Hepatology International (2019) 13:560–572
https://doi.org/10.1007/s12072-019-09952-5

ORIGINAL ARTICLE

The prognostic role of liver stiffness in patients with chronic liver disease: a systematic review and dose–response meta-analysis

Yue Shen[1] · Sheng-Di Wu[1] · Ling Wu[1] · Si-Qi Wang[1] · Yao Chen[2] · Li-Li Liu[1] · Jing Li[3] · Chang-Qing Yang[3] · Ji-Yao Wang[1] · Wei Jiang[1]

Received: 4 February 2019 / Accepted: 18 May 2019 / Published online: 4 July 2019

Abstract

Background and aims Liver stiffness measurement (LSM) by transient elastography (TE) has been assessed for the evaluation of clinically relevant outcomes in patients with chronic liver diseases (CLDs) while with variable results. This systematic review and meta–analysis aims to investigate the relationship between baseline LSM by TE and the development of clinically relevant outcomes.

Methods The systematic review identified eligible cohorts reporting the association between baseline LSM by TE and risk of hepatic carcinoma (HCC), hepatic decompensation (HD), all–cause and/or liver–related mortality and liver–related events (LREs) in CLD patients. Summary relative risks (RRs) with 95% confidence intervals (CIs) were estimated using a random–effect model. The dose–response association was evaluated by generalized least squares trend (Glst) estimation and restricted cubic splines. Commands of GLST, MKSPLINE, MVMETA were applied for statistical analysis.

Results 62 cohort studies were finally included, reporting on 43,817 participants. For one kPa (kilopascal) increment in baseline liver stiffness (LS), the pooled RR (95% CI) was 1.08 (1.05–1.11) for HCC, 1.08 (1.06–1.11) for all–cause mortality, 1.11 (1.05–1.17) for liver-related mortality, 1.08 (1.06–1.10) for HD and 1.07 (1.04–1.09) for LREs. Furthermore, the nonlinear dose–response analysis indicated that the significant increase in the risk of corresponding clinically relevant outcomes turned to a stable increase or a slight decrease with increasing baseline LS changing primarily in the magnitude of effect rather than the direction.

Conclusions The dose–response meta-analysis presents a combination between the levels of baseline LS and RRs for each clinically relevant outcome. TE, which is noninvasive, might be a novel strategy for risk stratification and identification of patients at high risk of developing these outcomes.

Keywords Fibroscan · Liver · Cancer · Cirrhosis · Outcome

Yue Shen and Sheng-Di Wu share co-first authorship.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s12072-019-09952-5) contains supplementary material, which is available to authorized users.

✉ Wei Jiang
jiang.wei@zs-hospital.sh.cn

[1] Department of Gastroenterology and Hepatology, Zhongshan Hospital, Fudan University, 180 Fenglin Road, Xuhui District, Shanghai, People's Republic of China

[2] Department of Emergency Medicine, Zhongshan Hospital, Fudan University, Shanghai, People's Republic of China

[3] Department of Gastroenterology, Tongji Hospital, Tongji University, Shanghai, People's Republic of China

Abbreviations

ALT	Alanine aminotransferase
CHB	Chronic hepatitis B
CIs	Confidence intervals
CLD	Chronic liver disease
Glst	Generalized least-squares trend
HBV	Hepatitis B virus
HCC	Hepatic carcinoma
HCV	Hepatitis C virus
HD	Hepatic decompensation
HIV	Human immunodeficiency virus
HR	Hazard ratio
kPa	Kilopascal
LREs	Liver-related events
LS	Live stiffness
LSM	Liver stiffness measurement

PH Portal hypertension
RRs Relative risks
TE Transient elastography
ULN Upper limit of normal

Introduction

Chronic liver disease, of which liver fibrosis is the key component, is an increasing health burden associated with significant morbidity. The progression of liver fibrosis characterized by excessive deposition of extracellular matrix is associated with an increased risk of cirrhosis [1]. Without proper management, subsequent liver structure distortion, function impairment, and hemodynamic deterioration may lead to portal hypertension (PH)-related complications as well as an increased tendency for hepatocarcinogenesis [2]. Therefore, reliable methods for the early detection of the presence and progression of liver fibrosis are indispensable for recognizing patients at high risk.

Considering the invasiveness, sampling variability, and diagnostic variability of liver biopsy [3, 4], tremendous efforts have been made to explore alternative, noninvasive markers of liver fibrosis. LSM by elastography has drawn great attention, from which TE has shown extraordinary performance in identifying the stage of liver fibrosis [5, 6]. Furthermore, LSM by TE has also been proposed for predicting complications of cirrhosis in patients with clinically significant PH [7]; assessing the risk of HCC [8], decompensated cirrhosis [9], mortality [10], LREs [11]. Novel LS-based models have also been developed to calculate the risk in CLD patients accurately [12]. Though increasing cross-sectional diagnostic studies and longitudinal prognostic studies assessing the relationship between baseline LSM and clinically relevant outcomes in CLD patients have emerged during past 5 years, its prognostic value remains inconclusive.

A systematic review and meta-analysis conducted in 2013 explored a linear association between the risk of clinically relevant outcomes and baseline LSM by TE and magnetic resonance elastography [13]. More evidence considering the prognostic role of LSM by TE has emerged over the past few years, and sufficient data is available for further nonlinear, dose–response analysis. Here, we conducted a systematic review and meta-analysis to assess the association comprehensively between baseline LSM by TE and the development of each clinically relevant outcomes, including HCC, HD (variceal bleeding, ascites, hepatic encephalopathy, hepatorenal syndrome, spontaneous bacterial peritonitis), all-cause and/or liver-related mortality, and a composite of these outcomes, which was defined as LREs in patients with CLD. Taking the distribution of cases, total number and the RRs with the variance estimates across each level of baseline LS into consideration, the dose–response meta-analysis is aimed to present a combination between the levels of baseline LS and RRs for each clinically relevant outcome, making the results more productive and persuasive in CLD epidemiology.

Materials and methods

We conducted this systematic review and meta-analysis under the guidance of the Cochrane Handbook of Interventions [14] and reported following the Meta-analysis of Observational Studies in Epidemiology (MOOSE) reporting guidelines [15] (Supplementary Table 1).

Literature search

First, Web of Science, PubMed, Embase, and Cochrane Library were systematically searched until November 15, 2018, under the direction of an expert librarian. The details of the search strategy are included in Supplementary Table 2. Second, reference of included studies and reviews were manually searched to identify potentially relevant publications. Third, to include potential studies from conference proceedings, the conference abstracts index from The Liver Meeting, The International Liver Congress and Digestive Disease Week were manually searched, respectively, between 2008 and 2018, when there was a rapid increase in the use of TE for the assessment of liver fibrosis.

Selection criteria

Considering the prognostic objective of this study, observational cohorts that met the following criteria were considered eligible: (1) The participants were free of reported outcomes at the time of cohort entry and within 6 months after enrollment. (2) LS values were measured using TE at baseline. (3) Selected studies systematically assessed the development of clinically relevant outcomes in CLD participants. (4) All the participants had a minimum follow-up period of 6 months. (5) Hazard ratio (HR), RR with 95% CI and incidence rate of clinically relevant outcomes for each category of baseline LS was reported, or sufficient data was available for calculation. (6) Inclusion was not otherwise restricted by study size, language, or publication type.

We excluded the following studies: (1) descriptive studies and case–control studies; (2) studies that evaluated the prognostic value of LS in patients without evidence of CLD at enrollment; (3) studies with LS measured by other quantitative elastography methods; (4) studies that evaluated the prognostic role of LS measured during follow-up or change of serial LS values; (5) studies without necessary data to calculate RR with 95% CI for further analysis.

