Journal Pre-proof

Association Between Fibrosis Stage and Outcomes of Patients with Non-Alcoholic Fatty Liver Disease: a Systematic Review and Meta-Analysis

Rod S. Taylor, Rebecca J. Taylor, Sue Bayliss, Hanes Hagstrom, Patrik Nasr, Jorn M. Schattenberg, Masatoshi Ishigami, Hidenori Toyoda, Vincent Wai-Sun Wong, Noam Peleg, Amir Shlomai, Giada Sebastiani, Yuya Seko, Neeraj Bhala, Zobair M. Younossi, Quentin M. Anstee, Stuart McPherson, Philip N. Newsome, Mattias Ekstedt, Per Stål, Rolf Hultcrantz, Stergios Kechagias

PII: S0016-5085(20)30137-2
DOI: https://doi.org/10.1053/j.gastro.2020.01.043
Reference: YGAST 63185

To appear in: *Gastroenterology*
Accepted Date: 22 January 2020

Please cite this article as: Taylor RS, Taylor RJ, Bayliss S, Hagstrom H, Nasr P, Schattenberg JM, Ishigami M, Toyoda H, Wai-Sun Wong V, Peleg N, Shlomai A, Sebastiani G, Seko Y, Bhala N, Younossi ZM, Anstee QM, McPherson S, Newsome PN, Association Between Fibrosis Stage and Outcomes of Patients with Non-Alcoholic Fatty Liver Disease: a Systematic Review and Meta-Analysis, *Gastroenterology* (2020), doi: https://doi.org/10.1053/j.gastro.2020.01.043.

Association Between Fibrosis Stage and Outcomes of Patients with Non-Alcoholic Fatty Liver Disease: a Systematic Review and Meta-Analysis

Short title: Stage of liver fibrosis in NAFLD and outcome.

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

[1] Institute of Health and Well Being, University of Glasgow, United Kingdom; rod.taylor@glasgow.ac.uk

[2] R[2] Consultancy, Glasgow, United Kingdom; rebeccatayloruk@hotmail.com

[3] Institute of Applied Health Research, University of Birmingham, United Kingdom; s.bayliss@bham.ac.uk

[4] Unit of Hepatology, Department of Upper GI Diseases, Karolinska University Hospital, Stockholm, Sweden; hannes.hagstrom@ki.se

[5] Department of Gastroenterology and Hepatology, Department of Medical and Health Sciences, Linköping University, Linköping, Sweden; patrik.nasr@liu.se

[6] University Medical Centre of the Johannes Gutenberg-University, Mainz, Germany; joern.schattenberg@unimedizin-mainz.de

[7] Department of Gastroenterology and Hepatology, Nagoya University Graduate School of Medicine, Nagoya, Japan; masaishi@med.nagoya-u.ac.jp

[8] Department of Gastroenterology and Hepatology, Ogaki Municipal Hospital, Ogaki, Japan; hmtoyoda@spice.ocn.ne.jp

[9] Department of Medicine and Therapeutics, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong; wongv@cuhk.edu.hk

[10] Department of Gastroenterology and Hepatology, Rabin Medical Center, Beilinson hospital, Petach-Tikva; Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; noampe1@clalit.org.il

[11] Department of Medicine D, Rabin Medical Center, Beilinson hospital, Petach-Tikva; Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; amirsh9@clalit.org.il

[12] Department of Medicine, McGill University Health Centre, Montreal, Quebec, Canada; giada.sebastiani@mcgill.ca
[13] Department of Molecular Gastroenterology and Hepatology, Kyoto Prefectural University of Medicine, Kyoto, Japan; yuyaseko@koto.kpu-m.ac.jp
[14] Institute of Applied Health Research, University Hospitals Birmingham NHS Foundation Trust and the University of Birmingham, United Kingdom; neeraj.bhala@uhb.nhs.uk
[16] Department of Medicine, Inova Fairfax Hospital, Falls Church, Virginia, United States; zobair.younossi@inova.org
[16] Institute of Clinical & Translational Research, Faculty of Medical Sciences, Newcastle University, Newcastle-upon-Tyne, United Kingdom; Newcastle NIHR Biomedical Research Centre & Liver Transplant Unit, The Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle-upon-Tyne, United Kingdom; quentin.anstee@newcastle.ac.uk
[17] Liver Transplant Unit, The Newcastle upon Tyne Hospitals NHS Foundation Trust; Institute of Clinical and Translational Research, Newcastle University, & Newcastle NIHR Biomedical Research Centre, Newcastle-upon-Tyne, United Kingdom; stuart.mcpherson@nuth.nhs.uk
[18] National Institute for Health Research Biomedical Research Centre at University Hospitals Birmingham NHS Foundation Trust and the University of Birmingham, UK. Centre for Liver and Gastrointestinal Research, Institute of Immunology and Immunotherapy, University of Birmingham, UK. Liver Unit, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK; p.n.newsome@bham.ac.uk

Collaborators

Mattias Ekstedt (mattias.ekstedt@liu.se), Per Stål (per.stal@ki.se), Rolf Hultcrantz (rolf.hultcrantz@ki.se), Stergios Kechagias (stergios.kechagias@liu.se)

Address for correspondence: Professor Rod Taylor MSc, PhD. Chair of Population Health Research, MRC/CSO Social and Public Health Sciences Unit, University of Glasgow, Top floor, 200 Renfield Street, Glasgow, G2 3AX; Mobile: + 44 (0)7968 152537; Email: rod.taylor@glasgow.ac.uk

Word count: 6,240 (including abstract, references and figures/table legends)

Grant support: Gilead funded this project. PNN was supported by the National Institute of Health Research (NIHR) Birmingham Biomedical Research Centre (BRC). QMA and SMcP are Newcastle NIHR Biomedical Research Centre investigators. GS is supported by a Junior 1 and 2 Salary Award from Fonds de Recherche du Québec - Santé (FRQS) (#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.

Conflict of interest: RST, RJT, SB, SMcP, and PNN all received funding from Gilead for their contribution to this project. Other authors received no funding for this project. Other research funding. HH: Consulting fees from Novo Nordisk, IQVIA, Gilead Inc. Research grants from Gilead Inc., Astra Zeneca, Intercept. Advisory Board Bristol-Myers Squibb; QMA: Abbvie, Allergan/Tobira, AstraZeneca, GlaxoSmithKline, Glympse Bio, Novartis Pharma AG, Pfizer Ltd., Vertex. Performs Consultancy for: Abbott Laboratories, Acuitas Medical, Allergan/Tobira, Blade, BNN Cardio, Cirius, CymaBay, EcoR1, E3Bio, Eli Lilly & Company Ltd., Galmed, Genfit SA, Gilead, Grunthal, HistoIndex, Indalo, Imperial Innovations, Intercept Pharma Europe Ltd., Inventiva, IQVIA, Janssen, Kenes, Madrigal, MedImmune, Metacrine, NewGene, NGMBio, North Sea Therapeutics, Novartis, Novo Nordisk A/S, Pfizer Ltd., Poxel, ProSciento, Raptor Pharma, Servier, Viking Therapeutics. Speaker for: Abbott Laboratories, Allergan/Tobira, BMS, Clinical Care Options, Falk, Fishawack, Genfit SA, Gilead, Integritas Communications, MedScape. Received Royalties from: Elsevier Ltd.; JMS received Consultancy: AbbVie, BMS, BBN Cardio, Boehringer Ingelheim, Galamed, Genfit, Gilead Sciences, Intercept Pharmaceuticals, IQVIA, MedImmune, Novartis, Pfizer. Research Funding: Gilead Sciences, Yakult Europe B.V. Lectures: Falk Foundation, Takeda, MSD; VWS 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; and a speaker for Bristol-Myers Squibb, Echosens, Gilead Sciences and Merck; he has also received an unrestricted grant from Gilead Sciences for fatty liver research; GS has acted as speaker for Merck, Gilead, Abbvie, ViiV, served as an advisory board member for Merck, Gilead and Novartis has

received research funding from Merck and Echosens. ZMY has receive research funding or is a consultant for Gilead Sciences, Intercept, BMS, Novartis, Novo Nordisk, Vicking, Terns and Siemens; SMcP 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 manufacturer.

Contributions: RST and SMcP conceived the idea for the paper. RST, RJT, SMcP, PNN and QMA contributed text to the first draft and revision of the paper. RST led the data analysis. HH, PN, JS, MI, HT, VW, NP, AS, GS, YS, NB, and ZMY contributed data and revised the manuscript. All authors contributed summary outcome data to the project and provided review and comment on a draft of the paper. All the authors approved the final version of the paper. RST is guarantor.

Abbreviations:

BMI: body mass index
CI: confidence interval
CLDQ: Chronic Liver Disease Questionnaire
CRN: NASH Clinical Research Network
EMA: European Medicines Agency
FDA: Food and Drug Administration
FLIP: fatty liver inhibition of progression
HTA: Health Technology Assessment
IQR: interquartile range
HRQoL: health-related quality of life
MCS: Mental Component Sore;
NAFLD: non-alcoholic fatty liver disease
NASH: non-alcoholic steatohepatitis
NICE: National Institute for Health and Care Excellence
NR: not reported
QUIPS: Quality in Prognosis Studies
PCS: Physical Component Score
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomised controlled trial
RR: relative risk

SD: standard deviation

SF-36: Short-Form 36

Abstract:

Background & Aims: Biopsy-confirmed liver fibrosis is a prognostic factor for patients with non-alcoholic 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 non-alcoholic 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% CI values for mortality and morbidities. We performed fixed-effect and random-effect model meta-analyses. Meta-regression 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 to 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 transplantation 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 following 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.

KEY WORDS: biomarker, disease progression, prognosis, liver disease

Introduction

Non-alcoholic 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 greater than 5% of hepatocytes, and has a spectrum of histological features ranging from steatosis without fibrosis to non-alcoholic steatohepatitis (NASH) with varying stages of fibrosis.[2] The Fatty Liver Inhibition of Progression Steatosis-Activity-Fibrosis (SAF) criteria and the NASH Clinical Research Network (CRN) NAFLD Activity Score are the most widely adopted semiquantitative scores used for assessing histological disease activity.[3] To sustain a diagnosis of NASH, both require histological 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 miscategorised 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 in with paired patient liver biopsies 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 (EMA) and payers typically only accept surrogate endpoints 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 (e.g. a reduction in biopsy-confirmed fibrosis stage) and a relevant clinical outcome (e.g. reduced liver-related mortality) in the setting of a single (or multiple) randomised controlled trial (RCT).[6,7]

A systematic review and meta-analysis including five observational cohort studies (1,495 NAFLD patients) assessed liver fibrosis as a prognostic marker of mortality.[8] The authors 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 considered the outcome of mortality; (3) the comparison between fibrosis stage and death did not account for the potential confounding by factors such as age, gender, and statin usage; (4) did not include analyses of the impact of liver fibrosis in the subgroup of NAFLD patients 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 (up to 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: (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 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 PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement.[11] The review was registered with PROSPERO - International prospective register of systematic reviews (CRD42019121054).

Identification of studies and searches

A detailed search strategy used both Indexing languages (e.g. MeSH 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 (e-Appendix 1). The following electronic databases were searched up to August 2018 by an experienced information specialist (SB): MEDLINE (Ovid), EMBASE (Ovid), Cochrane Library (Wiley), as were the trial registers - ClinicalTrials.gov, WHO ICTRP including ISRCTN and EU Clinical Trials Register. The search results were combined into an Endnote v.9 database to facilitate reference management. The reference lists of 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

In order to fully data extract and quality assess studies, we excluded studies only available as abstracts (authors were contacted to clarify the availability of full publication). We restricted inclusion to English language papers. We excluded studies reporting non-invasive 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 included studies: study design, participants' characteristics (i.e. number of patients with NAFLD, NASH and by fibrosis stage, and 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 through use of the QUIPS (QUality In Prognosis Studies) tool.[12] This prognostic risk of bias tool was adapted to suit the requirements of this review (e-Appendix 2). The tool has six domains:

1. Study participation
2. Study attrition
3. Prognostic factor measurement
4. Outcome measurement
5. Study confounding (research team clinicians (PNN, SMc) determined the key confounders: age, gender, 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 one or more domains are scored as partly or one domain is scored as 'no';
- High risk of bias: describes studies for which more than one 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 GRADE (Grading of Recommendations, Assessment, Development and Evaluations) approach applied to prognostic studies.[13]

Statistical analysis

Data were analysed 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), morbidity (cirrhosis, liver cancer, and all liver-related events) by stage of fibrosis for all patients with NAFLD. In addition, the number of events were also extracted separately in two groups of NAFLD patients: (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 – a RR > 1.00 indicated an increased risk of outcome with increasing fibrosis stage.

Whilst 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. We, therefore, also sought to identify the hazard ratios (HRs) (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, authors 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^2 test of heterogeneity and the Cochrane I^2 statistic cut-offs, i.e. 0% to 40%: heterogeneity might not be important; 30% to 60%: may represent moderate heterogeneity; 50% to 90%: may represent substantial heterogeneity; and 75% to 100%: considerable heterogeneity.[14] When pooling the results across studies, we employed a random-effects meta-analysis model where there was formal evidence of statistical heterogeneity (i.e. Chi^2 test p-value<0.10 and substantial heterogeneity as defined by an I^2 statistic ≥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 (≥7 studies)[14] small-study effects and publication bias were assessed using the funnel plot asymmetry and the Egger test.[15]

Meta-regression was used to examine the association between the pre-defined 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 was contributing data from ≥5 studies.[14] If there were suitable data (i.e. RCTs reporting change in fibrosis stage and outcomes of interest: mortality, liver-related morbidity and HRQoL) we planned to calculate and report two key indicators of surrogate endpoint validation.[16] Firstly, 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). Secondly, from the trial-based analysis, we would estimate the surrogate threshold effect, i.e. 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, Texas) software.

Results

Study selection

After de-duplication, our database searches identified a total of 6,083 titles/abstracts. A further 210 study titles were identified from trial registers. Following 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 summarised in the flow diagram in Figure 1. Citations and reasons for studies excluded on review of full publication are listed in e-Appendix 3.

