Received: 6 December 2022 | Accepted: 25 December 2022

DOI: 10.1097/HC9.0000000000000075

ORIGINAL ARTICLE

OPEN

The Steatosis-associated fibrosis estimator (SAFE) score: validation in the general US population

Laurens A. van Kleef | **Robert J. de Knegt** | **Ibrahim Ayada** | **Qiuwei Pan** | **Willem Pieter Brouwer**

Department of Gastroenterology and Hepatology, Erasmus MC University Medical Centre, Rotterdam, The Netherlands

Correspondence
Laurens A. van Kleef, Department of Gastroenterology and Hepatology, Erasmus MC University Medical Centre, Rotterdam, The Netherlands.
Email: l.vankleef@erasmusmc.nl

Abstract

Background: Noninvasive tests are important in the initial risk stratification of people at risk of fibrosis. The recently developed steatosis-associated fibrosis estimator (SAFE) score may have such potential but awaits external validation.

Methods: We analyzed 6973 participants aged 18–80 in the National Health and Nutrition Examination Survey 2017-2020 cycle with data on liver stiffness and SAFE score without prevalent heart failure. Fibrosis was defined as liver stiffness $\geq$8.0 kPa. Accuracy was evaluated by AUC and assessment of test characteristics at the prespecified cutoffs for ruling out/ruling in fibrosis.

Results: The SAFE score categorized 14.7% of the population as high risk, 30.4% as intermediate risk, and 54.9% as low risk for fibrosis. The actual fibrosis prevalence in these groups was 28.0%, 10.9%, and 4.0%, respectively, translating into a positive predictive value of 0.28 at the high-risk cutoff and a negative predictive value of 0.96 at the low-risk cutoff. The AUC of the SAFE score (0.748) was significantly higher than the fibrosis-4 index (0.619) or NAFLD fibrosis score (0.718). However, test performance strongly depended on age categories: 90% of participants aged 18–40 years were considered at low risk for fibrosis, including 89/134 (66%) of clinically significant fibrosis cases. In the oldest group (60–80 y), fibrosis could only be safely ruled out among 17%, corresponding to a high referral rate of up to 83%. The best SAFE score performance was found in the middle-aged group (40–60 y). The results were consistent in target populations with metabolic dysfunction or steatosis.

Abbreviations: AUC, area under the receiver characteristic operator curve; ALT, alanine aminotransferase; AST, aspartate aminotransferase; DOR, diagnostic OR; FIB-4, fibrosis-4 index; HDL-C, HDL-cholesterol; MAFLD, metabolic dysfunction associated fatty liver disease; NHANES, National Health and Nutrition Examination Survey; NHCS, National Center for Health Statistics; NFS, NAFLD fibrosis score; NPV, negative predictive value; P25-P75, 25th to 75th percentile; PPV, positive predictive value; SAFE, steatosis-associated fibrosis estimator.

Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website, www.hepcommjournal.com.

Conclusions: The SAFE score has overall good diagnostic accuracy in detecting fibrosis but was highly dependent on age. The SAFE score lacked sensitivity in younger populations and the ability to rule out fibrosis in older populations.

INTRODUCTION

Fatty liver disease is increasingly prevalent, estimated to affect 1 in 3 adults globally.[1] It has become one of the major causes of advanced liver disease, a leading cause of HCC also in noncirrhotic livers, and is already the second leading indication for liver transplantation in the US.[2] The rising obesity and fatty liver epidemic combined with an aging population may lead to a surge in patient referrals for specialist care and may lead to a dramatic increase in health care costs. Therefore, it is imperative to identify patients at high risk for advanced liver disease who may benefit from hepatologist consultation. Despite several noninvasive tests being available, it remains challenging to identify these patients either because of their poor performance, costs, limited availability, or dependence on age.[3,4]

Recently, the steatosis-associated fibrosis estimator (SAFE) score was published by Sripongpun et al[5], as a means to detect clinically significant fibrosis in a noninvasive manner. A defining feature of this score is the inclusion of globulins, which has previously been used together with platelets in the prediction of fibrosis in chronic hepatitis B patients.[6] The SAFE score, in addition to globulins, included age, body mass index (BMI), diabetes, aspartate aminotransferase, alanine aminotransferase, and platelets.

The SAFE-score yielded good performance in their training [area under the receiver characteristic operator curve (AUC) 0.79] and testing sets (AUC 0.80 and 0.83). However, this score was developed among a selected NAFLD population in which clinically significant fibrosis was highly prevalent (45%), contrasting the general population for which this score was actually designed for. Moreover, no subgroup analyses have been reported, probably because of the limited sample size of the training and testing sets (n = 130–676). In the National Health and Nutrition Examination Survey (NHANES) III, the authors promisingly demonstrated the SAFE score being a predictor for excess mortality. However, whether this was attributed to fibrosis is unclear, since no data were available for fibrosis.

Hence, we aimed to externally validate this SAFE score in a nonhospital-based, multiethnic community-dwelling population cohort using the 2017-2020 NHANES cycle (independent of NHANES III), which included transient elastography outcomes. Moreover, we aimed to validate this score across a range of clinically relevant subgroups.

