Gastroenterology 2020;158:1611–1625

Association Between Fibrosis Stage and Outcomes of Patients With Nonalcoholic Fatty Liver Disease: A Systematic Review and Meta-Analysis

Rod S. Taylor,[1,*] Rebecca J. Taylor,[2] Sue Bayliss,[3] Hannes Hagström,[4] Patrik Nasr,[5] Jorn M. Schattenberg,[6] Masatoshi Ishigami,[7] Hidenori Toyoda,[8] Vincent Wai-Sun Wong,[9] Noam Peleg,[10,11] Amir Shlomai,[12] Giada Sebastiani,[13] Yuya Seko,[14] Neeraj Bhala,[15] Zobair M. Younossi,[16] **Quentin M. Anstee**,[17,18,*] Stuart McPherson,[19,20,21] and **Philip N. Newsome**[22,23,24,*]

[1]Institute of Health and Well Being, University of Glasgow, United Kingdom; [2]R[2] Consultancy, Glasgow, United Kingdom; [3]Institute of Applied Health Research, University of Birmingham, United Kingdom; [4]Unit of Hepatology, Department of Upper GI Diseases, Karolinska University Hospital, Stockholm, Sweden; [5]Department of Gastroenterology and Hepatology, Department of Medical and Health Sciences, Linköping University, Linköping, Sweden; [6]University Medical Centre of the Johannes Gutenberg-University, Mainz, Germany; [7]Department of Gastroenterology and Hepatology, Nagoya University Graduate School of Medicine, Nagoya, Japan; [8]Department of Gastroenterology and Hepatology, Ogaki Municipal Hospital, Ogaki, Japan; [9]Department of Medicine and Therapeutics, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong; [10]Department of Gastroenterology and Hepatology, Rabin Medical Center, Beilinson Hospital, Petach-Tikva; Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; [11]Department of Medicine D, Rabin Medical Center, Beilinson hospital, Petach-Tikva; [12]Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; [13]Department of Medicine, McGill University Health Centre, Montréal, Quebec, Canada; [14]Department of Molecular Gastroenterology and Hepatology, Kyoto Prefectural University of Medicine, Kyoto, Japan; [15]Institute of Applied Health Research, University Hospitals Birmingham NHS Foundation Trust and the University of Birmingham, United Kingdom; [16]Department of Medicine, Inova Fairfax Hospital, Falls Church, Virginia; [17]Institute of Clinical and Translational Research, Faculty of Medical Sciences, Newcastle University, Newcastle-upon-Tyne, United Kingdom; [18]Newcastle National Institute of Health Research Biomedical Research Centre and Liver Transplant Unit, The Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle-upon-Tyne, United Kingdom; [19]Liver Transplant Unit, The Newcastle upon Tyne Hospitals NHS Foundation Trust; [20]Institute of Clinical and Translational Research, Newcastle University, Newcastle-upon-Tyne, United Kingdom; [21]Newcastle National Institute of Health Research Biomedical Research Centre, Newcastle-upon-Tyne, United Kingdom; [22]National Institute for Health Research Biomedical Research Centre at University Hospitals Birmingham NHS Foundation Trust and the University of Birmingham, United Kingdom; [23]Centre for Liver and Gastrointestinal Research, Institute of Immunology and Immunotherapy, University of Birmingham, United Kingdom; [24]Liver Unit, University Hospitals Birmingham NHS Foundation Trust, Birmingham, United Kingdom

BACKGROUND & AIMS: Biopsy-confirmed liver fibrosis is a prognostic factor for patients with nonalcoholic fatty liver disease (NAFLD). We performed a systematic review to quantify the prognostic value of fibrosis stage in patients with NAFLD and the subgroup of patients with nonalcoholic steatohepatitis (NASH) and to assess the evidence that change in fibrosis stage is a surrogate endpoint. **METHODS:** We searched the MEDLINE, Embase, Cochrane Library, and trial registry databases through August 2018 for prospective or retrospective cohort studies of liver-related clinical events and outcomes in adults with NAFLD or NASH. We collected data on mortality (all cause and liver related) and morbidity (cirrhosis, liver cancer, and all liver-related events) by stage of fibrosis, determined by biopsy, for patients with NAFLD or NASH. Using fibrosis stage 0 as a reference population, we calculated fibrosis stage-specific relative risk (RR) and 95% confidence interval (CI) values for mortality and morbidities. We performed fixed-effect and random-effect model meta-analyses. Metaregression was used to examine associations among study design (prospective vs retrospective cohort), overall risk of bias (medium or high), and mean duration of follow-up (in years). **RESULTS:** Our meta-analysis included 13 studies, comprising 4428 patients with NAFLD; 2875 of these were reported to have NASH. Compared with no fibrosis (stage 0), unadjusted risk increased with increasing stage of fibrosis (stage 0 vs 4): all-cause mortality RR, 3.42 (95% CI, 2.63–4.46); liver-related mortality RR, 11.13 (95% CI, 4.15–29.84); liver transplant RR, 5.42 (95% CI, 1.05–27.89); and liver-related events RR, 12.78 (95% CI, 6.85–23.85). The magnitude of RR did not differ significantly after adjustment for confounders, including age or sex in the subgroup of NAFLD patients with NASH. Three studies examined the effects of increasing fibrosis on quality of life had inconsistent findings. **CONCLUSIONS:** In a systematic review and meta-analysis, we found biopsy-confirmed fibrosis to be associated with risk of mortality and liver-related morbidity in patients with NAFLD, with and without adjustment for confounding factors and in patients with reported NASH. Further studies are needed to assess the association between fibrosis stage and patient quality of life and establish that change in liver fibrosis stage is a valid endpoint for use in clinical trials.

Keywords: Biomarker; Disease Progression; Prognosis; Liver Disease.

Nonalcoholic fatty liver disease (NAFLD) has become a major health problem because of its potential to evolve into cirrhosis, with consequential risks of death and morbidity, including hepatocellular carcinoma and liver transplantation.[1] NAFLD is defined as fatty change (steatosis) affecting more than than 5% of hepatocytes, and it has a spectrum of histologic features ranging from steatosis without fibrosis to nonalcoholic steatohepatitis (NASH) with varying stages of fibrosis.[2] The Fatty Liver Inhibition of Progression Steatosis–Activity–Fibrosis criteria and the NASH Clinical Research Network (CRN) NAFLD Activity Score are the most widely adopted semiquantitative scores used for assessing histologic disease activity.[3] To sustain a diagnosis of NASH, both require histologic evidence of the presence of steatosis, hepatocyte ballooning, and lobular inflammation. In patients with NAFLD, it is widely accepted that liver fibrosis develops as a result of liver injury secondary to steatohepatitis. Given that disease activity in NASH may fluctuate over time and liver biopsy may be subject to sampling variability, some patients with NASH may be miscategorized as not having NASH. Moreover, the fibrosis progression rate and the proportion of individuals with NAFLD having fibrosis progression was similar in a long-term study with paired patient liver biopsy samples according to whether or not they had NASH at baseline.[4]

Observational studies have shown biopsy-confirmed liver fibrosis to be a major prognostic predictor of liver-related and overall mortality in patients with NAFLD.[5] Thus, liver fibrosis is deemed a putative surrogate for disease outcome, and so reduction in fibrosis is commonly used as a primary endpoint in clinical trials of treatments for NASH.[6] Surrogate endpoints allow for the earlier assessment of the benefits of treatments without waiting for longer-term, final patient-relevant outcomes to accrue, such as hepatocellular cancer, cirrhosis, liver failure, liver transplant, or death. However, regulators such as the US Food and Drug Administration (FDA) and European Medicines Agency and payers typically accept surrogate endpoints only if their validity has been proven. In addition to evidence of their biological and pathophysiological plausibility, evidence of validation requires demonstration of the association between the treatment effect of the surrogate (eg, a reduction in biopsy-confirmed fibrosis stage) and a relevant clinical outcome (eg, reduced liver-related mortality) in the setting of a single (or multiple) randomized controlled trial (RCT).[6,7]

A systematic review and meta-analysis including 5 observational cohort studies (1495 patients with NAFLD) assessed liver fibrosis as a prognostic marker of mortality.[8] The researchers reported that patients with NAFLD and fibrosis were at increased risk of overall and liver-related mortality and that this risk was related to advancing fibrosis stage. However, this previous study was subject to a number of limitations: (1) a small number of studies and a sparse number of events (a total of 56 liver-related deaths) meant the meta-analysis results were potentially less precise and also subject to bias[9,10]; (2) only the outcome of mortality was considered; (3) the comparison between fibrosis stage and death did not account for the potential confounding by factors such as age, sex, and statin usage; (4) the study did not include analyses of the impact of liver fibrosis in the subgroup of patients with NAFLD with NASH; and (5) the study did not consider the question of change in liver fibrosis as a putative surrogate endpoint. Furthermore, we are aware of the publication of additional primary studies since the literature searches (through November 2016) of this prior review.

The overarching aim of this study was to undertake a systematic review and meta-analysis to assess the evidence for stage of liver fibrosis as a predictor for mortality, liver-related morbidity, and health-related quality of life (HRQoL) in patients with NAFLD and the subgroup with NASH. The specific research questions that we sought to address were as follows. (1) What is the evidence for liver fibrosis as a prognostic marker of mortality, morbidity, and HRQoL in NAFLD and NASH? (2) What is the evidence for the change

WHAT YOU NEED TO KNOW

BACKGROUND AND CONTEXT

The stage (or extent) of liver fibrosis, confirmed by biopsy, is believed to be a prognostic factor for risk death in people with non-alcoholic fatty liver disease (NAFLD).

NEW FINDINGS

This systematic review and meta-analysis of 4428 patients in 13 studies found that, with and without adjustments for potential confounding factors, fibrosis stage was associated with all-cause mortality, liver-related mortality, and morbidity in patients with NAFLD.

LIMITATIONS

This was a systematic review of previous publications. There was insufficient evidence to determine whether fibrosis stage associated with health-related quality of life or whether a change in fibrosis stage is associated with response to treatment.

IMPACT

It is important to monitor liver fibrosis stage in patients with NAFLD. Studies are needed to determine whether change in fibrosis stage can be used as an endpoint for treatment of NAFLD.

*** Authors share co-first authorship.**

Abbreviations used in this paper: **CI, confidence interval; CLDQ, chronic liver disease questionnaire; CRN, NASH Clinical Research Network; FDA, US Food and Drug Administration; FLIP, fatty liver inhibition of progression; HRQoL, health-related quality of life; NAFLD, nonalcoholic fatty liver disease; NASH, nonalcoholic steatohepatitis; RCT, randomized controlled trial; RR, relative risk; SD, standard deviation; SF-36, Short-Form 36.**

Most current article

0016-5085
https://doi.org/10.1053/j.gastro.2020.01.043

in liver fibrosis as a valid surrogate endpoint for mortality, morbidity, and HRQoL in NAFLD and NASH?

Methods

This systematic review was conducted and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement.[11] The review was registered with the PROSPERO international prospective register of systematic reviews (CRD42019121054).

Identification of Studies and Searches

A detailed search strategy used both indexing languages (Medical Subject Headings and EmTree) and free text terms for NAFLD or NASH. These terms were combined with a second set of terms (for fibrosis) and liver-related clinical events or patient-related outcomes. A copy of the search strategy is available (Supplementary Table 1). The following electronic databases were searched through August 2018 by an experienced information specialist (SB): MEDLINE (Ovid), Embase (Ovid), and Cochrane Library (Wiley), as were the trial registers ClinicalTrials.gov, the World Health Organization International Clinical Trials Registry Platform including International Standard Randomized Controlled Trial Number and the European Union Clinical Trials Register. The search results were combined into an Endnote (Clarivate Analytics, Philadelphia, PA), version 9, database to facilitate reference management. The reference lists all eligible studies, and identified systematic reviews were checked for additional studies.

Study Selection

Studies were included in this review if they met the following criteria:

- Study design: prospective or retrospective cohort studies, RCTs or non-RCTs.
- Population: adult (≥18 years) patients with biopsy-proven NAFLD with or without the presence of NASH
- Exposure: biopsy-confirmed liver fibrosis stage
- Outcomes: all-cause and liver-related mortality, liver-related morbidity, and HRQoL

To fully data extract and quality assess studies, we excluded studies available only as abstracts (study investigators were contacted to clarify the availability of full publication). We restricted inclusion to English language papers. We excluded studies reporting noninvasive indices of liver fibrosis (e.g. fibrosis-4 index, NAFLD fibrosis score).

Data Extraction and Risk of Bias Assessment

The following information was extracted from the included studies: study design, participants' characteristics (ie, number of patients with NAFLD and NASH and by fibrosis stage, as well as key confounders [see below]), method of NAFLD and NASH diagnosis and liver fibrosis assessment, final outcomes reported, length of follow-up, and outcome results.

Study risk of bias was assessed with the Quality In Prognosis Studies tool.[12] This prognostic risk of bias tool was adapted to suit the requirements of this review (Supplementary Table 2).

The tool has 6 domains:

1. Study participation
2. Study attrition
3. Prognostic factor measurement
4. Outcome measurement
5. Study confounding (research team clinicians [PNN, SM] determined the key confounders: age, sex, diabetes mellitus, hypertension, statin use, and smoking at cohort baseline)
6. Statistical analysis and reporting

For each domain, the adequacy of reporting by a study was assessed as *yes*, *partly*, or *no*. Based on domain assessments, studies were assigned to the following overall categories of risk of bias:

- Low risk of bias: describes studies for which all domains are scored as *yes*
- Moderate risk of bias: describes studies for which 1 or more domains are scored as partly or 1 domain is scored as *no*
- High risk of bias: describes studies for which more than 1 domain is scored as *no*

The rating of the overall quality of the evidence from this review was undertaken in consideration of current guidance on the use of the Grading of

Recommendations, Assessment, Development and Evaluations (GRADE)) approach applied to prognostic studies.[13]

Statistical Analysis

Data were analyzed in accordance with the Cochrane Handbook for Systematic Reviews of Interventions.[14] We extracted the number of patients who experienced mortality (all cause and liver related) and morbidity (cirrhosis, liver cancer, and all liver-related events) by stage of fibrosis for all patients with NAFLD. In addition, the number of events was also extracted separately in 2 groups of patients with NAFLD: (1) those reported to have NASH and (2) those reported not to have NASH. Using fibrosis stage 0 as a reference population, fibrosis stage-specific relative risk (RR) and 95% confidence interval (CI) for mortality and morbidity outcomes were estimated within the study; an RR of >1.00 indicated an increased risk of outcome with increasing fibrosis stage.

Although this crude (or unadjusted) RR compares risk by stage of liver fibrosis, it does not consider the potential variability in the duration of follow-up between studies and potential differences in patient characteristics between each of the fibrosis strata, which could confound the comparison. Therefore, we also sought to identify the hazard ratios (and their standard error) for change in fibrosis stage adjusted for confounders.

Using fibrosis stage 0 as reference, the continuous outcome of HRQoL was extracted as a mean and standard deviation (or equivalent) for each fibrosis stage. Where not reported in publications, investigators were contacted for summary outcome data.