Study and patient characteristics

The included 13 studies recruited a total of 4,428 patients with biopsy confirmed NAFLD and subgroup of patients 2,875 (65%) had a histological proven diagnosis of NASH. Trial and study characteristics are presented in Table 1. Twelve were observational cohort studies (seven retrospective, five prospective) and one was a RCT. The median average age across studies was 51.0 years and 51% were male. 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 (BMI) of 31.3 kg/m^2). Fibrosis staging was confirmed by liver biopsy and centrally assessed in the majority of multi-centre studies. The distribution of NAFLD patients by fibrosis stage was: stage 0 - 1,040 (23%); stage 1 - 1,094 (25%); stage 2 - 602 (14%); stage 3 - 922 (21%); stage 4 - 770 (17%). Bhala et al (2011)[18] and Vilar-Gomez (2018)[29] included only patients with stage 3 and 4 and were therefore not includable in meta-analyses.

The method of NASH diagnosis was poorly described but was judged to be adequately defined in seven studies.[20,22,23,24,25,28,30] The two most common diagnostic metrics were fatty liver inhibition (FLIP) criteria or NASH Clinical Research Network (CRN) (i.e., presence of steatosis, ballooning and lobular inflammation). The median average duration of study follow up was 6.2 years, ranging from seven 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 (2017)[24] which was deemed to be at high risk of bias and Vilar-Gomez (2018)[29] judged to be at low risk of bias (see Table 2). The QUIPS 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 (2011)[18] and Vilar-Gomez (2018)[29] provided a sufficiently detailed description of loss to follow-up to assess risk of attrition, while the studies of Leung et al (2017)[24] and Younossi et al (2018)[30] provided no information on loss to follow-up. Angulo et al (2015)[3] and Vilar-Gomez (2018)[29] were the only studies to report all key confounders (i.e., age, gender, diabetes mellitus, hypertension, statin use and smoking) and adjust for them all in their data analysis. Leung et al (2017)[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 NAFLD without adjustment for covariates

Across the ten studies reporting clinical events, a total of 591 out of 3,338 (17.7%) patients with NAFLD died over the period of follow-up, eight studies reported 95 liver-related deaths in 2,729 patients (3.5%). Seven studies reported 52 out of 2,510 (2.1%) NAFLD patients experienced a liver transplantation. Events due to liver morbidity were reported in 362 out of 3,125 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 to 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 transplantation, and all event liver morbidity and this risk was incremental according to fibrosis stage (see Table 3, Figures 2 & 3). No statistical heterogeneity ($I^2 = 0\%$) was observed for the comparison of fibrosis stages 1-4 versus stage 0 across the four event outcomes.

Fibrosis related event outcomes in all patients with NAFLD after adjustment for confounding covariates

A sub-group of six studies provided hazard ratios for events comparing mild to moderate fibrosis (stages 0 to 2) to advanced fibrosis (stage 3 to 4) based on multivariable Cox regression models that adjusted for potential key confounding covariates.[21,23-25,26,27,29] All studies adjusted their analyses for age, gender, diabetes and hypertension with exception of Seko et al (2015)[27] that adjusted for age, gender, 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 e-appendix 4). In those studies that provided both an adjusted and unadjusted risk ratio, the magnitude of increased risk with advanced fibrosis appeared to similar as indicated by overlapping 95% confidence intervals. These conclusions remained consistent when Seko et al (2015) was removed from the meta-analysis.

Impact of the presence of NASH on fibrosis related event outcomes without adjustment for covariates

Four studies reported fibrosis related event outcomes in a cohort of NAFLD patients reported to have either NASH or not have NASH.[20,23,24,25] A low level of statistical heterogeneity (I^2 = 0%) was seen with the exception of liver transplantation events for stage 0 vs 4 in the subgroup without NASH where there was evidence of substantial heterogeneity (I^2 = 56%) and 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% confidence of RR estimates (see Table 4 and e-Appendix 5).

Fibrosis related HRQoL outcomes

Three studies (1,089 NAFLD and 718 NASH patients) reported HRQoL using either the generic measure of SF-36 or the liver-specific measure of the 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 appropriate and instead numeric results were summarised across individual studies (see e-Appendix 6).

The cross-sectional analysis of David et al (2009)[19] used the generic Short Form-36 to report that in a total of 713 NAFLD patients, those with stage 4 fibrosis (‘cirrhosis’) had significantly ($p<0.001$) worse physical health as assessed by SF-36 Physical Component Score (PCS) compared with patients with NAFLD and fibrosis stages 0-3 (median 37 vs. median 47-50, $p<0.001$). This finding remained after adjustment for

potential confounders (i.e. age, sex, race, marital status, education, annual household income, BMI, type 2 diabetes). The study authors reported no significant difference across fibrosis stage for SF-36 Mental Component Score (MCS) (data not reported). Those with NASH reported significantly poorer physical health compared with those with no NASH (median 22.5 vs. 47.1, p<0.02).

The prospective cohort of Huber et al (2018)[22] found no difference in unadjusted total CLDQ score comparing a total of 304 patients with NAFLD stage 3-4 and stage 0-2 fibrosis (mean 4.9, SD 1.2 vs. 5.1, SD 1.3; p=0.07). NASH was associated with a significant lower HRQoL compared to patients with NAFL (mean: 4.85, SD: 1.3 vs. 5.31, 1.1, p<0.01).

In an RCT with 72 NASH patients, Younossi et al (2018)[31] found no difference in unadjusted baseline HRQoL between stage 2 and 3 fibrosis in either SF-36 (PCS - mean 45.0, SD 9.6 vs. 43.4, 10.3 & MCS - 51.0, 9.6 vs. 50.6, 12.7, both p>0.05) or total CLDQ score (mean 4.83, SD 1.10 vs. 4.91, 1.25, p>0.05).[30]

Meta-regression

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 (i.e. study design, overall risk of bias or follow up) on the impact of stage of fibrosis for either of these outcomes (see e-Appendix 7).

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 event for comparison of fibrosis stage 0 vs 2 (Egger test, P=0.05) (see e-Appendix 8). This asymmetry appeared to be due to an absence of small- to medium-sized studies with a RR<1.0.

Quality of evidence

Based on the GRADE approach[13] we found the quality of evidence for fibrosis in NAFLD as prognostic predictor of all-cause mortality to be high and for liver-mortality to be moderate (see e-Appendix 9). The quality of evidence for liver-related mortality. liver transplantation, 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 mortality and morbidity in NAFLD (with a five to 12 increase in relative risk 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 NAFLD patients with NAFL or NASH. There was, however, limited and contradictory evidence of the impact of stage of fibrosis on the HRQoL, primarily due to the small number of studies, heterogeneity of the study subjects and lack of data from NASH-specific HRQoL instruments, such as CLDQ-NASH.[32]

In context of current evidence

This study demonstrates 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 methodological rigour of this study now provides the confidence to support the conclusion of previous studies and recommendations of clinical guidelines. With advancing fibrosis, there is a stepwise increase in relative risk for liver morbidity, liver mortality and all-cause mortality.

Our review also extends previous findings to the subset of patients who have reported histological 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 categorised as currently having histological 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 United States FDA have recently published a table of surrogate endpoints that have either been already used in their development programs for drugs that have been approved, or surrogate endpoints that 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. 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. Whilst 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 as regulatory bodies and payers, who are responsible for healthcare reimbursement decisions and typically more stringent in their evidence requirements, prefer evidence from final patient-relevant outcomes and will only accept surrogate endpoints if based on formal evidence of validation.[5,6] The importance of the link between putative surrogates to clinically meaningful outcomes is recognised 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 4,428 NAFLD patients and 591 all-cause deaths. In contrast, the systematic review and meta-analysis of Dulai et al 2017[8] included 1,495 NAFLD patients and 348 deaths. We extended the scope of this previous study to consider: the potential impact of key confounder variables, the subgroup of NAFLD patients with NASH, 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 recognise 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) i.e. estimation of pooled relative risk, 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 (e.g. age, gender, 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. 3-4) was similar when compared to the results of the simple (unadjusted) pooled RR approach to pooling studies using hazard ratios and following adjustment for potential key confounders. Secondly, whilst 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-recognised controversy of current clinical practice.[37] In order 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 recognise that some patients with NASH (steatosis, ballooning and lobular inflammation) may be miscategorised as not having NASH because of sampling error on the biopsy. Moreover, a liver biopsy only represents 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 towards cirrhosis some features of NASH, such as steatosis and hepatocyte ballooning, may become less prominent and thus a patient may be categorised as not having 'active' NASH, yet NASH was clearly the causative factor in their liver fibrosis. Thirdly, 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. Fourthly, whilst we sought to include a range of clinical outcomes, the wide meta-analysis confidence intervals 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 methodological quality – seven out of 13 studies were retrospective in design and three were overall judged to be at high risk of bias. Nevertheless, our meta-regression 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. (1) We need to better understand the association between fibrosis stage and patient reported well-being. Future outcome NAFLD and NASH studies therefore need to consistently

collect patient HRQoL using generic (such as EQ-5D-5L) and disease-specific measures (such as CLDQ-NASH[32]). (2) 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 (e.g. REGENERATE, REVERSE, RESOLVE-IT, AURORA).[42-45] (3) 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 non-invasive alternative biomarkers as prognostic markers or validated surrogate endpoints, an issue that is currently being explored by two large international multi-stakeholder consortia in Europe (IMI2 LITMUS) and the USA (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 NAFLD patients with and without reported NASH; increasing fibrosis stage being associated with a five to 12-fold increase in the relative risk of liver-related events. Further evidence from well reported studies is needed in order 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.

Acknowledgements: We thank Dr Juliana Bottomley, Gilead for her comments on the review protocol and draft of the manuscript.

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;60:565-75.
4. McPherson S, Hardy T, Henderson E, Burt AD, Day CP, Anstee QM. Evidence of NAFLD. Progression from steatosis to fibrosing-steatohepatitis using paired biopsies: implications for prognosis & clinical management. *J Hepatol.* 2015; 62:1148-55.
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;149:389-97.
6. Hannah WN, Jr, Torres DM, Harrison SA. Nonalcoholic steatohepatitis and endpoints in clinical trials. *Gastroenterol & Hepatol*. 2016;12: 756-63.
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, et al. 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. *Brit Med J.* 2016;352:i1981. doi: 10.1136/bmj.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;62:1006-12.

12. Hayden JA, Côté P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. *Ann Inter Med*. 2006;144:427-37.
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*. Brit Med J.* 2015;350:h870.
14. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.0 (updated August 2019). Cochrane, 2019. www.training.cochrane.org/handbook (last accessed 30th December 2019)
15. Egger M, Davey Smith G, Schneiger M, et al. Bias in meta-analysis detected by a simple, graphical test. *Brit Med J.* 1997;315:629-34.
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;30:13.
17. Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. *Pharm Stat*. 2006;5:173-86.
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;54:1208-16.
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;49:1904-12.
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.* 2018 Dec 20. pii: S1542-3565(18)31392-2. doi: 10.1016/j.cgh.2018.12.016. [Epub ahead of print]

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: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: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;45:728-38.
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;155:443-57
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;53:1874-82.
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, et al. Validation of Chronic Liver Disease Questionnaire for Nonalcoholic Steatohepatitis in patients with biopsy-proven nonalcoholic steatohepatitis. *Clin Gastroenterol Hepatol*. 2019. pii: S1542-3565

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:e0175717.
34. Sanyal AJ, Chalasani N, Kowdley KV, et al. Pioglitazone, vitamin E, or placebo for nonalcoholic steatohepatitis. *N Engl J Med.* 2010;362:1675-85.
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;385:956-65.
36. US Federal Drug Agency (FDA). *Table of Surrogate Endpoints That Were the Basis of Drug Approval or Licensure*. Available at: https://www.fda.gov/drugs/development-resources/table-surrogate-endpoints-were-basis-drug-approval-or-licensure (last accessed 22nd July 2019)
37. US Federal Drug Agency (FDA). *Noncirrhotic Nonalcoholic Steatohepatitis With Liver Fibrosis: Developing Drugs for Treatment Guidance for Industry* – draft, June 2019. Available at: https://www.federalregister.gov/documents/2019/06/07/2019-11951/nonalcoholic-steatohepatitis-with-compensated-cirrhosis-developing-drugs-for-treatment-draft (last accessed 22nd July 2019)
38. Rinella ME, Tacke F, Sanyal AJ, et al. Report on the AASLD/EASL joint workshop on clinical trial endpoints in NAFLD. *J Hepatol.* 2019 pii: S0168-8278(19)30283-1.
39. Rinella ME, Loomba R, MD, Caldwell SH, et al. Controversies in the Diagnosis and Management of NAFLD and NASH. *Gastroenterol Hepatol (NY).* 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):iii-iv, 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 (last accessed 22nd July 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 (last accessed 22nd July 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 (last accessed 22nd July 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 (last accessed 22nd July 2019)
46. Innovative Medicines Initiative. LITMUS. Liver Investigation: Testing marker utility in steatohepatitis. https://www.imi.europa.eu/projects-results/project-factsheets/litmus (last accessed 22nd July 2019)
47. Foundation for National Institutes for Health Biomarkers Consortium. Non-Invasive Biomarkers of Metabolic Liver Disease (NIMBLE) https://fnih.org/news/announcements/nimble (last accessed 22nd July 2019)

Author names in bold designate shared co-senior authorship:

Rod S Taylor, Stuart McPherson, Philip N Newsome, Quentin M Anstee

FIGURE LEGENDS

Figure 1. Summary of study selection process

Figure 2. Meta-analysis: Unadjusted relative risk of all-cause mortality by fibrosis stage (vs. stage 0) in all NAFLD patients

Figure 3. Meta-analysis: Unadjusted relative risk of liver events by fibrosis stage (vs. stage 0) in all NAFLD patients

Table 1. Characteristics of included studies.