METHODS

Study population

This study was a secondary analysis of existing data performed within the NHANES study. The NHANES was designed to investigate individuals' health and nutritional status throughout the US. In short, across all study participants, extensive data on health and nutrition were collected by interview, physical examination, and a range of clinical measurements and tests (including transient elastography). Detailed information regarding the procedures and rationale has been described elsewhere.[7,8] For the specific purpose of this study, we excluded participants with incomplete data on the SAFE score components or with a history of heart failure and, therefore, potentially unreliable liver stiffness measurement.[9] Data are publicly available from the NHANES database (https://www.cdc.gov/nchs/nhanes/index.htm).

Biomarker-based noninvasive scores

We calculated the SAFE, NAFLD fibrosis score (NFS), and Fibrosis-4 index (FIB-4) according to their formulas.[5,10,11] Of note, BMI levels > 40 were truncated at 40, aligning with the formula instructions.

$$\begin{aligned}\text{SAFE} = {} & 2.97 * \text{Age} + 5.99 * \text{BMI} \\ & + 154.85 * \ln(\text{AST}) - 58.23 * \ln(\text{ALT}) \\ & + 195.48 * \ln(\text{globulin}) - 141.61 * \ln(\text{platelets}) \\ & + 62.85 \text{ IF diabetes} - 75\end{aligned}$$

$$\begin{aligned}\text{NFS} = {} & -1.675 + 0.037 * \text{Age} + 0.094 * \text{BMI} \\ & + 1.13 \text{ IF (pre)diabetes} + 0.99 * \frac{\text{AST}}{\text{ALT}} \\ & - 0.013 * \text{platelets} - 0.66 * \text{albumin}\end{aligned}$$

$$\text{FIB4} = \frac{\text{Age} * \text{AST}}{\text{platelets} * \sqrt{\text{ALT}}}$$

Age was expressed in years, BMI in kg/m^2, aspartate aminotransferase and alanine aminotransferase in U/L, platelets as 10^9/L, and albumin and globulin in g/dL.

Transient elastography

Participants underwent transient elastography using the FibroScan model 502 V2 Touch equipped with an M and XL probe (FibroScan, Echosens, Paris) to assess liver stiffness. Participants were requested to fast for at least 3 hours. Measurements were considered valid if at least 10 measurements were obtained with an IQR <30%. Clinically significant fibrosis was defined as liver stiffness ≥8.0 kPa.[12] Steatosis was assessed by a same-session controlled attenuation parameter (CAP) measurement. Controlled attenuation parameter levels ≥275 dB/m were used to diagnose steatosis since it had over 90% of sensitivity.[13] Additional details about the procedures are described.[14,15]

Covariates

Research assistants systematically collected data among all participants, including age, race, and anthropometrics (length, height, and waist circumference). Questionnaires included questions on the presence of heart failure and alcohol consumption. Excessive alcohol consumption was defined as >10 g/d in women and >20 g/d in men. This cutoff was already associated with increased all-cause mortality in the NHANES III data set.[16] Blood samples were taken and analyzed for among others alanine aminotransferase, aspartate aminotransferase, albumin, globulin, platelets, triglycerides, and HDL-C. Metabolic dysfunction was defined according to the metabolic dysfunction–associated fatty liver disease (MAFLD) criteria: overweight, diabetes, or at least 2 minor criteria (eg, hypertension, high waist circumference, and dyslipidemia).[17]

Statistical analysis

First, we assessed the diagnostic accuracy for the presence of clinically significant fibrosis according to the SAFE score in the overall population. Besides visualizing the clinically significant fibrosis rate per SAFE score category, this assessment included an AUC analysis. Next, we investigated the diagnostic accuracy across subgroups, which were based on age (18–40, 40–60, and 60–80 y), sex (male/female), race (White, Black, or Hispanic), excessive alcohol consumption (Yes/No), steatosis (Yes/No), and metabolic dysfunction (Yes/No).

Second, we analyzed the SAFE score performance at the prespecified cutoffs (low risk of fibrosis with a SAFE score of <0 or high risk with a score of ≥100 and intermediate for 0–100), according to the previous publication.[5] Using these prespecified cutoffs, we investigated the pretest and posttest prevalence of clinically significant fibrosis across the entire population and the aforementioned clinically relevant subgroups. In addition, we assessed the sensitivity, specificity, positive predictive value, negative predictive value (NPV), and diagnostic odds ratio at the outer borders of the SAFE score (0 and 100), according to the prespecified cutoffs. Sensitivity analyses included an assessment of SAFE score performance among the age categories in target populations: (1) metabolic dysfunction and (2) steatosis.

Third, we visualized the distribution of low risk, intermediate, and high risk of clinically significant fibrosis, according to the SAFE score per age category (of 5 y), which was then plotted relative to the rate of clinically significant fibrosis as assessed by liver stiffness ≥8.0 kPa. The additional analysis included the same visualization for the FIB-4 and the NFS.

Finally, we compared the performance of the SAFE score in the overall population with the performance of the FIB-4 and NFS. AUCs were assessed for significant differences using the DeLong test. We did not use the provided weights of the NHANES, as this study was designed not to study prevalences but to test the SAFE score performance across subgroups. Analyses were performed in R version 4.0.4 (Foundation for Statistical Computing, Vienna, Austria. p-values < 0.05 were considered statistically significant).

