to screen for an increased risk of sarcopenia. Since it requires self-assessment, it is not appropriate for use in patients with impaired judgment or severe communication problems. Whether it can be reliably administered to proxies on behalf of patients is an issue that needs to be clarified. We aimed to study the reliability of SARC-F by proxy and examine the agreement between patient and proxy responses.
Methods: This is a cross-sectional study conducted between September 2019 and October 2021. Patients were recruited from 2 settings: outpatients and nursing home residents. Proxies were relatives/caregivers in community-dwelling setting and nurses in nursing home. We transformed SARC-F to SARC-F by proxy and studied its reliability with interrater and test–retest reliability analyses in the first phase. In the second phase, we examined the concordance between patient and proxy responses in total and item by item.
Results: Total sample size was 279 (172 patients and 107 proxies). Community-dwelling older adults made up 58.1% of the older adult population. Median age of older adults was 72 (60-93), and 44.8% were female. SARC-F by proxy showed an excellent interrater and test–retest reliability, with intraclass correlation coefficients of 0.91 and 0.90, respectively ($P < .001$). It also demonstrated a high level of internal consistency, with a Cronbach's alpha value of 0.82. The total scores of SARC-F by patient and SARC-F by proxy showed a moderate correlation ($r = 0.635$; $P < .001$). The fourth item demonstrated the highest, and the fifth item showed the lowest correlation (r values = 0.591 and 0.443, respectively).
Conclusion: According to our study, SARC-F by proxy can be reliably administered to proxies on behalf of older adults when conditions that prevent reliable judgment or communication exist. Further validity studies of SARC-F by proxy are needed to verify whether it will work well in identifying sarcopenia cases in older adults.

Keywords: Geriatric assessment, older adults, patient-reported outcome measures, proxy, sarcopenia

INTRODUCTION

Sarcopenia has become a hot topic that healthcare professionals have shown more interest in recent years since plenty of reports have been published on its close relationship with adverse outcomes like falls, disabilities, hospitalizations, and mortality in older adults.[1] Recent guides on sarcopenia have recommended the use of SARC-F questionnaire in case finding, with a score of ≥4 meaning that certain individual has an increased risk of sarcopenia.[2,3]

Although SARC-F showed low-to-moderate level of sensitivity, it demonstrated a high level of specificity in identifying sarcopenia, ending up mostly detecting severe cases.[4] Several reports tried to increase its sensitivity by coming up with different thresholds[5] or modifying it by implementing certain measurements [like SARC-CalF (SARC-F and calf circumference)].[6] Whether it can also be used in other conditions (like identifying physical frailty)[7] has been another point to be addressed. Adding to its ease of use and practicality, it has gained an undeniable interest in

*The content of this manuscript was not presented in any scientific event before.

Corresponding author: Gülistan Bahat, e-mail: gbahatozturk@yahoo.com

Received: November 21, 2022 Accepted: December 12, 2022

sarcopenia practice in recent years. However, besides the features that favor its common use, it also embodied some limitations coming from its self-reported nature, as it should not be preferred in the presence of dementia or other serious neuropsychiatric problems that may impair judgment or communication problems that make administration of the questionnaire impossible.

In routine practice, healthcare professionals sometimes apply self-reported assessments to caregivers, on behalf of patients. However, the judgments of proxies (relatives, caregivers, or sometimes healthcare staff) may not always be realistic or properly reflect the actual situation, and they can differ depending on several factors like the proximity of the relationship, education level, mood, or caregiver burden of the proxy.[8] On the contrary, sometimes, the perspective of proxies may be more objective and closer to the truth. Hence, the ideal thing is that the reliability and validity of using proxy-reported questionnaires on behalf of patients should be studied before and applied accordingly.

In the literature, there are various reliability studies of proxy-reported questionnaires, mostly assessing the quality of life (QoL) in different patient populations.[9-11] As a self-reported questionnaire, whether it is reliable and valid to apply SARC-F to the proxies on behalf of patients is an issue that needs to be clarified. In order to evaluate how well a test can predict a certain outcome, it is first necessary to study whether the use of that test is reliable. Reliability refers to the degree to which the results obtained by measurement can be replicated. Lack of reliability is expected to affect the validity of certain measurements invariably and can arise from divergence between observers or instruments of measurements.[12] Therefore, this study aims to analyze the reliability of SARC-F questionnaire applied to proxies (namely, *SARC-F by proxy*) on behalf of older patients and study the concordance between SARC-F results obtained from patients and their proxies.

Main Points

- SARC-F is a 5-item screening tool that is recommended for case finding in sarcopenia. As a self-reported questionnaire, it embodies a limitation as it should not be applied to individuals with impaired judgment and communication problems.
- In older adults with dementia or communication difficulties, SARC-F by proxy may be applied to the relatives/caregivers on behalf of the patients, demonstrating excellent reliability.
- SARC-F responses obtained from patients and their proxies showed moderate level of agreement, with the highest agreement on the fourth (climbing stairs) and the lowest agreement on the fifth item (falls in the past year).

METHODS

This study consisted of 2 phases: namely, "adaptation of Turkish SARC-F into *SARC-F by proxy* and reliability analysis" constituted the first and assessment of correlation between "*SARC-F by patient*" and "*SARC-F by proxy*" constituted the second phase. We conducted the study between September 2019 and October 2021, among geriatric outpatients admitted to a tertiary health clinic and residents living in the largest nursing home (NH) in the city that was subordinated to metropolitan municipality. The study was approved by the local ethics committee (reference: 1503/2020, date of approval: October 26, 2020). We received informed consent from all of the participants. We received legal permission from NH administration.