Author (year) Country	Study design Time period Sampling frame	Population diagnosis	Population demographics	Fibrosis staging	Outcomes reported	Follow up
Angulo (2015) Australia, Denmark, Iceland, Thailand, UK, US NR centres	Retrospective cohort study 1975-2005[2] Consecutive patients	619 patients with liver biopsy confirmed NAFLD 284 with NASH, method of confirmation not reported	Age: median 49 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 transplantation, Liver events[1]	Median 12.6 yrs (range 0.3-35.1)
Bhala (2011) Australia, Italy, UK, USA, Thailand 4 centres	Prospective cohort study 1984-2006[2] 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 DM: 50.6% White: 91.5% Male: 39.5% HTN: 44.1% Statin use: 21.5% Smoking: Not reported	Biopsy reviewed independently and reported as stage 3 & 4	Overall mortality, Liver-related mortality, Overall vascular events, Myocardial infarction, Total liver events[3], Varices, Ascites, Encephalopathy	Mean: 7.1 yrs (range: 0.5-24.75)
David (2009) US	Cross-sectional study (based on NAFLD	713 patients with liver biopsy confirmed NAFLD	Age: mean 48 DM: not reported White: 76.2%	Biopsy centrally confirmed and reported as stage	Health-related quality of life (SF-36)	Not applicable

(NASH CRN Research Group) 8 centres	prospective cohort & PIVENS RCT) 2004-2007[2] Not reported	436 with NASH, method of confirmation not reported	Male: 37.7% HTN: 27% Statin use: not reported Smoking: not reported	0-4		
Hagstrom (2017i,ii) Sweden 2 centres	Retrospective cohort study 1971-2009[4] All patients	646 patients with liver biopsy confirmed NAFLD 383 with NASH, defined by fatty liver inhibition of progression (FLIP) algorithm	Age: mean 48 DM: 14.4% White: not reported Male: 62.2% HTN: 30.3% Statin use: not reported Smoking: 24.0%	Biopsy centrally confirmed and reported as stage 0-4	Overall mortality, Severe liver disease[5]	Mean: 19.9 yrs (range: 0.4-40)
Huber (2018) Germany, Spain, UK (European NAFLD registry) 3 centres	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 DM: 51.3% [T2] White: not reported Male: 53.3% HTN: 66.8% Statin use: not reported Smoking: not reported	Biopsy centrally confirmed and reported as stage 0-4	Health-related quality of life Chronic Liver Disease Questionnaire (CLDQ)	Up to 6-months post biopsy

Ito (2019) Japan 2 centres	Retrospective cohort study 1999-2014[5] All patients	246 patients with liver biopsy confirmed NAFLD 156 with NASH, defined by FLIP criteria	Age: median 55 DM: 45.1% White: not reported Male: 52% HTN: 41.6% Statin use: no reported Smoking: not reported	Biopsy centrally confirmed and reported as stage 0-4	Overall mortality, Liver cirrhosis, Liver cancer, extra-hepatic cancer, cardiovascular disease	Median: 7.0 yrs (range 4.4-10.0)
Leung (2017) Hong Kong 1 centre	Prospective cohort study 2006-2015[6] Consecutive patients	300^ patients with liver biopsy confirmed NAFLD 151 with NASH, defined by FLIP criteria	Age: mean 51 DM: 55.4% White: not reported Male: 55.7% HTN: 55.4% Statin use: not reported Smoking: not reported	Biopsy centrally confirmed and reported as stage 0-4	Overall mortality, liver-related events[7], non-hepatic cancer, cardiovascular disease	Median: 4.1 yrs (range NR)
Peleg (2018) Israel 1 centre	Retrospective cohort study 2005-2012[6] All patients	153 patients with liver biopsy confirmed NAFLD 27 with NASH, defined by the presence of steatosis, ballooning and	Age: mean 49.5 DM: 63.4% [T] White: not reported Male: 55.5% HTN: 41.1% Statin use: 53.8% Smoking: not reported	Biopsy confirmed and reported as stage 0-4	Overall mortality, malignancies, liver events[8], hospital admissions	Mean: 8.3 yrs (range 5.1-12.0)

		lobular inflammation				
Sebastiani (2015) Canada Single centre	Retrospective cohort study 2004-2013[9] Consecutive patients	148 patients with liver biopsy confirmed NAFLD, 148 with NASH, definition not specified	Age: mean 49.5 DM: 33.1% White: not reported Male: 69.6% HTN: 39.2% Statin use: not reported Smoking: no reported	Biopsy confirmed and reported as stage 0-4	Clinical outcomes[10]	Median: 5 yrs (interquartile range: 3-8)
Seko (2015) Japan 1 centre	Retrospective cohort study 1999-2013[6] All patients	312 patients with liver biopsy confirmed NAFLD 176 with NASH, defined by Younossi criteria	Age: median 59 DM: 35% White: not reported Male: 51% HTN: not reported Statin use: 40.3% Smoking: not reported	Biopsy confirmed and reported as stage 0-4	Overall mortality, malignancies	Median: 4.8 yrs (0.3-15.7)
Vilar-Gomez (2018) Australia, Cuba, Hong Kong, Spain 5 centres	Prospective cohort study 1995-2016[11] Consecutive patients	458 patients with liver biopsy confirmed NAFLD 458 with assumed to be NASH by nature of stage 3-	Age: mean 55.9 DM: 67% White: 81% Male: 48% HTN: 61% Statin use: 24% Smoking: 17%	Biopsy reviewed independently and reported as stage 3 & 4	Overall mortality, major clinical events[12]	Mean: 5.5 yrs (2.7-8.2)

		4 fibrosis				
Younossi (2011/17) USA 3 centres	Retrospective cohort study Not reported Not reported	210[+] patients with liver biopsy confirmed NAFLD 131 with NASH, defined by the presence of steatosis, ballooning and lobular inflammation	Age: mean 49* DM: 20.5% [T2] White: not reported Male: 37.8% HTN: not reported Statin use: not reported Smoking: not reported	Biopsy confirmed, NAS and Brunt 0-4 fibrosis	Liver-related mortality	Median: 12.1 yrs (interquartile range: 4.9-15.5)
Younossi (2018) USA/Canada 23 centres	Randomised controlled trial 2015-2017[11] 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 DM: 70.8% White: 90.3% Male: 30.6% HTN: 66.7% Statin use: not reported Smoking: not reported	Biopsy confirmed stage 2-3 fibrosis	Health-related quality of life (SF-36 & Chronic Liver Disease Questionnaire (CLDQ))	Up to 24 weeks

Footnotes:

1. Gastroesophageal varices/bleeding, ascites, portosystemic encephalopathy, spontaneous bacterial peritonitis, hepatocellular cancer, hepatopulmonary syndrome, hepatorenal syndrome
2. Year of recruitment

3. Liver failure, gastroesophageal varices, ascites, encephalopathy, hepatopulmonary syndrome, HCC
4. Year of diagnosis
5. Acute and subacute liver failure, ascites, oesophageal varices, hepatorenal syndrome, chronic liver failure, cirrhosis NUD, hepatic encephalopathy, liver failure NUD, portal hypertension, hepatocellular carcinoma
6. Year of biopsy
7. Hepatocellular carcinoma, ascites, spontaneous bacterial peritonitis, hepatorenal syndrome, variceal bleeding, hepatic encephalopathy, liver transplantation
8. Oesophageal varices, hepatic encephalopathy, ascites and TIPS.
9. Year of study visits
10. Death, liver transplantation, cirrhosis complications
11. Years of study

*: weighted mean; +: based on number reported in Dulai et al 2017 review [8]; ^: 307 patients reported in paper but data provided by research groups included only 300

Abbreviations: BMI – body mass index; CRN – NASH Clinical Research Network; DM – diabetes mellitus; FLIP: fatty liver inhibition of progression; HTN; hypertension NAFLD – Non-alcoholic fatty liver disease; NASH; NR – not reported;

Table 2. Assessment of risk of bias of included studies - based on QUIPs tool.

Author (year) Country	Study Population	Study Attrition	Prognostic factor measurement	Outcome measurement	Confounding assessment & account	Data analysis & reporting	Overall assessment*
Angulo (2015)	Yes	Partly	Yes	Partly	Yes	Yes	Moderate risk of bias
Bhala (2011)	Yes	Yes	Yes	Yes	Yes	Partly	Moderate risk of bias
David (2009)	Partly	No	Yes	Yes	Partly	Partly	Moderate risk of bias
Hagstrom (2017i,ii)	Yes	Partly	Yes	Yes	Partly	Partly	Moderate risk of bias
Huber (2018)	Partly	No	Yes	Yes	Partly	Partly	Moderate risk of bias
Ito (2019)	Yes	Partly	Yes	Yes	Partly	Partly	Moderate risk of bias
Leung (2017)	Yes	Partly	Partly	Yes	No	No	High risk of bias
Peleg (2018)	Yes	Partly	Yes	Yes	Partly	Partly	Moderate risk of bias
Sebastiani	Yes	Partly	Yes	Yes	Partly	Partly	Moderate risk

(2015)							of bias
Seko (2015)	No	Partly	Partly	Partly	Partly	Partly	Moderate risk of bias
Vilar-Gomez (2018)	Yes	Yes	Yes	Yes	Yes	Yes	Low risk of bias
Younossi (2011/17)	Partly	Partly	Yes	Yes	Partly	No	Moderate risk of bias
Younossi (2018)	Partly	Yes	Yes	Yes	Yes	No	Moderate 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 one or more domains are scored as partly or one domain is scored as “no”; High risk of bias: describes studies for which more than one domain is scored as “no”.

Table 3. Meta-analysis: Pooled unadjusted relative risk by fibrosis stage (relative to stage 0) for all patients with NAFLD

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

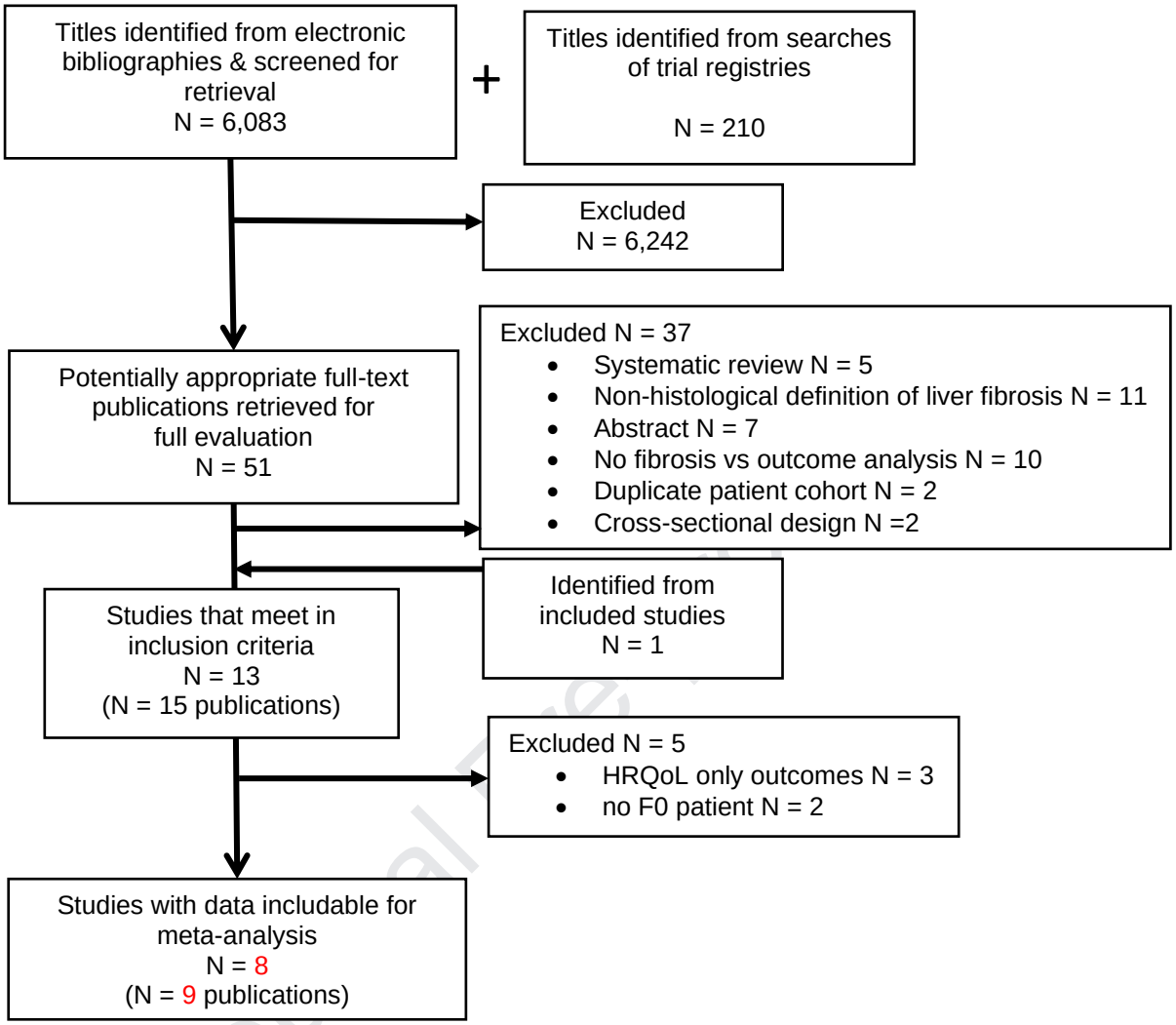
All meta-analyses fixed effect

Table 4. Stratified meta-analysis: Pooled unadjusted relative risk by fibrosis stage (relative to stage 0) for NAFLD patients with reported NASH versus NALFD patients with no reported NASH (N=4 studies)

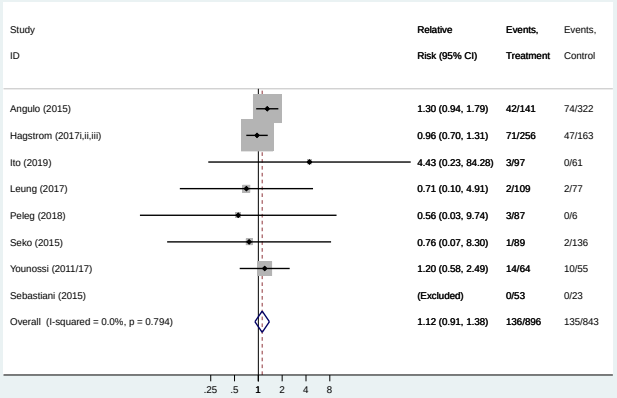
	Stage 0 versus 1 **Relative Risk (95% CI),** **P-value** n/N vs n/N, I^2 statistic	**Stage 0 versus 2** **Relative Risk (95% CI),** **P-value** n/N vs n/N, I^2 statistic	**Stage 0 versus 3** **Relative Risk (95% CI),** **P-value** n/N vs n/N, I^2 statistic	**Stage 0 versus 4** **Relative Risk (95% CI),**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