Ethics

NHANES was approved by the National Centre for Health Statistics research ethics review board, and all the participants provided informed consent. This study was conducted according to the principles as set forth in the Declaration of Helsinki and Istanbul. All authors had access to the study data and reviewed and approved the final manuscript.

RESULTS

Participants

We included 7768 adults of the NHANES 2017-2020 cycle, with available data on liver stiffness. Of them, 202 were excluded for the presence of heart failure and 593 for missing data on the individual components of the SAFE score, leaving 6973 participants for analysis. The median age was 49 years (33–63), 49.0% were male, and 73.1% had a BMI ≥25 kg/m^2. The median liver stiffness was 5.0 kPa (4.1–6.1) and exceeded ≥8.0 kPa in 9.6% (n = 672). The SAFE score was categorized according to the prespecified cutoffs (≤0, 0–100, and ≥100) and indicated low risk in 3.831 (54.9%), intermediate risk in 2.119 (30.4%) and high risk in 1.023 (14.7%) participants. Clinically significant fibrosis prevalence in these groups was 4.0%, 10.9%, and 28.0%, respectively. Detailed baseline characteristics for these subgroups are available in Table 1.

TABLE 1 Participants' characteristics stratified for SAFE score categories

	All n = 6973	SAFE < 0 n = 3831	SAFE 0–100 n = 2119	SAFE ≥ 100 n = 1023
Demographics				
Age (y)	49 (33, 63)	36 (26, 48)	60 (51, 68)	67 (59, 75)
Male	3418 (49.0)	1793 (46.8)	1056 (49.8)	569 (55.6)
Race				
Mexican-American	1634 (23.4)	954 (24.9)	460 (21.7)	220 (21.5)
Non-Hispanic Black	1729 (24.8)	812 (21.2)	578 (27.3)	339 (33.1)
Non-Hispanic White	2401 (34.4)	1315 (34.3)	728 (34.4)	358 (35.0)
Other	1209 (17.3)	750 (19.6)	353 (16.7)	106 (10.4)
College	3858 (58.1)	2162 (61.6)	1192 (56.4)	504 (49.4)
Current smoking	1198 (17.2)	740 (19.3)	316 (14.9)	142 (13.9)
Metabolic health				
BMI ≥ 25	5094 (73.1)	2472 (64.5)	1735 (81.9)	887 (86.7)
Diabetes	1224 (17.6)	147 (3.8)	527 (24.9)	550 (53.8)
Biochemistry				
AST (U/L)	19 (16, 24)	18 (15, 22)	20 (17, 25)	23 (19, 32)
ALT (U/L)	18 (13, 26)	17 (12, 24)	18 (14, 26)	21 (15, 32)
HDL-C (mmol/L)	1.4 (0.4)	1.4 (0.4)	1.4 (0.4)	1.4 (0.4)
Triglycerides (mmol/L)	1.3 (0.9, 1.8)	1.1 (0.8, 1.7)	1.4 (1.0, 2.0)	1.4 (1.0, 2.0)
Globulin (g/L)	30.9 (4.3)	30.1 (3.8)	31.1 (4.2)	32.9 (5.2)
Albumin (g/L)	40.8 (3.3)	41.4 (3.3)	40.3 (3.0)	39.8 (3.4)
Platelets (10^9/L)	247 (65)	267 (64)	236 (56)	197 (51)
Transient elastography				
Liver stiffness (kPa)	5.0 (4.1, 6.1)	4.7 (3.9, 5.6)	5.2 (4.3, 6.4)	6.0 (4.7, 8.3)
Clinically significant fibrosis	672 (9.6)	154 (4.0)	232 (10.9)	286 (28.0)
CAP (dB/m)	263 (62)	247 (59)	278 (60)	290 (62)
Steatosis	2899 (41.6)	1226 (32.0)	1066 (50.3)	607 (59.3)

Note: Data are presented as mean (SD), median (P25–P75) or n and percentage. Standard International (SI) units are used, which may differ from the units used in the formulas for the noninvasive tests. Fibrosis was defined as liver stiffness ≥ 8.0 kPa and steatosis as CAP ≥ 275.
Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; CAP, controlled attenuation parameter; HDL-C, high density lipoprotein cholesterol; SAFE, steatosis-associated fibrosis estimator.

Overall SAFE score performance

In the overall population, the SAFE score yielded an AUC of 0.748 (95% CI, 0.728–0.768) in the detection of clinically significant fibrosis. Similar performance was obtained in the target populations with steatosis (AUC: 0.704) or metabolic dysfunction (AUC: 0.728). Higher SAFE scores were increasingly associated with a higher prevalence of liver stiffness ≥ 8.0 kPa among all participants, as shown in Figure 1. Interestingly, scores ≥ 200 were associated with 40%–70% risk of clinically significant fibrosis but were relatively uncommon (2.5%).