First Phase (Adaptation of Turkish SARC-F into SARC-F by Proxy and Reliability Analysis)

This phase included 7 steps based on the guidelines put forward by World Health Organization (WHO).[13] We organized an expert panel including 2 bilingual experts (one having English and the other having Turkish as their mother tongue) and 2 bilingual geriatricians. In the first step, an expert panel composed of 2 geriatricians transformed the Turkish-validated SARC-F into *SARC-F by proxy*: We converted the items originally filled out by patients into the items asked to caregivers/relatives for the assessment on behalf of patients. In this way, proxies would elicit substituted judgment, where they projected themselves into the *body and mind* of the patients when answering the questions (i.e., proxy-patient perspective[14]). In step 2, a bilingual translator having English as her mother tongue and blinded to the original questionnaire back-translated the Turkish *SARC-F by proxy* into the English *SARC-F by proxy*. In step 3, 2 geriatricians and 2 bilingual experts had a meeting and reviewed 2 forms of the test in terms of conceptual and cultural equivalence to reach a satisfactory version. Later at step 4, we e-mailed the 2 agreed-upon versions (Turkish *SARC-F by proxy* and the translated version) to one of the authors of the original English SARC-F, John Morley, and got approval. In step 5 (pre-test phase), we administered *SARC-F by proxy* to 5 female and 5 male participants face-to-face in order to get their opinions, suggestions, and doubts about the comprehension and cultural relevance of the test. World Health Organization guidelines on the process of adaptation and validation of instruments recommended that at least 10 subjects should participate in the pre-test step.[13] At pre-test step, we excluded participants with any cognitive dysfunction or severe hearing impairment that would affect comprehension of the questionnaire. In step 6, 2 geriatricians administered *SARC-F by proxy* to 21 proxies in different rooms on the same day, in order to assess inter-rater reliability. In the final step, we applied *SARC-F*

by proxy face-to-face or by phone to these 21 participants 14 days later in order to evaluate test–retest reliability. We specified a time interval of "14 days," because we considered that this interval would be long enough to prevent recall bias of previous answers and short enough for significant changes in physical capacity to develop.[15]

Second Phase (Assessment of Concordance Between "SARC-F by Patient" and "SARC-F by Proxy")

Since the secondary aim of the study was to find out whether *SARC-F by proxy* would demonstrate a high level of agreement with patient responses, we had to assess the correlation between *SARC-F by patient* and *SARC-F by proxy* results (obtained from their caregivers/relatives). Hence, there were 2 study populations: patients and their proxies. Patient population also included older adults from 2 different settings: community-dwelling (CD) older adults and NH residents. Inclusion criteria for older individuals were being older than 60 years of age and having a concurrent caregiver or a relative who knew and observed the patient well enough to reliably answer the questions. Exclusion criteria for older individuals were having moderate or severe dementia (already diagnosed or recent diagnosis through Clinical Dementia Rating Scale (CDR)), severe hearing or visual impairment, severe depression, admitting to the outpatient clinic alone or with a companion who cannot make a reliable assessment about the patient, and refusal to participate. Inclusion criteria for proxy group were being older than 18 years of age and having frequent contact (contact on a weekly basis, at least via telephone) with the patients or residents that allows the proxy insight into the individual's situation,[8] at least for the last year. Exclusion criteria were having diagnosis of dementia, depression, hearing impairment, having insufficient contact with patient to meet the above-mentioned criteria for proxy assessments, and refusal to participate.

SARC-F is a 5-item self-questionnaire recommended for sarcopenia screening and case finding by the European Working Group on Sarcopenia in Older People 2 guide (EWGSOP2).[2] It evaluates **S**trength, **A**ssistance in walking, **R**ising from a chair/bed, **C**limbing stairs, and Falls. A score of ≥4 means the patient has an increased risk of sarcopenia. *SARC-F by proxy* is the transformed version of SARC-F in which 5 items of SARC-F were converted to the questions directed to the caregivers/relatives of the patients and asked for the answers given on behalf of the patient with a proxy–patient perspective. Proxy assessments can be performed by asking a proxy to assess the patient as they think the patient would respond (i.e., proxy–patient perspective) or to provide their own judgment (i.e., proxy–proxy perspective) on the patients' health status. In proxy–patient perspective, proxy is instructed to "try to view the situation as the patient would" or "think as the patient would."[14] The difference between self-report of the patients' and proxies' perspectives is called "inter-rater gap," and it was hypothesized that this gap was smaller for proxy–patient perspective than proxy–proxy perspective.[8] Therefore, we decided to use this perspective in applying the questionnaire to the proxies.

In the original SARC-F, strength is assessed by asking how much difficulty the patient have in carrying/lifting 10 pounds. However, in Turkish SARC-F validation study, 10 pounds was adapted as 5 kg instead of the precise calculation (4.54 kg) in order to ease the understanding of the question in daily practice, as suggested by EuGMS Sarcopenia Special Interest Group.[15] English and Turkish forms of *SARC-F by proxy* and their scoring system can be found in Supplementary Table 1.

Guidelines for calculation of minimum sample size generally recommend a respondent-to-item ratio ranging from 5 : 1 to 30 : 1.[16] As *SARC-F by proxy* is a 5-item questionnaire, we decided that we would reach an adequate sample size with at least 150 older adults and their proxies. A geriatrician performed face-to-face interviews with older individuals and their proxies in different rooms on the same day.