Where data was appropriately reported, we sought to undertake meta-analysis. Statistical heterogeneity between studies was assessed using the chi-squared test of heterogeneity and the Cochrane I^2 statistic cutoffs: ie, 0%–40%: heterogeneity might not be important; 30%–60%: may represent moderate heterogeneity; 50%–90%: may represent substantial heterogeneity; and 75%–100%: considerable heterogeneity.[14] When pooling the results across studies, we used a random-effects meta-analysis model where there was formal evidence of statistical heterogeneity (ie, chi-squared test P value $< .10$ and substantial heterogeneity as defined by an I^2 statistic $\geq$ 50%). For outcomes with lower levels of statistical heterogeneity, we applied both fixed-effect and random-effect models and reported where there was a discrepancy in model finding. Where there was an adequate number of studies ($\geq$7 studies),[14] small-study effects and publication bias were assessed with funnel plot asymmetry and the Egger test.[15]

Meta-regression was used to examine the association between the predefined study level variables: study design (prospective vs retrospective cohort), overall risk of bias (medium or high), and mean duration of follow-up (in years). This regression analysis was limited to those outcomes for which there were contributing data from $\geq$5 studies.[14] If there were suitable data (ie, RCTs reporting change in fibrosis stage and outcomes of interest: mortality, liver-related morbidity, and HRQoL), we planned to calculate and report 2 key indicators of surrogate endpoint validation.[16] First, we would calculate correlation coefficient and the R^2 for the trial-level relationship between intervention–control differences in fibrosis and each of the final outcomes using weighting by the inverse of the variance (for the treatment effect on final outcomes). Second, from the trial-based analysis, we would estimate the surrogate threshold effect, that is, the intercept of the prediction band of the regression line with zero effect on the final outcome.[17]

All data analyses were conducted using Stata, version 16.0 (Stata Corp, College Station, TX) software.

Results

Study Selection

After de-duplication, our database searches identified a total of 6083 titles/abstracts. A further 210 study titles were identified from trial registers. After review of all titles and full study publications, a total of 13 studies (15 publications) were judged to meet the inclusion criteria for this review.[3,18–31] The study selection process is summarized in Figure 1. Citations and reasons why studies were excluded on review of the full publication are listed in Supplementary Table 3.

Study and Patient Characteristics

The included 13 studies recruited a total of 4428 patients with biopsy-confirmed NAFLD, and a subgroup of 2875 patients (65%) had a histologically proven diagnosis of NASH. Trial and study characteristics are presented in Table 1. Twelve were observational cohort studies (7 retrospective, 5 prospective), and 1 was an RCT. The median average age across studies was 51.0 years, and 51% of participants were men. Populations were multimorbid, with a high prevalence of hypertension (median, 41.6%), diabetes mellitus (median, 47.8%), treatment with statins (median, 24.0 %), and overweight (median average body mass index, 31.3 kg/m^2). Fibrosis staging was confirmed by liver biopsy and centrally assessed in the majority of multicenter studies. The distribution of patients with NAFLD by fibrosis stage was as follows: stage 0: 1040 (23%); stage 1: 1094 (25%); stage 2: 602 (14%); stage 3: 922 (21%); and stage 4: 770 (17%). Bhala et al[18] and Vilar-Gomez[29] included only patients with stage 3 and 4 and were therefore not included in the meta-analyses.

The method of NASH diagnosis was poorly described but was judged to be adequately defined in 7 studies.[20,22–25,28,30] The 2 most common diagnostic metrics were fatty liver inhibition of progression (FLIP) criteria or NASH CRN (ie, presence of steatosis, ballooning, and lobular inflammation). The median average duration of study follow-up was 6.2 years, ranging from 7 months to 19.9 years.

Risk of Bias Assessment

All studies were judged to have a moderate risk of bias, with the exception of Leung et al,[24] which was deemed to be at high risk of bias, and Vilar-Gomez,[29] judged to be at low risk of bias (see Table 2). The Quality in Prognosis Studies criteria of study population, prognostic factor measurement, and outcome measurement were generally well met (*yes* or *partly*); however, there was limited consideration of criteria of attrition, confounding measurement, and data analysis. Only Bhala et al[18] and Vilar-Gomez[29] provided a sufficiently detailed description of loss to follow-up to assess risk of attrition, whereas the studies of Leung et al[24] and Younossi et al[30] provided no information on loss to follow-up. Angulo et al[5] and Vilar-Gomez[29] were the only studies to report all key confounders (ie, age, sex, diabetes mellitus, hypertension, statin use, and smoking) and adjust for them all in their data analysis. Leung et al[24] failed to report either how confounders were taken into account or how they were included in their data analysis.

Outcomes

Fibrosis Stage Outcomes in All Patients With Nonalcoholic Fatty Liver Disease Without Adjustment for Covariates. Across the 10 studies reporting clinical events, a total of 591 out of 3338 (17.7%) patients with NAFLD died over the period of follow-up, and 8 studies reported 95 liver-related deaths in 2729 patients (3.5%). Seven studies reported 52 out of 2510 (2.1%) patients NAFLD who experienced a liver transplant. Events due to liver morbidity were reported in 362 out of 3125 patients (11.5%) across 8 studies based on combinations of events that included liver failure, ascites, encephalopathy, and liver cancer. Meta-analysis showed that, compared with patients with NAFLD and no fibrosis (stage 0), patients with fibrosis were at an increased unadjusted RR of all-cause mortality, liver-related mortality, liver transplant, and all-event liver morbidity, and this risk was incremental according to fibrosis stage (see Table 3 and Figures 2 and 3). No statistical heterogeneity (I^2 = 0%) was observed for the comparison of fibrosis stages 1–4 vs stage 0 across the 4 event outcomes.

Fibrosis-Related Event Outcomes in All Patients With Nonalcoholic Fatty Liver Disease After

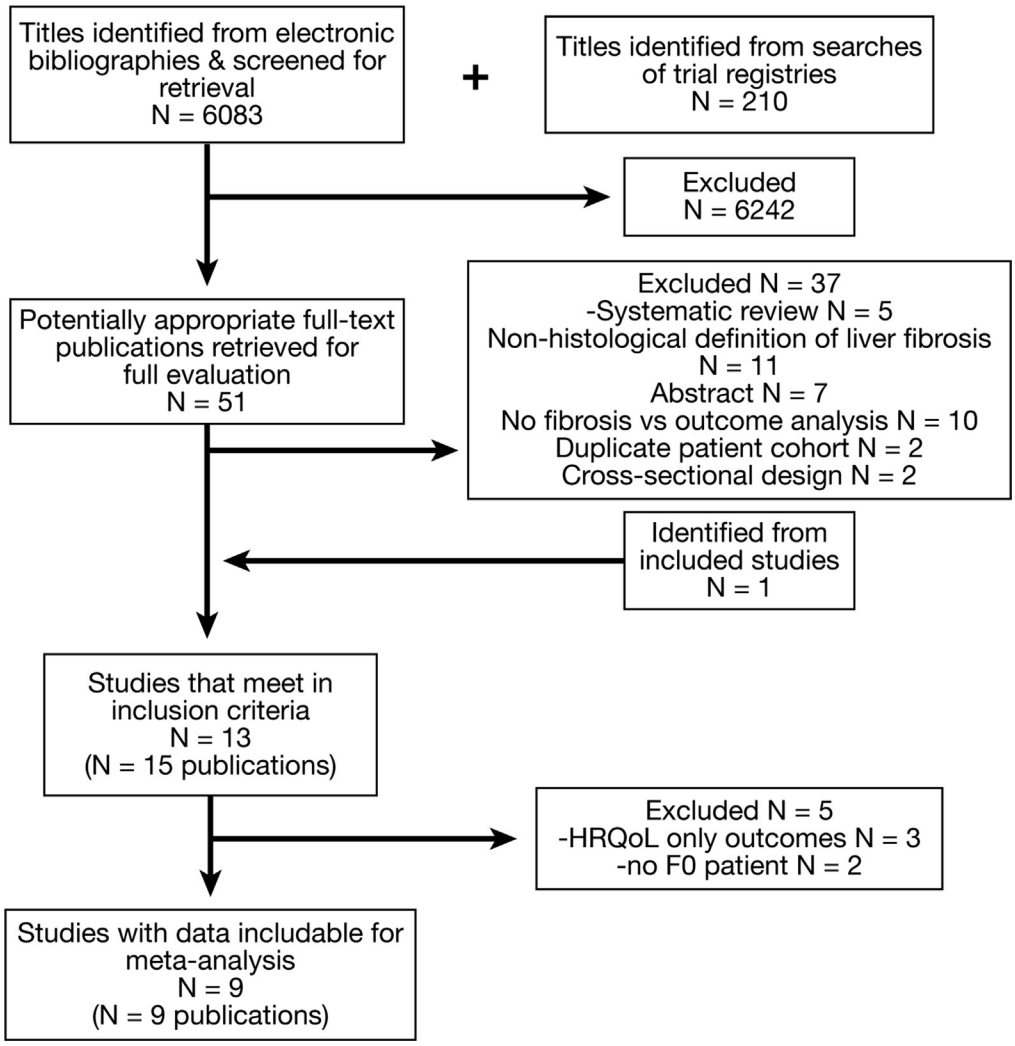


Figure 1. Summary of the study selection process.

Adjustment for Confounding Covariates. A subgroup of 6 studies provided hazard ratios for events comparing mild to moderate fibrosis (stages 0–2) to advanced fibrosis (stage 3 or 4) based on multivariable Cox regression models that adjusted for potential key confounding covariates.[21,23–27,29] All studies adjusted their analyses for age, sex, diabetes, and hypertension, with exception of Seko et al,[27] who adjusted for age, sex, diabetes, and statin use. No studies included adjustment for both smoking and statin use. Although not all studies reported data on event outcomes, there was a clear incremental risk with advanced fibrosis across all event outcomes, as shown by a pooled hazard ratio of >1.0 (see Supplementary Table 4). In those studies that provided both an adjusted and unadjusted risk ratio, the magnitude of increased risk with advanced fibrosis appeared to be similar, as indicated by overlapping 95% CIs. These conclusions remained consistent when the Seko et al study was removed from the meta-analysis.

Impact of the Presence of Nonalcoholic Steatohepatitis on Fibrosis-Related Event Outcomes Without Adjustment for Covariates. Four studies reported fibrosis-related event outcomes in a cohort of patients with NAFLD reported to either have NASH or not have NASH.[20,23–25] A low level of statistical heterogeneity (I^2 = 0%) was seen, with the exception of liver transplant events for stage 0 vs 4 in the subgroup without NASH, where there was evidence of substantial heterogeneity (I^2 = 56%) and was pooled using a random-effects meta-analysis (see Table 4).

There was an increase in the unadjusted risk of events with increasing stage of fibrosis for patients with NAFLD irrespective of the presence of NASH. The magnitude of increasing unadjusted risk appeared similar between patients with NAFLD with/without reported NASH, with overlapping 95% CI of RR estimates (see Table 4 and Supplementary Figure 1).

Fibrosis-Related Health-Related Quality of Life Outcomes. Three studies (1089 patients with NAFLD and 718 patients with NASH) reported HRQoL using either the generic measure, Short Form–36 (SF-36), or the liver-specific measure, Chronic Liver Disease Questionnaire (CLDQ). Given the heterogeneity of outcomes (generic instruments and liver-specific instrument but no NASH-specific instrument), meta-analysis was not deemed

Table 1. Characteristics of the Included Studies

First author (year) and country	Study design, time period, and sampling frame	Population diagnosis	Population demographics	Fibrosis staging	Outcomes reported	Follow-up
Angulo et al (2015)[5] Australia, Denmark, Iceland, Thailand, United Kingdom, United States NR centers	Retrospective cohort study 1975–2005[a] Consecutive patients	619 patients with liver biopsy–confirmed NAFLD 284 with NASH, method of confirmation not reported	Age: median, 49 y DM: 37.5% White: 88% Male: 37.5% HTN: 30.7% Statin use: 63% Smoking: 8.7%	Biopsy centrally confirmed and reported as stage 0–4	Overall mortality, liver transplant, liver events[b]	Median: 12.6 y Range: 0.3–35.1 y
Bhala et al (2011)[18] Australia, Italy, United Kingdom, United States, Thailand 4 centers	Prospective cohort study 1984–2006[a] Consecutive patients	247 patients with liver biopsy–confirmed NAFLD with advanced fibrosis or cirrhosis 247 with NASH, all with advanced fibrosis or cirrhosis	Age: mean, 55 y DM: 50.6% White: 91.5% Male: 39.5% HTN: 44.1% Statin use: 21.5% Smoking: NR	Biopsy reviewed independently and reported as stage 3 and 4	Overall mortality, liver-related mortality, overall vascular events, myocardial infarction, total liver events,[c] varices, ascites, encephalopathy	Mean: 7.1 y Range: 0.5–24.75 y
David et al (2009)[19] United Kingdom (NASH CRN Research Group) 8 centers	Cross-sectional study (based on NAFLD prospective cohort and PIVENS RCT) 2004–2007[a] Not reported	713 patients with liver biopsy–confirmed NAFLD 436 with NASH, method of confirmation not reported	Age: mean, 48 y DM: NR White: 76.2% Male: 37.7% HTN: 27% Statin use: NR Smoking: NR	Biopsy centrally confirmed and reported as stage 0–4	HRQoL (SF-36)	Not applicable
Hagström (2017)[20,21] Sweden 2 centers	Retrospective cohort study 1971–2009[d] All patients	646 patients with liver biopsy–confirmed NAFLD 383 with NASH, defined by FLIP algorithm	Age: mean, 48 y DM: 14.4% White: NR Male: 62.2% HTN: 30.3% Statin use: NR Smoking: 24.0%	Biopsy centrally confirmed and reported as stage 0–4	Overall mortality, severe liver disease[e]	Mean: 19.9 y Range: 0.4–40
Huber et al (2019)[22] Germany, Spain, United Kingdom (European NAFLD registry) 3 centers	Prospective cohort study Not reported Not reported	304 patients with liver biopsy–confirmed NAFLD 210 with NASH, defined by the presence of steatosis, ballooning, and lobular inflammation	Age: median, 54 y DM: 51.3% [T2] White: NR Male: 53.3% HTN: 66.8% Statin use: NR Smoking: NR	Biopsy centrally confirmed and reported as stage 0–4	HRQoL, CLDQ	Up to 6 months after biopsy
Ito et al (2019)[23] Japan 2 centers	Retrospective cohort study 1999–2014[e] All patients	246 patients with liver biopsy–confirmed NAFLD 156 with NASH, defined by FLIP criteria	Age: median, 55 y DM: 45.1% White: NR Male: 52% HTN: 41.6%	Biopsy centrally confirmed and reported as stage 0–4	Overall mortality, liver cirrhosis, liver cancer, extrahepatic cancer,	Median: 7.0 y Range: 4.4–10.0