NAFLD without NASH	1.08 (0.45 to 2.58) 8/230 vs 11/268, 0%	2.85 (1.12 to 7.24) 8/230 vs 11/71, 0%	4.56 (1.64 to 12.60) 8/230 vs 7/36, 0%	9.80 (3.12 to 30.76) 8/230 vs 15/28, 0%

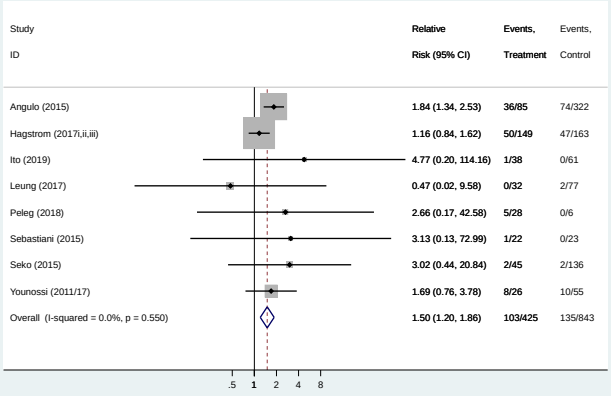
NA: not applicable; *: random effects meta-analyses. All other meta-analyses fixed effect


Titles identified from electronic bibliographies & screened for retrieval
N = 6,083
+
Titles identified from searches of trial registries
N = 210
Excluded
N = 6,242
Potentially appropriate full-text publications retrieved for full evaluation
N = 51
Excluded N = 37
Systematic review N = 5
Non-histological definition of liver fibrosis N = 11
Abstract N = 7
No fibrosis vs outcome analysis N = 10
Duplicate patient cohort N = 2
Cross-sectional design N =2
Identified from included studies
N = 1
Studies that meet in inclusion criteria
N = 13
(N = 15 publications)
Excluded N = 5
HRQoL only outcomes N = 3
no F0 patient N = 2
Studies with data includable for meta-analysis
N = 8
(N = 9 publications)

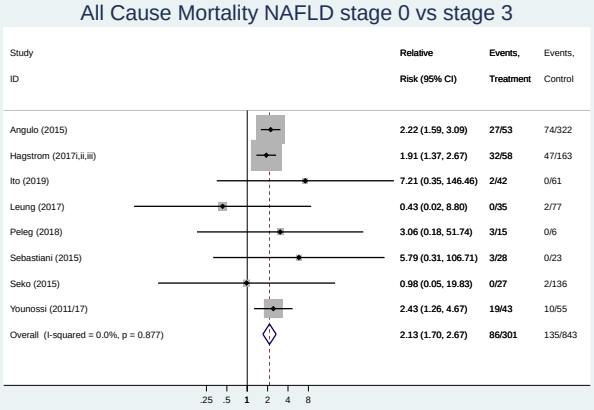
All Cause Mortality NAFLD stage 0 vs stage 1



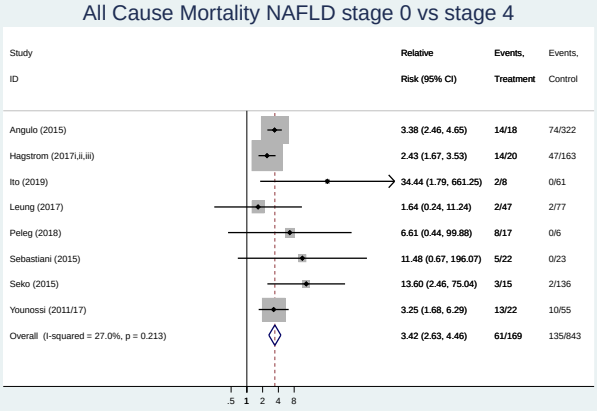
All Cause Mortality NAFLD stage 0 vs stage 2



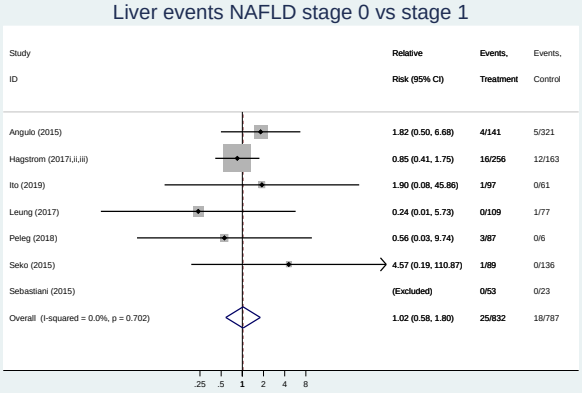
All Cause Mortality NAFLD stage 0 vs stage 3



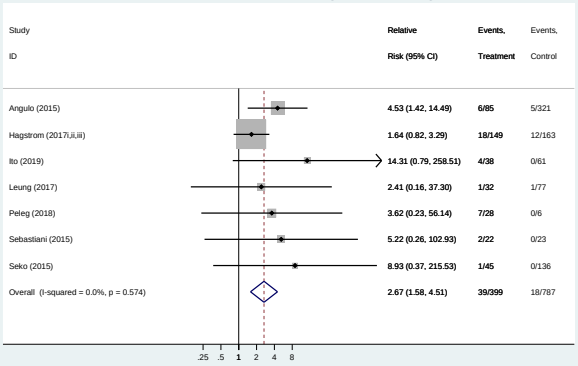
All Cause Mortality NAFLD stage 0 vs stage 4



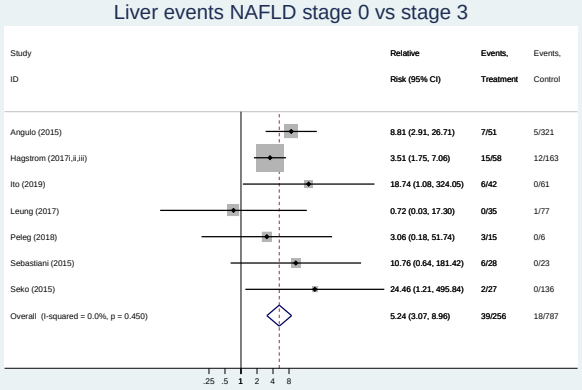
Liver events NAFLD stage 0 vs stage 1



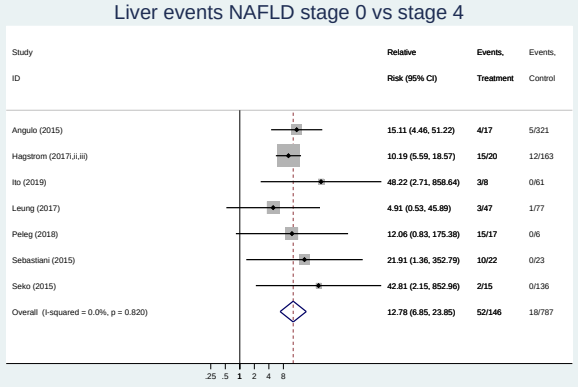
Liver events NAFLD stage 0 vs stage 2



Liver events NAFLD stage 0 vs stage 3



Liver events NAFLD stage 0 vs stage 4



SUPPLEMENTAL MATERIAL

eAppendix 1. Search strategy

Database: Ovid MEDLINE(R) 1946 to October week 3 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

eAppendix 2. QUIPS (QUality In Prognosis Studies) tool

Potential bias (circle one)	Items considered for assessment of potential opportunity for bias	Yes response	No response	Study 1	Study 2	Study 3	Study 4
Study Population The study sample represents the population on key characteristics sufficient to limit potential bias to the observed relationship between BMI and mortality 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.	AKI adult (>18 yrs) patients AND AKI criterion reported (e.g. creatinine clearance +/- urine output, staging) AND patient characteristics include ethnicity, underlying condition AND no major exclusions AND treatment method for AKI (if applicable) described e.g. RRT, mechanical ventilation	Non-adults included, patient characteristics not adequately described, underlying condition unknown/ not detailed, major exclusions involved and treatment method not detailed				
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 verses 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 BMI/ body mass is adequately measured in study participants to sufficiently limit bias Yes / Partly / No	Body mass measured at time of AKI presentation. Clear definition of BMI given and BMI category ranges provided with number of participants in each BMI category reported OR sufficient data provided to determine BMI categories. Adequate proportion of the study sample has complete data.	Data collection is prospective and recorded on presentation of AKI AND BMI criteria defined	Definition of BMI not clear or sufficiently detailed (e.g. BMI > 30 is categorised as obese with no other obese categories included or clear data on the range of BMIs > 30)				
Outcome measurement	Clear definition of mortality measurement provided, including duration of	Mortality incidence recorded AND time-	mortality incidence and				

Mortality incidence is adequately measured in study participants to sufficiently limit potential bias Yes / Partly / No	follow-up. Mortality risk measured prior to outcome occurring.	frame for mortality reported AND data collection for mortality risk is prospective	timeframe for follow-up not reported				
Confounding measurement and account Important potential confounders are appropriately accounted for, limiting potential bias with respect to body mass Yes / Partly / No	Important potential confounders are accounted for in the study design (e.g. patients separated into different BMI categories if ethnic origin differs i.e. afro-Caribbean, south Asian) and analysis. Measurement of all important confounders is adequately valid and reliable (e.g. accounting for hydration status when measuring weight -- under-hydrated or fluid overloaded patients). The method and setting of confounding measurement are the same for all participants (e.g. same scales). Appropriate imputation method is used for missing confounder data. Appropriate adjustment used and clearly outlined. Interventions do not confound body mass results or mortality outcome.	BMI categories adequately defined AND based on accurate, reliable weights. Adjusters, if used, are appropriate and clearly outlined. Intervention method does not impact on mortality outcome	BMI or body mass measurements are not clearly or adequately defined. Validity of body mass measurements not reported on				
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. mortality incidence risk). 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 strategy or results AND inappropriate statistical model AND selective reporting of results				

eAppendix 3. Excluded studies

1. Dam-Larsen S, Becker U, Franzmann MB, Larsen K, Christoffersen P, Bendtsen F. Final results of a long-term, clinical follow-up in fatty liver patients. *Scand J Gastroenterol.* 2009;44:1236-43 [no fibrosis by outcome analysis]
2. Dulai PS 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 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 Hagstrom 2017]
4. Golabi P et al. Non-alcoholic steatofibrosis (NASF) can independently predict mortality in patients with non-alcoholic fatty liver disease (NAFLD). *BMJ Open Gastroenterology* 2018; 5 e000198. [not association between fibrosis and mortality]
5. Hagström H 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 Hagstrom 2017]
6. Hashimoto E, Yatsuji S, Kaneda H, Yoshioka Y, Taniai M, Tokushige K, Shiratori K. The characteristics and natural history of Japanese patients with nonalcoholic fatty liver disease. *Hepatol Res.* 2005;33:72-6. [no fibrosis by outcome analysis]
7. Huber Y et al. Health-related quality of life in patients with non-alcoholic fatty liver disease. *J Hepatology* 2017;66 (Supplement 1):S597-S598. [abstract only and non-invasive fibrosis]
8. Huber Y et al. Health-related quality of life correlates with histological severity in non-alcoholic fatty liver disease. *J Hepatology* 2018;68 (Supplement 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 Hepatology* 2018;68 (Supplement 1):S561. [abstract full paper included]
10. Jaruvongvanich V et al. 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. *Clinics & Res Hepatology Gastroenterol* 2017;41:629-634. [systematic review]

11. Kim D et al. 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 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:e0173499 [non-invasive fibrosis only]
13. Lee TY 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 et al. Nonalcoholic fatty liver disease: a spectrum of clinical and pathological severity. *Gastroenterol.* 1999;116:1413-1419. [no biopsy fibrosis-outcome analysis]
15. Miyake N et al. Progression of liver fibrosis is associated with non-liver-related mortality in patients with nonalcoholic fatty liver disease. *Hepatology* 2018;68 (Supplement 1):974A. [abstract & non-invasive fibrosis score – forward search no paper]
16. Patel JR et al. Risk of mortality by fibrosis stage in NAFLD: a systematic review and meta-analysis. *Hepatology.* 2017;64:1095A [abstract & systematic review]
17. Renelus BD 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 non-invasive fibrosis only]
18. Sayiner M, Stepanova M, Pham H, Noor B, Walters M, Younossi ZM. Assessment of health utilities and quality of life in patients with non-alcoholic fatty liver disease. *BMJ Open Gastroenterol.* 2016;3: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: 219. [systematic review]
20. Sanyal A 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 Hepatology.* 2017 (Supplement): S2–S3 [abstract and no fibrosis-outcome analysis]

21. Sebastiani G 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 (Supplement 1):S491-S492.[abstract and non-invasive 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 (Supplement 1):S563. [abstract - non-invasive fibrosis]
24. Stepanova M et al. Predictors of all-cause mortality and liver-related mortality in patients with non-alcoholic fatty liver disease (NAFLD). *Digest Dis & Sci.* 2013;58:3017-3023. [no fibrosis by outcome analysis]
25. Stepanova M et al. Pathologic features of non-alcoholic steatohepatitis (NASH) as independent predictors of liver-related mortality. *J Hepatology* 2011;S25. [no fibrosis by outcome analysis]
26. Strasser M et al. SAF score effectively identifies NAFLD subjects at high risk of subsequent liver related but not cardiovascular or malignancyassociated mortality and morbidity. *J Hepatology* 2017;66 (Supplement 1):S425-S426. [no fibrosis by outcome analysis]
27. Sun 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. *Hepatology Res.* 2016;46: 862-870. [systematic review]
28. Tada T et al. Progression of liver fibrosis is associated with non-liver-related mortality in patients with nonalcoholic fatty liver disease. *Hepatology Comm.* 2017;1:899-910. [non-invasive fibrosis vs non-hepatic mortality]
29. Treeprasertsuk S et al. NAFLD fibrosis score: a prognostic predictor for mortality and liver complications among NAFLD patients. *World J Gastroenterol.* 2013;19:1219-1229. [non-invasive fibrosis score]