Comprehensive Geriatric Assessment

We collected the demographic and clinical data of older adults including age, gender, education level, assistance in walking, living alone (yes/no), tobacco and alcohol use, and number of illnesses and regular drugs. We assessed functionality via Katz's basic activities of daily living (ADL) and Lawton's instrumental activities of daily living (IADL) scales. Katz ADL scores range between 0 and 6, and Lawton IADL between 0 and 8 (a score of 0 means complete dependency and full points mean complete independency, for both tests).[17,18] We assessed the cognitive status of the participants via CDR. The clinical Dementia Rating Scale evaluates patients' cognitive and functional performance in 6 areas: memory, orientation, judgment and problem solving, community affairs, home, hobbies, and personal care. Scores from each area are combined to obtain a composite score ranging from 0 to 3. A score of 0 indicates normal cognitive functions; however, a score of 0.5 indicates very mild/questionable dementia; 1 indicates mild, 2: moderate, and 3: severe dementia.[19] We assessed frailty through **F**atigue, **R**esistance, **A**mbulation, **I**llnesses, and **L**oss of weight (FRAIL) index: A subject with a score of ≥3 was considered as frail, 1-2 points as pre-frail, and 0 as robust.[20] We evaluated nutritional status via Mini-Nutritional Assessment-Short Form, with a score of <12 points interpreted as undernutrition, and <8 points as malnutrition.[21] We asked whether older adults experienced any falls during the past year. We defined polypharmacy as using ≥5 medications per day.

Statistical Analysis

We analyzed the normality of the numerical variables with histograms, probability plots, and Kolmogorov–Smirnov tests. We presented numerical variables as mean ± standard deviation or median (minimum-maximum) and categorical variables as numbers and frequencies. We compared 2 groups with an independent sample *t*-test or Mann–Whitney *U* test according to their normality analysis. We compared categorical variables using chi-square test with Yates correction and Fisher's exact test. We assessed the reliability of *SARC-F by proxy* by internal consistency, inter-rater, test–retest reliability, and concordance analyses. We tested inter-rater and test–retest reliability by intra-class correlation coefficient (ICC). We calculated ICC estimates and their 95% CI based on a single measurement, absolute agreement 2-way mixed-effects model. We defined reliability by ICC estimates as: ICC estimate [0.90: excellent reliability, between 0.75 and 0.9: good reliability, 0.5-0.75: moderate reliability, <0.5: poor reliability]. We tested internal consistency by Cronbach's alpha coefficient, with a value of >0.70 indicating a high level of internal consistency. We analyzed the correlation between *SARC-F by patient* and *SARC-F by proxy* in total and item-by-item by Spearman's rho correlation test. Alternatively, we defined *SARC-F by patient* and *SARC-F by proxy* results categorically (≥4 as positive screening) and studied the overall concordance rate. We reported the Cohen kappa coefficient (κ). κ values between 0.81 and 1 were considered as perfect, 0.6-0.8 indicated strong, 0.4-0.6 indicated moderate, 0.20-0.4 indicated low, between 0 and 0.20 indicated very slight agreement, and less than 0 indicated disagreement. We accepted a *P* value of less than .05 as significant. We performed statistical analyses by Statistical Package for Social Sciences (SPSS) v 21.0 (SPSS Statistics; IBM, Armonk, NY, USA) and MedCalc Statistical Software v 15.2 (MedCalc Software, Ostend, Belgium).

RESULTS

First Phase (Adaptation of SARC-F into SARC-F by Proxy and Reliability Analysis)

We assessed whether *SARC-F by proxy* was easy to understand in pretest step and included 5 men and 5 women, with a mean age of 42.3 ± 8.9. For both sexes, 3 of the participants were primary school graduates; 1 male and 2 female participants were high school graduates, and 1 male participant was a college graduate. They did not report any problems in comprehension of the items. Interrater and test–retest reliability steps included 21 proxies with a mean age of 56.2 ± 15.0. Inter-rater reliability analysis showed excellent reliability, with an ICC of 0.91 (0.80-0.97) ($P < .001$). Likewise, test–retest reliability was excellent with an ICC of 0.90 ($P < .001$). Detailed findings of inter-rater and test–retest reliability analyses are given in Tables 1 and 2. Internal consistency analysis of *SARC-F by proxy* showed a high level of consistency, with a Cronbach's alpha value of 0.82.

Table 1. Interrater Reliability Analysis of SARC-F by Proxy

	ICC	95% CI
SARC-F by proxy total	0.919	0.796-0.967
SARC-F by proxy; strength item	0.862	0.666-0.944
SARC-F by proxy; assistance item	0.745	0.364-0.897
SARC-F by proxy; rising item	0.604	0.071-0.836
SARC-F by proxy; climbing item	0.915	0.794-0.965
SARC-F by proxy; falls item	0.750	0.378-0.899

ICC, intraclass correlation coefficient.
We calculated ICC estimates and their 95% CI based on a single measurement, absolute agreement 2-way mixed-effects model.

Second Phase (Assessment of Concordance Between "SARC-F by Patient" and "SARC-F by Proxy")

Demographic and Clinical Characteristics of Older Adults and Their Proxies

Total sample consisted of 279 participants, with 172 older adults (100 for CD group and 72 for NH group) and 107 proxies (98 for CD group and 9 for NH group). Median age of older adults was 72 (min-max: 60-93), and the number of female participants was 77 (44.8%). Median age of the proxies was 52 (min-max: 20-85), with 67 (63.2%) being female. Proxies of CD group were mostly their adult children (54.6%) and proxies of NH group consisted entirely of nurses giving close care to the residents. The cause of the mismatch in patient–proxy numbers is that while 96 proxies in CD group performed *SARC-F by proxy* on behalf of only 1 older individual, 2 proxies administered the test on

Table 2. Test–Retest Reliability Analysis of SARC-F by Proxy

	ICC	95% CI
SARC-F by proxy total	0.952	0.883-0.980
SARC-F by proxy; strength item	0.837	0.596-0.934
SARC-F by proxy; assistance item	0.868	0.682-0.946
SARC-F by proxy; rising item	0.848	0.634-0.938
SARC-F by proxy; climbing item	0.934	0.837-0.973
SARC-F by proxy; falls item	0.800	0.500-0.919

ICC, intraclass correlation coefficient.
*We calculated ICC estimates and their 95% CI based on a single measurement, absolute agreement 2-way mixed-effects model.