Two independent authors (S.Y. and W.L.) reviewed the titles and abstracts of studies identified in the literature search. Studies that did not focus on the research question of interest were excluded. After reading and evaluating the full text of the remaining studies, two authors further determined whether to include them based on the selection criteria independently. There was a pre-test for study selection based on the full-text screening form (Supplementary Table 3). Disagreements were resolved by discussion with a third member (W.S.D.) if necessary. The screening process is illustrated in a flow chart (Fig. 1).

Data abstraction

Two authors (S.Y., W.L.) independently extracted data for the following six categories: study characteristics, patient characteristics, exposure assessment, outcomes reported, statistical analysis, and confounding variables (Supplementary Table 4). If multiple publications were from the same cohort, data from the most recent and comprehensive report was included, and data regarding the same cohort with different outcomes reported was pooled to minimize the risk of double-counting as well as maximize available data for different analysis. Reasonable estimation or contact with

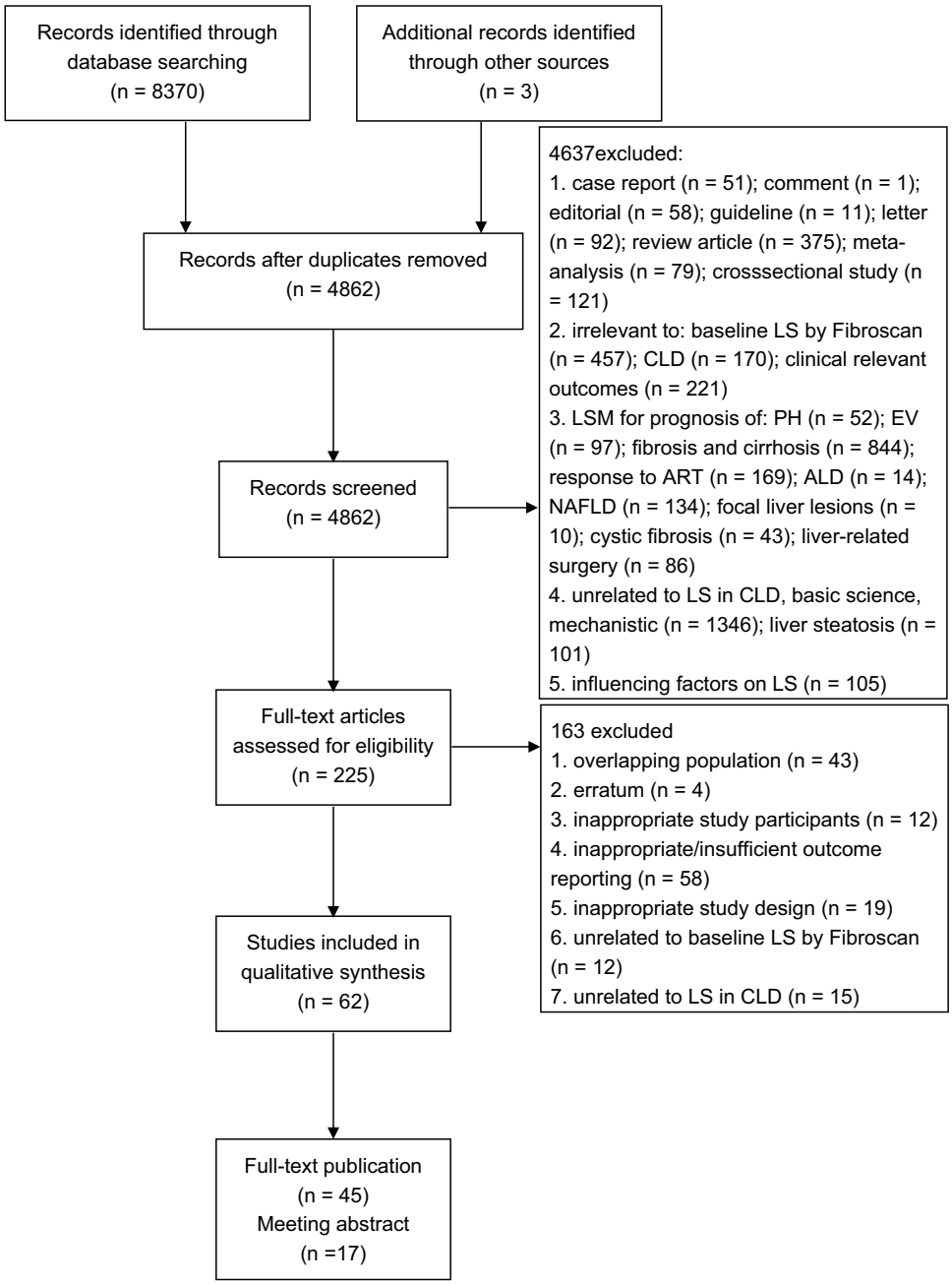

Fig. 1 Flowchart of study identification and selection process. *ALD* alcoholic liver disease, *ART* anti-viral treatment, *CLD* chronic liver disease, *EV* esophageal varices, *LS(M)* liver stiffness (measurement), *PH* portal hypertension, *NAFLD* non–alcoholic fatty liver disease

original authors was made when relevant information was not clear or insufficient in original studies, and the study was excluded if the corresponding author failed to reply. Extracted data were checked carefully by another author (W.S.D.), and disagreements were resolved by discussion and consensus.

Quality assessment

The quality of the prognostic studies included was assessed by two authors (S.Y., W.L.) independently, using the Quality In Prognosis Studies Tool based on six bias domains [16]. Issues to consider for judging overall rating of risk of bias are summarized in Supplementary Table 5.

Statistical analysis

The RRs and 95% CI were considered the effect size of all studies. We assume the reported HR to be equivalent to the RR.

Pooled RRs and 95% CI were calculated using the random–effects model, according to the DerSimonian–Laird method [17] in highest versus lowest analysis, which is the prerequisite for further dose–response analysis.

We estimated the linear dose–response relationship for each study with RRs reported across levels of LS based on the Glst estimation [18, 19]. The midpoint between the upper and lower boundary of each level was assigned as the dose value for the corresponding risk estimation. When the lowest category was open–ended, the lower boundary was considered to be zero, whereas for the open–ended highest category, the dose value was assumed to be the lower boundary multiplied by 1.5. The pooled RR and 95% CI were also calculated for per-kPa analysis, using a random-effects model.

For potential nonlinear associations from studies reporting RRs with corresponding 95% CI for at least three quantitative exposure categories, a two-stage, dose–response meta-analysis was performed [20]. First, study-specific estimations, using a restricted cubic spline model with four knots at the 5th, 35th, 65th, and 95th centiles of the levels of baseline LS took the relationship within each set of published RRs into account. Second, study-specific estimates were pooled using the best-fitting model. Departure from linearity was assessed by testing the null hypothesis that the coefficients of the splines were equal to zero.

Between-study heterogeneity was assessed using Cochrane's Q ($p < 0.10$ was suggestive of significant heterogeneity) test, and an I^2 statistic with a value greater than 50% was suggestive of considerable heterogeneity [21]. Meta-regression was performed to explore the proportion of total variation due to between-study heterogeneity. ($p < 0.05$ suggested that the variable was a significant source of heterogeneity). Subgroup analysis was conducted by study type and design, location, follow-up, number of patients, age, etiology, alanine aminotransferase (ALT) level, stage, treatment, and adjustment. To test whether the results could have been driven by a single study, influence analysis was done by omitting one study at a time. Publication bias was evaluated using a contour-enhanced funnel plot, Egger's asymmetry test [22], and Begg's test [23]. $p < 0.10$ was considered to indicate significant publication bias, and a further trim and fill method was performed for adjustment. All the statistical analyses were performed with Stata version 13 (Stata Corp, College Station, Texas, USA) and RevMan 5.3 (The Cochrane Collaboration, Oxford, UK).