Table 1. Continued

First author (year) and country	Study design, time period, and sampling frame	Population diagnosis	Population demographics	Fibrosis staging	Outcomes reported	Follow-up
			Statin use: NR Smoking: NR		cardiovascular disease	
Leung et al (2017)[24] Hong Kong 1 center	Prospective cohort study 2006–2015[f] Consecutive patients	300[g] patients with liver biopsy–confirmed NAFLD 151 with NASH, defined by FLIP criteria	Age: mean, 51 y DM: 55.4% White: NR Male: 55.7% HTN: 55.4% Statin use: NR Smoking: NR	Biopsy centrally confirmed and reported as stage 0–4	Overall mortality, liver-related events,[h] nonhepatic cancer, cardiovascular disease	Median: 4.1 y Range: NR
Peleg et al (2018)[25] Israel 1 center	Retrospective cohort study 2005–2012[f] All patients	153 patients with liver biopsy–confirmed NAFLD 27 with NASH, defined by the presence of steatosis, ballooning, and lobular inflammation	Age: mean, 49.5 y DM: 63.4% [T2] White: NR Male: 55.5% HTN: 41.1% Statin use: 53.8% Smoking: NR	Biopsy confirmed and reported as stage 0–4	Overall mortality, malignancies, liver events,[i] hospital admissions	Mean: 8.3 y Range: 5.1–12.0 y
Sebastiani et al (2015)[26] Canada Single center	Retrospective cohort study 2004–2013[j] Consecutive patients	148 patients with liver biopsy–confirmed NAFLD, 148 with NASH, definition not specified	Age: mean, 49.5 y DM: 33.1% White: NR Male: 69.6% HTN: 39.2% Statin use: NR Smoking: NR	Biopsy confirmed and reported as stage 0–4	Clinical outcomes[k]	Median: 5 y Interquartile range: 3–8 y
Seko et al (2015)[27] Japan 1 center	Retrospective cohort study 1999–2013[f] All patients	312 patients with liver biopsy–confirmed NAFLD 176 with NASH, defined by Younossi criteria[29]	Age: median, 59 y DM: 35% White: NR Male: 51% HTN: NR Statin use: 40.3% Smoking: NR	Biopsy confirmed and reported as stage 0–4	Overall mortality, malignancies	Median: 4.8 y Range: 0.3–15.7 y
Vilar-Gomez (2018)[28] Australia, Cuba, Hong Kong, Spain 5 centers	Prospective cohort study 1995–2016[l] Consecutive patients	458 patients with liver biopsy–confirmed NAFLD 458 with assumed-to-be NASH by nature of stage 3–4 fibrosis	Age: mean, 55.9 y DM: 67% White: 81% Male: 48% HTN: 61% Statin use: 24% Smoking: 17%	Biopsy reviewed independently and reported as stage 3 and 4	Overall mortality, major clinical events[m]	Mean: 5.5 y Range: 2.7–8.2

Table 1. Continued

First author (year) and country	Study design, time period, and sampling frame	Population diagnosis	Population demographics	Fibrosis staging	Outcomes reported	Follow-up
Younossi et al (2011, 2017)[29,30] United States 3 centers	Retrospective cohort study Not reported Not reported	210[n] patients with liver biopsy–confirmed NAFLD 131 with NASH, defined by the presence of steatosis, ballooning, and lobular inflammation	Age: mean, 49 y[o] DM: 20.5% [T2] White: NR Male: 37.8% HTN: NR Statin use: NR Smoking: NR	Biopsy-confirmed NAS and Brunt 0–4 fibrosis	Liver-related mortality	Median: 12.1 y IQR: 4.9–15.5
Younossi et al (2018)[31] United States/Canada 23 centers	Randomized controlled trial 2015–2017[f] Not reported	72 patients with liver biopsy–confirmed NAFLD 72 with NASH, defined by the presence of steatosis, ballooning, and lobular inflammation	Age: mean, 54 y DM: 70.8% White: 90.3% Male: 30.6% HTN: 66.7% Statin use: NR Smoking: NR	Biopsy-confirmed stage 2 or 3 fibrosis	Health-related quality of life (SF-36 and CLDQ)	Up to 24 weeks

DM, diabetes mellitus; HTN, hypertension; PIVENS, Pioglitazone, Vitamin E or Placebo for Nonalcoholic Steatohepatitis trial; NAS, NAFLD Activity Score; NR, not reported, T2, Type 2 diabetes mellitus.
[a]Year of recruitment.
[b]Gastroesophageal varices/bleeding, ascites, portosystemic encephalopathy, spontaneous bacterial peritonitis, hepatocellular cancer, hepatopulmonary syndrome, hepatorenal syndrome.
[c]Liver failure, gastroesophageal varices, ascites, encephalopathy, hepatopulmonary syndrome, hepatocellular carcinoma.
[d]Year of diagnosis.
[e]Acute and subacute liver failure, ascites, esophageal varices, hepatorenal syndrome, chronic liver failure, cirrhosis non-ulcer dyspepsia, hepatic encephalopathy, liver failure NUD, portal hypertension, hepatocellular carcinoma.
[f]Year of biopsy.
[g]There were 307 patients reported in the article, but data provided by research groups included only 300.
[h]Hepatocellular carcinoma, ascites, spontaneous bacterial peritonitis, hepatorenal syndrome, variceal bleeding, hepatic encephalopathy, liver transplant.
[i]Esophageal varices, hepatic encephalopathy, ascites, and transjugular intrahepatic portosystemic shunting.
[j]Year of study visits.
[k]Death, liver transplant, cirrhosis complications.
[l]Years of study.
[m]First event of hepatic decompensation, hepatic chronic cirrhosis, major vascular events, and non-hepatic malignancies.
[n]Based on number reported in Dulai et al review.[8]
[o]Weighted mean.

Table 2. Assessment of Risk of Bias of Included Studies, Based on Quality in Prognosis Studies Tool

First author (year)	Study population	Study attrition	Prognostic factor measurement	Outcome measurement	Confounding assessment and account	Data analysis and reporting	Overall assessment[a]
Angulo et al (2015)[5]	Yes	Partly	Yes	Partly	Yes	Yes	Moderate risk of bias
Bhala et al (2011)[18]	Yes	Yes	Yes	Yes	Yes	Partly	Moderate risk of bias
David et al (2009)[19]	Partly	No	Yes	Yes	Partly	Partly	Moderate risk of bias
Hagström (2017)[20,21]	Yes	Partly	Yes	Yes	Partly	Partly	Moderate risk of bias
Huber et al (2019)[22]	Partly	No	Yes	Yes	Partly	Partly	Moderate risk of bias
Ito et al (2019)[23]	Yes	Partly	Yes	Yes	Partly	Partly	Moderate risk of bias
Leung et al (2017)[24]	Yes	Partly	Partly	Yes	No	No	High risk of bias
Peleg et al (2018)[25]	Yes	Partly	Yes	Yes	Partly	Partly	Moderate risk of bias
Sebastiani et al (2015)[26]	Yes	Partly	Yes	Yes	Partly	Partly	Moderate risk of bias
Seko et al (2015)[27]	No	Partly	Partly	Partly	Partly	Partly	Moderate risk of bias
Vilar-Gomez et al (2018)[28]	Yes	Yes	Yes	Yes	Yes	Yes	Low risk of bias
Younossi et al (2011, 2017)[29,30]	Partly	Partly	Yes	Yes	Partly	No	Moderate risk of bias
Younossi et al (2018)[31]	Partly	Yes	Yes	Yes	Yes	No	Moderate risk of bias

[a]Low risk of bias describes studies for which all domains are scored as *yes*. Moderate risk of bias describes studies for which 1 or more domains are scored as *partly* or 1 domain is scored as *no*. High risk of bias describes studies for which more than 1 domain is scored as *no*.

appropriate, and instead, numeric results were summarized across individual studies (see Supplementary Table 5).

The cross-sectional analysis of David et al[19] used the generic SF-36 to report that in a total of 713 patients with NAFLD, those with stage 4 fibrosis (cirrhosis) had significantly ($P < .001$) worse physical health as assessed by SF-36 Physical Component Score compared with patients with NAFLD and fibrosis stages 0–3 (median, 37 vs 47–50; $P < .001$). This finding remained after adjustment for potential confounders (ie, age, sex, race, marital status, education, annual household income, body mass index, type 2 diabetes). The study investigators reported no significant difference across fibrosis stage for SF-36 Mental Component Score (data not reported). Those with NASH reported significantly poorer physical health compared with those with no NASH (median, 22.5 vs. 47.1; $P < .02$).

The prospective cohort of Huber et al[22] found no difference in unadjusted total CLDQ score comparing a total of 304 patients with NAFLD stage 3 or 4 and stage 0–2 fibrosis (mean [SD], 4.9 [1.2] vs 5.1 [1.3]; $P = .07$). NASH was associated with a significantly lower HRQoL compared with patients with NAFL (mean [SD], 4.85 [1.3] vs 5.31 [1.1]; $P < .01$).

In an RCT with 72 patients with NASH, Younossi et al[31] found no difference in unadjusted baseline HRQoL between stage 2 and 3 fibrosis in either SF-36 (Physical Component Score: mean [SD], 45.0 [9.6] vs 43.4 [10.3]; Mental Component Score: 51.0 [9.6] vs 50.6 [12.7]; both $P > .05$) or total CLDQ score (mean [SD], 4.83 [1.10] vs 4.91 [1.25], $P > .05$).[30]

Metaregression

Given the number of studies reporting clinical outcome data, we were able to undertake univariate meta-regression for RR analysis for all-cause mortality and all liver events for patients with NAFLD. There was no evidence of a differential effect of study-level characteristics (ie, study design, overall risk of bias, or follow-up) on the impact of stage of fibrosis for either of these outcomes (see Supplementary Table 6).

Small Study Bias

We were able to assess small study bias for the relative outcomes of all-cause mortality and all liver events in patients with NAFLD. There was no formal evidence of funnel plot asymmetry, except for all liver events for comparison of fibrosis stage 0 vs 2 (Egger test, $P = .05$) (see Supplementary Figure 2). This asymmetry appeared to be due to an absence of small- to medium-sized studies with an RR of <1.0.

Quality of Evidence

Based on the Grading of Recommendations, Assessment, Development, and Evaluations (GRADE) approach,[13] we found the quality of evidence for fibrosis in NAFLD as a prognostic predictor of all-cause mortality to be high and for liver mortality to be moderate (see Supplementary Table 7). The quality of evidence for liver-related mortality. liver transplant, and HRQoL for both NAFLD and NASH were all judged to be low due to the sparse number of events or small number of studies. The outcome of all liver events was also judged to be of low quality because of inconsistency in its definition across studies. Given the smaller amount of evidence (studies and events), evidence quality for all outcomes for NASH was low.

Discussion

This systematic review and meta-analysis identified a substantive and consistent body of international observational evidence that showed that stage of biopsy-confirmed liver fibrosis is a strong predictor of future all-cause

Table 3. Meta-analysis: Pooled Unadjusted Relative Risk by Fibrosis Stage (Relative to Stage 0) for All Patients With NAFLD

Number of studies	Stage 0 vs 1 RR (95% CI), *P* value n/N vs n/N, I^2 statistic	Stage 0 vs 2 RR (95% CI), *P* value n/N vs n/N, I^2 statistic	Stage 0 vs 3 RR (95% CI), *P* value n/N vs n/N, I^2 statistic	Stage 0 vs 4 RR (95% CI), *P* value n/N vs n/N, I^2 statistic
All-cause mortality				
8	1.12 (0.91–1.38) 135/843 vs 136/896, 0%	1.50 (1.20–1.86) 135/843 vs 103/425, 0%	2.13 (1.70–2.67) 135/843 vs 86/301, 0%	3.42 (2.63–4.46) 135/843 vs 61/169, 27%
Liver-related mortality				
7	1.05 (0.35–3.16) 3/521 vs 7/755, 0%	2.53 (0.88–7.27) 3/521 vs 10/340, 0%	6.65 (1.99–22.25) 3/521 vs 12/248, 0%	11.13 (4.15–29.84), 0% 3/521 vs 22/151
Liver transplantation				
6	0.40 (0.02–7.50) 0/466 vs 2/691, 0%	1.98 (0.24–16.10) 0/466 vs 3/314, 0%	RR not calculable 0/466 vs 0/205, 0%	5.42 (1.05–27.89) 0/466 vs 6/129, 0%
All liver events				
7	1.02 (0.58–1.89) 18/787 vs 25/823, 0%	2.67 (1.58–4.51) 19/787 vs 39/399, 0%	5.24 (3.97–8.98) 19/787 vs 39/256, 0%	12.78 (6.85–23.85) 19/787 vs 52/156, 0%

NOTE. All meta-analyses were fixed effect.

mortality and morbidity in NAFLD (with a 5–12 fold increase in RR of death and liver-related events, including liver failure, transplantation, and liver cancer). Beyond the increased risk associated with fibrosis, the available data do not provide evidence for additional differential risk between the reported subgroups of patients with NAFLD with NAFL or NASH. There was, however, limited and contradictory evidence of the impact of stage of fibrosis on the HRQoL,

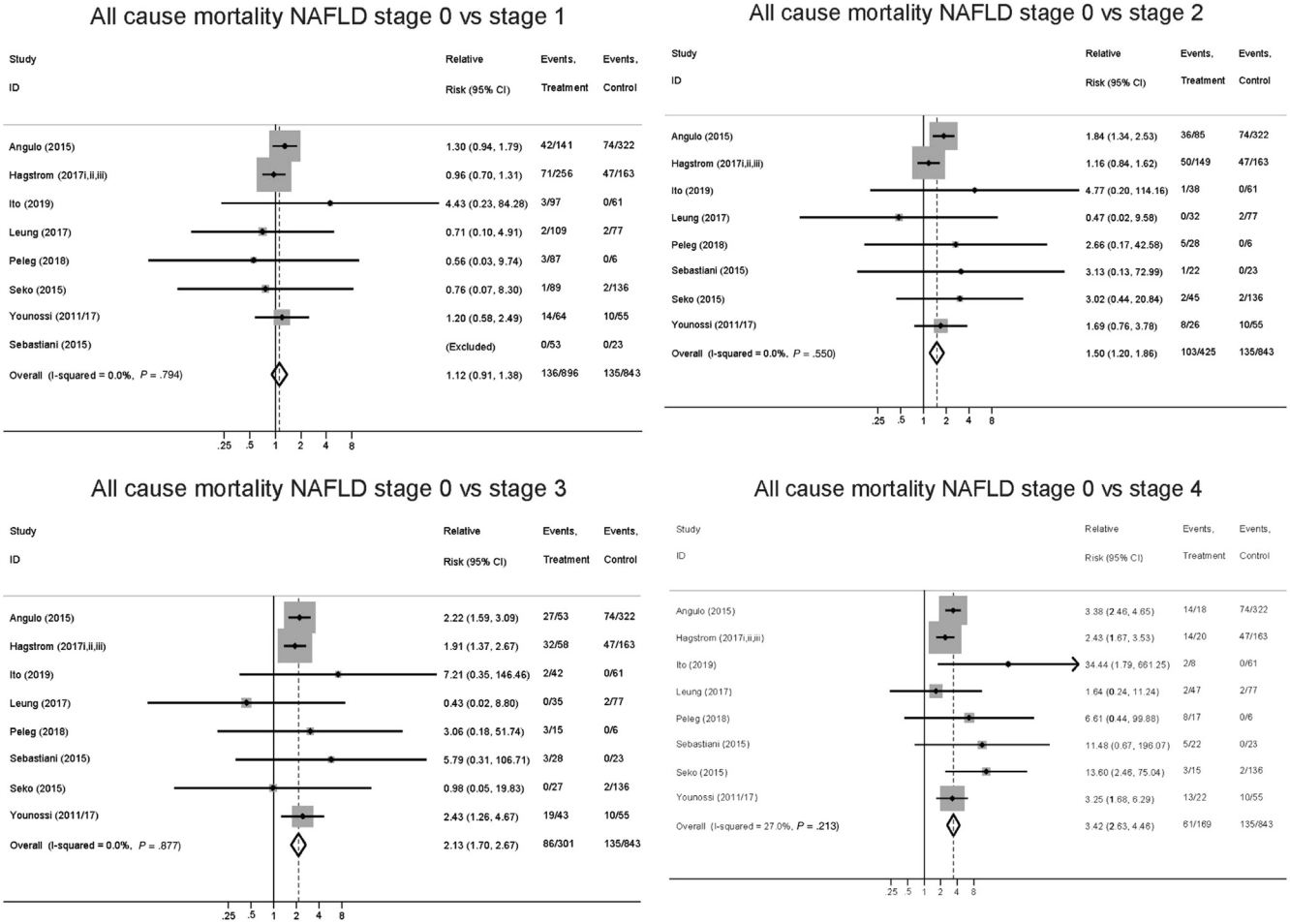


Figure 2. Meta-analysis: unadjusted RR of all-cause mortality by fibrosis stage (vs stage 0) in all patients with NAFLD.