Table 3. The Demographical and Clinical Characteristics of the Study Population (N=279)				
	Total	**CD Group**	**NH Group**	*P*
Older individuals				
Number (%)	172 (100)	100 (58.1)	72 (41.9)	
Age (years)*	72 (60-93)	72.5 (60-93)	71 (60-85)	.08
Gender (female)¶	77 (44.8%)	65 (65%)	12(16.7%)	**<.001**
Education level¶				.48
Illiterate	24 (14.3%)	14 (14%)	10 (14.7%)	
Primary school	93 (55.4%)	52 (52%)	41 (60.3%)	
Secondary school	10 (6%)	5 (5%)	5 (7.4%)	
High school	24 (14.3%)	16 (16%)	8 (11.8%)	
University	17 (10.1%)	13 (13%)	4 (5.9%)	
Assistance in walking¶	37 (21.5%)	20 (20%)	17 (23.6%)	.57
Living alone¶	83 (48.2%)	83 (16.2%)	-	N/A
Smoking¶	44 (25.6%)	4 (4%)	40 (55.6%)	**<.001**
Alcohol¶	4 (2.3%)	3 (3%)	1 (1.4%)	.64
Number of illnesses*	4 (0-10)	4 (0-10)	3.5 (0-7)	**.02**
Number of regular drugs*	7 (0-17)	6 (0-15)	7.5 (0-17)	**.03**
Chronic illnesses¶				
Diabetes mellitus	67 (39%)	47 (47%)	20 (27.8%)	**0.01**
Hypertension	112 (65.1%)	81 (81%)	31 (43.1%)	**<.001**
Dyslipidemia	80 (46.5%)	59 (59%)	21 (29.2%)	**<.001**
Ischemic heart disease	63 (36.6%)	50 (50%)	13 (18.1%)	**<.001**
Heart failure	23 (13.4%)	5 (5%)	18 (25.4%)	**<.001**
COPD	33 (19.2%)	5 (5%)	28 (38.9%)	**<.001**
Hypothyroidism	18 (10.5%)	12 (12%)	6 (8.3%)	.43
Proxies				
Number (%)¥	107 (100)	98 (91.6)	9 (8.4)	
Age*	52 (20-85)	55 (20-85)	31(24-47)	<.001
Gender (female)¶	67 (63.2%)	62 (63.9%)	5 (55.6%)	.72
Education level¶				.06
Illiterate	2 (1.9%)	2 (2.1%)	-	
Primary school	24 (22.6%)	24 (24.7%)	-	
Secondary school	6 (5.7%)	6 (6.2%)	-	
High school	29 (27.4%)	23 (23.7%)	6 (66.7%)	
University	45 (42.5%)	42 (43.3%)	3 (33.3%)	
Relationship to the older participant¶				
Spouse		31 (32%)	-	N/A
Child		53 (54.6%)	-	
Daughter-son in law		2 (2.1%)	-	
Sibling		6 (6.2%)	-	
Caregiver		2 (2.1%)	-	
*Nurse^*		-	9 (100%)	
Others#		3 (3.1%)	-	

¶Number (percentage);*Median (minimum-maximum);¥Exceptionally, there were 2 proxies from CD group who answered the questions on behalf of 2 different older individuals. One of them was the daughter of an old couple, and the other was the niece of 2 sisters;^All nurses were responsible for medical treatments of more than 1 resident. The number of residents per nurse ranged between 4 and 17; Niece, nephew, and neighbor.
CD, community-dwelling; COPD, chronic obstructive pulmonary disease; NH, nursing home.

behalf of their 2 different relatives. In NH, 9 nurses from different wards of NH took part in the study. Hence, each of them responded to *SARC-F by proxy* on behalf of the residents from the wards for which they were responsible. The number of residents per nurse ranged between 4 and 17. The demographical characteristics of older individuals and their proxies are given in detail in Table 3.

Median number of chronic diseases was 4 (min-max: 0-10), with hypertension being the most prevalent (65.1%; n=112). When 2 groups of older participants were compared in terms of their clinical characteristics, CD group had significantly more female participants, higher total number of chronic diseases and higher hypertension, dyslipidemia, and ischemic heart disease prevalence. Nursing home group had higher prevalence of tobacco use, chronic obstructive pulmonary disease (COPD), and heart failure and higher number of regular medications. The clinical characteristics of the study population are given in Table 3.

The median total *SARC-F by patient* score was 2 (0-10), and 54 (31.4%) of the participants had an increased risk of sarcopenia. Community-dwelling and nursing home groups showed no significant difference in terms of *SARC-F by patient* results. Other CGA findings also showed that CD and NH groups were similar in terms of geriatric syndromes, except for CD group being more frequent fallers and NH group being more dependent in terms of ADL. Findings of CGA are given in Table 4.