Results

Of the 8373 articles identified through the systematic search, a total of 62 studies with 43,817 participants were finally included in our meta–analysis. The coefficient of agreement between two reviewers for the article selection was excellent (Cohen $\kappa = 88.24\%$). The detailed reasons for study exclusion are shown in Fig. 1. We classified all 62 studies according to different outcomes reported and performed highest versus lowest, per-kPa, as well as further nonlinear dose–response analysis according to the prerequisites within each analysis (Supplementary Table 6). The general characteristics and references of all included 62 studies are listed in Supplementary Table 7, and the general baseline information is shown in Supplementary Tables 8, 9, and 10.

Quality of included studies

Considering the six bias domains in the quality assessment for prognosis analysis, the majority of the included studies were at low to moderate risk of bias in study participation, attrition, prognostic factor, and outcome measurement. However, some studies were at high risk of bias in study confounding because important confounders (MELD score, Child–Pugh score and/or their components, treatment for underlying CLD) had not been adjusted sufficiently or with clear statements. Studies with effect size reported only in univariate analysis or calculated by number of cases and total subjects were at high risk of bias in the domain of statistical analysis and reporting. The overall quality assessment is shown in Fig. 2 and Supplementary Table 11.

Risk of hepatic carcinoma

Highest versus lowest analysis

Fourteen studies with a total of 13,489 participants, were included, of which 502 developed HCC. The risk of HCC

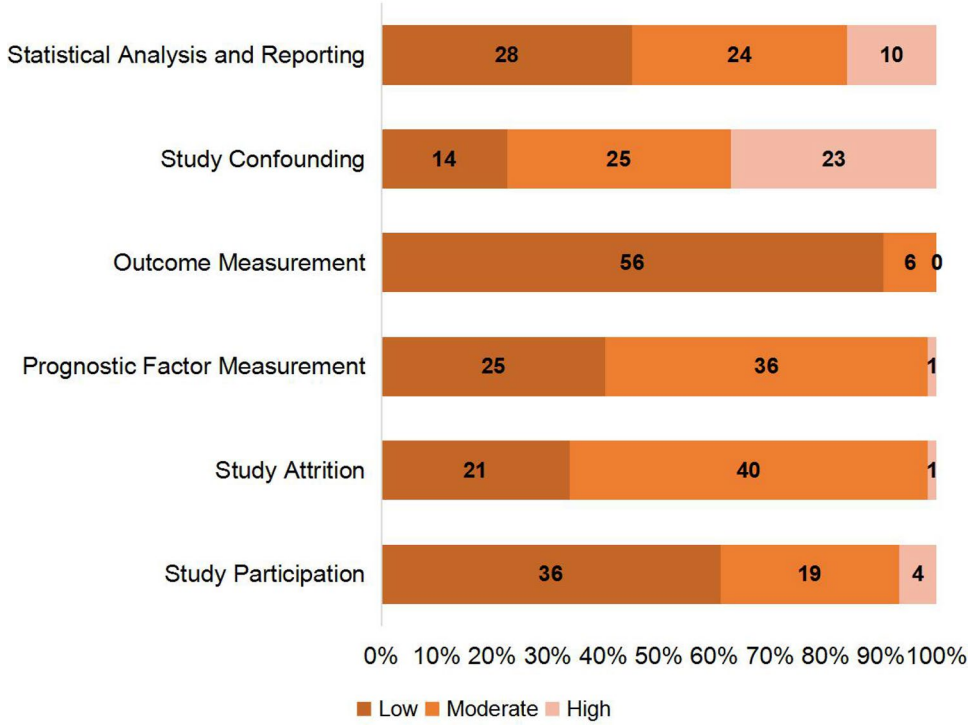

Fig. 2 Quality assessment of included studies using the quality in prognosis studies tool. The quality of all included studies was scores as low, moderate or high risk of bias based on six domains including study participation, study attrition, prognostic factor measurement, outcome measurement, study confounding and statistical analysis and reporting

was significantly higher in groups with the highest category of baseline LS compared with the lowest category (RR 5.31, 95% CI 3.76–7.51) (Fig. 3a). The positive associations in subgroup analysis all remained statistically significant (Supplementary Fig. 1A1). The year of publication accounted for 62.9% between-study heterogeneity with τ^2 decreasing from 0.205 to 0.076 in the univariate meta-regression (Supplementary Table 12A). In the sensitivity analysis, the pooled RR ranged from 4.78 (95% CI 3.48–6.56) to 5.79 (95% CI 3.88–8.64). Exclusion of the study by Masuzaki and Fung et al. yielded a low heterogeneity (I^2 41.5%, $P_{\text{heterogeneity}}$: 0.006) with RR of 4.43 (95% CI 3.19–6.15) (Supplementary Fig. 2A1).

Dose–response analysis

A linear dose–response model for the association between baseline LS and risk of HCC was assumed for 16 single studies with a total of 12,070 participants, of which 620 developed HCC. The risk of HCC was increased by 8% for per-kPa increase of baseline LS (RR 1.08, 95% CI 1.05–1.11) (Fig. 4a). A subgroup analysis by ALT level showed no positive association of statistical significance among studies with baseline ALT < 1*ULN (RR 1.129, 95% CI 0.995–1.282) or ALT > 2*ULN (RR 1.042, 95% CI 0.995–1.091) (Supplementary Fig. 1A2). No significant covariate contributing to considerable between-study heterogeneity was found in univariate meta-regression analysis (Supplementary Table 12A). In sensitivity analysis, the positive association remained stable, varying between 1.07 (95% CI 1.05–1.09) and 1.09 (95% CI 1.06–1.11). There was a significant decrease (I^2 44.4%) by omitting the studies of both Li et al. and Fung et al., with the pooled RR of 1.06 (95% CI 1.04–1.07) (Supplementary Fig. 2A2). A 5 kPa increment in baseline LS was associated with a 32% higher risk of developing HCC (RR 1.32, 95% CI 1.25–1.40).

Six studies, with 5566 participants, were eligible for further nonlinear dose–response analysis, of which 249 participants developed HCC. Evidence of a nonlinear association was found between baseline LS and risk of HCC (χ^2: 131.59, $P_{nonlinearity}$:0.000). With 5 kPa as a reference, the risk of HCC increased with increasing baseline LS, and the RR with 95% CI was 1.80 (1.49–2.18) for 7.2 kPa, 5.38 (3.38–8.56) for 12.5 kPa, 9.05 (5.78–14.17) for 19 kPa, and 14.36 (9.10–22.67) for 35 kPa. The nonlinear trend showed a steep increase in the risk of HCC with baseline LS below 16.7 kPa, followed by a relatively stable growth (Fig. 5a).