SAFE score performance according to age

Next, the performance of the SAFE score was analyzed according to the prespecified cutoffs: <0 for ruling out, 0–100 meaning intermediate, and ≥ 100 for ruling in liver clinically significant fibrosis. It was observed that the distribution of the SAFE score categories differed substantially across the different age groups (Table 2). Clinically significant fibrosis was ruled out in 91.3% of young participants, while in contrast, this was only 16.5% among participants aged 60–80. Importantly, the posttest clinically significant fibrosis prevalence within the low-risk (3.6%–5.3%), intermediate (9.0%–19.4%), and high-risk groups (24.3%–43.5%) varied significantly for the age subgroups. Moreover, of the total clinically significant fibrosis cases (n = 134) in the young group, 89 (66%) resided in the SAFE < 0 (ruled out) group and were thus missed. Depicted in Table 3 are additional markers for diagnostic accuracy at the prespecified cutoffs. Aligning previous outcomes, the sensitivity among young participants was low (34% low cut-off, 7% high cut-off), whereas, in the elderly, the specificity was poor (18% low cutoff and 71% high cutoff). The

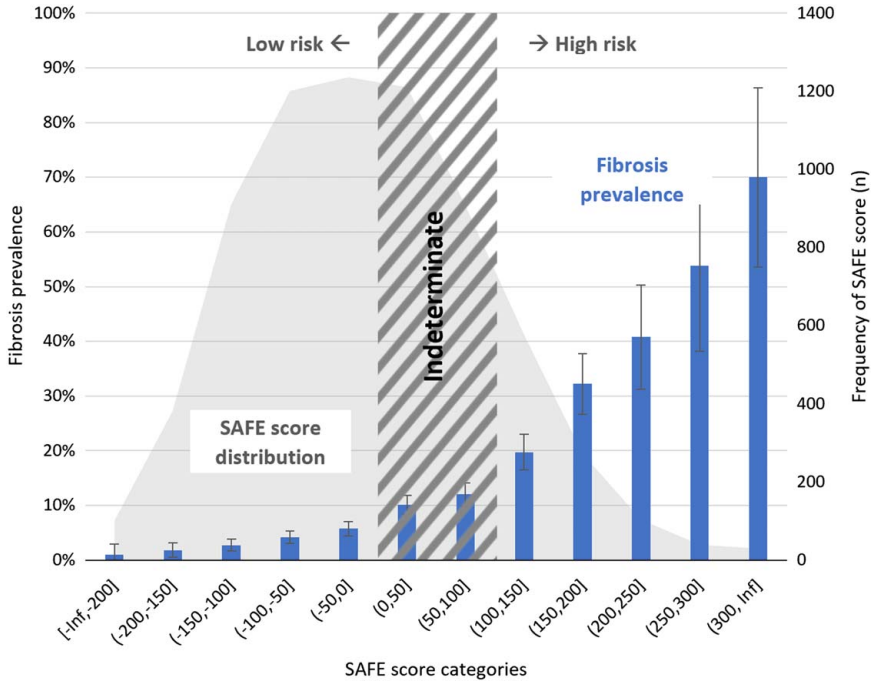

FIGURE 1 Observed clinically significant fibrosis prevalence per SAFE score group. The bars represent clinically significant fibrosis prevalence with 95% CI as assessed by liver stiffness ≥8.0 kPa. The gray area is the distribution of participants within the SAFE score category. The shaded bar covering SAFE scores between 0 and 100, illustrates the area in which the SAFE score remains intermediate according to these prespecified cutoffs. Abbreviations: SAFE, steatosis-associated fibrosis estimator.

results were consistent when diagnostic performance was investigated across age subgroups among participants with metabolic dysfunction or steatosis (Supplemental Table 1, http://links.lww.com/HC9/A169).

Focusing on age as an important factor in the SAFE score performance, we assessed the distribution of this score according to age groups. As shown in Figure 2, the proportion of SAFE scores predicting a low risk of clinically significant fibrosis attenuates from >95% in the youngest group (18–25 y) to < 10% in the oldest subgroup (75–80 y). It should be noted that this subsequent major increase for the "at high risk for fibrosis" subgroup by increasing age does not correspond with the prevalence of high liver stiffness for the same age groups (illustrated by the green line).

SAFE score performance according to different subgroups

The performance was good in participants with metabolic dysfunction, but among those without metabolic dysfunction (that had a low clinically significant fibrosis prevalence of 3.4%), the sensitivity was poor (26% high cutoff, 52% low cutoff; Table 3). Interestingly, the ability to rule out clinically relevant fibrosis was especially good in patients with excessive alcohol consumption (NPV 0.98). In the other clinically relevant subgroups (race and sex), the posttest clinically significant fibrosis prevalence in the indeterminate group across the remaining subgroups was similar to the pretest prevalence (Table 2). In addition, the performance of the SAFE score was more comparable to the overall performance (Table 3). Importantly, the performance of the SAFE score was not dependent on the prevalence of metabolic dysfunction or steatosis, illustrated by similar results across subgroups after excluding participants without metabolic dysfunction or steatosis (Supplemental Table 1, http://links.lww.com/HC9/A169).