30. Unalp-Arida A et al. Liver fibrosis scores predict liver disease mortality in the United States population. *Hepatology* 2017;66: 84-95. [non-invasive fibrosis only]
31. Wijarnpreecha K et al. Non-invasive fibrosis markers are independent predictors of mortality among U.S. adults with nonalcoholic fatty liver disease. *J Hepatology* 2017;66 (Supplement 1):S662-S663. [non-invasive fibrosis only]
32. Xun Y H 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. [non-invasive fibrosis only]
33. Yoshihisa A et al. Liver fibrosis score predicts mortality in heart failure patients with preserved ejection fraction. *ESC Heart Failure* 2018;5:262-270. [non-invasive 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. [non-invasive fibrosis only]
35. Younossi Z M 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, Racila A, Lam B, Pham HT, Hunt S. 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 Z M et al. Validation of chronic liver disease questionnaire for non-alcoholic steatohepatitis in patients with biopsy-proven non-alcoholic steatohepatitis. *Clin Gastroenterol Hepatol.* 2019 pii: S1542-3565(19)30021 [no fibrosis by outcome analysis]

e-Appendix 4. Meta-analysis: Pooled hazard ratio (adjusted) and pooled relative risk (unadjusted) by fibrosis stage 0-2 vs 2-4 for all patients with NAFLD

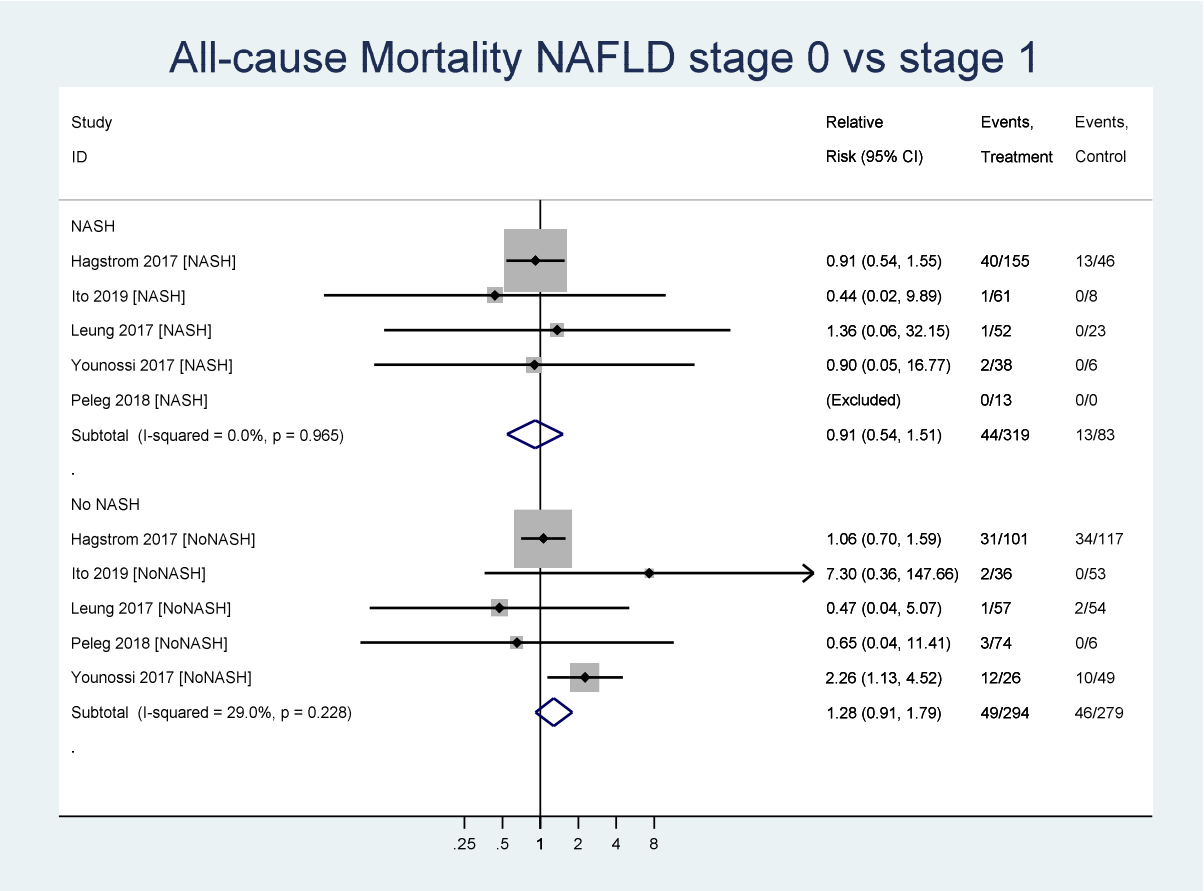
	Stage 0-2 versus 3-4 **Adjusted hazard ratio (95% CI) P-value, I^2 statistic**	**Stage 0-2 versus 3-4** **Unadjusted relative risk (95% CI) P-value, I^2 statistic**
All-cause mortality		
N=5 studies	2.24 (1.48 to 3.39), 31%	2.25 (1.85 to 2.72), 35%
Liver-related mortality		
N= 4 studies	5.12 (2.48 to 10.55), 0%	6.42 (3.45 to 11.95), 0%
Liver transplantation		
N=2 studies	10.89 (2.01 to 58.99), 0%	3.40 (0.96 to 12.00), 0%
All liver events		
N=6 studies	5.58 (3.70 to 8.40), 0%	6.31, (4.60 to 8.65), 0%

All results are fixed effect meta-analysis

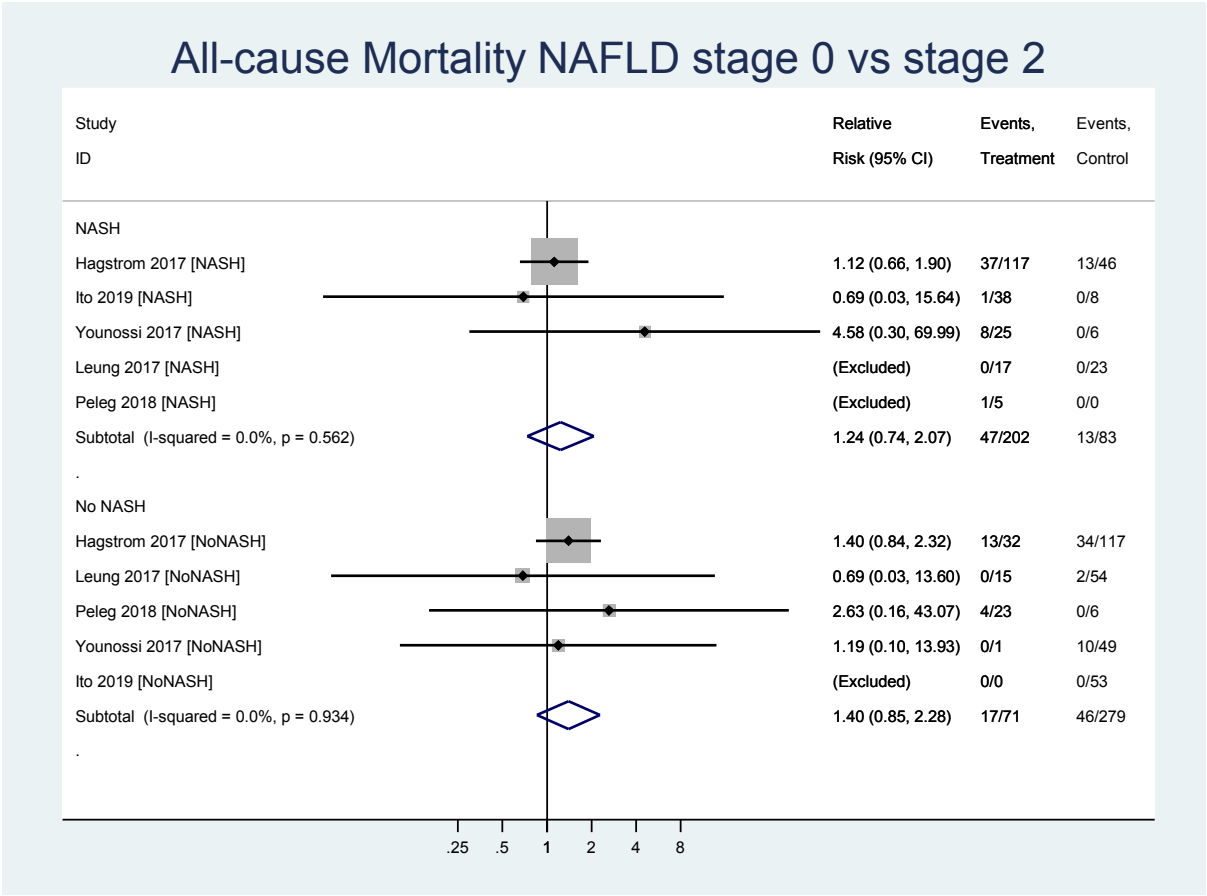
e-Appendix 5. Meta-analysis of events by NAFLD with NASH vs NALFD without NASH