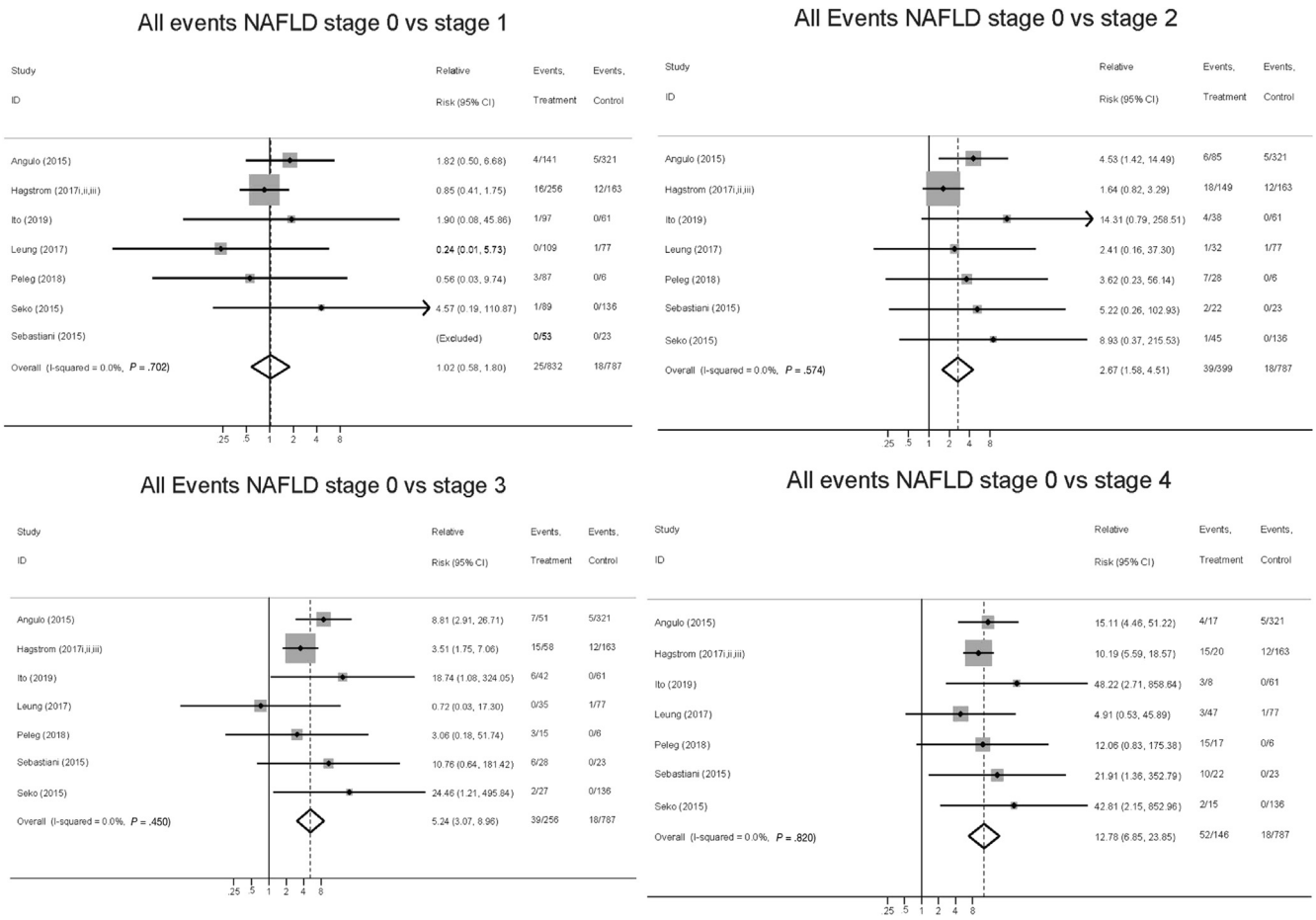


Figure 3. Meta-analysis: unadjusted RR of liver events by fibrosis stage (vs stage 0) in all patients with NAFLD.

primarily due to the small number of studies, heterogeneity of the study participants, and lack of data from NASH-specific HRQoL instruments, such as the CLDQ-NASH.[32]

In the Context of Current Evidence

This study shows that both with and without adjustment for key potential confounding variables, biopsy-confirmed fibrosis is a key prognostic marker for all-cause and liver-related mortality in patients with NAFLD.[8] The size and methodologic rigor of this study now provides the confidence to support the conclusions of previous studies and recommendations of clinical guidelines. With advancing fibrosis, there is a stepwise increase in RR for liver morbidity, liver mortality, and all-cause mortality.

Our review also extends previous findings to the subset of patients who have reported histologic evidence of NASH, showing that the risk of mortality and morbidity of increasing fibrosis stage appears be similar in magnitude to that seen for the whole cohort of patients with NAFLD, which includes patients categorized as currently having histologic evidence of NASH or non-NASH. This is particularly important given the increasing focus of clinical trials on interventions on the inclusion of patients with NASH and the focus of these trials on a primary outcome that includes biopsy-confirmed fibrosis.[33–35]

The FDA recently published a table of surrogate endpoints that either have been already used in their development programs for drugs that have been approved or are surrogate endpoints that the FDA has indicated acceptance of in their guidance or other documents.[36,37] The FDA table of surrogate endpoints currently lists an "improvement of fibrosis with no worsening of steatohepatitis" as a surrogate endpoint for clinical trials in NASH.[36] Notably, our review did not identify strong evidence from RCTs that have reported an association between treatment-related improvement of stage of fibrosis and mortality, morbidity, or HRQoL. Therefore, currently, there appears to be no direct scientific evidence to validate fibrosis improvement as an established and validated surrogate endpoint of long-term outcomes. Although surrogacy of fibrosis is biologically plausible, and stage of fibrosis is a strong prognostic marker, making fibrosis improvement a reasonable endpoint for granting provisional regulatory approval, there is ultimately a need to generate robust data to support this based on regulatory treatment trials in this field. This is important because regulatory bodies and payers, who are responsible for health care reimbursement decisions and are typically more stringent in their evidence requirements, prefer evidence from final patient-relevant outcomes and will accept surrogate endpoints only if they are based on formal evidence of validation.[5,6] The importance of the link between putative

Table 4. Stratified Meta-analysis: Pooled Unadjusted RR by Fibrosis Stage (Relative to Stage 0) for Patients With NAFLD With Reported NASH vs Patients With NALFD With No Reported NASH (n = 4 Studies)

Mortality and clinical events	Stage 0 vs 1 RR (95% CI), *P* value n/N vs n/N, I^2 statistic	Stage 0 vs 2 RR (95% CI), *P* value n/N vs n/N, I^2 statistic	Stage 0 vs 3 RR (95% CI), *P* value n/N vs n/N, I^2 statistic	Stage 0 s 4 RR (95% CI), *P* value n/N vs n/N, I^2 statistic
All-cause mortality				
NAFLD with NASH	0.91 (0.54–1.51) 13/83 vs 44/319, 0%	1.24 (0.74–2.07) 13/83 vs 47/202, 0%	1.99 (1.17–3.41) 13/83 vs 45/155, 0%	3.26 (1.78–5.98) 13/83 vs 31/90, 0%
NAFLD without NASH	1.15 (0.87–1.52) 46/279 vs 49/294, 29%	1.40 (0.85–2.28) 46/279 vs 17/71, 0%	2.60 (1.64–4.09) 46/279 vs 11/38, 0%	2.91 (1.08–7.87) 46/279 vs 8/23, 0%
Liver-related mortality				
NAFLD with NASH	0.35 (0.07–1.77) 2/83 vs 3/319, 0%	0.78 (0.21–2.92) 2/83 vs 6/201, 0%	1.24 (0.31–4.93) 2/83 vs 10/155, 0%	3.74 (0.83–16.83) 2/83 vs 13/90, 0%
NAFLD without NASH	1.10 (0.40–3.04) 1/279 vs 3/291, 0%	7.31 (0.68–78.10) 1/279 vs 2/72, NA	26.0 (2.60–260.04) 1/279 vs 2/38, NA	8.17 (1.27–52.58) 1/279 vs 18/114, 0%
Liver transplantation				
NAFLD with NASH	RR not estimable 0/62 vs 0/281, NA	RR not estimable 0/62 vs 0/176, NA	RR not estimable 0/62 vs 0/114, NA	RR not estimable 0/62 vs 1/69, NA
NAFLD without NASH	0.47 (0.02–8.79) 0/245 vs 2/268, NA	3.50 (0.52–23.69) 0/245 vs 3/71, 0%	RR not estimable 0/245 vs 0/36, NA	15.07 (0.63–359.22)[a] 0/245 vs 3/23, 56%
All liver events				
NAFLD with NASH	0.47 (0.17–1.29) 5/77 vs 9/281, 0%	1.21 (0.51–2.91) 5/77 vs 19/176, 0%	2.16 (0.85–4.47) 5/77 vs 17/114, 0%	6.48 (2.89–14.85) 5/77 vs 23/69, 0%
NAFLD without NASH	1.08 (0.45–2.58) 8/230 vs 11/268, 0%	2.85 (1.12–7.24) 8/230 vs 11/71, 0%	4.56 (1.64–12.60) 8/230 vs 7/36, 0%	9.80 (3.12–30.76) 8/230 vs 15/28, 0%

NA, not applicable.
[a]Random-effects meta-analyses. All other meta-analyses were fixed effect.

surrogates to clinically meaningful outcomes is recognized in the recent publication from an international workshop on clinical trial endpoints.[38]

Strengths and Limitations

We believe this to be the most contemporary, comprehensive, and methodologically robust assessment of the literature to date, including 4428 patients with NAFLD and 591 all-cause deaths. In contrast, the systematic review and meta-analysis of Dulai et al[8] included 1495 patients with NAFLD and 348 deaths. We extended the scope of this previous study to consider the potential impact of key confounder variables, the subgroup of patients with NAFLD with NASH, the impact on liver-related morbidity and patient HRQoL, and the evidence base for change in stage of fibrosis as surrogate endpoint. Eleven out of 13 research teams of the included studies provided additional quantitative outcome data (not reported in their original published papers). As a result, we were able to ensure the inclusiveness of our meta-analysis. The comprehensiveness of data capture is supported by our finding of little or no publication bias.

However, we recognize that our review has some limitations that largely reflect the nature and reporting of included studies. First, our primary analysis (and where we had most available data)—that is, estimation of pooled RR—was based on a simple comparison of the risk of outcomes in patients according to their stage of fibrosis (fibrosis stage 0 as reference). Given the fact that the demographic and clinical characteristics of patients (eg, age, sex, diabetes status) for the fibrosis stage categories is likely to be different, this crude (or unadjusted) analysis of RRs is likely to be prone to confounding. However, our adjusted analysis showed that the magnitude of outcome risk with increased fibrosis stage (fibrosis stage 0–2 vs 3 or 4) was similar when compared with the results of the simple (unadjusted) pooled RR approach to pooling studies using hazard ratios and following adjustment for potential key confounders. Second, although we sought to extend our review to include data on NASH, included studies often did not provide a clear definition or explanation of how NASH was diagnosed. Differential diagnosis of NAFLD and NASH is a well-recognized controversy of current clinical practice.[39] To make our findings as robust as possible, we limited our meta-analysis to the subgroup of studies that had a clear definition of NASH, such as the FLIP or NASH CRN score. However, even when selecting studies with a clear definition of NASH, we recognize that some patients with NASH (steatosis, ballooning, and lobular inflammation) may be miscategorized as not having NASH because of sampling error on the biopsy. Moreover, a liver biopsy represents only a single point in time, and steatohepatitis may fluctuating over time due to complex gene–environment interactions and in response to weight loss. Furthermore, as fibrosis progresses toward cirrhosis, some features of NASH, such as steatosis and hepatocyte ballooning, may become less prominent and, thus, a patient may be categorized as not having active NASH, yet NASH was clearly the causative

factor in the liver fibrosis. Third, studies reported liver-related morbidity based on differing combinations of liver-related clinical events. Therefore, there is a need for caution in the interpretation of the meta-analysis pooling of this composite outcome across studies. Fourth, although we sought to include a range of clinical outcomes, the wide meta-analysis CIs for some fibrosis stage outcome comparisons indicate the relatively sparse number of events available, especially liver transplantation. However, we also found no evidence of publication bias. Finally, included studies were of mixed methodologic quality—7 out of 13 studies were retrospective in design, and 3 were overall judged to be at high risk of bias. Nevertheless, our metaregression analysis showed that our findings were insensitive to either study design or overall study risk of bias.

Our review has identified several important areas for future research. First, we need to better understand the association between fibrosis stage and patient-reported well-being. Future outcomes for NAFLD and NASH studies, therefore, need to consistently collect patient HRQoL using generic (such as the 5 level EuroQoL) and disease-specific measures (such as the CLDQ-NASH[32]). Second, formal scientific validation of fibrosis as an acceptable surrogate endpoint is needed. Accepted statistical methods for surrogate validation include demonstration of a surrogate–final outcome association based on patient-level data from a single RCT or from meta-analyses of multiple RCTs.[16,40,41] This evidence need is being addressed through long-term follow-up capturing hard clinical outcomes in all NASH phase 3 trials that are currently recruiting (eg, REGENERATE, REVERSE, RESOLVE-IT, AURORA).[42–45] Third, biopsy is an invasive procedure that limits clinical applicability in routine screening for NASH, and there is a need, therefore, to investigate the suitability of other noninvasive alternative biomarkers as prognostic markers or validated surrogate endpoints, an issue that is currently being explored by 2 large international multi-stakeholder consortia in Europe (IMI2 LITMUS) and the United States (FNIH NIMBLE).[46,47]

In conclusion, our study shows that with and without adjustment of key confounders, biopsy-confirmed fibrosis is a key prognostic marker of both mortality and liver-related morbidity in NAFLD and the subgroups of patients with NAFLD with and without reported NASH, with increasing fibrosis stage being associated with a 5- to 12-fold increase in the RR of liver-related events. Further evidence from well-reported studies is needed to clarify the impact of fibrosis stage on patient well-being (including NASH-specific HRQoL instruments) and to confirm change in biopsy-confirmed fibrosis as a valid surrogate endpoint in the context of RCTs of treatments for NAFLD and NASH.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at https://doi.org/10.1053/j.gastro.2020.01.043.