Performance of SAFE score versus FIB-4 or NFS

Finally, we compared the SAFE score to the FIB-4 and NFS. In the overall population, the SAFE score (AUC 0.748) significantly outperformed the FIB-4 (AUC: 0.619) and NFS (AUC: 0.718) based on the DeLong test (both $p < 0.001$). The performance across subgroups has been depicted in Supplemental Table 2, http://links.lww.com/HC9/A169. In general, the FIB-4 has low predictive value, even among the middle aged (40–60 y) for which it was designed, whereas the NFS had better performance across subgroups. However, the NFS had (similar to the SAFE score) great differences in test characteristics across age groups while maintaining reasonable AUCs. Moreover, the NFS, like the SAFE score, had a large proportion (30%) in which the fibrosis risk remained indeterminate using the

TABLE 2 Pretest clinically significant fibrosis prevalence in the entire population and posttest clinically significant fibrosis prevalence in SAFE score risk categories

	Pretest		SAFE < 0		SAFE 0–100		SAFE ≥100	
	n	Fibrosis	%	Fibrosis	%	Fibrosis	%	Fibrosis
All	6973	9.6	54.9	4.0	30.4	10.9	14.7	28.0
Age								
18–40	2416	5.6	91.3	4.0	7.7	19.4	1.0	43.5
40–60	2281	10.0	54.7	3.6	34.8	11.7	10.5	38.1
60+	2276	13.5	16.5	5.3	50.0	9.0	33.4	24.3
Race								
White	2401	9.9	54.8	4.5	30.3	11.3	14.9	26.8
Black	1729	10.5	47.0	4.6	33.4	9.7	19.6	26.0
Hispanic	1634	10.1	58.4	3.9	28.2	12.8	13.5	31.4
Sex								
Male	3418	11.2	52.5	4.9	30.9	12.7	16.6	28.6
Female	3555	8.1	57.3	3.3	29.9	9.2	12.8	27.1
Excessive alcohol								
Yes	747	9.0	57.3	2.1	25.7	9.4	17.0	31.5
No	5898	9.8	54.5	4.4	31.0	11.0	14.5	27.4
Metabolic dysfunction								
Yes	5720	11.0	48.4	4.8	34.5	11.2	17.1	28.1
No	1253	3.4	84.6	1.9	11.8	7.4	3.6	24.4
Steatosis								
Yes	2899	16.7	42.3	7.8	36.8	16.9	20.9	34.3
No	4073	4.6	63.9	2.2	25.9	4.9	10.2	18.8

Note: Data are presented as n or percentage. The posttest fibrosis prevalence is depicted for SAFE <0 (low risk), SAFE 0–100 (indeterminate, and SAFE ≥100 (high risk). Clinically significant fibrosis was defined as liver stiffness ≥8.0 kPa and steatosis as controlled attenuation parameter ≥275.
Abbreviations: SAFE, steatosis-associated fibrosis estimator.

prespecified cutoffs of <−1.455 for ruling out and >0.675 for ruling in fibrosis. Finally, similar to the SAFE score, the high-risk group, according to the FIB-4 and NFS, increased substantially with age, disconcordant with the clinically significant fibrosis prevalence.

DISCUSSION

We investigated the performance of the SAFE score in the detection of patients at high risk of clinically significant fibrosis in the general population and demonstrated that the SAFE score has an overall good performance, especially among middle-aged individuals (aged 40–60) but lacks performance in younger (aged 18–40) and the ability to rule out clinically significant fibrosis in older (aged 60–80) populations. The test characteristics were comparable in target populations with metabolic dysfunction or steatosis, in which noninvasive assessment of fibrosis is indicated.

Identifying patients at risk of clinically significant fibrosis remains challenging. There is an urgent need for readily available, inexpensive, and reliable noninvasive tests in the assessment of liver health, which is recommended in individuals with metabolic risk factors.[13,18] Our comprehensive external validation of the newly presented SAFE score indicates that the AUC derived in the training and small testing sets (AUC: 0.79–0.83) was significantly higher than what we obtained in this large population-based cohort (AUC: 0.748, 95% CI, 0.728–0.768).[5] Nonetheless, the SAFE score yielded significantly better AUCs than their direct competitors, such as the FIB-4 and NFS, indicating the good performance of the SAFE score in the overall population.

The discriminative value of the SAFE score across subgroups was good (as illustrated by the AUC) and not negatively affected by no strict exclusion criteria for alcohol use. However, with increasing age, there was a gradual decrease in the proportion categorized as low risk, starting at >95% in participants aged 18–25 years to <10% in participants aged 75–80 years, similar to other age-dependent indices, as shown in this cohort and by previous studies.[19,20] Hence, it is not surprising that the SAFE score was associated with mortality in the NHANES III.[5] This dependence on age resulted in a poor 7% sensitivity at the high cutoff and 34% at the low cutoff in the youngest subgroup (aged 18–40). Consequently, 93% of young participants with clinically

TABLE 3 Diagnostic accuracy of the SAFE score in ruling out and ruling in clinically significant fibrosis

	Ruled out: SAFE score <0						Ruled in: SAFE score ≥100					
	n	Fibrosis	Sens	Spec	NPV	DOR	n	Fibrosis	Sens	Spec	PPV	DOR
All	3831	154	77	58	0.96	0.21	1023	286	43	88	0.28	5.59
Age groups												
18–40	2207	89	34	93	0.96	0.15	23	10	7	99	0.43	13.96
40–60	1248	45	80	59	0.96	0.17	239	91	40	93	0.38	8.48
60–80	376	20	94	18	0.95	0.31	761	185	60	71	0.24	3.63
Race												
White	1315	59	75	58	0.96	0.24	358	96	41	88	0.27	4.94
Black	812	37	80	50	0.95	0.26	339	88	49	84	0.26	4.89
Hispanic	954	37	78	62	0.96	0.17	220	69	42	90	0.31	6.27
Sex												
Male	1793	87	77	56	0.95	0.23	569	163	42	87	0.29	4.77
Female	2038	67	77	60	0.97	0.20	454	123	43	90	0.27	6.61
Excessive alcohol												
Yes	428	9	87	62	0.98	0.10	127	40	60	87	0.31	10.10
No	3216	142	75	58	0.96	0.24	854	234	41	88	0.27	5.17
Metabolic dysfunction												
Yes	2771	134	79	52	0.95	0.25	978	275	44	86	0.28	4.83
No	1060	20	52	86	0.98	0.15	45	11	26	97	0.24	12.28
Steatosis												
Yes	1226	96	80	47	0.92	0.28	607	208	43	83	0.34	3.81
No	2604	58	69	66	0.98	0.23	416	78	41	91	0.19	7.44