All -cause mortality

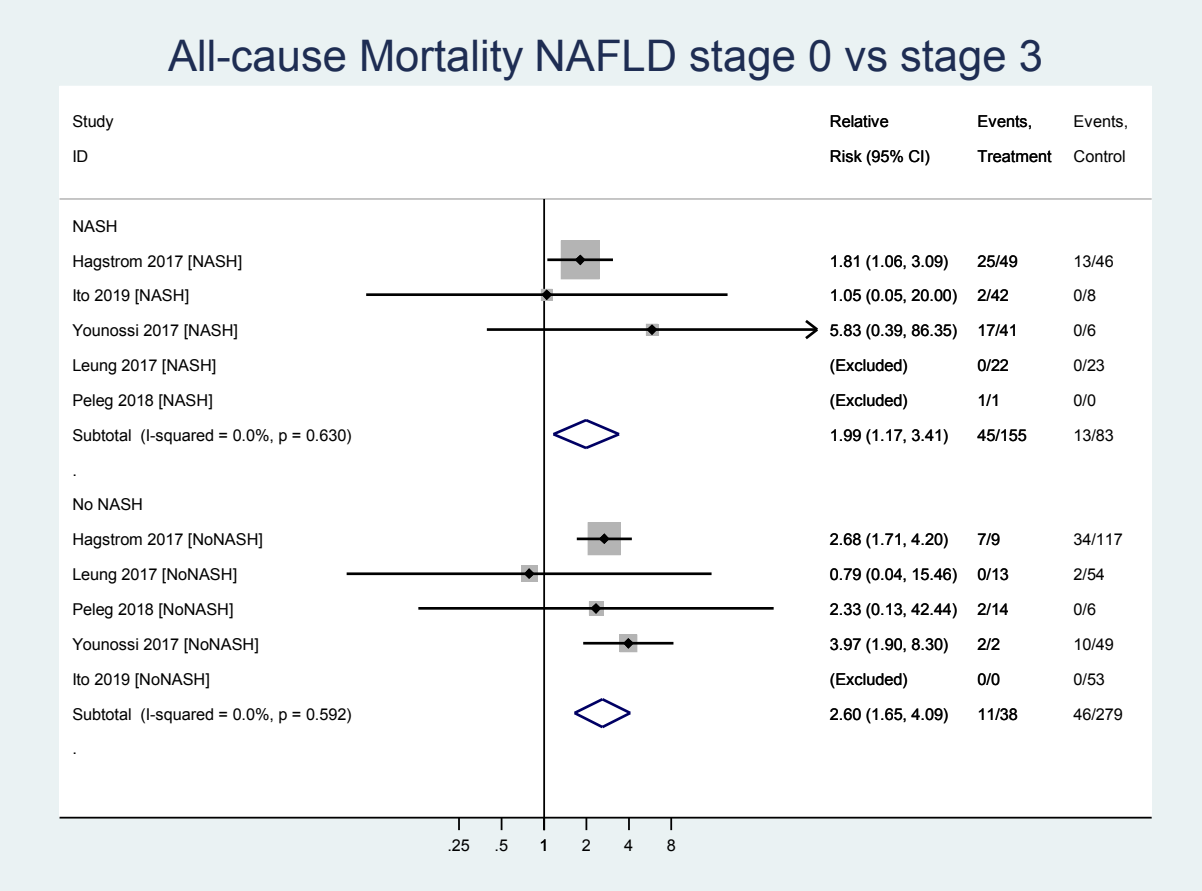
Stage 0 vs 1



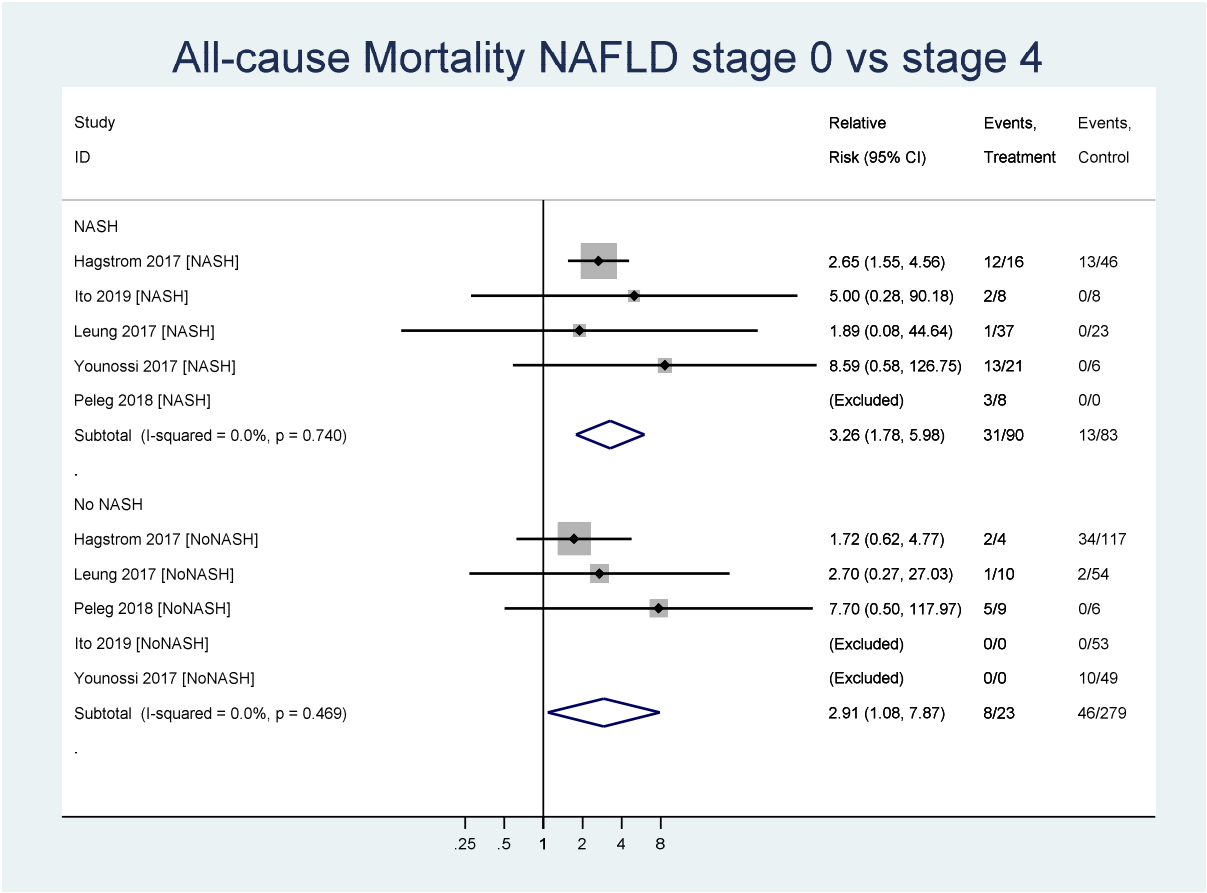
Stage 0 vs 2


All-cause Mortality NAFLD stage 0 vs stage 2
Study
ID
Relative
Risk (95% CI)
Events,
Treatment
Events,
Control
NASH
Hagstrom 2017 [NASH] 1.12 (0.66, 1.90) 37/117 13/46
Ito 2019 [NASH] 0.69 (0.03, 15.64) 1/38 0/8
Younossi 2017 [NASH] 4.58 (0.30, 69.99) 8/25 0/6
Leung 2017 [NASH] (Excluded) 0/17 0/23
Peleg 2018 [NASH] (Excluded) 1/5 0/0
Subtotal (I-squared = 0.0%, p = 0.562) 1.24 (0.74, 2.07) 47/202 13/83
.
No NASH
Hagstrom 2017 [NoNASH] 1.40 (0.84, 2.32) 13/32 34/117
Leung 2017 [NoNASH] 0.69 (0.03, 13.60) 0/15 2/54
Peleg 2018 [NoNASH] 2.63 (0.16, 43.07) 4/23 0/6
Younossi 2017 [NoNASH] 1.19 (0.10, 13.93) 0/1 10/49
Ito 2019 [NoNASH] (Excluded) 0/0 0/53
Subtotal (I-squared = 0.0%, p = 0.934) 1.40 (0.85, 2.28) 17/71 46/279
.
.25
.5
1
2
4
8

Stage 0 vs 3

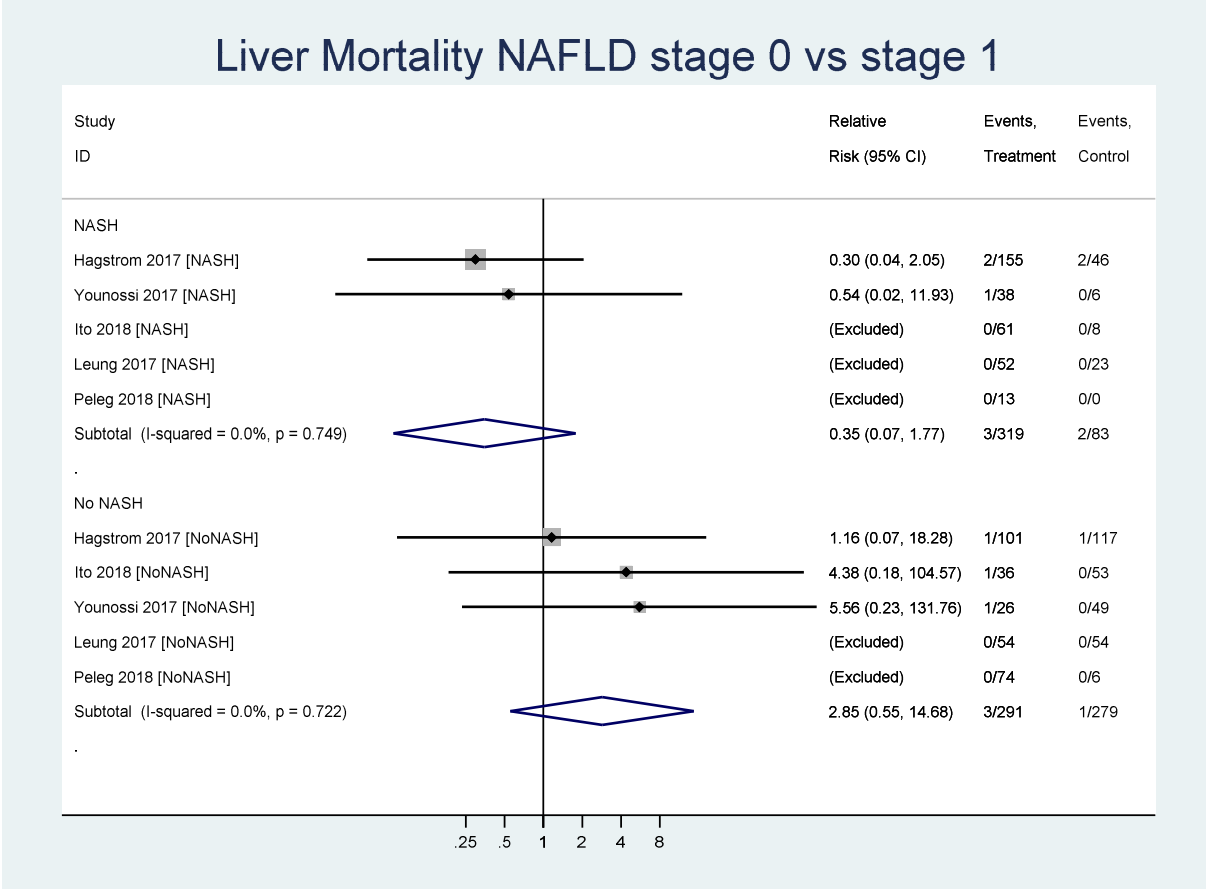

All-cause Mortality NAFLD stage 0 vs stage 3
Study ID
Relative Risk (95% CI)
Events, Treatment
Events, Control
NASH
Hagstrom 2017 [NASH] 1.81 (1.06, 3.09) 25/49 13/46
Ito 2019 [NASH] 1.05 (0.05, 20.00) 2/42 0/8
Younossi 2017 [NASH] 5.83 (0.39, 86.35) 17/41 0/6
Leung 2017 [NASH] (Excluded) 0/22 0/23
Peleg 2018 [NASH] (Excluded) 1/1 0/0
Subtotal (I-squared = 0.0%, p = 0.630) 1.99 (1.17, 3.41) 45/155 13/83
No NASH
Hagstrom 2017 [NoNASH] 2.68 (1.71, 4.20) 7/9 34/117
Leung 2017 [NoNASH] 0.79 (0.04, 15.46) 0/13 2/54
Peleg 2018 [NoNASH] 2.33 (0.13, 42.44) 2/14 0/6
Younossi 2017 [NoNASH] 3.97 (1.90, 8.30) 2/2 10/49
Ito 2019 [NoNASH] (Excluded) 0/0 0/53
Subtotal (I-squared = 0.0%, p = 0.592) 2.60 (1.65, 4.09) 11/38 46/279
.25 .5 1 2 4 8

Stage 0 vs 4

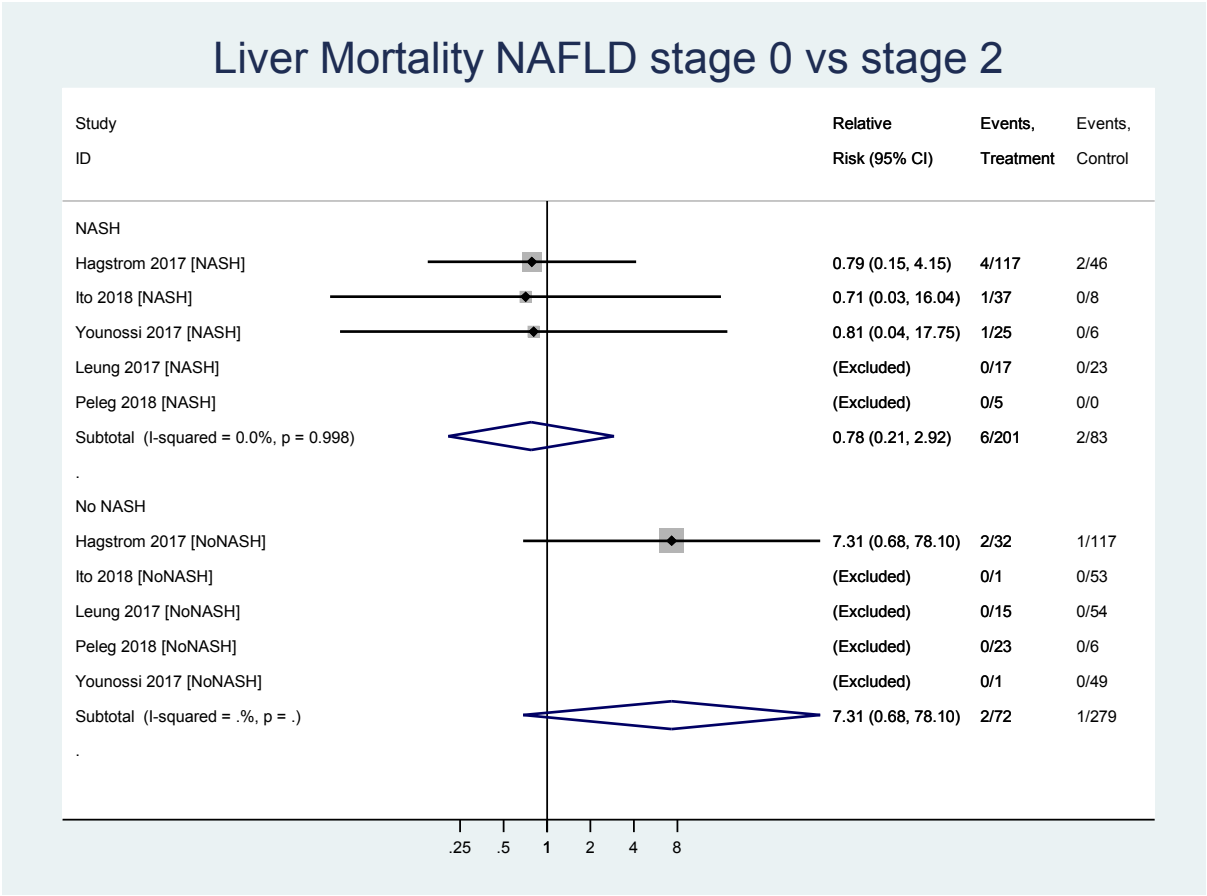

All-cause Mortality NAFLD stage 0 vs stage 4
Study
ID
Relative
Risk (95% CI)
Events,
Treatment
Events,
Control
NASH
Hagstrom 2017 [NASH] 2.65 (1.55, 4.56) 12/16 13/46
Ito 2019 [NASH] 5.00 (0.28, 90.18) 2/8 0/8
Leung 2017 [NASH] 1.89 (0.08, 44.64) 1/37 0/23
Younossi 2017 [NASH] 8.59 (0.58, 126.75) 13/21 0/6
Peleg 2018 [NASH] (Excluded) 3/8 0/0
Subtotal (I-squared = 0.0%, p = 0.740) 3.26 (1.78, 5.98) 31/90 13/83
.
No NASH
Hagstrom 2017 [NoNASH] 1.72 (0.62, 4.77) 2/4 34/117
Leung 2017 [NoNASH] 2.70 (0.27, 27.03) 1/10 2/54
Peleg 2018 [NoNASH] 7.70 (0.50, 117.97) 5/9 0/6
Ito 2019 [NoNASH] (Excluded) 0/0 0/53
Younossi 2017 [NoNASH] (Excluded) 0/0 10/49
Subtotal (I-squared = 0.0%, p = 0.469) 2.91 (1.08, 7.87) 8/23 46/279
.
.25
.5
1
2
4
8