References

1. Younossi Z, Anstee QM, Marietti M, et al. Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. Nat Rev Gastroenterol Hepatol 2018; 15:11–20.
2. Farrell GC, Larter CZ. Nonalcoholic fatty liver disease: from steatosis to cirrhosis. Hepatology 2006;43:S99–S112.
3. Bedossa P, FLIP Pathology Consortium. Utility and appropriateness of the fatty liver inhibition of progression (FLIP) algorithm and steatosis, activity, and fibrosis (SAF) score in the evaluation of biopsies of nonalcoholic fatty liver disease. Hepatology 2014;60:565–575.
4. McPherson S, Hardy T, Henderson E, et al. Evidence of NAFLD. Progression from steatosis to fibrosing-steatohepatitis using paired biopsies: implications for prognosis & clinical management. J Hepatol 2015; 62:1148–1155.
5. Angulo P, Kleiner DE, Dam-Larsen S, et al. Liver fibrosis, but no other histologic features, is associated with long-term outcomes of patients with nonalcoholic fatty liver disease. Gastroenterology 2015;149:389–397.
6. Hannah WN Jr, Torres DM, Harrison SA. Nonalcoholic steatohepatitis and endpoints in clinical trials. Gastroenterol & Hepatol 2016;12:756–763.
7. Ciani O, Buyse M, Drummond M, et al. Use of surrogate end points in healthcare policy: a proposal for adoption of a validation framework. Nat Rev Drug Discov 2016; 15:516.
8. Dulai PS, Singh S, Patel J, et al. Increased risk of mortality by fibrosis stage in nonalcoholic fatty liver disease: systematic review and meta-analysis. Hepatology 2017; 65:1557–1565.
9. Safiri S, Khazaei S, Mansori K, Ayubi E. Comments on "Increased Risk of Mortality by Fibrosis Stage in Nonalcoholic Fatty Liver Disease: Systematic Review and Meta-Analysis". Hepatology 2017;66:1358–1359.
10. Greenland S, Mansournia MA, Altman DG. Sparse data bias: a problem hiding in plain sight. BMJ 2016; 352:i1981.
11. Moher DA, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol 2009;62:1006–1012.
12. Hayden JA, Côté P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. Ann Inter Med 2006;144:427–437.
13. Iorio A, Spencer FA, Falavigna M, et al. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. BMJ 2015;350:h870.
14. Higgins JPT, Thomas J, Chandler J, et al, eds. Cochrane handbook for systematic reviews of interventions,

version 6.0. https://training.cochrane.org/handbook. Updated August 2019. Accessed December 30, 2019.
15. Egger M, Davey Smith G, Schneiger M, et al. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997;315:629–634.
16. Ciani O, Davis S, Tappenden P, et al. Validation of surrogate endpoints in advanced solid tumours: systematic review of statistical methods, results, and implications for policy makers. Int J Technol Assess Health Care 2014;30:312–324.
17. Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. Pharm Stat 2006;5:173–186.
18. Bhala N, Angulo P, van der Poorten D, et al. The natural history of nonalcoholic fatty liver disease with advanced fibrosis or cirrhosis: an international collaborative study. Hepatology 2011;54:1208–1216.
19. David K, Kowdley KV, Unalp A, et al. Quality of life in adults with nonalcoholic fatty liver disease: baseline data from the nonalcoholic steatohepatitis clinical research network. Hepatology 2009;49:1904–1912.
20. Hagström H, Nasr P, Ekstedt M, et al. Fibrosis stage but not NASH predicts mortality and time to development of severe liver disease in biopsy-proven NAFLD. J Hepatol 2017;67:1265–1273.
21. Hagström H, Nasr P, Ekstedt M, et al. Risk for development of severe liver disease in lean patients with nonalcoholic fatty liver disease: a long-term follow-up study. Hepatol Commun 2017;2:48–57.
22. Huber Y, Boyle M, Hallsworth K, et al. Health-related quality of life in non-alcoholic fatty liver disease associates with hepatic inflammation. Clin Gastroenterol Hepatol 2019;17:2085–2092.
23. Ito T, Ishigami M, Ishizu Y, et al. Utility and limitations of noninvasive fibrosis markers for predicting prognosis in biopsy-proven Japanese non-alcoholic fatty liver disease patients. J Gastroenterol Hepatol 2019;34:207–214.
24. Leung JC, Loong TC, Wei JL, et al. Histological severity and clinical outcomes of nonalcoholic fatty liver disease in nonobese patients. Hepatology 2017;65:54–64.
25. Peleg N, Sneh Arbib O, Issachar A, et al. Noninvasive scoring systems predict hepatic and extra-hepatic cancers in patients with nonalcoholic fatty liver disease. PLoS One 2018;13(8):e0202393.
26. Sebastiani G, Alshaalan R, Wong P, et al. Prognostic value of non-invasive fibrosis and steatosis tools, hepatic venous pressure gradient (HVPG) and histology in nonalcoholic steatohepatitis. PLoS One 2015;10(6):e0128774.
27. Seko Y, Sumida Y, Tanaka S, et al. Predictors of malignancies and overall mortality in Japanese patients with biopsy-proven non-alcoholic fatty liver disease. Hepatol Res 2015;45:728–738.
28. Vilar-Gomez E, Calzadilla-Bertot L, Wai-Sun Wong V, et al. Fibrosis severity as a determinant of cause-specific mortality in patients with advanced nonalcoholic fatty liver disease: a multi-national cohort study. Gastroenterol 2018;155:443–457.
29. Younossi ZM, Stepanova M, Rafiq N, et al. Pathologic criteria for nonalcoholic steatohepatitis: interprotocol agreement and ability to predict liver-related mortality. Hepatology 2011;53:1874–1882.
30. Younossi ZM, Stepanova M, Rafiq N, et al. Nonalcoholic steatofibrosis independently predicts mortality in nonalcoholic fatty liver disease. Hepatol Commun 2017;1:421–428.
31. Younossi ZM, Stepanova M, Lawitz E, et al. Improvement of hepatic fibrosis and patient-reported outcomes in non-alcoholic steatohepatitis treated with selonsertib. Liver Int 2018;38:1849–1859.
32. Younossi ZM, Stepanova M, Younossi I, Racila A. Validation of Chronic Liver Disease Questionnaire for Nonalcoholic Steatohepatitis in patients with biopsy-proven nonalcoholic steatohepatitis. Clin Gastroenterol Hepatol 2019;17:2093–2100.
33. McPherson S, Wilkinson N, Tiniakos D, et al. A randomised controlled trial of losartan as an anti-fibrotic agent in non-alcoholic steatohepatitis. PLoS One 2017;12(4):e0175717.
34. Sanyal AJ, Chalasani N, Kowdley KV, et al. Pioglitazone, vitamin E, or placebo for nonalcoholic steatohepatitis. N Engl J Med 2010;362:1675–1685.
35. Neuschwander-Tetri BA, Loomba R, Sanyal AJ, et al. Farnesoid X nuclear receptor ligand obeticholic acid for non-cirrhotic, non-alcoholic steatohepatitis (FLINT): a multicentre, randomised, placebo-controlled trial. Lancet 2015;385(9972):956–965.
36. US Food & Drug Administration. Table of surrogate endpoints that were the basis of drug approval or licensure. https://www.fda.gov/drugs/development-resources/table-surrogate-endpoints-were-basis-drug-approval-or-licensure. Accessed July 22, 2019.
37. US Food & Drug Administration. Noncirrhotic nonalcoholic steatohepatitis with liver fibrosis: developing drugs for treatment; draft guidance for industry; availability. https://www.federalregister.gov/documents/2019/06/07/2019-11951/nonalcoholic-steatohepatitis-with-compensated-cirrhosis-developing-drugs-for-treatment-draft. Accessed July 22, 2019.
38. Rinella ME, Tacke F, Sanyal AJ, Anstee QM. Report on the AASLD/EASL joint workshop on clinical trial endpoints in NAFLD. J Hepatol 2019;71:823–833.
39. Rinella ME, Loomba R, Caldwell SH, et al. Controversies in the diagnosis and management of NAFLD and NASH. Gastroenterol Hepatol (N Y) 2014;10:219–227.
40. Ciani O, Piepoli M, Smart N, et al. Validation of exercise capacity as a surrogate endpoint in exercise-based rehabilitation for heart failure: a meta-analysis of randomized controlled trials. JACC Heart Fail 2018;6:596–604.
41. Pavey T, Hoyle M, Ciani O, et al. Dasatinib, nilotinib and standard-dose imatinib for the first-line treatment of chronic myeloid leukaemia: systematic reviews and economic analyses. Health Technol Assess 2012;16(42):1–277.
42. Randomized global phase 3 study to evaluate the impact on NASH with fibrosis of obeticholic acid treatment (REGENERATE): NCT02548351. https://clinicaltrials.gov/ct2/show/NCT02548351. Accessed July 22, 2019.

43. Study evaluating the efficacy and safety of obeticholic acid in subjects with compensated cirrhosis due to nonalcoholic steatohepatitis (REVERSE): NCT03439254. https://clinicaltrials.gov/ct2/show/NCT03439254. Accessed July 22, 2019.
44. Phase 3 study to evaluate the efficacy and safety of elafibranor versus placebo in patients with nonalcoholic steatohepatitis (NASH) (RESOLVE-IT): NCT03028740. https://clinicaltrials.gov/ct2/show/NCT02704403. Accessed July 22, 2019.
45. AURORA: phase 3 study for the efficacy and safety of CVC for the treatment of liver fibrosis in adults with NASH: NCT03028740. https://clinicaltrials.gov/ct2/show/NCT03028740. Accessed July 22, 2019.
46. Innovative Medicines Initiative. LITMUS. Liver investigation: testing marker utility in steatohepatitis. https://www.imi.europa.eu/projects-results/project-factsheets/litmus. Accessed July 22, 2019.
47. Foundation for the National Institutes for Health. FNIH Biomarkers Consortium launches NIMBLE to replace invasive and painful biopsy with non-invasive biomarkers for liver disease. https://fnih.org/news/announcements/nimble. Published June 11, 2019. Accessed July 22, 2019.

Received August 29, 2019. Accepted January 22, 2020.

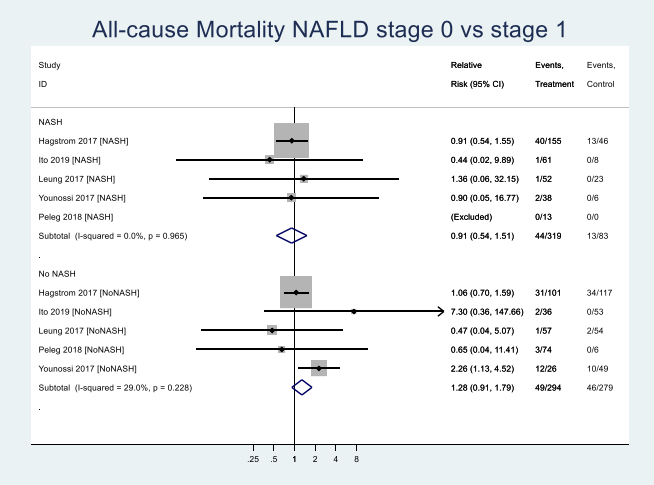
Correspondence
Address correspondence to: Rod S. Taylor MSc, PhD, MRC/CSO, Social and Public Health Sciences Unit, University of Glasgow, top floor, 200 Renfield St, Glasgow, G2 3AX. e-mail: rod.taylor@glasgow.ac.uk.

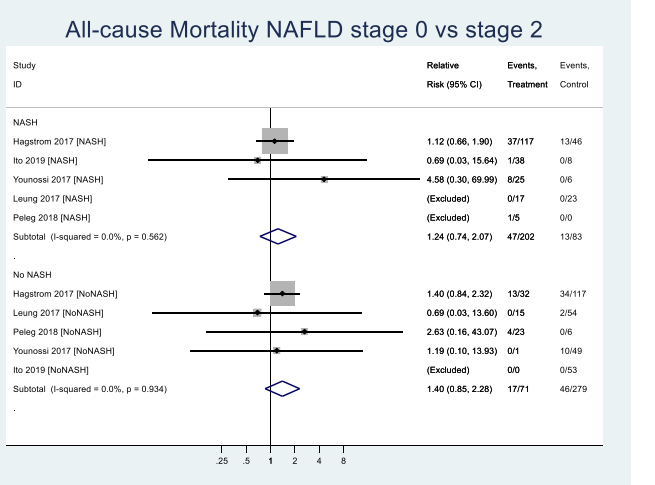
Acknowledgments
We thank Dr Juliana Bottomley, Gilead, for her comments on the review protocol and draft of the manuscript. Collaborators: Mattias Ekstedt, Per Stål, Rolf Hultcrantz, and Stergios Kechagias.

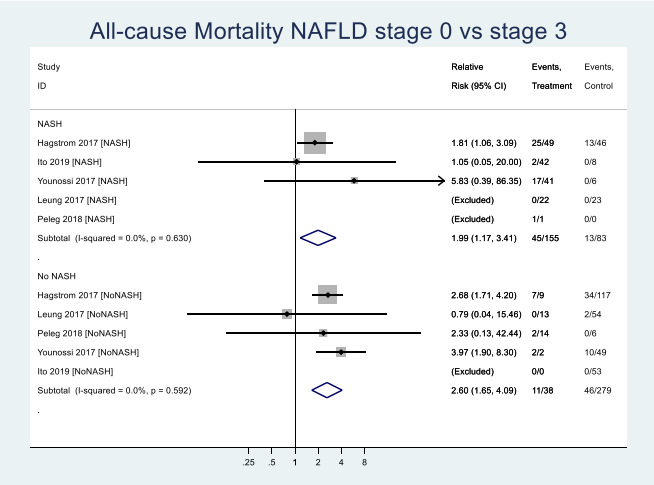
CRediT Authorship Contributions
Rod S. Taylor, PhD (Conceptualization: Lead; Data curation: Lead; Formal analysis:
Lead; Funding acquisition: Lead; Methodology: Lead; Project administration: Lead;
Supervision: Lead; Writing – original draft: Lead; Writing – review & editing: Lead). Rebecca J. Taylor, MSc (Conceptualization: Lead; Data curation: Lead). Sue Bayliss, MSc (Data curation: Lead; Writing – review & editing: Equal). Hanes Hagström, MD (Data curation: Equal; Writing – review & editing: Equal). Patrik Nasr, MD (Data curation: Equal; Writing – review & editing: Equal). Joern Schattenberg, MD (Data curation: Equal; Writing – review & editing: Equal). Quentin Anstee, MD (Data curation: Equal; Writing – original draft: Lead; Writing – review & editing: Lead). Masatoshi Ishigami, MD (Data curation: Equal; Writing – review & editing: Equal). Hidenori Toyoda, MD (Data curation: Equal; Writing – review & editing: Equal). Vincent Wai-Sun Wong, MD (Data curation: Equal; Writing – review & editing: Equal). Noam Peleg, MD (Data curation: Equal; Writing – review & editing: Equal). Amir Shlomai, MD (Data curation: Equal; Writing – review & editing: Equal). Giada Sebastiani, MD (Data curation: Equal; Writing – review & editing: Equal). Yuya Seko, MD (Data curation: Equal; Writing – review & editing: Equal). Neeraj Bhala, MD (Data curation: Equal; Writing – review & editing: Equal). Zobair Younossi, MD (Data curation: Equal; Writing – review & editing: Equal). Stuart McPherson, MD (Conceptualization: Lead; Funding acquisition: Lead; Writing – original draft: Lead; Writing – review & editing: Lead). Philip Newsome, PhD (Conceptualization: Lead; Funding acquisition: Lead; Writing – original draft: Lead; Writing – review & editing: Lead).