Risk of mortality

Highest versus lowest analysis

Five prospective cohorts with 8756 participants, and three retrospective studies with 1751 participants, were included. The risk of all-cause mortality was significantly higher in the highest category of baseline LS (RR 4.30, 95% CI 2.85–6.50) (Fig. 3b). We did not find a positive association among patients with a follow–up period of more than 5 years

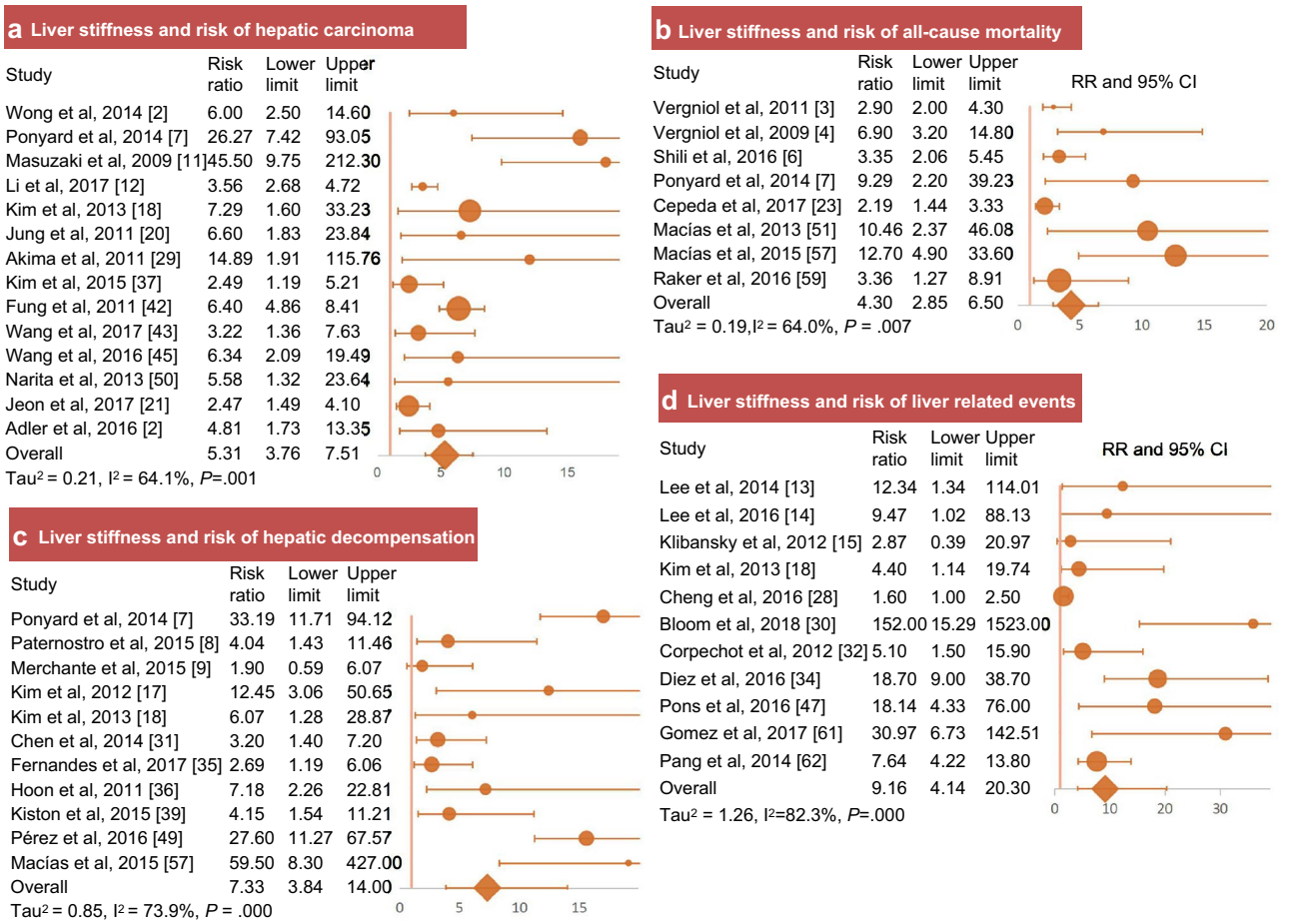

Fig. 3 Forest plot. Association between baseline liver stiffness and risk of **a** hepatic carcinoma, **b** all-cause mortality, **c** hepatic decompensation, **d** liver-related events in highest versus lowest analysis

(RR 3.80, 95% CI 0.96–15.05) (Supplementary Fig. 1B1). No covariate was found to have a significant impact on between–study heterogeneity (Supplementary Table 12B). The positive association was stable, and no single study was found to have an excessive impact on the pooled RR in the influence analysis. In the sensitivity analysis, the heterogeneity decreased to 49.4% after the exclusion of the study by Macías et al. (2015), which was conducted in Spain (RR 3.6, 95% CI 2.52–5.13) (Supplementary Fig. 2B1).

Considering that only two studies provided sufficient information for further nonlinear dose–response analysis, we did not perform the highest versus lowest analysis in studies reporting liver-related mortality.

Dose–response analysis

Seven studies with 8803 participants and 418 cases were included in the analysis for risk estimation of all–cause mortality. Similar to the outcome of HCC, the pooled RR was 1.08 (95% CI 1.06–1.11) (Fig. 4b). All the studies were conducted in Europe, especially in Spain, and were published with the full-text, of which participants included were all at different stages of fibrosis or cirrhosis, with the mean age under 55. In the subgroup analysis, studies with follow–up periods less than 3 years (RR 1.03, 95% CI 0.81–1.31) or fewer than 1000 participants (RR 1.05, 95% CI 0.98–1.13) did not show a statistically significant positive association (Supplementary Fig. 1B2). The univariate meta-regression analysis yielded follow-up period as the major source of heterogeneity (p: 0.04) and ALT level as a covariate with borderline statistical significance (p: 0.058) (Supplementary Table 12B). The pooled RR remained stable, ranging from 1.07 (95% CI 1.05–1.09) to 1.09 (95% CI 1.06–1.11); no significant change was found in the heterogeneity. After excluding the studies by both Vergniol et al. [33] and Macías et al., the heterogeneity decreased significantly (I^2: 26.0%), with RR at 1.06 (95% CI 1.05–1.08), and the latter explained much of the heterogeneity (I^2: 56.8%, $P_{\text{heterogeneity}}$: 0.041) (Supplementary Fig. 2B2). The risk of all-cause mortality increased by