Note: Diagnostic accuracy for the SAFE score at the thresholds to rule out (SAFE <0) or rule in (SAFE ≥100) fibrosis. Fibrosis was defined as liver stiffness ≥8.0 kPa and steatosis as controlled attenuation parameter ≥275.
Abbreviations: DOR, diagnostic OR; NPV, negative predictive value; PPV, positive predictive value; SAFE, steatosis-associated fibrosis estimator.

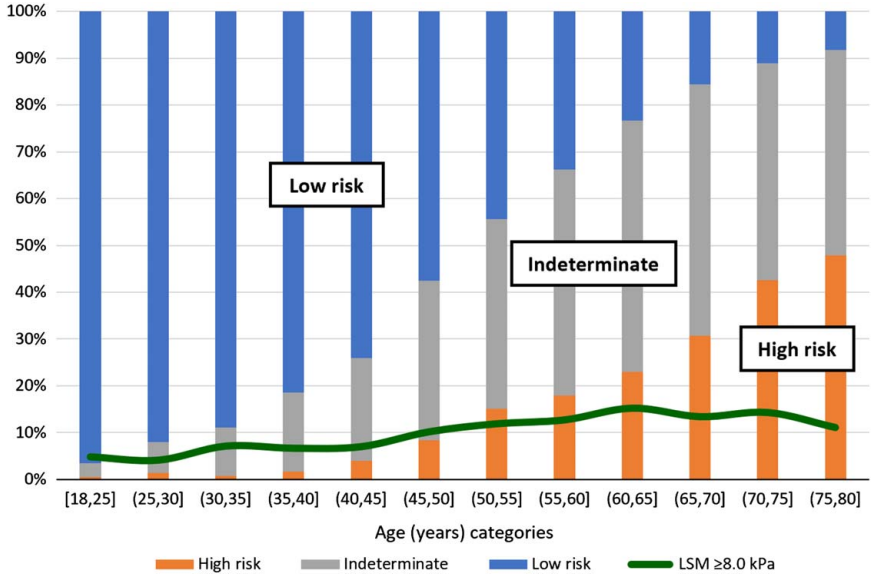

FIGURE 2 SAFE score predicted intermediate-risk/high-risk groups increase with older age, in contrast to observed clinically significant fibrosis prevalence. Blue reflects predicted low risk (SAFE <0), gray intermediate (SAFE 0-100), and orange high risk (SAFE ≥100) for the presence of fibrosis. The green line reflects the observed clinically significant fibrosis prevalence assessed by liver stiffness ≥8.0 kPa.
Abbreviations: LSM, Liver stiffness measurement; SAFE, steatosis-associated fibrosis estimator.

significant fibrosis were not classified as high risk for clinically significant fibrosis. Of note, the poor performance among young participants was not explained by the prevalence of metabolic dysfunction or steatosis. Missing out on clinically significant fibrosis in this young group is worrisome because they might benefit most from early detection, given the life span ahead.

Another concern is that clinically significant fibrosis was only considered low risk (and could thus be ruled out) in 17% of the participants aged 60–80, resulting in high referral rates (up to 83%) for additional tests among the elderly. This poor performance in the oldest subgroup is a significant problem, as there is an increasingly aging population with steatosis wrongly referred leading to increasing and potentially high health care costs, whereas the clinical relevance of (advanced) fatty liver disease is disputed.[21,22] Hence, this category of patients should better be stratified according to their risk of fibrosis and health care use and possible preventable societal costs. Therefore, although there is an apparent benefit of using this SAFE score over the FIB-4 and NFS in terms of the whole range of participants, further optimization and/or development of risk scores with consistent performance across age groups is of utmost importance. However, regardless of which score is being used, a holistic assessment of individual patients should drive referral decisions.

The SAFE score yields high NPVs (96% for the entire population) and is $>90\%$ across all subgroups. However, one should take into consideration that the NPV for relatively rare diseases is always high. In fact, the NPV for a test that randomly selects patients, like tossing a coin, is 100% minus the disease prevalence. This is especially important for population-based studies, given the lower prevalence of disease compared with hospital populations. For example, when fibrosis has a prevalence of ~10% as in this study, a nondiscriminative test like tossing a coin would result in a high NPV of 90%. Hence, it is important to go beyond the NPV to assess test performance, for example, focusing on referral rates and clinical utility in different subgroups. Nonetheless, the NPV of the SAFE score (96%) still convincingly illustrates the rather good ability to rule out clinically significant fibrosis among the 60% of the population with a SAFE score of <0, which exceeded the NPV of the FIB-4 or NFS in this cohort (data not shown).