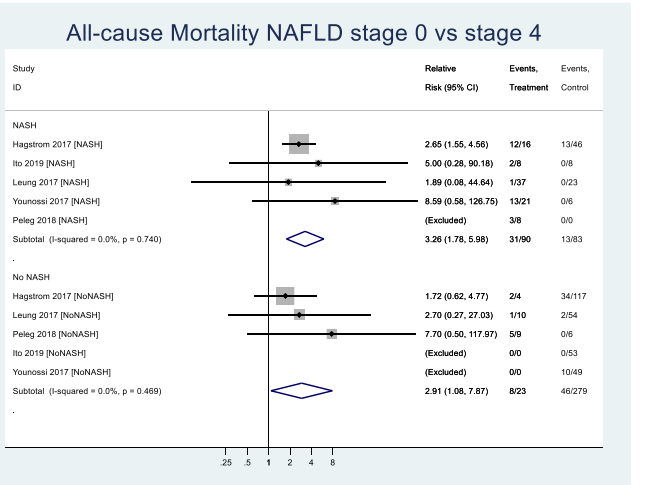
Conflicts of interest
These authors disclose the following: Rod S. Taylor, Rebecca J. Taylor, Sue Bayliss, Stuart McPherson, and Philip N. Newsome have received funding from Gilead for their contributions to this project. Hannes Hagström has received consulting fees from Novo Nordisk, IQVIA, and Gilead Inc; has received research grants from Gilead Inc, AstraZeneca, and Intercept; has served on the advisory board for Bristol-Myers Squibb; has received QMA from AbbVie, Allergan/Tobira, AstraZeneca, GlaxoSmithKline, Glympse Bio, Novartis Pharma AG, Pfizer Ltd, and Vertex; performs consultancy for Abbott Laboratories, Acuitas Medical, Allergan/Tobira, Blade, BNN Cardio, Cirius, CymaBay, EcoR1, E3Bio, Eli Lilly and Company Ltd, Galmed, GENFIT SA, Gilead, Grunthal, HistoIndex, Indalo, Imperial Innovations, Intercept Pharma Europe Ltd, Inventiva, IQVIA, Janssen, Kenes, Madrigal, MedImmune, Metacrine, NewGene, NGM Bio, North Sea Therapeutics, Novartis, Novo Nordisk A/S, Pfizer Ltd, Poxel, ProSciento, Raptor Pharma, Servier, and Viking Therapeutics; has been a speaker for: Abbott Laboratories, Allergan/Tobira, Bristol-Myers Squibb, Clinical Care Options, Falk, Fishawack, GENFIT SA, Gilead, Integritas Communications, Medscape; and has received royalties from Elsevier Ltd. Jorn M. Schattenberg has received consultancy fees from AbbVie, Bristol-Myers Squibb, BBN Cardio, Boehringer Ingelheim, Gala Medical, GENFIT, Gilead Sciences, Intercept Pharmaceuticals, IQVIA, MedImmune, Novartis, and Pfizer; has received research funding from Gilead Sciences and Yakult Europe BV; and has given lectures for Falk Foundation, Takeda, Merck Sharp Dohme; Vincent Wai-Sun Wong has served as a consultant or advisory board member for 3V-BIO, AbbVie, Allergan, Boehringer Ingelheim, Echosens, Gilead Sciences, Janssen, Novartis, Novo Nordisk, Perspectum Diagnostics, Pfizer, and Terns; has been a speaker for Bristol-Myers Squibb, Echosens, Gilead Sciences, and Merck; and has received an unrestricted grant from Gilead Sciences for fatty liver research. Giada Sebastiani has acted as a speaker for Merck, Gilead, AbbVie, ViiV; has served as an advisory board member for Merck, Gilead, and Novartis; and has received research funding from Merck and Echosens. Zobair M. Younossi has received research funding or is a consultant for Gilead Sciences, Intercept, Bristol-Myers Squibb, Novartis, Novo Nordisk, Vicking, Terns, and Siemens. Stuart McPherson has acted as a speaker or advisory board/consultancy for AbbVie, Allergan, Gilead, Intercept, Merk, Sequana. The conclusions are those of the authors and not the manufacturers. The remaining authors disclose no conflicts.

Funding
Gilead funded this project. Quentin M. Anstee and Stuart McPherson are Newcastle NIHR Biomedical Research Centre investigators. Philip N. Newsome was supported by the National Institute of Health Research (NIHR) Birmingham Biomedical Research Centre. Giada Sebastiani is supported by a Junior 1 and 2 Salary Award from Fonds de Recherche du Québec–Santé (#27127 and #267806) and research salary from the Department of Medicine of McGill University. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health.

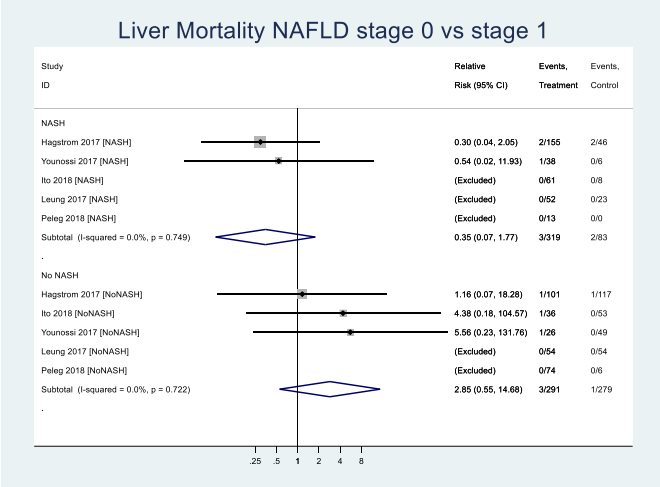


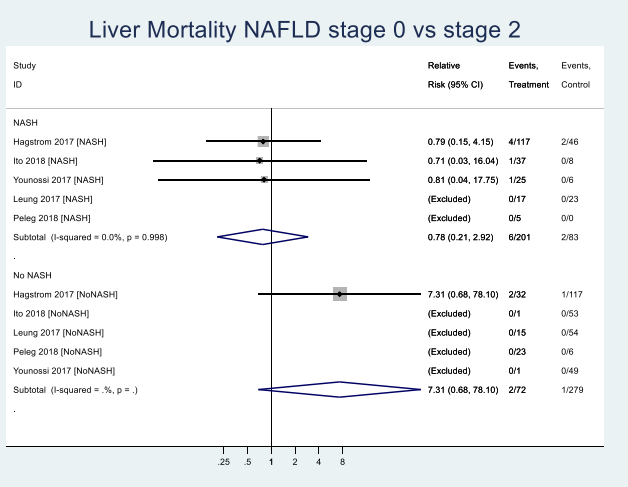


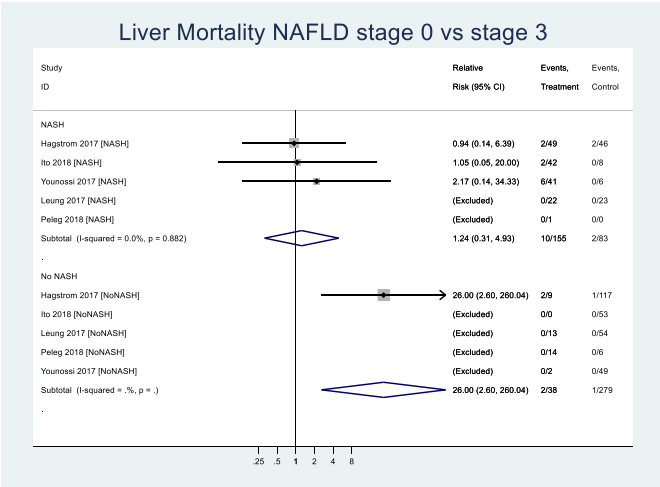


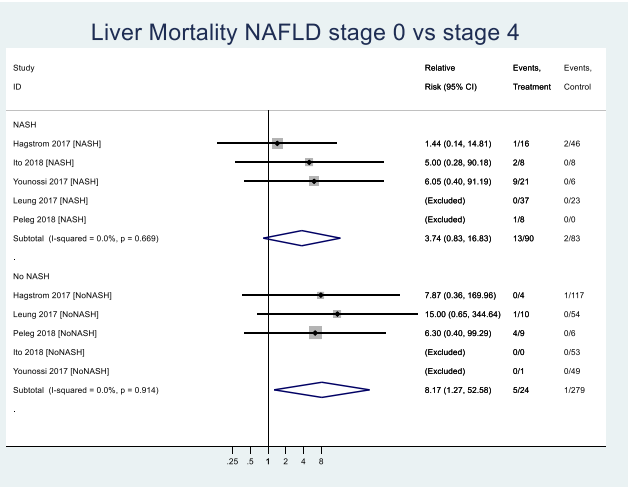


Supplementary Figure 1. Meta-analysis of Events by NAFLD With NASH vs NALFD Without NASH

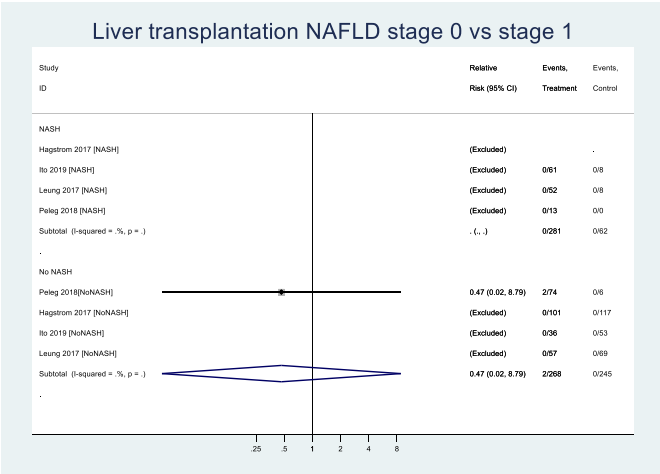


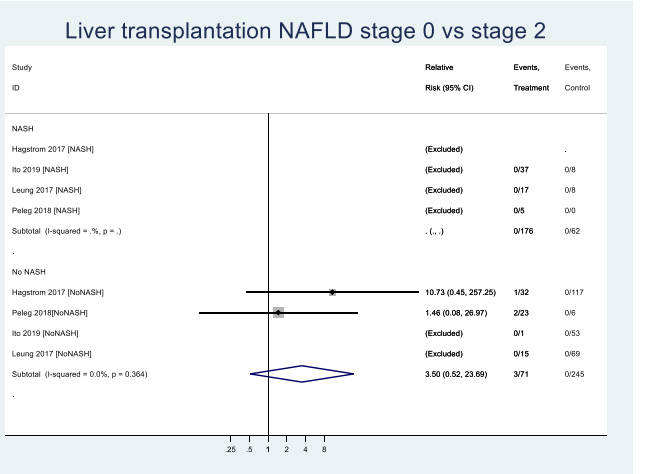


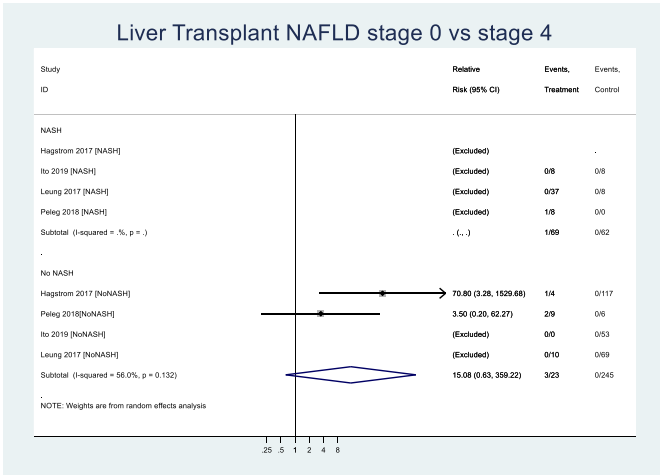




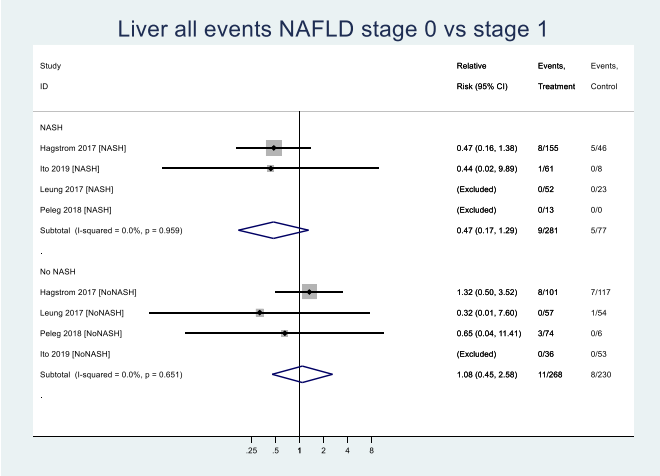
Supplementary Figure 1. (*continued*).