Liver mortality

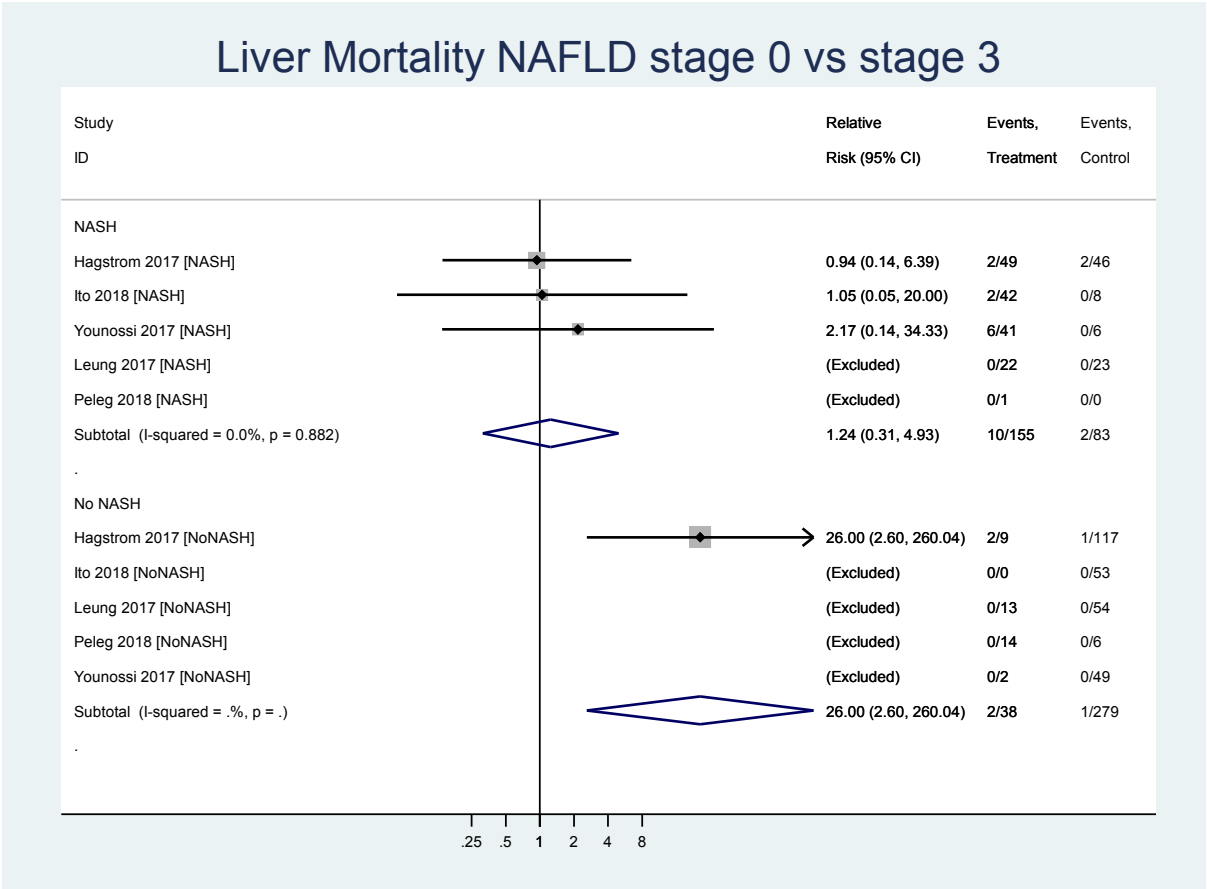
Stage 0 vs 1



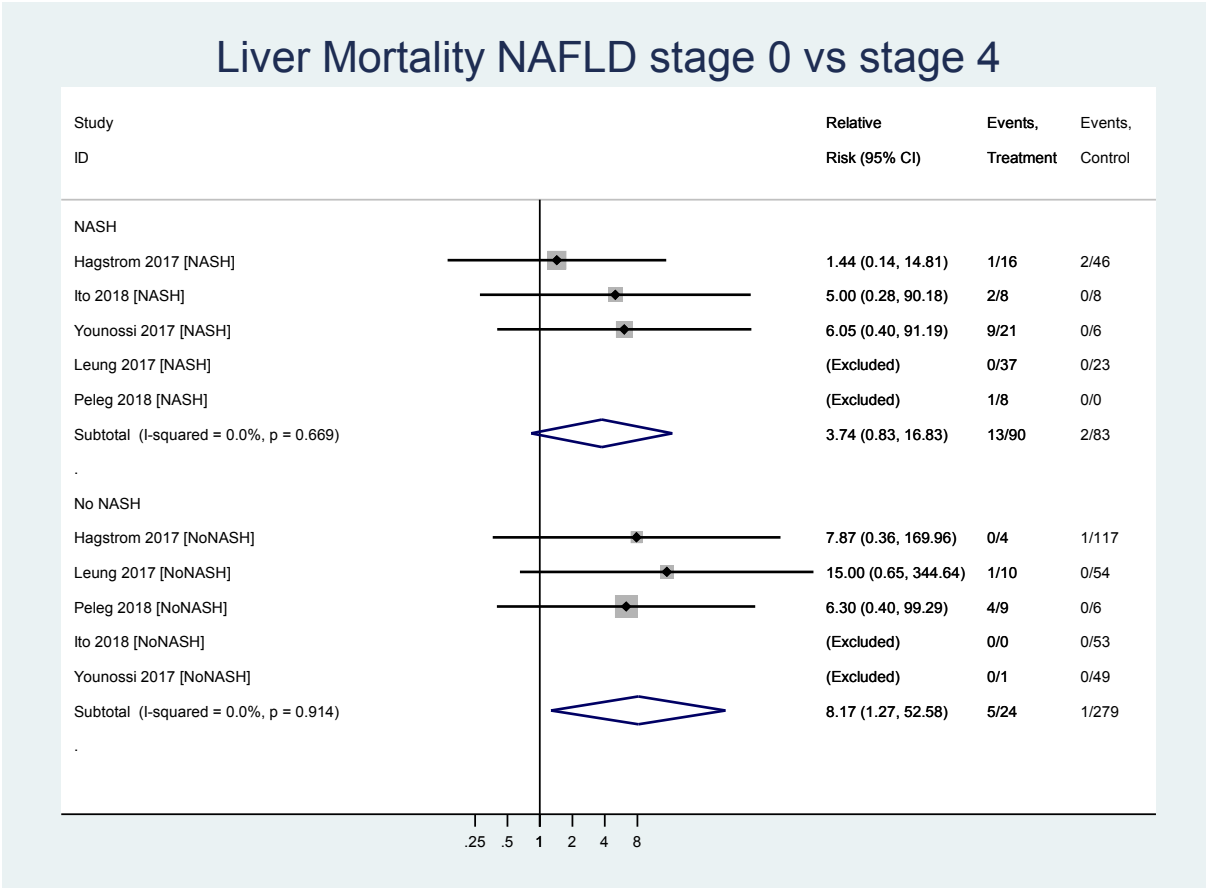
Stage 0 vs 2


Liver Mortality NAFLD stage 0 vs stage 2
Study
ID
Relative
Risk (95% CI)
Events,
Treatment
Events,
Control
NASH
Hagstrom 2017 [NASH] 0.79 (0.15, 4.15) 4/117 2/46
Ito 2018 [NASH] 0.71 (0.03, 16.04) 1/37 0/8
Younossi 2017 [NASH] 0.81 (0.04, 17.75) 1/25 0/6
Leung 2017 [NASH] (Excluded) 0/17 0/23
Peleg 2018 [NASH] (Excluded) 0/5 0/0
Subtotal (I-squared = 0.0%, p = 0.998) 0.78 (0.21, 2.92) 6/201 2/83
.
No NASH
Hagstrom 2017 [NoNASH] 7.31 (0.68, 78.10) 2/32 1/117
Ito 2018 [NoNASH] (Excluded) 0/1 0/53
Leung 2017 [NoNASH] (Excluded) 0/15 0/54
Peleg 2018 [NoNASH] (Excluded) 0/23 0/6
Younossi 2017 [NoNASH] (Excluded) 0/1 0/49
Subtotal (I-squared = .%, p = .) 7.31 (0.68, 78.10) 2/72 1/279
.
.25
.5
1
2
4
8

Stage 0 vs 3

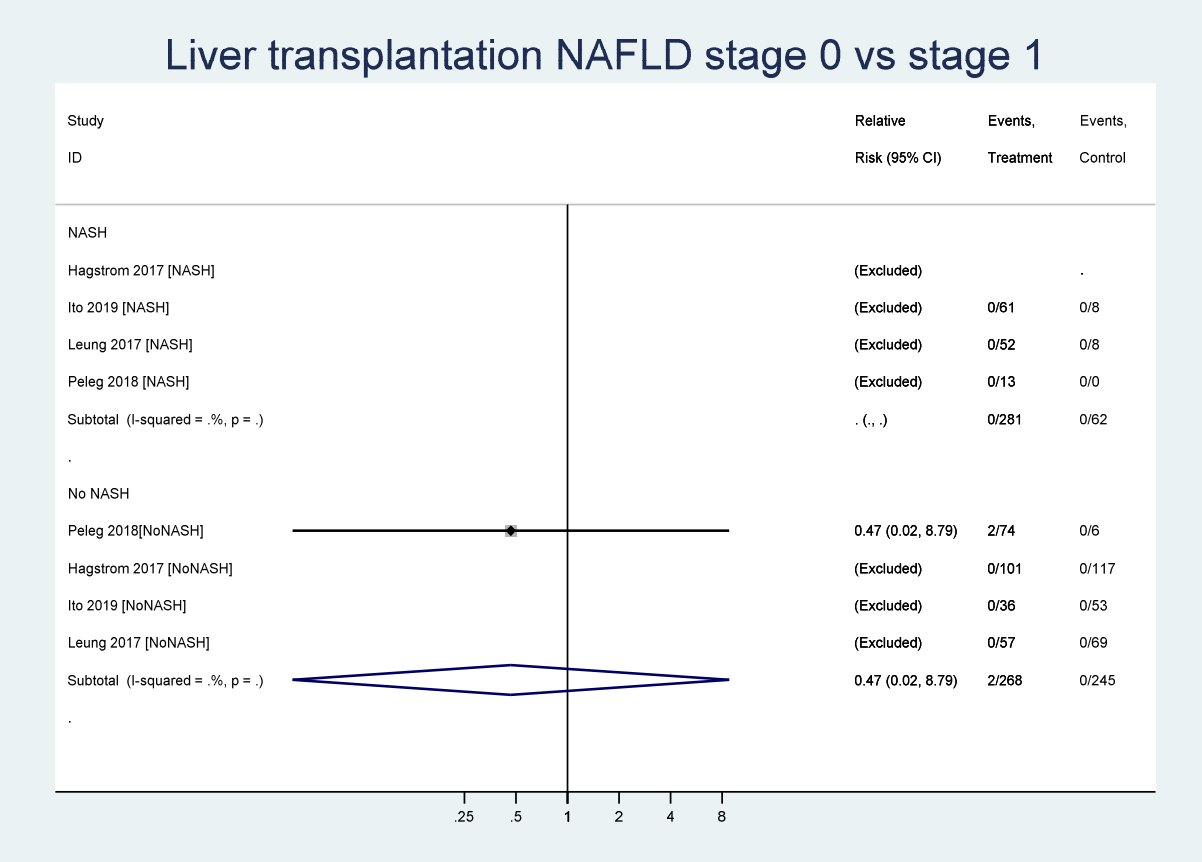

Liver Mortality NAFLD stage 0 vs stage 3
Study
ID
Relative
Risk (95% CI)
Events,
Treatment
Events,
Control
NASH
Hagstrom 2017 [NASH]
0.94 (0.14, 6.39)
2/49
2/46
Ito 2018 [NASH]
1.05 (0.05, 20.00)
2/42
0/8
Younossi 2017 [NASH]
2.17 (0.14, 34.33)
6/41
0/6
Leung 2017 [NASH]
(Excluded)
0/22
0/23
Peleg 2018 [NASH]
(Excluded)
0/1
0/0
Subtotal (I-squared = 0.0%, p = 0.882)
1.24 (0.31, 4.93)
10/155
2/83
.
No NASH
Hagstrom 2017 [NoNASH]
26.00 (2.60, 260.04)
2/9
1/117
Ito 2018 [NoNASH]
(Excluded)
0/0
0/53
Leung 2017 [NoNASH]
(Excluded)
0/13
0/54
Peleg 2018 [NoNASH]
(Excluded)
0/14
0/6
Younossi 2017 [NoNASH]
(Excluded)
0/2
0/49
Subtotal (I-squared = .%, p = .)
26.00 (2.60, 260.04)
2/38
1/279
.
.25
.5
1
2
4
8