Concordance Analysis Between SARC-F by Patient and SARC-F by Proxy

The median total *SARC-F by proxy* score was 2 (0-10), and according to *SARC-F by proxy*, 52 (30.2%) of the participants had positive sarcopenia screening. Although *SARC-F by patient* results did not show significant difference between settings, positive screening for *SARC-F by proxy* was significantly higher in community setting compared to NH (38.0% vs. 19.4%, *P*=.009). Median scores for *SARC-F by proxy* were 3 (0-9) and 1 (0-10) in

Table 4. Comprehensive Geriatric Assessment Findings of the Older Adults

	Total	Community-Dwelling	Nursing Home	*P*
ADL*	6 (1-6)	6 (1-6)	6 (2-6)	**.001**
IADL*	8 (0-8)	8 (0-8)	8 (2-8)	.7
CDR¶				**<.001**
Normal	65 (48.9%)	39 (63.9%)	26 (36.1%)	
MCI	57 (42.9%)	14 (23.0%)	43 (59.7%)	
Early dementia	11 (8.3%)	8 (13.1%)	3 (4.2%)	
Falls in the past year¶	51 (29.7%)	36 (36%)	15 (20.8%)	**.03**
FRAIL*	1 (0-5)	1 (0-5)	1 (0-4)	.05
FRAIL¶				.07
Robust	54 (31.4%)	27 (27%)	27 (37.5%)	
Pre-frail	72 (41.9%)	40 (40%)	32 (44.4%)	
Frail	46 (26.7%)	33 (33%)	13 (18.1%)	
MNA-SF*	13 (5-14)	13 (6-14)	13 (5-14)	.55
Under nutrition¶	40 (23.7%)	24 (24.5%)	16 (22.5%)	.76
Malnutrition¶	7 (4.1%)	5 (5.1%)	2 (2.8%)	.76
Polypharmacy¶	123 (71.9%)	69 (69.7%)	54 (75%)	.45
SARC-F by patient*	2 (2-10)	2 (0-9)	2 (2-10)	.42
SARC-F by patient ≥4¶	54 (31.4%)	35 (35%)	19 (26.4%)	.23

*Median (minimum-maximum);¶Numbers (percentage).
ADL, activities of daily living; CDR, Clinical Dementia Rating Scale; IADL, instrumental activities of daily living; MCI, mild cognitive impairment; MNA-SF, Mini-Nutritional Assessment-Short Form.

CD and NH settings, respectively ($P=.002$). The scores of *SARC-F by patient* and *SARC-F by proxy* showed a moderate correlation, with a correlation coefficient of 0.635 ($P < .001$). The fourth item demonstrated the highest correlation ($r=0.591$) and the lowest was shown for the fifth item ($r=0.443$). We alternatively defined *SARC-F by proxy* and *SARC-F by patient* results categorically (as SARC-F ≥ 4 being positive sarcopenia screening), and studied the agreement between 2 tests. We obtained a κ value of 0.482 (0.341-0.623), which again showed a moderate agreement between patient and proxy results. When we performed a further analysis to study whether the agreement level differed between settings, we found out that results from 2 different settings individually showed moderate agreement, with κ value of community setting being slightly higher than NH setting [0.504 (0.330-0.678) versus 0.414 (0.171-0.657)]. Detailed findings of the results of both tests and their correlation analysis are given in Table 5.

Table 5. Concordance Analyses of SARC-F by Patient and SARC-F by Proxy

	SARC-F by Patient	SARC-F by Proxy	Correlation Coefficient
Total score	2 (0-10)	2 (0-10)	**.635***
1. Strength			
None	54.1%	47.1%	**.454***
Some	25.6%	39.1%	
A lot	20.3%	14.0%	
2. Assistance in walking			
None	72.5%	73.3%	**.446***
Some	15.2%	18.6%	
A lot	12.3%	8.1%	
3. Rise from a chair			
None	64.5%	61%	**.503***
Some	22.7%	33.1%	
A lot	12.8%	5.8%	
4. Climbing stairs			
None	38.4%	46.5%	**.591***
Some	38.4%	32.0%	
A lot	23.3%	21.5%	
5. Falls			
None	70.3%	71.5%	**.443***
Some	23.8%	23.8%	
A lot	5.8%	4.7%	

*P value <.001.

DISCUSSION

In this study, we adapted the Turkish SARC-F to *SARC-F by proxy* with the aim of studying whether it could be applied to their caregivers/relatives on behalf of older individuals who are unable to cooperate or make reliable judgments on their clinical conditions. We found out that *SARC-F by proxy* had an excellent inter-rater and test–retest reliability, and it demonstrated a moderate level of concordance with patient-reported SARC-F.

In routine practice, it is a common method to refer to the statements of the patients' relatives/caregivers on behalf of the patients who cannot make judgments about their own health status or have difficulty establishing reliable communication. However, whether this method is reliable and valid enough should be examined and well-demonstrated in order to use it as a substitute for patients' self-reports. In this context, several proxy-reported questionnaires (mostly QoL assessment tools) have been studied in different patient populations like patients with dementia, stroke, cancer, Parkinson's disease, or other neurological disorders,[9-11,22] with mixed results about their reliability and concordance with patient reports. In the case of sarcopenia, there are insufficient data on how valid or reliable the screening tool SARC-F is when administered to proxies on behalf of patients. Hence, this study serves to fill the gap in the literature on this particular issue.