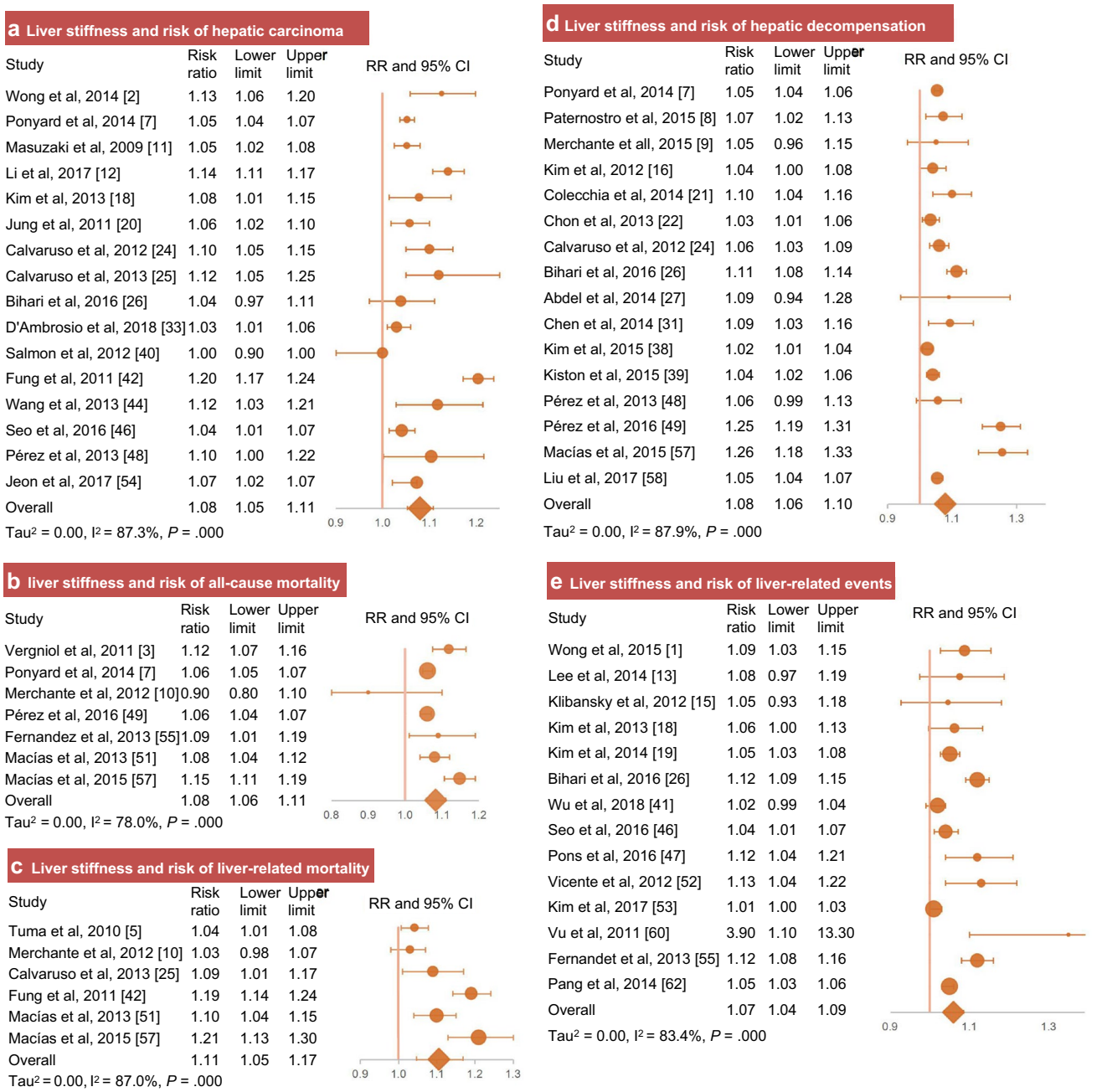

Fig. 4 Forest plot. Association between baseline liver stiffness and risk of **a** hepatic carcinoma, **b** all-cause mortality, **c** liver–related mortality, **d** hepatic decompensation, **e** liver-related events in per-kPa analysis

60% with 5-kPa increments in baseline LS (RR 1.60, 95% CI 1.34–1.92).

We also performed a linear dose–response model for assessing the association between baseline LS and risk of liver-related mortality. Taking six studies with a total of 2455 participants and 87 cases into consideration, 1 kPa increment of baseline LS was associated with a 6% higher risk of liver-related mortality. (RR 1.11, 95% CI 1.05–1.17) (Fig. 4c).

Three studies reporting sufficient information were eligible for the nonlinear dose–response meta–analysis. Data from 4374 participants, of which 194 developed all-cause mortality, was collected. Study-specific estimates were pooled using the fixed-effects model, because no obvious heterogeneity was observed (Q: 3.15, p: 0.96) among three studies and showed significant evidence of departure of linearity (χ^2: 187.53, $P_{\text{nonlinearity}}$: 0.000). Compared with the reference of 5 kPa, the pooled RR with 95% CI was

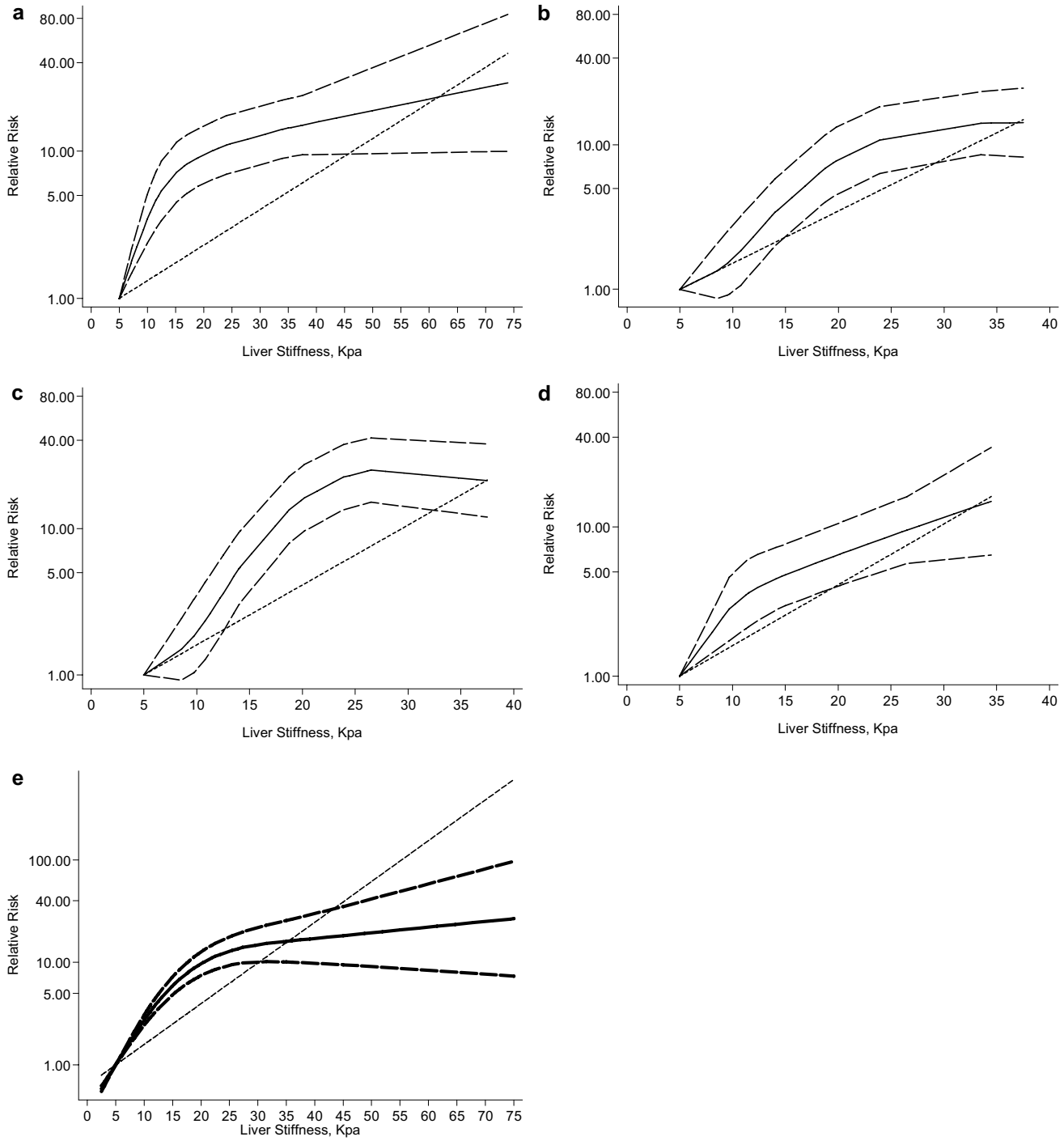

Fig. 5 Dose–response analysis. The dose–response relationship between baseline liver stiffness and risk of **a** hepatic carcinoma, **b** all-cause mortality, **c** hepatic decompensation, **d** liver-related events, **e** composite outcomes. The short dash line represents the linear relationship. The solid line and the long dash line represent the estimated RR and its 95% CI for the spline model. The value of 5 kPa served as referent. The relative risks are plotted on the log

1.34 (0.86–2.07) for 8.5 kPa, 3.25 (1.90–5.56) for 13.5 kPa, 7.72 (4.51–13.22) for 19.8 kPa, and 14.25 (8.22–24.73) for 37.5 kPa, respectively. The plot showed a steep increase in the risk of all-cause mortality for baseline LS below

24.1 kPa and then turned to a relatively stable increase (Fig. 5b).