Another significant drawback of the SAFE score (similar to other non-invasive tests like the NFS) is the large proportion (30%) of this community-dwelling population scoring between 0 and 100, which thus fell in the intermediate group. This was even higher among participants with metabolic dysfunction (35%) or steatosis (37%), in which noninvasive tests are recommended to assess liver health.[13] Given the large proportion of the population that fell in this intermediate group, this may result in a tremendous amount of referrals (up to 83% in the elderly), likely exceeding the current health care capacity in already strained systems.

Future studies should evaluate whether the SAFE score could be improved by accounting for age in a different manner. For example, potential improvements may lay in (1) recalibrating the weight attributed to age and considering the use of natural cubic splines to account for the different impact of age on fibrosis risk across age categories; (2) truncating age at, for example, a minimum of 40 years and a maximum of 60 years, likewise was done for BMI in the SAFE score formula; (3) introducing age-specific cutoffs such as previously suggested for the FIB-4 and NFS, which also included age.[19] Although beyond the aim of this study, overcoming the differences in performance and interpretation of the SAFE score across age strata could further improve its utility in referral strategies and decision-making, avoiding potentially increasing health care costs.

Limitations

The following limitations need to be mentioned. First, validating noninvasive tests for fibrosis using transient elastography instead of histology may, at first glance, seem suboptimal. However, current referral strategies rely on transient elastography as a validation test in case FIB-4 or NFS are suggestive of fibrosis.[13,18] Thus, in clinical practice, an initial test result indicating high risk for fibrosis will likely result in validation with transient elastography and not directly in performing a liver biopsy. Therefore, it makes sense to validate noninvasive tests designed to select participants for referral using transient elastography. However, we concede that liver stiffness can be increased attributable to causes other than fibrosis, such as venous congestion.[9] Hence, we ruled out participants with a history of heart failure in this external validation. However, residual impact of venous congestion on liver stiffness could not be ruled out. Second, our study aimed to investigate the SAFE score performance in specific subgroups, and therefore, the weights to simulate the US population were not applied. As a consequence, the overall performance, when applied to the general population, might be slightly different. Finally, it should be noted that the FIB-4 and NFS are actually designed to detect F3 fibrosis unlike the SAFE score, which was designed to detect F2 fibrosis. This may have favored the SAFE score performance over the FIB-4 and NFS.

CONCLUSIONS

In conclusion, in this external validation study, the SAFE score has good diagnostic accuracy in identifying individuals with high liver stiffness, suggestive of fibrosis, especially among individuals aged 40–60 years. However, the SAFE score lacked the ability to detect clinically significant fibrosis in younger populations and the ability to rule out clinically significant fibrosis in older

populations, potentially leading to a high but not-justified referral rate in the elderly with a negative impact on health care costs. Nonetheless, the performance of the SAFE score exceeded other commonly used noninvasive tests and should, in our opinion, better be used instead of the FIB-4 and NFS as a tool to determine who requires further liver health evaluation.

AUTHOR CONTRIBUTIONS

Laurens A. van Kleef and Willem Pieter Brouwer: collection of data; Laurens A. van Kleef and Willem Pieter Brouwer: study design, data analysis, and writing of the manuscript;

ACKNOWLEDGMENTS

The authors thank the contribution of the participants of the NHANES cohort, research assistants, and facilitating personnel.

FUNDING INFORMATION

Financial support was provided by the Foundation for Liver and Gastrointestinal Research, Rotterdam, the Netherlands. The funding source did not influence study design, data collection, analysis and interpretation of the data, nor the writing of the report and decision to submit for publication.

CONFLICTS OF INTEREST

Robert J. de Knegt is a speaker for Echosens, a consultant for AbbVie, and received grants from Abbvie, Gilead, and Janssen. The remaining authors have no conflicts to report.

ORCID

Laurens A. van Kleef https://orcid.org/0000–0002–2333–1182
Robert J. de Knegt https://orcid.org/0000–0003–0934–6975
Ibrahim Ayada https://orcid.org/0000–0002–8197–3458
Qiuwei Pan https://orcid.org/0000–0001–9982–6184
Willem Pieter Brouwer https://orcid.org/0000–0001–8713–1481