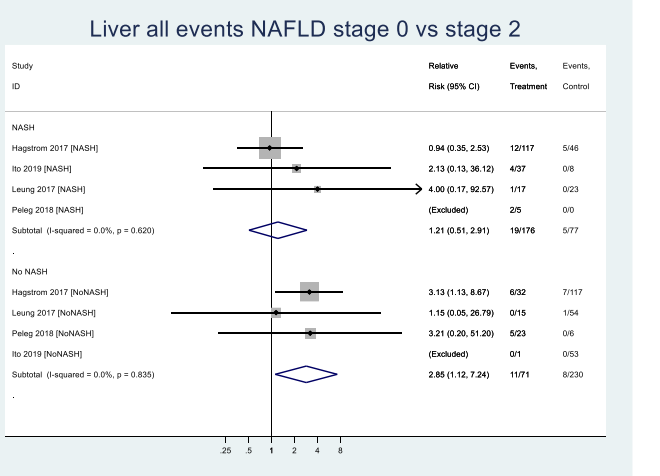


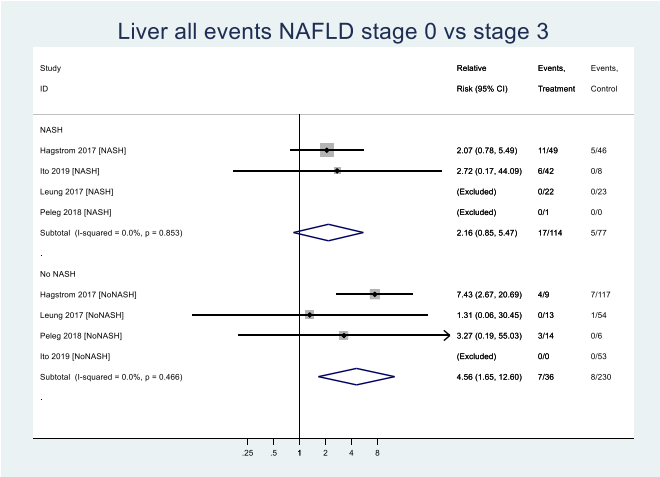


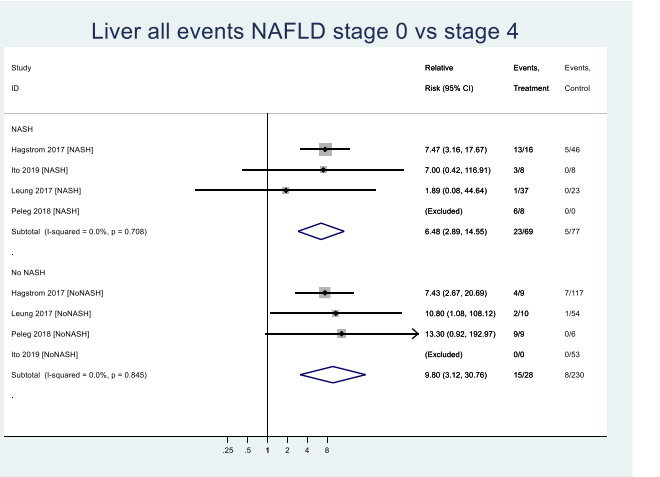


Supplementary Figure 1. (*continued*).

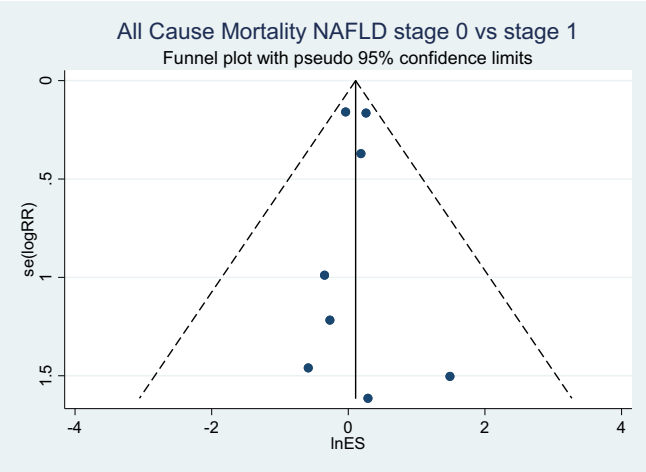




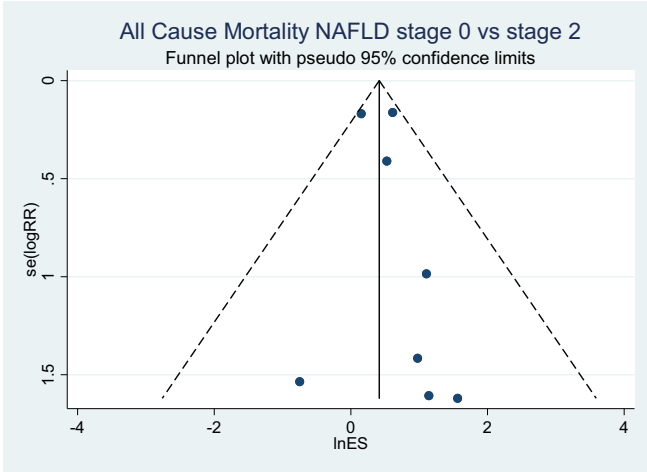




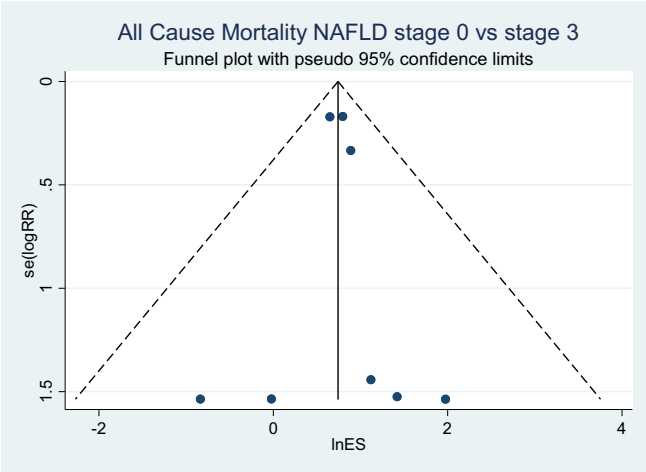
Supplementary Figure 1. (*continued*).



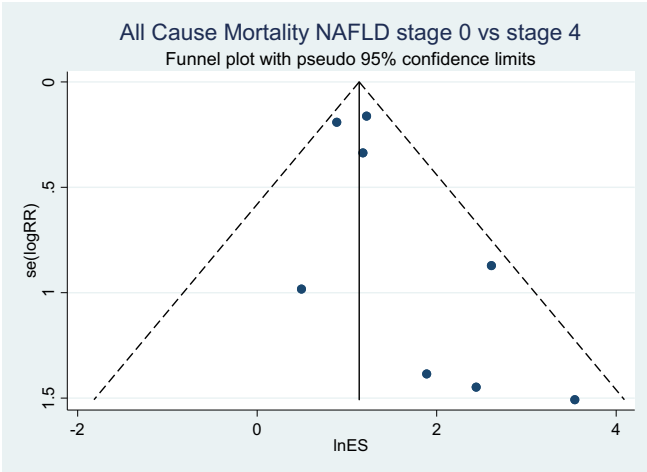
Egger test *P* value = .996.



Egger test *P* value = .485.

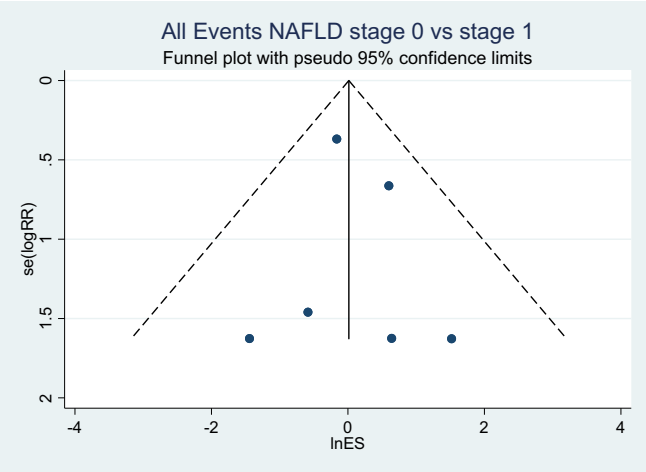


Egger test *P* value = .89.

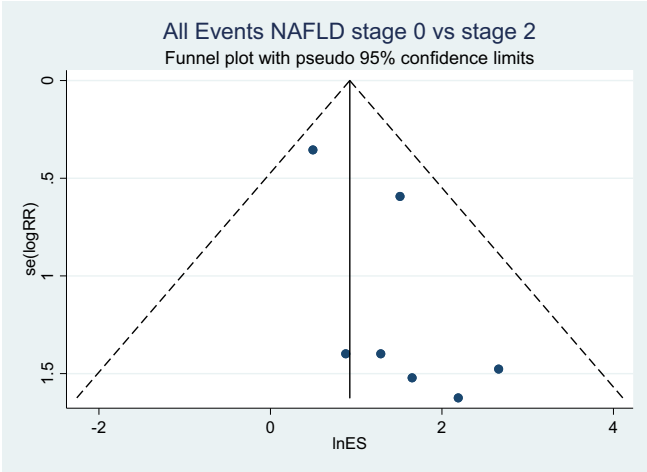


Egger test *P* value = .11.

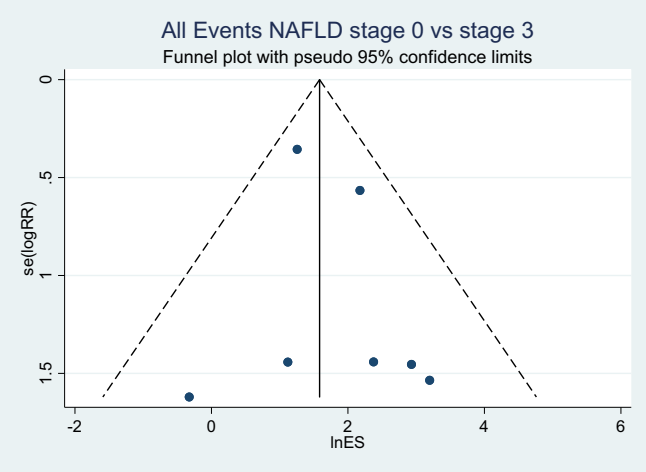
Supplementary Figure 2. Assessment of Small Study Bias



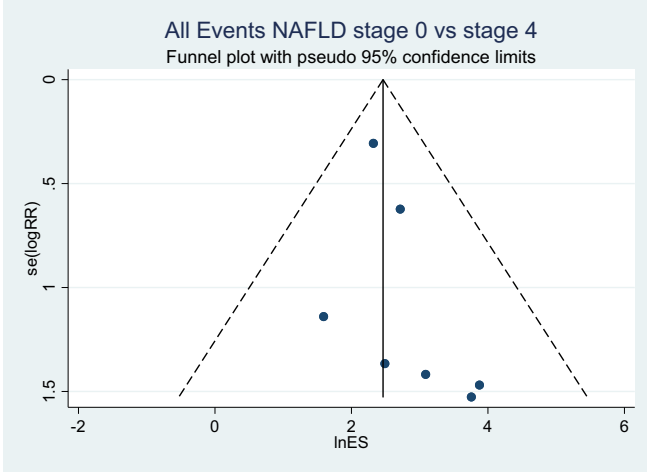
Egger test *P* value = .75.



Egger test *P* value = .05



Egger test *P* value = .49



Egger test *P* value = .20.

Supplementary Figure 2. (*continued*).

Supplementary Table 1. Search Strategy

Database: Ovid MEDLINE: 1946 to week 3 of October 2018	
1	(NAFLD or NASH).mp. or non-alcoholic fatty liver.ti,ab.
2	non-alcoholic steatohepatitis.ti,ab.
3	Non-alcoholic Fatty Liver Disease/
4	exp Fatty Liver/
5	or/1-4
6	fibrosis.ti,ab.
7	fibrosis/
8	cirrhosis or cirrhoses.ti,ab.
9	or/6-8
10	surrogate$.ti,ab.
11	variceal bleed$.ti,ab.
12	decompensat$.ti,ab.
13	(scar$ adj2 liver$).ti,ab.
14	ascites.ti,ab.
15	outcome$.ti,ab.
16	disease progress$.ti,ab,
17	(patient adj3 outcome$) or PROM$.ti,ab
18	((liver) adj2 (cancer or transplant$ or carcinoma$ or failure)).ti,ab
19	death$.mp. or mortality.ti,ab
20	hepatocellular cancer.ti,ab.
21	hepatic encephalopathy.ti,ab.
22	hepatoencephalopathy.ti,ab.
23	exp liver neoplasms/
24	or/10-23
25	5 and 9
26	24 and 25
27	(pre-clinical or rat or rats or mouse or mice or animal) or animals.ti,ab
28	26 not 27

Supplementary Table 2. Quality In Prognosis Studies Tool

Potential bias (highlight one)	Items considered for assessment of potential opportunity for bias	Yes response	No response	Evidence for response (include extracts from paper to illustrate judgements)	Comments
Study Population The study sample represents the population on key characteristics sufficient to limit potential bias to the observed relationship between fibrosis stage and all-cause mortality, liver-related morbidity, and HRQoL. Yes Partly No	The source population of interest is adequately described for key characteristics and the study setting supports the applicability of results. Eligibility criteria and recruitment are adequately described and the inclusion/exclusion criteria applied uniformly to all screened for eligibility. There is adequate participation in the study by eligible individuals and sufficient information was given about those who did not participate. The baseline characteristics or participants included is adequately described for characteristics and representative of the population of interest.	Adult (>18 yrs) participants with confirmed (ultrasonography or biopsy-proven) NASH or NAFLD AND source of patients and their characteristics reported including gender, age, presence of type-2 diabetes, hypertension, statin usage, and smoking status AND no major exclusions AND representative population (e.g. all consecutive patients recruited, or random sample of patients selected)	Non-adults included, patient characteristics not adequately described, unconfirmed disease, major exclusions involved and recruitment method not reported or likely to be unrepresentative of patient population.		
Study Attrition Loss to follow-up (from sample to study population) is not associated with key characteristics (i.e. the study data represent the sample), sufficient to limit potential bias Yes Partly No	Attempts to collect information on participants who dropped out of the study are described. Reasons for loss to follow-up are provided. There are no important differences between key characteristics (e.g. ethnicity, underlying condition, age, treatment method) and outcomes in participants who completed the study and those who did not.	Reasons lost to follow-up reported with numbers AND comparison of lost versus not lost to follow-up with no important differences, or if important differences found addressed in the analysis	Attrition/ denominators not reported/ accounted for		
Prognostic factor measurement Liver fibrosis is adequately measured in study participants to sufficiently limit bias Yes Partly No	Clear definition of fibrosis staging given, report that biopsy confirmed and method/setting of assessment same across participants, and fibrosis stages (0 to 4) provided with number of participants in each category reported. Adequate proportion of the study sample has complete data on fibrosis stage. Appropriate methods are used if imputation is used for missing prognostic factor data.	Data collection is prospective AND fibrosis biopsy confirmed AND number of participants in each fibrosis stage reported.	Definition of fibrosis stage not clear or sufficiently detailed.		