Risk of hepatic decompensation

Highest versus lowest analysis

The pooled RR based on 11 cohorts was 7.33 (95% CI 3.84–14.00), indicating that the risk of HD was significantly higher among cases with a higher baseline LS value (Fig. 3c). The positive association remained statistically significant among all subgroup analyses except for patients with hepatitis C virus (HCV) (Supplementary Fig. 1C1). Besides, the univariate meta-regression analysis identified the design of the study, number of patients, and treatment as main sources of heterogeneity (Supplementary Table 12C). The sensitivity analysis indicated that the pooled RR varied between 6.15 (95% CI 3.32–11.39) and 8.37 (95% CI 4.30–16.29). Significant change in heterogeneity (I^2: 41.40) was found only when studies by both Ponyard and Pérez et al. were omitted, and the pooled RR was 4.68 (95% CI 2.86–7.65) (Supplementary Fig. 2C1).

Dose–response analysis

Each kPa increase in baseline LS was associated with an 8% higher risk of developing HD in the meta–analysis of 16 studies (RR 1.08, 95% CI 1.06–1.10) (Fig. 4d). We did not find a positive association among studies with a follow-up period of more than 5 years (RR 1.15, 95% CI 0.97–1.36) (Supplementary Fig. 1C2). The design of the included studies was found to be a significant potential source for between-study heterogeneity (p: 0.03) (Supplementary Table 12C). A sensitivity analysis showed that the pooled RR varied from 1.07 (95% CI 1.05–1.08) to 1.08 (95% CI 1.06–1.11), indicating that the effect size was relatively stable. No individual study was found to have an excessive impact on between-study heterogeneity (Supplementary Fig. 2C2). Our analysis also indicated that the RR of HD was 2.15 (95% CI 1.37–3.37) for a 5-kPa increment in baseline LS.

Furthermore, data from four studies with 6368 participants, of which 280 developed HD, was adopted in the nonlinear dose–response analysis. Evidence of a nonlinear association between the risk of HD and baseline LS was found (χ^2: 295.20, $P_{\text{nonlinearity}}$: 0.000). There was a steep increase in the risk of HD, with an increasing baseline LS, followed by a slight decrease for baseline LS up to 26.5 kPa. Compared to a reference level of 5 kPa, the RR (95% CI) was 1.50 (0.92–2.44) for 8.6 kPa, 4.69 (2.63–8.37) for 13.5 kPa, 16.23 (9.63–27.35) for 20.2 kPa, and 21.29 (11.98–37.83) for 37.5 kPa (Fig. 5c).

Risk of liver-related events

Highest versus lowest analysis

Eight prospective studies with 4160 participants, two retrospective studies with 437 participants, and one retrospective–prospective study with 2052 participants were included. The RR comparing the highest-with the lowest-category of baseline LS was 9.16 (95% CI 4.14–20.30) (Fig. 3d). The positive association also remained statistically significant in subgroup analysis except among four studies with stage unknown (RR 4.88, 95% CI 0.46–52.30) (Supplementary Fig. 1D1). The univariate meta-regression analysis yielded location and treatment as potential sources of the heterogeneity (Supplementary Table 12D). The sensitivity analysis indicated that the positive association was stable, with RR ranging from 7.56 (95% CI 3.48–16.44) to 11.25 (95% CI 6.49–19.52), suggesting further dose–response analysis. The latter was also accompanied by a significant decrease in heterogeneity, with I^2 decreasing from 82.3 to 42.7% (Supplementary Fig. 2D1).

Dose–response analysis

Data from 14 studies with 9060 participants, of which 721 participants developed LREs, was collected for the analysis. One kPa increment in baseline LS was associated with a 7% higher risk of developing LREs (RR 1.07, 95% CI 1.04–1.09) (Fig. 4e). The pooled RR was 1.78 (95% CI 0.55–5.76) for two studies published in abstract, whereas it was 1.06 (1.04–1.09) for the other 12 studies published in full-text. The subgroup analysis by study location did not show a positive association among three studies conducted in America (RR 1.06, 95% CI 0.95–1.17). There was also no evidence of statistically significant positive association among studies with a follow-up period more than 5 years (RR 1.07, 95% CI 0.99–1.16) as well as studies with a mean baseline ALT level of more than 2*ULN (RR 1.07, 95% CI 1.00–1.15). The positive association persisted in studies adjusted for main confounders but not in unadjusted studies (RR 1.06, 95% CI 0.98–1.15) (Supplementary Fig. 1D2). No significant covariate contributing to considerable between-study heterogeneity was found (Supplementary Table 12D). The estimated RR in sensitivity analysis was stable, ranging from 1.06 (95% CI 1.04–1.08) to 1.07 (95% CI 1.05–1.10). We did not find that any single study had an excessive impact on between-study heterogeneity (Supplementary Fig. 2D2). A 5-kPa increment in baseline LS was associated with a 60% higher risk of LREs (RR 1.60, 95% CI 1.43–1.79).

Five studies with a total of 3108 participants were included in the nonlinear dose–response analysis. We found strong evidence of a nonlinear association when taking 5 kPa as a reference ($\chi^2 = 73.33$, $P_{\text{nonlinearity}}$: 0.01) and found a more

pronounced increase in the risk of LREs at a lower range of LS. The pooled RR (95% CI) of LREs was 2.83 (1.73–4.62) for 9.7 kPa, 4.49 (2.77–7.29) for 14 kPa, 6.72 (4.13–10.91) for 20.5 kPa, and 14.88 (6.49–34.12) for 34.5 kPa (Fig. 5d).

Nonlinear dose–response analysis for a composite of outcomes

All 18 studies, including a total of 19,416 participants, with the distribution of cases in total subjects, and adjusted RR with its 95% CI for at least three quantitative exposure categories known, were considered. Because considerable heterogeneity was found (Q: 74.78, p: 0.03), we performed a two-stage dose–response, random-effects analysis. A 58% higher risk of composite outcomes with a 5-kPa increment in baseline LS was found. Furthermore, the nonlinear trend ($P_{\text{nonlinearity}}$: 0.000) of the risk showed a most pronounced increase with baseline LS below 20 kPa, and then the slope of the RR curve stabilized at approximately the value of 30 kPa. The pooled RR with 95% CI was 1.23 (1.20–1.27) for 6 kPa, 1.97 (1.82–2.14) for 8.3 kPa, 4.17 (3.54–4.92) for 12.5 kPa, 10.06 (7.68–13.17) for 20.5 kPa, and 15.30 (10.17–23.03) for 31.5 kPa, respectively, indicating a significant positive association between baseline LS and the risk of clinically relevant outcomes (Fig. 5e).