Screening tools have to be valid enough to predict the conditions or outcomes that they are used for, in order to be recommended for use in routine practice. However, besides validity, another related, equally important concept is the reliability of that certain test since first of all, a test needs to be reliable in order to be valid. It should be able to produce consistent results regardless of the tester and time.[12] In order to check whether *SARC-F by proxy* was a reliable screening tool, we had to adopt the original SARC-F to *SARC-F by proxy*, by transforming the questions directed to the relatives/caregivers rather than patients. We conducted the process as if we were studying the reliability of a questionnaire that was developed for the first time and followed the recommended steps to be taken in reliability studies. We found out that *SARC-F by proxy* demonstrated excellent reliability, with considerably high ICC values for interrater and test–retest reliability analysis (0.91 and 0.90, respectively).

In the second phase of the study, we examined the concordance between patient and proxy reports. Median

values of *SARC-F by patient* and *SARC-F by proxy* were both 2 (0-10), and the patient and proxy scores showed a moderate correlation, like most of the proxy assessments in the literature.[23-25] An important challenge about the proxy-reported questionnaires is that proxy reports are prone to demonstrate systematic differences and hence may not be interchangeable with self-reports all the time. This proxy bias (in other words, inter-rater gap) was associated with different factors. First of all, the nature of the relationship of the proxy, the frequency of the contact, and intimacy are highly important factors that influence the concordance between scores. In order to obtain concordant results, we specified an inclusion criterion of having frequent contact (contact on a weekly basis at least) with the patients or residents that allows the proxy insight into the individual's situation, for at least 1 year. Proxies of CD older adults were mostly their children having at least weekly contact, with spouses coming after, who are expected to give more consistent answers with the patients since they spend more time together. In fact, there is no standard threshold or definition for "frequent contact;" therefore, we adapted this definition from similar proxy-based reliability studies.[8] However, we necessarily elongated the period of contact to at least 1 year since the fifth item questions fall in the previous year, which is very prone to recall bias. In line with this, the item with least concordance between patients and proxies was found to be the fifth item.

Apart from intimate relationship, proxy respondents can also be selected for their professional capacity or skills to make judgments on behalf of patients. However, ratings from different types of proxies may not be interchangeable. In the studies assessing QoL, nurses or clinicians were found to overestimate QoL mostly, unlike family proxies who had a tendency of underestimating.[8] Similarly, although *SARC-F by patient* responses did not differ statistically, nurses scored significantly lower than family members/caregivers in community in our study, suggesting that healthcare professionals may really have a tendency to see the situation more positively than reality. Still, it is vague whether healthcare professionals' point of view is unrealistic or closer to the truth, while judgments of nurses might be more valid as they were probably more objective, they might also be insufficient since they (nurses) are not expected to have closer relationships with proxies than family members. It was also reported that there was a higher patient–proxy concordance for family members compared to healthcare professionals.[10] Supportive of this, the correlation coefficient was higher for the agreement between patients and proxies in community setting, compared to residents and nurses in NH, although responses from the proxies of different settings both demonstrated moderate agreement with patients in our study.

Another important factor influencing the concordance between patient–proxy results is the educational level of the respondents since higher educational level was associated with better agreement.[9] In our study, although proxies were mostly graduate, the education level of most of the patients was primary school, and this may also have prevented a stronger agreement. In addition, although we excluded the patients with moderate to severe dementia, and it is accepted that individuals with mild form of dementia can preserve their reasoning,[26] including patients with cognitive impairment, may have still affected the results since SARC-F also requires respondent to memorize falling episodes in previous year. Furthermore, it was also reported that whether proxy or patient, respondents' mood was also one of the determinants of proxy–patient correlation.[8] Likewise, caregiver burden may have also affected the results. A study analyzing the agreement of 135 dyads of patients and caregivers on QoL of patients with Alzheimer's disease reported that caregivers' burden and depression were 2 of the major factors associated with discrepancy in the results.[27] Unfortunately, we did not perform any assessment on depression, anxiety, or caregiver burden in our study.

Another possible reason for patient–proxy responses showing not strong but moderate agreement may be explained by the theory of "U-shaped relationship between self-proxy agreement and patients' health status." [8] In other words, the self-proxy agreement is estimated to be generally higher for patients in very good or very poor health status. It was reported that the interrater gap was smaller when the patients were more independent in ADL and had fewer neuropsychiatric symptoms, in a study assessing QoL in patients with Alzheimer's disease.[27] The middle part of the curve is estimated to be composed of individuals whose health status is not very bad but who can adapt to bearable adversities and thus ensure their well-being. However, an objective and well-observed proxy may detect and interpret certain findings as signs of poor health condition. In our study, although most of the patients were cognitively intact and independent in ADL and IADL, almost half of the patient population was pre-frail and had considerable rates of chronic diseases and polypharmacy. Therefore, this group of patients seems to fit more somewhere in between rather than at the ends of the above-mentioned curve and hence ending with not strong but moderate agreement between patient–proxy reports.

Another factor that is considered as an important determinant of correlation between patient and proxy reports is the internal consistency of the instrument used. It was reported that a study using an instrument with low internal consistency would not end up with high levels

of concordance between patient and proxy reports.[8] In our study, SARC-F by proxy demonstrated a high level of internal consistency, with a Cronbach's alpha value of 0.82. Hence, this feature of the test seems to be one of the factors that strengthened the agreement between patient and proxy reports.