REFERENCES

1. Le MH, Yeo YH, Li X, Li J, Zou B, Wu Y, et al. 2019 Global NAFLD prevalence: a systematic review and meta-analysis. Clin Gastroenterol Hepatol. 2021. doi:10.1016/j.cgh.2021.12.002. Epub ahead of print.
2. Burra P, Becchetti C, Germani G. NAFLD and liver transplantation: disease burden, current management and future challenges. JHEP Rep. 2020;2:100192.
3. Graupera I, Thiele M, Serra-Burriel M, Caballeria L, Roulot D, Lai-Hung Wong G, et al. Low accuracy of FIB-4 and NAFLD fibrosis scores for screening for liver fibrosis in the population. Clin Gastroenterol Hepatol. 2021. doi:10.1016/j.cgh.2021.12.034. Epub ahead of print.
4. van Kleef LA, Sonneveld MJ, de Man RA, de Knegt RJ. Poor performance of FIB-4 in elderly individuals at risk for chronic liver disease—implications for the clinical utility of the EASL NIT guideline. J Hepatol. 2022;76:2526–8.
5. Sripongpun P, Kim WR, Mannalithara A, Charu V, Vidovszky A, Asch S, et al. The steatosis-associated fibrosis estimator (SAFE) score: a tool to detect low-risk NAFLD in primary care. Hepatology. 2022. doi:10.1002/hep.32545.
6. Liu XD, Liu X-D, Wu J-L, Liang J, Zhang T, Sheng Q-S. Globulin-platelet model predicts minimal fibrosis and cirrhosis in chronic hepatitis B virus infected patients. World J Gastroenterol. 2012;18:2784–92.
7. Johnson CL, Paulose-Ram R, Ogden CL, et al. National health and nutrition examination survey: analytic guidelines, 1999-2010. Vital Health Stat 2. 2013:1–24.
8. Akinbami LJ, Chen T-C, Davy O, Ogden CL, Fink S, Clark J, et al. National Health and Nutrition Examination Survey, 2017-March 2020 Prepandemic File: sample design, estimation, and analytic guidelines. Vital and Health statistics Ser 1, Programs and Collection Procedures. 2022:1–36. doi:10.15620/cdc:106273
9. van Kleef LA, Lu Z, Ikram MA, de Groot NMS, Kavousi M, de Knegt RJ. Liver stiffness not fatty liver disease is associated with atrial fibrillation: the Rotterdam study. J Hepatol. 2022. doi:10.1016/j.jhep.2022.05.030. Epub ahead of print.
10. Vallet-Pichard A, Mallet V, Nalpas B, Verkarre V, Nalpas A, Dhalluin-Venier V, et al. FIB-4: an inexpensive and accurate marker of fibrosis in HCV infection. comparison with liver biopsy and fibrotest. Hepatology. 2007;46:32–6.
11. Treeprasertsuk S, Björnsson E, Enders F, Suwanwalaikorn S, Lindor KD. NAFLD fibrosis score: a prognostic predictor for mortality and liver complications among NAFLD patients. World J Gastroenterol. 2013;19:1219.
12. Roulot D, Costes JL, Buyck JF, Warzocha U, Gambier N, Czernichow S, et al. Transient elastography as a screening tool for liver fibrosis and cirrhosis in a community-based population aged over 45 years. Gut. 2011;60:977–84.
13. Berzigotti A, Tsochatzis E, Boursier J, Castera L, Cazzagon N, Friedrich-Rust M, et al. EASL Clinical Practice Guidelines on non-invasive tests for evaluation of liver disease severity and prognosis—2021 update. J Hepatol. 2021;75:659–89.
14. Zhang X, Heredia NI, Balakrishnan M, Thrift AP. Prevalence and factors associated with NAFLD detected by vibration controlled transient elastography among US adults: results from NHANES 2017–2018. PLoS One. 2021;16:e0252164.
15. National Center for Health Statistics. National Health and Nutrition Examination Survey 2017-2020. Liver ultrasound transient elastography procedures manual. 2022. Accessed October 3, 2022. https://wwwn.cdc.gov/Nchs/Nhanes/2017-2018/P_LUX.htm
16. van Kleef LA, de Knegt RJ, Brouwer WP. Metabolic dysfunction-associated fatty liver disease and excessive alcohol consumption are both independent risk factors for mortality. Hepatology. 2022. doi:10.1002/hep.32642. Epub ahead of print.
17. Eslam M, Newsome PN, Anstee QM, Sarin SK, Targher G, Romero-Gomez M, et al. A new definition for metabolic associated fatty liver disease: an international expert consensus statement. J Hepatol. 2020. doi:10.1016/j.jhep.2020.03.039
18. Karlsen TH, Sheron N, Zelber-Sagi S, Carrieri P, Dusheiko G, Bugianesi E, et al. The EASL–Lancet Liver Commission: protecting the next generation of Europeans against liver disease complications and premature mortality. The Lancet. 2022;399:61–116.
19. Sugiyama A, Kurisu A, E B, Ouoba S, Ko K, Rakhimov A, et al. Distribution of FIB-4 index in the general population: analysis of 75,666 residents who underwent health checkups. BMC Gastroenterol. 2022;22:241.
20. McPherson S, Hardy T, Dufour J-F, Petta S, Romero-Gomez M, Allison M, et al. Age as a confounding factor for the accurate non-invasive diagnosis of advanced NAFLD fibrosis. Am J Gastroenterol . 2017;112:740–51.

21. Shang Y, Nasr P, Widman L, Hagström H. Risk of cardiovascular disease and loss in life expectancy in NAFLD. Hepatology. 2022. doi:10.1002/hep.32519. Epub ahead of print.
22. van Kleef LA, Sonneveld MJ, Kavousi M, Ikram MA, de Man RA, de Knegt RJ. Fatty liver disease is not associated with increased mortality in the elderly—a prospective cohort study. Hepatology. 2022. doi:10.1002/hep.32635. Epub ahead of print.

How to cite this article: van Kleef LA, de Knegt RJ, Ayada I, Pan Q, Brouwer WP. The Steatosis-associated fibrosis estimator (SAFE) score: validation in the general US population. Hepatol Commun. 2023;7:e0075. https://doi.org/10.1097/HC9.0000000000000075