Publication bias (Supplementary Fig. 3)

Publication bias was found in highest versus lowest analysis for the risk estimation of HCC (A1) and LREs (C1) but not of HD (B1). A further trim-and-fill method was performed to recalculate pooled RR, which continued to reveal a statistically significant association, with four missing studies imputed to HCC (RR 4.28, 95% CI 2.97–6.17) and five missing studies imputed to LREs (RR 3.6, 95% CI 1.60–7.91). We also found publication bias in per–kPa analysis for the risk estimation of HD (B2) and LREs (C2), though not of HCC (A2). The positive association remained stable, with six studies imputed to HD (RR 1.05, 95% CI 1.02–1.07) and six studies imputed to LREs (RR 1.04, 95% CI 1.01–1.06). The estimation of publication bias was not performed for risk of mortality because of the limited number of studies.

Discussion

The present study of 62 cohort studies of 43,817 CLD participants with multiple etiologies and different stages of CLD provides newly summarized evidence regarding the association between baseline LS and risk of clinically relevant outcomes. All of the analysis remained statistically significant when comparing the highest with the lowest category of baseline LS, thus suggesting further dose–response analysis.

In per-kPa analysis, our results show that a higher baseline LS is associated with an increased risk of HCC (by 8%), all-cause mortality (by 8%), liver-related mortality (by 11%), HD (by 8%), and LREs (by 6%), indicating that LSM by TE may be a promising tool for risk stratification and identification of CLD patients at high risk of disease progression.

In per-kPa analysis of the risk of HCC and LREs, the subgroup analysis by ALT level showed no significant positive association among studies with a mean baseline ALT level < 1*ULN or > 2*ULN. The latter was probably because of the elevation of ALT, even when mild to moderate, could lead to the discrepancies between TE values and the actual degree of fibrosis [24]. High ALT level has a confounding effect on TE values, which is in parallel with histologically necrotic inflammation activity. In addition, antiviral treatment with pegylated interferon in combination with ribavirin for HCV infection [25] and nucleos(t)ide analogues for hepatitis B virus (HBV) infection [26] has been reported to be associated with the regression of fibrosis. Thus, the prognostic role of baseline LSM might be different between patients with or without corresponding treatment. It should be noted that patients included in our analysis were under treatment in whole or in part at baseline or during follow-up. It is not possible and reasonable to explore the effect of treatment on the prognostic performance of baseline LSM simply in patients with or without treatment since treatment was mainly determined by guidelines instead of study design. Still, people with CLD undergoing antiviral treatment may be at risk of disease progression [27]. Our results also indicated that one kPa increment in baseline LS was associated with a 7% higher risk of HCC, 4% higher risk of HD, and 5% higher risk of LREs, among studies with patients all under treatment for corresponding viral hepatitis. Moreover, the RRs of each clinically relevant outcome were lower for patients under treatment in whole than in part who have the same baseline LS value. In addition, in the subgroup analysis stratified by etiology, patients co-infected with HCV and human immunodeficiency virus (HIV) were at a higher risk of HD (RR 1.10, 95% CI 1.09–1.31) and LREs (RR 1.13, 95% CI 1.05–1.21) than patients with chronic hepatitis B (CHB) or mixed etiology in per-kPa analysis. The results were consistent with previous studies showing that the progression of liver fibrosis to end-stage liver disease was accelerated, and hepatic complications were more frequent in patients co–infected with HCV and HIV [28]. Furthermore, in the per-kPa analysis of the risk of HCC, CHB patients (RR 1.10, 95% CI 1.05–1.14) were at a higher risk than chronic hepatitis C patients (RR 1.06, 95%CI 1.04–1.08), which was consistent with the subgroup analysis by location, considering that HBV infection is the main etiology for CLD

in Asia, where the incidence of HCC is the highest in the world [29], whereas HCV infection or co–infection of HCV and HIV weighs much more in Europe.

Strong evidence of flexible, nonlinear association was found between the risk of corresponding clinically relevant outcomes and baseline LS. Because the normal LS value that has been investigated in healthy subjects is around 5 kPa [30, 31], we took it as reference in the nonlinear dose–response analysis. All the figures revealed the similar trend, that the RRs increased significantly at a lower range of baseline LS, followed by a transition to a stable increase or slight decrease at different cut-offs. In the analysis of HD, the risk showed a slight decrease with baseline LS above 25 kPa. This inconsistency finding may generally be due to a fewer number of patients with baseline LS above 25 kPa in the included studies. The result considering the pooled risk of a composite of all these clinically relevant outcomes, using the random-effects model, was similar to the study conducted before revealing a 58% higher risk with a 5-kPa increment in baseline LS [32]. Considering the efficacy of antiviral treatment in ameliorating CLD progression, we further performed the two–stage dose–response, random-effects analysis by treatment. There was no significant difference in RRs for a composite of all clinically relevant outcomes between 7 studies with all patients under treatment for CLD and 9 studies with a proportion of patients under treatment (data not shown).

The strength of our study stems from a comprehensive search and rigorous text screening, enabling us to include a large number of cohort studies. We went to great lengths in data abstraction. We integrated data regarding the same cohort with different outcomes reported as to make the most and to minimize the selection bias. We performed the analysis on longitudinal cohorts with a follow-up period of more than 6 months to explore the prognostic performance of baseline LS in predicting clinically relevant outcomes systematically. Taking heterogeneity into consideration, we also explored the nonlinear relationship by choosing the best-fitting model, respectively. The positive associations were all robust in subgroup analysis and sensitivity analysis of both highest versus lowest analysis and per-kPa analysis. Although showing considerable heterogeneity, this mainly affected the magnitude of effect rather than the direction of effect, because the pooled RR, after exclusion of the studies with obvious heterogeneity, was stable. Subsequent subgroup analysis, meta–regression, and sensitivity analysis were performed to determine that the heterogeneity could be explained by publication year, location, treatment, and number of patients in highest versus lowest analysis, follow-up period, and study design in per-kPa analysis.

We acknowledge that there are both inherent and unavoidable limitations in our meta-analysis. First, although being as rigorous as possible in literature inclusion, we could not decide for certain whether data from multiple publications was regarding the same cohort according to the basic characteristics of the studies. This was particularly a problem in studies conducted in Severance Hospital, Korea. Second, we could not ignore that heterogeneity of patients existed in each single study. Variability appeared in stage of fibrosis determined by different diagnostic methodologies and cut-offs, quality of TE assessment and adjustment for confounding factors in results analysis. In addition, considerable heterogeneity in both the category and proportion of etiologies and corresponding treatment for CLD makes subsequent meta-regression and subgroup analysis by specific regimen or dichotomy of with or without treatment not optimal. For more convincing results, individual meta-analysis is preferred while the access to comprehensive data remains unsolved. Third, there was considerable heterogeneity among unknown confounders and publication bias in our study. It is noticeable, however, the profound discrepancy of individualized antiviral treatment generated heterogeneity partly in the highest versus lowest analysis of HD and LREs while it had no association with heterogeneity in the other analysis. Moreover, the positive association remained after the trim-and-fill method was performed, the imputed studies were mainly located in areas with no statistical significance in the contour-enhanced funnel plot, indicating significant publication bias. Fourth, the utility of serial LS values to track dynamic change in fibrotic burden has been evaluated in both continuous and categorical LS values [33]. Because there was great variability in results reporting and a limited number of studies, the prognostic role of serial LS values remains to be further elucidated. Fifth, we assumed a linear dose–response model for studies with RRs reported for each category of baseline LS in per-kPa analysis. Together with RRs reported for each kPa increment directly, we incorporated all the results using the random-effects model. If individual data were available, a more precise outcome by direct GLST command can be expected. Sixth, in nonlinear dose–response analysis, after the centralization of LS value by subtracting the respective reference value in each study, certain error is unavoidable when taking 5 kPa as the same reference. In addition, studies reporting RRs for per kPa increment directly were not eligible for further dose–response analysis. Since most patients had a baseline LSM under 40 kPa while RRs cannot be inferred beyond the range between minimal and maximal LS values from original studies, further studies are warranted to validate the sharp increase of RRs at a lower range of baseline LS and estimate RRs for CLD patients with baseline LS above 40 kPa.