Supplementary Table 2. Continued

Potential bias (highlight one)	Items considered for assessment of potential opportunity for bias	Yes response	No response	Evidence for response (include extracts from paper to illustrate judgements)	Comments
Outcome measurement Outcomes (mortality, liver-related events, and health-related quality of life) are adequately measured in study participants to sufficiently limit potential bias Yes Partly No	Clear definition of outcome (mortality, liver-related events, and health-related quality of life) measurement provided, including duration of follow-up. Outcomes measured prior to outcome occurring.	Number of mortality, liver-related events and health-related quality of life recorded AND time-frame for outcomes reported AND data collection is prospective	outcome and time-frame for follow-up not reported		
Confounding measurement and account Important potential confounders are appropriately accounted for, limiting potential bias with respect to fibrosis stage Yes Partly No	Important confounders are accounted for in the study design (i.e. gender, age, and type 2 diabetes status) and analysis. Measurement of all important confounders is adequately valid and reliable. The method and setting of confounding measurement are the same for all participants. Appropriate imputation method is used for missing confounder data. Appropriate adjustment used and clearly outlined. Interventions do not confound outcomes (mortality, liver-related events, and health-related quality of life).	Confounders (i.e. gender, age, presence of type-2 diabetes, statin usage, and smoking status) are reported AND analysis of impact of fibrosis stage on outcomes adjusts for confounders and clearly described.	Confounders (i.e. gender, age, and type 2 diabetes status) not reported and no statement of adjustment for confounders in data analysis or not clear.		
Analysis and reporting The statistical analysis is appropriate for the design of the study, limiting potential for presentation of invalid results Yes Partly No	There is sufficient presentation of data to assess the adequacy of the analysis. The selected statistical method of analysis is adequate for the design of the study (e.g. time to event analysis for mortality and liver related events). There is no selective reporting of results.	Statistical model used appropriate for the study design and type of data AND strategy and results clearly reported AND completeness of reporting of results	Unclear reporting of results, inappropriate statistical model, and selective reporting of results		

Supplementary Table 3. Excluded Studies

1. Dam-Larsen S, Becker U, Franzmann MB, et al. Final results of a long-term, clinical follow-up in fatty liver patients. Scand J Gastroenterol 2009;44:1236–1243. **No fibrosis by outcome analysis**
2. Dulai PS, Singh S, Patel J, et al. Increased risk of mortality by fibrosis stage in nonalcoholic fatty liver disease: systematic review and meta-analysis. Hepatology 2017;65:1557–1565. **Systematic review**
3. Ekstedt M, Hagström H, Nasr P, et al. Fibrosis stage is the strongest predictor for disease-specific mortality in NAFLD after up to 33 years of follow-up. Hepatology 2015;61:1547–1554. **Subset of Hagström 2017.**[20,21]
4. Golabi P, Stepanova M, Pham HT, et al. Non-alcoholic steatofibrosis (NASF) can independently predict mortality in patients with non-alcoholic fatty liver disease (NAFLD). BMJ Open Gastroenterol 2018;5(1):e000198. **No association between fibrosis and mortality**
5. Hagström H, Nasr P, Ekstedt M, et al. SAF score and mortality in NAFLD after up to 41 years of follow-up. Scand J Gastroenterol 2017;52:87–91. **Subset of Hagström 2017.**[20,21]
6. Hashimoto E, Yatsuji S, Kaneda H, et al. The characteristics and natural history of Japanese patients with nonalcoholic fatty liver disease. Hepatol Res 2005;33:72–76. **No fibrosis by outcome analysis**
7. Huber Y, Labenz C, Ganter M et al. Health-related quality of life in patients with non-alcoholic fatty liver disease. J Hepatol 2017;66(Suppl 1):S597–S598. **Abstract only and noninvasive fibrosis**
8. Huber Y et al. Health-related quality of life correlates with histological severity in non-alcoholic fatty liver disease. J Hepatol 2018;68(Suppl 1):S831–S832. **Abstract and full paper included**
9. Ito T et al. Utility and limitation of non-invasive fibrosis markers for predicting the prognosis in biopsy-proven Japanese NAFLD patients. J Hepatol 2018;68(Suppl 1):S561. **Abstract and full paper included**
10. Jaruvongvanich V, Wijarnpreecha K, Ungprasert P. The utility of NAFLD fibrosis score for prediction of mortality among patients with nonalcoholic fatty liver disease: a systematic review and meta-analysis of cohort study. Clin Res Hepatol Gastroenterol 2017;41:629–634. **Systematic review**
11. Kim D, Kim WR, Kim HJ, Therneau TM. Association between noninvasive fibrosis markers and mortality among adults with nonalcoholic fatty liver disease in the United States. Hepatology 2013;57:1357–1365. **NFD score, not liver biopsy**
12. Le MH, Devaki P, Ha NB, et al. Prevalence of non-alcoholic fatty liver disease and risk factors for advanced fibrosis and mortality in the United States. PLoS One 2017;12(3):e0173499. **Noninvasive fibrosis only**
13. Lee T-Y, Wu J-C, Yu S-H, et al. The occurrence of hepatocellular carcinoma in different risk stratifications of clinically noncirrhotic nonalcoholic fatty liver disease. Int J Cancer 2017;141:1307–1314. **No inclusion of fibrosis**
14. Matteoni CA, Younossi ZM, Gramlich T, et al. Nonalcoholic fatty liver disease: a spectrum of clinical and pathological severity. Gastroenterology 1999;116:1413–1419. **No biopsy fibrosis outcome analysis**
15. Miyake N, Tada T, Kobayashi N et al. Progression of liver fibrosis is associated with non-liver-related mortality in patients with nonalcoholic fatty liver disease. Hepatology 2018;68(Suppl 1):974A. **Abstract and noninvasive fibrosis score—forward search no paper**
16. Patel JR, Dulai PS , Younossi ZM et al. Risk of mortality by fibrosis stage in NAFLD: a systematic review and meta-analysis. Hepatology 2017;64:1095A. **Abstract and systematic review**
17. Renelus BD, Fengxia Yan F, Flood MC et al. Comparison of noninvasive fibrosis scores and association with mortality in adults with moderate to severe hepatic steatosis and NAFLD. Hepatology 2015:603A. **Abstract and noninvasive fibrosis only**
18. Sayiner M, Stepanova M, Pham H, et al. Assessment of health utilities and quality of life in patients with non-alcoholic fatty liver disease. BMJ Open Gastroenterol 2016;3(1):e000106. **Cross-sectional study**
19. Salomone F, Micek A, Godos J. Simple scores of fibrosis and mortality in patients with NAFLD: a systematic review with meta-analysis. J Clin Med 2018;7(8):219. **Systematic review**
20. Sanyal A, Harrison S, Ratziu V et al. Changes in fibrosis, but not the NAFLD activity score (NAS), are associated with disease progression in patients with nonalcoholic steatohepatitis (NASH) and advanced fibrosis. J Hepatol 2017;(Suppl):S2–S3. **Abstract and no fibrosis-outcome analysis**
21. Sebastiani G, Deschenes M, Alshaalan R, et al. Prediction of 10-year clinical outcomes in NASH by non-invasive fibrosis and steatosis tools, hepatic venous pressure gradient (HVPG) and liver histology. Hepatology 2014;(1):597A. **Abstract— full paper included**
22. Singh A et al. Validity of non-invasive fibrosis scores to detect advanced fibrosis in patients with type 2 diabetes with suspected non-alcoholic fatty liver disease. Am J Gastroenterol 2017;112(Suppl 1):S491–S492. **Abstract and noninvasive fibrosis**
23. Stauber RE et al. Enhanced liver fibrosis (ELF) score accurately detects advanced fibrosis in nonalcoholic fatty liver disease (NAFLD). J Hepatology 2018;68(Suppl 1):S563. **Abstract—noninvasive fibrosis**
24. Stepanova M, Rafiq N, Makhlouf H, et al. Predictors of all-cause mortality and liver-related mortality in patients with non-alcoholic fatty liver disease (NAFLD). Dig Dis Sci 2013;58:3017–3023. **No fibrosis by outcome analysis**
25. Stepanova M, Rafiq N, Makhlouf H, et al. Pathologic features of non-alcoholic steatohepatitis (NASH) as independent predictors of liver-related mortality. J Hepatol 2011;54:S25–S44. **No fibrosis by outcome analysis**
26. Strasser M, Feldman A, Schranz M, et al. SAF score effectively identifies NAFLD subjects at high risk of subsequent liver related but not cardiovascular or malignancy-associated mortality and morbidity. J Hepatol 2017;66(1 Suppl):S425–S426. **No fibrosis by outcome analysis**
27. Sun W, Cui H, Li N, et al. Comparison of FIB-4 index, NAFLD fibrosis score and BARD score for prediction of advanced fibrosis in adult patients with non-alcoholic fatty liver disease: a meta-analysis study. Hepatol Res 2016;46:862–870. **Systematic review**
28. Tada T, Kumada T, Toyoda H, et al. Progression of liver fibrosis is associated with non-liver-related mortality in patients with nonalcoholic fatty liver disease. Hepatol Commun 2017;1:899–910. **Noninvasive fibrosis vs nonhepatic mortality**
29. Treeprasertsuk S, Björnsson E, Enders F, et al. NAFLD fibrosis score: a prognostic predictor for mortality and liver complications among NAFLD patients. World J Gastroenterol 2013;19:1219–1229. **Noninvasive fibrosis score**
30. Unalp-Arida A, Ruhl CE. Liver fibrosis scores predict liver disease mortality in the United States population. Hepatology 2017;66:84–95. **Noninvasive fibrosis only**

Supplementary Table 3. Continued

31. Wijarnpreecha K, Scribani C, Ungprasert P et al. Non-invasive fibrosis markers are independent predictors of mortality among U.S. adults with nonalcoholic fatty liver disease. J Hepatol 2017;66(Suppl 1):S662–S663. **Noninvasive fibrosis only**
32. Xun Y-H, Guo J-C, Lou G-Q, et al. Non-alcoholic fatty liver disease (NAFLD) fibrosis score predicts 6.6-year overall mortality of Chinese patients with NAFLD. Clin Exper Pharmacol Physiol 2014;41:643–649. **Noninvasive fibrosis only**
33. Yoshihisa A, Sato Y, Yokokawa T, et al. Liver fibrosis score predicts mortality in heart failure patients with preserved ejection fraction. ESC Heart Fail 2018;5:262–270. **Noninvasive fibrosis only**
34. Younossi ZM et al. Non-alcoholic fatty liver disease fibrosis score (NFS) is an independent predictor of mortality in patients with sNAFLD. Hepatology 2013;(1):511A. **Noninvasive fibrosis only**
35. Younossi ZM, Otgonsuren M, Henry L, et al. Association of nonalcoholic fatty liver disease (NAFLD) with hepatocellular carcinoma (HCC) in the United States from 2004 to 2009. Hepatology 2015;62:1723–1730. **No fibrosis by outcome analysis**
36. Younossi ZM, Stepanova M, Henry L, et al. A disease-specific quality of life instrument for non-alcoholic fatty liver disease and non-alcoholic steatohepatitis: CLDQ-NAFLD. Liver Int 2017;37:1209–1218. **Cross-sectional study**
37. Younossi ZM, Stepanova M, Younossi I, Racila A. Validation of chronic liver disease questionnaire for nonalcoholic steatohepatitis in patients with biopsy-proven nonalcoholic steatohepatitis. Clin Gastroenterol Hepatol 2019;17:2093–2100. **No fibrosis by outcome analysis**

Supplementary Table 4. Meta-analysis: Pooled Hazard Ratio (Adjusted) and Pooled RR (Unadjusted) by Fibrosis Stage 0–2 vs 2–4 for All Patients With NAFLD

Number of Studies	Stage 0–2 vs 3–4 Adjusted hazard ratio (95% CI), I^2 statistic	Stage 0–2 vs 3–4 Unadjusted relative risk (95% CI), I^2 statistic
All-cause mortality (n = 5)	2.24 (1.48–3.39), 31%	2.25 (1.85–2.72), 35%
Liver-related mortality (n =4)	5.12 (2.48–10.55), 0%	6.42 (3.45–11.95), 0%
Liver transplant (n = 2)	10.89 (2.01–58.99), 0%	3.40 (0.96–12.00), 0%
All liver events (n = 6)	5.58 (3.70–8.40), 0%	6.31, (4.60–8.65), 0%

NOTE. All results are fixed-effect meta-analysis.

Supplementary Table 5. HRQoL by Fibrosis Stage Across Individual Studies

Study (Year)	Stage 0	Stage 1	Stage 2	Stage 3	Stage 4
David (2009)[19]	N, median (IQR)	N, median (IQR)	N, median (IQR)	N, median (IQR)	N, median (IQR)
SF-36 PCS	167, 50 (42.5–56)	211, 50 (39–54)	138, 47 (36–54)	131, 48 (37–53)	66, 37 (31–48)
SF-36 MCS	NR	NR	NR	NR	NR
Huber (2018)[22]	N, mean (SD)	N, mean (SD)	N, mean (SD)	N, mean (SD)	N, mean (SD)
CLDQ total	36, 4.76 (NR)	74, 5.23 (NR)	67, 5.10 (NR)	82, 4.90 (NR)	N, mean (SD)
Younossi (2018)[31]	N, mean (SD)	N, mean (SD)	N, mean (SD)	N, mean (SD)	N, mean (SD)
SF-36 PCS	—	—	25, 45.0 (8.7)	47, 43.4 (10.3)	—
SF-36 MCS	—	—	25, 51.0 (9.6)	47, 50.6 (12.7)	—
CLDQ total	—	—	25, 4.83 (1.10)	47, 4.91 (1.25)	—

MCS, Mental Component Score; NR, not reported; PCS, Physical Component Score; SD, standard deviation.

Supplementary Table 6. Univariate Meta-regression Analysis, *P* Values

Study level covariate	Fibrosis stage 0 vs 1	Fibrosis stage 0 vs 2	Fibrosis stage 0 vs 3	Fibrosis stage 0 vs 4
All-cause mortality				
Retrospective vs prospective study design	.774	.47	.48	.20
High vs moderate risk of bias	.92	0.67	.67	.46
Duration of follow up, *y*	.36	.11	.46	.12
All-liver related events				
Retrospective vs prospective study design	.41	.97	.28	.47
High vs moderate risk of bias	.670	.80	.73	.98
Duration of follow-up, *y*	.61	.13	.28	.51

Supplementary Table 7. GRADE Assessment of Quality of Evidence for Liver Fibrosis as a Prognostic Marker for NAFLD

	Study design	Risk of bias	Indirectness	Imprecision	Additional considerations[a]	Quality
All-cause mortality	Observational	Serious[b]	Not serious[c]	Not serious[d]	Large effect, no publication bias, confounder adjusted	High
Liver-related mortality	Observational	Serious[b]	Not serious[c]	Serious[e]	Large effect, no publication bias, confounder adjusted	Moderate
Liver transplantation	Observational	Serious[b]	Not serious[c]	Serious[e]	Large effect, no publication bias, confounder adjusted	Low
All liver events	Observational	Serious[b]	Serious[f]	Not serious[d]	Large effect, no publication bias, confounder adjusted	Low
HRQoL	Observational	Serious[b]	Not serious[c]	Serious[g]	Large effect, no publication bias, confounder adjusted	Low

NOTE. From Iorio et al[13]: High indicates very confident that the true prognosis (probability of future events) lies close to that of the estimate. Moderate indicates moderately confident that the true prognosis (probability of future events) is likely to be close to the estimate, but there is a possibility that it is substantially different. Low indicates that confidence in the estimate is limited: the true prognosis (probability of future events)

[a]Positive considerations can allow upgrading of GRADE rating.
[b]Risk of bias for individual studies judged to be moderate or high (see Table 2).
[c]Appropriate population and outcomes.
[d]Sufficient number of events.
[e]Insufficient number of events (especially for stage 0 fibrosis).
[f]Inconsistent definition of composite outcome of all liver events across studies.
[g]Small number of studies.