In the literature, there is only 1 study reporting *SARC-F by proxy*'s reliability in older adults.[28] Maurus et al[28] included older adults from 2 different patient populations: (i) patients undergoing an inpatient geriatric rehabilitation for diverse medical conditions, and (ii) outpatients under surveillance for a rheumatological or hematological disease. Authors explained the reason for recruiting samples from 2 different settings as they intended to create a representative sample of older adults with different levels of functional impairments. They chose the proxies in at least weekly contact with the patient during the last 6 months and also included meeting the patient in person at least twice during the last 6 months. They defined 2 cohorts: proxies in cohort A responded SARC-F by proxy ad hoc (T1) and after 3 months (T2) (by making retrospective judgments about patients' condition at T1) and proxies in cohort B responded to the questions only at T2 (again, retrospective evaluation of patients' status at T1), in order to examine potential recall bias. Patients responded to *SARC-F by patient* only at admission (T1). They stated that they excluded patients with at least moderate cognitive impairment but did not mention whether they performed any examination of cognitive status of the patients. In total, they included 104 patients and 135 proxies and reported the interrater reliability between patient and proxy reports as substantial, with a κ value of 0.79. They also examined the agreement between patient and retrospective proxy reports, and they detected a substantial agreement in cohort A ($\kappa=0.61$) and a moderate agreement in cohort B ($\kappa=0.42$). Although 2 studies had methodological differences, reliability analyses of both studies show that SARC-F by proxy is a reliable tool for use in older adults and mainly shows moderate level of agreement between proxies and patient populations consisting of individuals from different settings (outpatients, hospitalized patients, and NH residents).

This study has certain limitations. Although we recruited participants from both CD and NH to increase the number of study population, the study population is not representative of whole older adult community, and findings cannot be generalized. Since we included older adults without cognitive impairment or at least mild dementia to ensure the reliability of the answers, we may have made a selection bias by creating a healthier population with mostly preserved functionality. Including participants from both settings can be considered as a limitation since the population is not homogenous but also a strength because it included older individuals from different functionality and comorbidity profiles. We tried to implement a proxy–patient perspective and asked the proxies to try to view the situation as the patient would (except for the fifth item of SARC-F by proxy) answer if they were the patients. However, this instruction may not have been fully understood by the proxies, and they may have just simply presented their own judgments (proxy–proxy perspective). In addition, we did not assess the proxies' cognitive status or mood, which could affect the reliability of the responses. One of the major strengths of this study is that it had sufficient number of participants for assessment of the test's reliability. Furthermore, we implemented proxy–patient perspective, which could promise a smaller interrater gap than proxy–proxy perspective.[14] To the best of our knowledge, this is one of the 2 studies in the literature examining the reliability of *SARC-F by proxy* in older adult population.

CONCLUSION

SARC-F is an important tool in sarcopenia case finding but has a limitation in application on patients with dementia or communication problems since it is a self-reported questionnaire. According to this study, SARC-F may be reliably applied to relatives or caregivers on behalf of the patients, in the name of *SARC-F by proxy*. Furthermore, *SARC-F by proxy* results showed a moderate correlation with *SARC-F by patient* scores. How well it can predict sarcopenia and other adverse outcomes will be revealed by future validity studies of *SARC-F by proxy*.

Ethics Committee Approval: The study was approved by the Istanbul University Istanbul Medical School ethical committee (Date: October 26, 2020, Decision No: 1503/2020).

Informed Consent: Written informed consent was obtained from all of the patients, nursing home residents and the proxies who participated in this study.

Peer-review: Externally peer-reviewed.

Author Contributions: Concept – S.Ö, G.B.; Design – S.Ö., G.B.; Supervision – H.D., M.A.K., G.B.; Resources – H.D., M.A.K., G.B.; Materials – S.Ö., C.Ö.A., H.Ö., C.K., Y.K., H.D.; Data collection and processing – S.Ö., C.Ö.A., H.Ö., C.K., Y.K., H.D.; Analysis and/or interpretation – S.Ö., M.M.Ö.; Literature search – S.Ö.; Writing manuscript – S.Ö.; Critical review – M.A.K., G.B.

Acknowledgements: The authors would like to warmly thank John E. Morley, Özlem Yılmaz, Kadriye Gümüş and İstanbul Kayısdağı Nursing Home staff for their contributions to this study.

Declaration of Interests: The authors declare that they have no conflicts of interest.

Funding: The authors declared that this study has received no financial support.