Our study has revealed a precious combination between the levels of continuous baseline LSM and RRs for each clinically relevant outcome, not only confirming the utility of baseline LSM by TE for predicting long-term prognosis

and monitoring the clinical course in CLD patients but also providing more productive and persuasive information for its application in CLD epidemiology. The corresponding RR for per-kPa increment and nonlinear dose–response relationship is slightly different within the risk estimation of each clinical outcome, which requires more sufficient data for further validation. Our study also suggests that noninvasive LSM by TE can be applied for risk stratification to identify patients at high risk of clinically relevant outcomes, and thus help physicians establish tailored management strategies by providing more detailed prognostic information. Further longitudinal cohorts with long-term monitoring of LS change are warranted to confirm the clinical practice and accuracy of baseline LS or serial change in LS by TE as indicators of the risk of clinically relevant outcomes.

Acknowledgements The study was supported by the National Nature Science Foundation, No. 81670541; and National Science and Technology Major Project of China, No. 2013ZX10002004 and No. 2017ZX10203202.

Compliance with ethical standards

Conflict of interest Yue Shen, Sheng-Di Wu, Ling Wu, Si-Qi Wang, Yao Chen, Li–Li Liu, Jing Li, Chang-Qing Yang, Ji-Yao Wang, Wei Jiang declare that they have no conflict of interest.

References

1. Bataller R, Brenner DA. Liver fibrosis. J Clin Invest. 2005;115:209–18.
2. Pungpapong S, Kim WR, Poterucha JJ. Natural history of hepatitis B virus infection: an update for clinicians. Mayo Clin Proc. 2007;82:967–75.
3. Maharaj B, Maharaj RJ, Leary WP, et al. Sampling variability and its influence on the diagnostic yield of percutaneous needle biopsy of the liver. Lancet. 1986;1:523–5.
4. Regev A, Berho M, Jeffers LJ, et al. Sampling error and intraobserver variation in liver biopsy in patients with chronic HCV infection. Am J Gastroenterol. 2002;97:2614–8.
5. Martinez SM, Crespo G, Navasa M, et al. Noninvasive assessment of liver fibrosis. Hepatology. 2011;53:325–35.
6. Ziol M, Handra-Luca A, Kettaneh A, et al. Noninvasive assessment of liver fibrosis by measurement of stiffness in patients with chronic hepatitis C. Hepatology. 2005;41:48–54.
7. Castera L, Pinzani M, Bosch J. Non invasive evaluation of portal hypertension using transient elastography. J Hepatol. 2012;56:696–703.
8. Masuzaki R, Tateishi R, Yoshida H, et al. Prospective risk assessment for hepatocellular carcinoma development in patients with chronic hepatitis C by transient elastography. Hepatology. 2009;49:1954–61.
9. Calvaruso V, Bronte F, Simone F, et al. Liver stiffness at baseline predicts decompensation and hepatocellular carcinoma in patients with compensated HCV cirrhosis. Hepatology. 2012;56:930A–1A.
10. Vicente FJ, Barreiro P, Vispo E, et al. Baseline liver stiffness and achievement of sustained virological response predict liver complications and death in HIV/HCV coinfected patients receiving peginterferon/ribavirin therapy. Hepatology. 2012;56:650A–1A.
11. Klibansky DA, Mehta SH, Curry M, et al. Transient elastography for predicting clinical outcomes in patients with chronic liver disease. J Viral Hepatitis. 2012;19:e184–93.
12. Wong GLH, Chan HLY, Wong CKY, et al. Liver stiffness–based optimization of hepatocellular carcinoma risk score in patients with chronic hepatitis B. J Hepatol. 2014;60:339–45.
13. Singh S, Fujii LL, Murad MH, et al. Liver stiffness is associated with risk of decompensation, liver cancer, and death in patients with chronic liver diseases: a systematic review and meta–analysis. Clin Gastroenterol H. 2013;11:1573–84.
14. Higgins JPT, Green S (eds) Cochrane Handbook for systematic reviews of interventions version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from http://handbook.cochrane.org. Accessed 16 May 2013.
15. Stroup DF, Berlin JA, Morton SC, et al. Meta–analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis of observational studies in epidemiology (MOOSE) group. JAMA. 2000;283:2008–12.
16. Hayden JA, van der Windt DA, Cartwright JL, et al. Assessing bias in studies of prognostic factors. Ann Intern Med. 2013;158:280–6.
17. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986;7:177–88.
18. Berlin JA, Longnecker MP, Greenland S. Meta-analysis of epidemiologic dose-response data. Epidemiology. 1993;4:218–28.
19. Orsini N, Bellocco R, Greenland S. Generalized least squares for trend estimation of summarized dose–response data. Stata J. 2006;6:40–57.
20. Orsini N, Li R, Wolk A, et al. Meta-analysis for linear and nonlinear dose-response relations: examples, an evaluation of approximations, and software. Am J Epidemiol. 2012;175:66–73.
21. Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. BMJ. 2003;327:557–60.
22. Egger M, Davey SG, Schneider M, et al. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315:629–34.
23. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. Biometrics. 1994;50:1088–101.
24. Song ZZ. Acute viral hepatitis increases liver stiffness values measured by transient elastography. Hepatology. 2008;48(349–350):350.
25. Zeuzem S, Feinman SV, Rasenack J, et al. Peginterferon alfa–2a in patients with chronic hepatitis C. N Engl J Med. 2000;343:1666–72.
26. Chang TT, Liaw YF, Wu SS, et al. Long-term entecavir therapy results in the reversal of fibrosis/cirrhosis and continued histological improvement in patients with chronic hepatitis B. Hepatology. 2010;52:886–93.
27. Fattovich G, Giustina G, Schalm SW, et al. Occurrence of hepatocellular carcinoma and decompensation in western European patients with cirrhosis type B. The EUROHEP Study Group on Hepatitis B Virus and Cirrhosis. Hepatology. 1995;21:77–82.
28. Fernández-Montero JV, Barreiro P, Vispo E, et al. Liver stiffness predicts liver-related complications and mortality in HIV patients with chronic hepatitis C on antiretroviral therapy. Aids. 2013;27:1129–34.
29. Wong VW, Chan HL. Prevention of hepatocellular carcinoma: a concise review of contemporary issues. Ann Hepatol. 2012;11:284.
30. Bazerbachi F, Haffar S, Wang Z, et al. Range of normal liver stiffness and predictors of suspected advanced fibrosis in apparently healthy individuals: a pooled analysis of 16,082 participants. Clin Gastroenterol Hepatol. 2018. https://doi.org/10.1016/j.cgh.2018.08.069.
31. Roulot D, Czernichow S, Le Clésiau H, et al. Liver stiffness values in apparently healthy subjects: Influence of gender and metabolic syndrome. J Hepatol. 2008;48:606–13.

32. Wang J, Li J, Zhou Q, et al. Liver stiffness measurement predicted liver-related events and all-cause mortality: a systematic review and nonlinear dose-response meta-analysis. Hepatol Commun. 2018;2:467–76.
33. Vergniol J, Boursier J, Coutzac C, et al. Evolution of noninvasive tests of liver fibrosis is associated with prognosis in patients with chronic hepatitis C. Hepatology. 2014;60:65–76.