REFERENCES

1. Beaudart C, Zaaria M, Pasleau F, Reginster JY, Bruyère O. Health outcomes of sarcopenia: a systematic review and meta-analysis. *PLoS One*. 2017;12(1):e0169548. [CrossRef]
2. Cruz-Jentoft AJ, Bahat G, Bauer J, et al. Sarcopenia: revised European consensus on definition and diagnosis. *Age Ageing*. 2019;48(1):16-31. [CrossRef]
3. Chen LK, Woo J, Assantachai P, et al. Asian Working Group for Sarcopenia: 2019 Consensus update on sarcopenia diagnosis and treatment. *J Am Med Dir Assoc*. 2020;21(3):300-307.e2. [CrossRef]
4. Bahat G, Erdoğan T, İlhan B. SARC-F and other screening tests for sarcopenia. *Curr Opin Clin Nutr Metab Care*. 2022;25(1):37-42. [CrossRef]
5. Erbas Sacar D, Kilic C, Karan MA, Bahat G. Ability of SARC-F to find probable sarcopenia cases in older adults. *J Nutr Health Aging*. 2021;25(6):757-761. [CrossRef]
6. Voelker SN, Michalopoulos N, Maier AB, Reijnierse EM. Reliability and concurrent validity of the SARC-F and its modified versions: a systematic review and meta-analysis. *J Am Med Dir Assoc*. 2021;22(9):1864-1876.e16. [CrossRef]
7. Bahat G, Ozkok S, Kilic C, Karan MA. SARC-F questionnaire detects frailty in older adults. *J Nutr Health Aging*. 2021; 25(4):448-453. [CrossRef]
8. Rand S, Caiels J. *Using proxies to assess quality of life: A review of the issues and challenges*. 2015 [access date: 2022 November, 8]; Available from: https://www.pssru.ac.uk/pub/4980.pdf.
9. Arons AM, Krabbe PF, Schölzel-Dorenbos CJ, van der Wilt GJ, Rikkert MG. Quality of life in dementia: a study on proxy bias. *BMC Med Res Methodol*. 2013;13:110. [CrossRef]
10. Jones JM, McPherson CJ, Zimmermann C, Rodin G, Le LW, Cohen SR. Assessing agreement between terminally ill cancer patients' reports of their quality of life and family caregiver and palliative care physician proxy ratings. *J Pain Symptom Manage*. 2011;42(3):354-365. [CrossRef]
11. Lewis CJ, Maier F, Eggers C, et al. Parkinson's disease patients with subthalamic stimulation and carers judge quality of life differently. *Parkinsonism Relat Disord*. 2014; 20(5):514-519. [CrossRef]
12. Bolarinwa OA. Principles and methods of validity and reliability testing of questionnaires used in social and health science researches. *Niger Postgrad Med J*. 2015;22(4):195-201. [CrossRef]
13. Ngwayi JRM, Obie KU, Tan J, Xu J, Alizada M, Porter DE. Process of translation and adaptation of instruments. *J Orthop Surg Res*. 2022;17(1):508. [CrossRef]
14. Pickard AS, Knight SJ. Proxy evaluation of health-related quality of life: a conceptual framework for understanding multiple proxy perspectives. *Med Care*. 2005;43(5):493-499. [CrossRef]
15. Bahat G, Yilmaz O, Oren MM, et al. Cross-cultural adaptation and validation of the SARC-F to assess sarcopenia: methodological report from European Union Geriatric Medicine Society Sarcopenia Special Interest Group. *Eur Geriatr Med*. 2018;9(1):23-28. [CrossRef]
16. Tsang S, Royse CF, Terkawi AS. Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. *Saudi J Anaesth*. 2017;11(suppl 1):S80-S89. [CrossRef]
17. Katz S. Assessing self-maintenance: activities of daily living, mobility, and instrumental activities of daily living. *J Am Geriatr Soc*. 1983;31(12):721-727. [CrossRef]
18. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist*. 1969;9(3):179-186. [CrossRef]
19. Morris JC. Clinical dementia rating: a reliable and valid diagnostic and staging measure for dementia of the Alzheimer type. *Int Psychogeriatr*. 1997;9(suppl 1):173-176. [CrossRef]
20. Morley JE, Malmstrom TK, Miller DK. A simple frailty questionnaire (FRAIL) predicts outcomes in middle aged African Americans. *J Nutr Health Aging*. 2012;16(7):601-608. [CrossRef]
21. Rubenstein LZ, Harker JO, Salvà A, Guigoz Y, Vellas B. Screening for undernutrition in geriatric practice: developing the short-form mini-nutritional assessment (MNA-SF). *J Gerontol A Biol Sci Med Sci*. 2001;56(6):M366-M372. [CrossRef]
22. Hocaoglu MB, Gaffan EA, Ho AK. Health-related quality of life in Huntington's disease patients: a comparison of proxy assessment and patient self-rating using the disease-specific Huntington's disease health-related quality of life questionnaire (HDQoL). *J Neurol*. 2012;259(9):1793-1800. [CrossRef]
23. Kutner JS, Bryant LL, Beaty BL, Fairclough DL. Symptom distress and quality-of-life assessment at the end of life: the role of proxy response. *J Pain Symptom Manage*. 2006; 32(4):300-310. [CrossRef]
24. Sneeuw KCA, Sprangers MAG, Aaronson NK. The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease. *J Clin Epidemiol*. 2002;55(11):1130-1143. [CrossRef]
25. Tang ST. Concordance of quality-of-life assessments between terminally ill cancer patients and their primary family caregivers in Taiwan. *Cancer Nurs*. 2006;29(1):49-57. [CrossRef]
26. Hegde S, Ellajosyula R. Capacity issues and decision-making in dementia. *Ann Indian Acad Neurol*. 2016;19(suppl 1): S34-S39. [CrossRef]
27. Zucchella C, Bartolo M, Bernini S, Picascia M, Sinforiani E. Quality of life in Alzheimer disease: a comparison of patients' and caregivers' points of view. *Alzheimer Dis Assoc Disord*. 2015;29(1):50-54. [CrossRef]
28. Maurus J, Terzer T, Benner A, et al. Validation of a proxy-reported SARC-F questionnaire for current and retrospective screening of sarcopenia-related functional impairments. *J Cachexia Sarcopenia Muscle*. 2022;13(1):264-275. [CrossRef]

Supplementary Table 1. SARC-F by proxy		
Component	**Question**	**Score**
1-Strength	Does your patient have difficulty in lifting and carrying a 5 kg weight?	No=0 Some=1 A lot or not able=2
2-Assistance in walking	Does your patient have difficulty in walking in a room?	No=0 Some=1 A lot/ with assistance/not able=2
3-Stand up from a chair	Does your patient have difficulty while standing up from a chair or a bed?	No=0 Some=1 A lot or not able without help=2
4-Climbing up the stairs	Does your patient have difficulty while climbing up a 10 stairs?	No=0 Some=1 A lot or not able=2
5-Falls	How many times has your patient fallen in the last year?	None=0 1-3 times=1 4 times or more=2
Screening score	Score ≥4 suggests sarcopenia	