Stage 0 vs 4

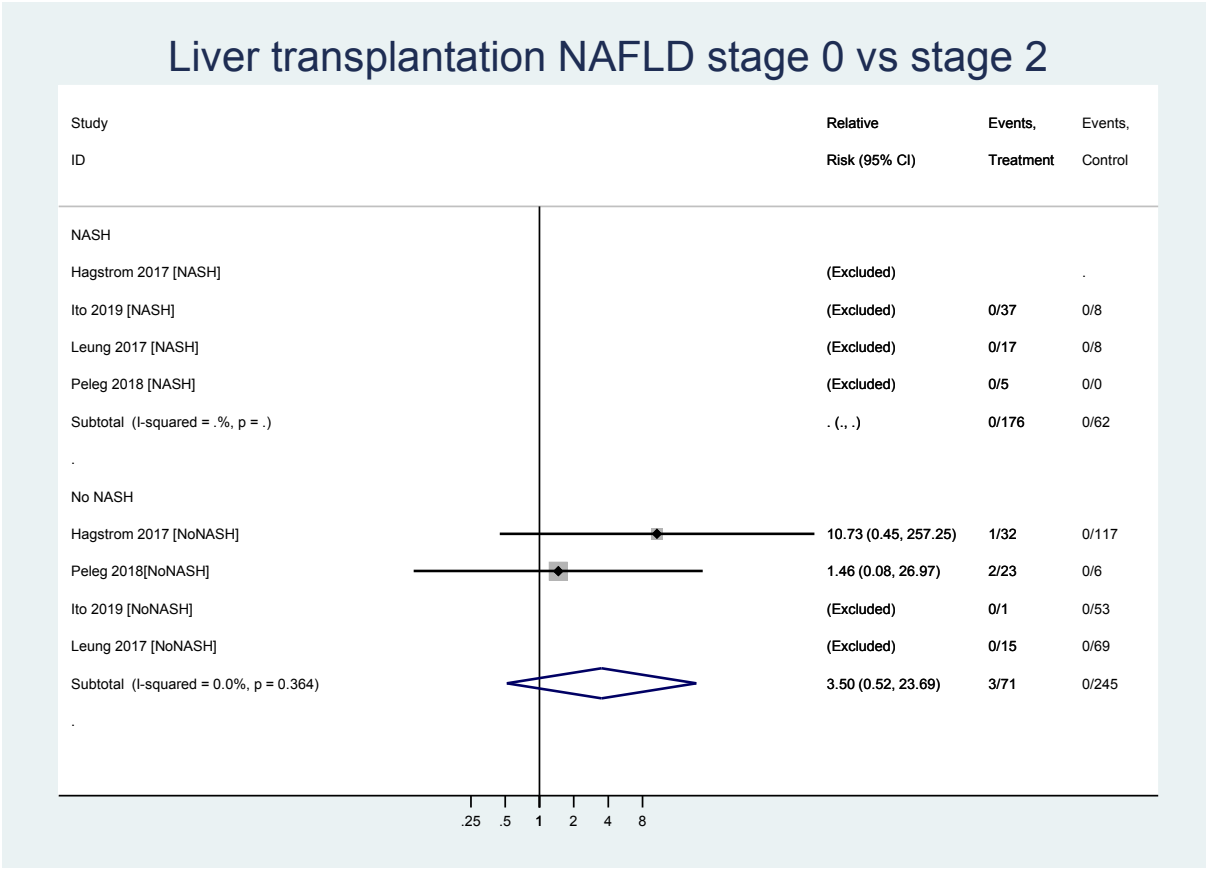

Liver Mortality NAFLD stage 0 vs stage 4
Study
ID
Relative
Risk (95% CI)
Events,
Treatment
Events,
Control
NASH
Hagstrom 2017 [NASH] 1.44 (0.14, 14.81) 1/16 2/46
Ito 2018 [NASH] 5.00 (0.28, 90.18) 2/8 0/8
Younossi 2017 [NASH] 6.05 (0.40, 91.19) 9/21 0/6
Leung 2017 [NASH] (Excluded) 0/37 0/23
Peleg 2018 [NASH] (Excluded) 1/8 0/0
Subtotal (I-squared = 0.0%, p = 0.669) 3.74 (0.83, 16.83) 13/90 2/83
.
No NASH
Hagstrom 2017 [NoNASH] 7.87 (0.36, 169.96) 0/4 1/117
Leung 2017 [NoNASH] 15.00 (0.65, 344.64) 1/10 0/54
Peleg 2018 [NoNASH] 6.30 (0.40, 99.29) 4/9 0/6
Ito 2018 [NoNASH] (Excluded) 0/0 0/53
Younossi 2017 [NoNASH] (Excluded) 0/1 0/49
Subtotal (I-squared = 0.0%, p = 0.914) 8.17 (1.27, 52.58) 5/24 1/279
.
.25 .5 1 2 4 8

Liver transplant

Stage 0 vs 1



Stage 0 vs 2

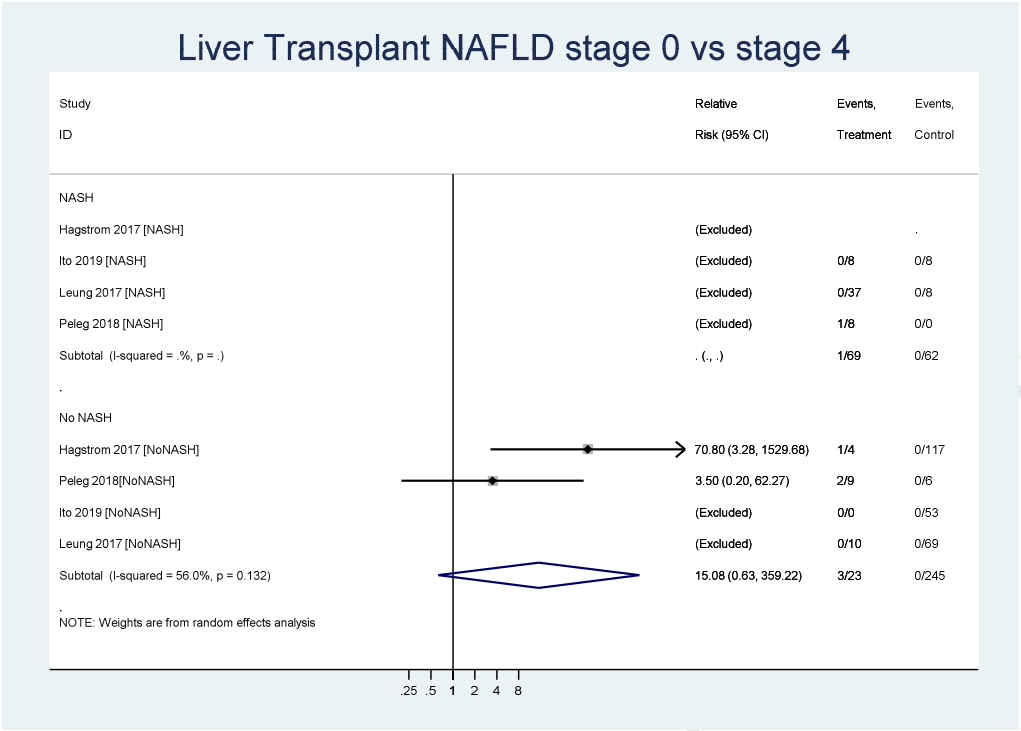

Liver transplantation NAFLD stage 0 vs stage 2
Study
ID
Relative
Risk (95% CI)
Events,
Treatment
Events,
Control
NASH
Hagstrom 2017 [NASH] (Excluded) .
Ito 2019 [NASH] (Excluded) 0/37 0/8
Leung 2017 [NASH] (Excluded) 0/17 0/8
Peleg 2018 [NASH] (Excluded) 0/5 0/0
Subtotal (I-squared = .%, p = .) . (., .) 0/176 0/62
.
No NASH
Hagstrom 2017 [NoNASH] 10.73 (0.45, 257.25) 1/32 0/117
Peleg 2018[NoNASH] 1.46 (0.08, 26.97) 2/23 0/6
Ito 2019 [NoNASH] (Excluded) 0/1 0/53
Leung 2017 [NoNASH] (Excluded) 0/15 0/69
Subtotal (I-squared = 0.0%, p = 0.364) 3.50 (0.52, 23.69) 3/71 0/245
.
.25 .5 1 2 4 8

Stage 0 vs 3

. metan n_3 n_3_no n_0 n_0_no, label(namevar= var1) nowt xlabel(0.25,0.5,1,2,4,8) counts texts(140) effect (Relative Risk) title (Liver transplant NAFLD stage 0 vs stage 2) by(NASH)

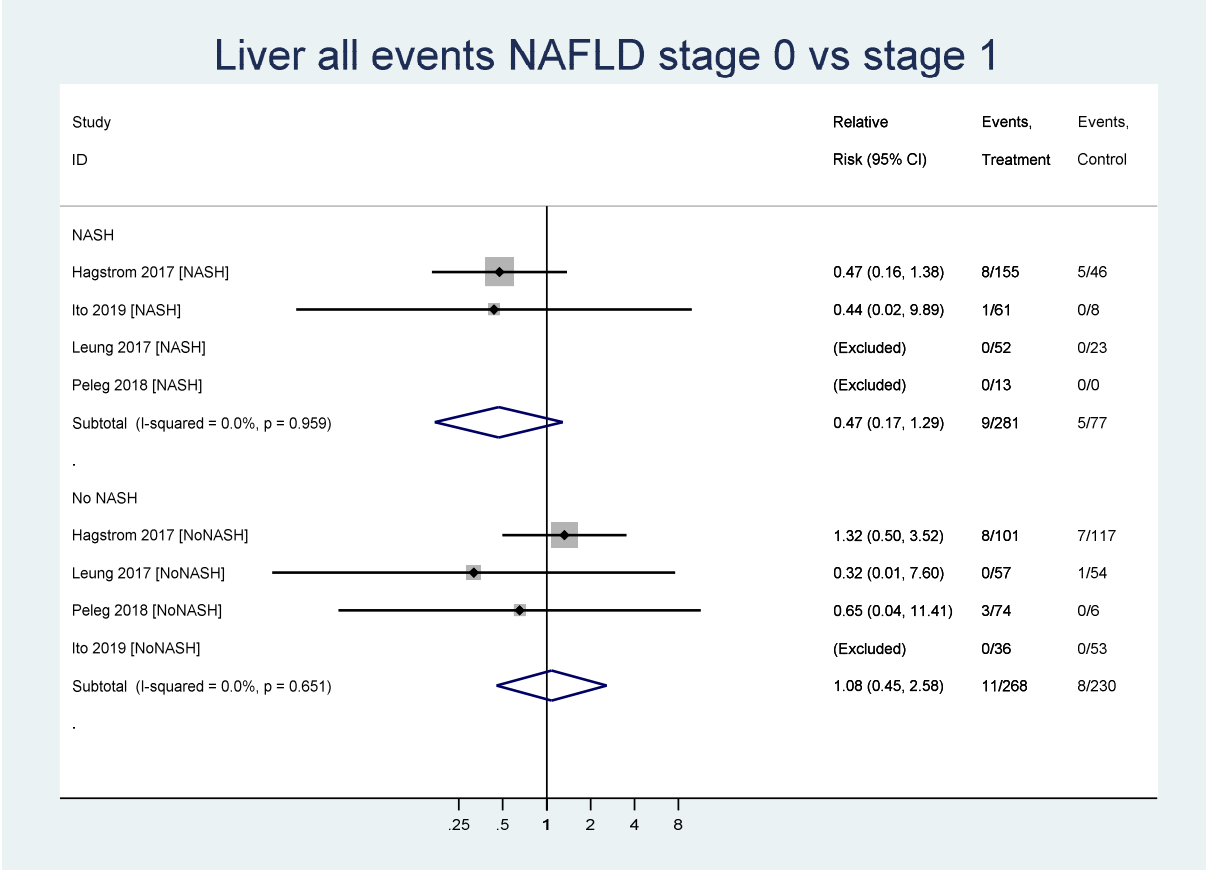
Insufficient data to perform this meta-analysis

Stage 0 vs 4

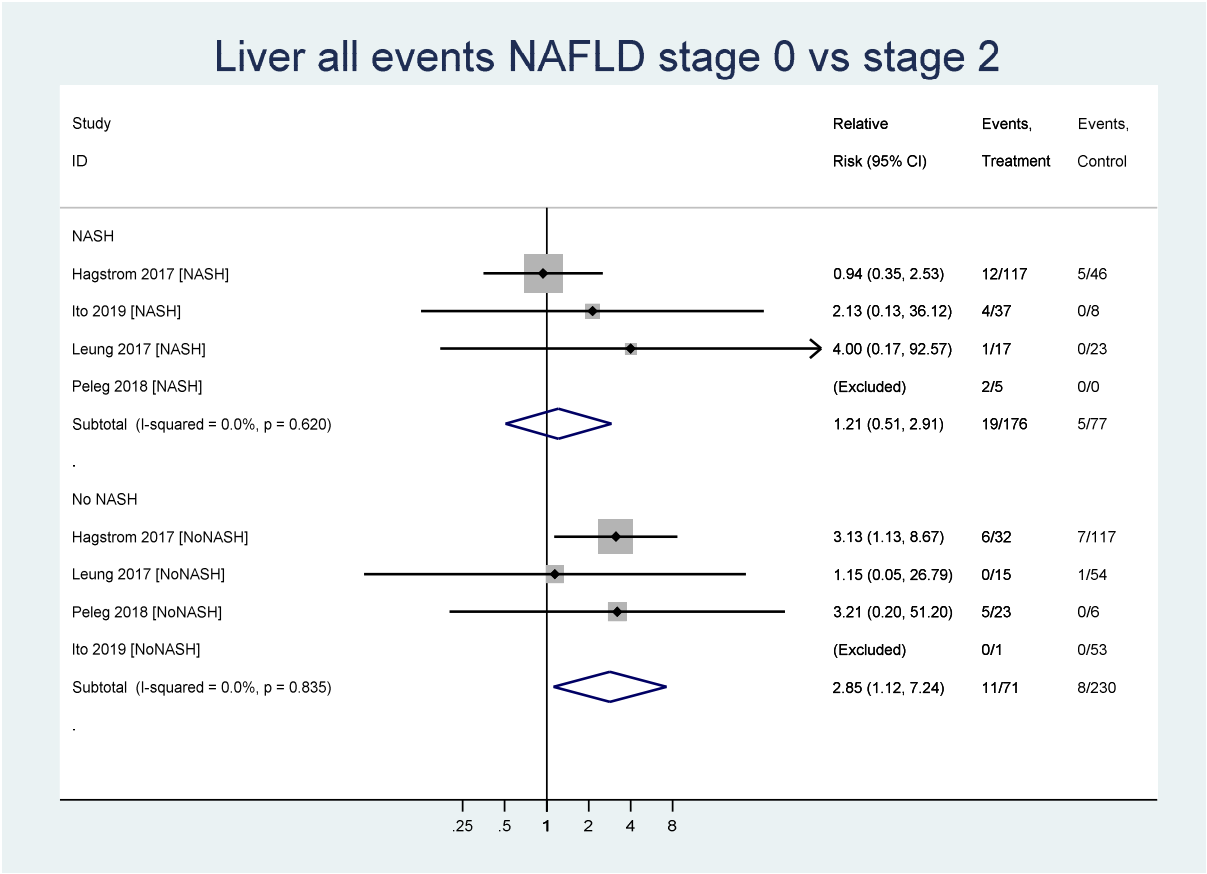

Liver Transplant NAFLD stage 0 vs stage 4
Study ID
Relative Risk (95% CI)
Events, Treatment
Events, Control
NASH
Hagstrom 2017 [NASH] (Excluded) .
Ito 2019 [NASH] (Excluded) 0/8 0/8
Leung 2017 [NASH] (Excluded) 0/37 0/8
Peleg 2018 [NASH] (Excluded) 1/8 0/0
Subtotal (I-squared = .%, p = .) . (., .) 1/69 0/62
No NASH
Hagstrom 2017 [NoNASH] 70.80 (3.28, 1529.68) 1/4 0/117
Peleg 2018[NoNASH] 3.50 (0.20, 62.27) 2/9 0/6
Ito 2019 [NoNASH] (Excluded) 0/0 0/53
Leung 2017 [NoNASH] (Excluded) 0/10 0/69
Subtotal (I-squared = 56.0%, p = 0.132) 15.08 (0.63, 359.22) 3/23 0/245
NOTE: Weights are from random effects analysis
.25 .5 1 2 4 8

All liver events

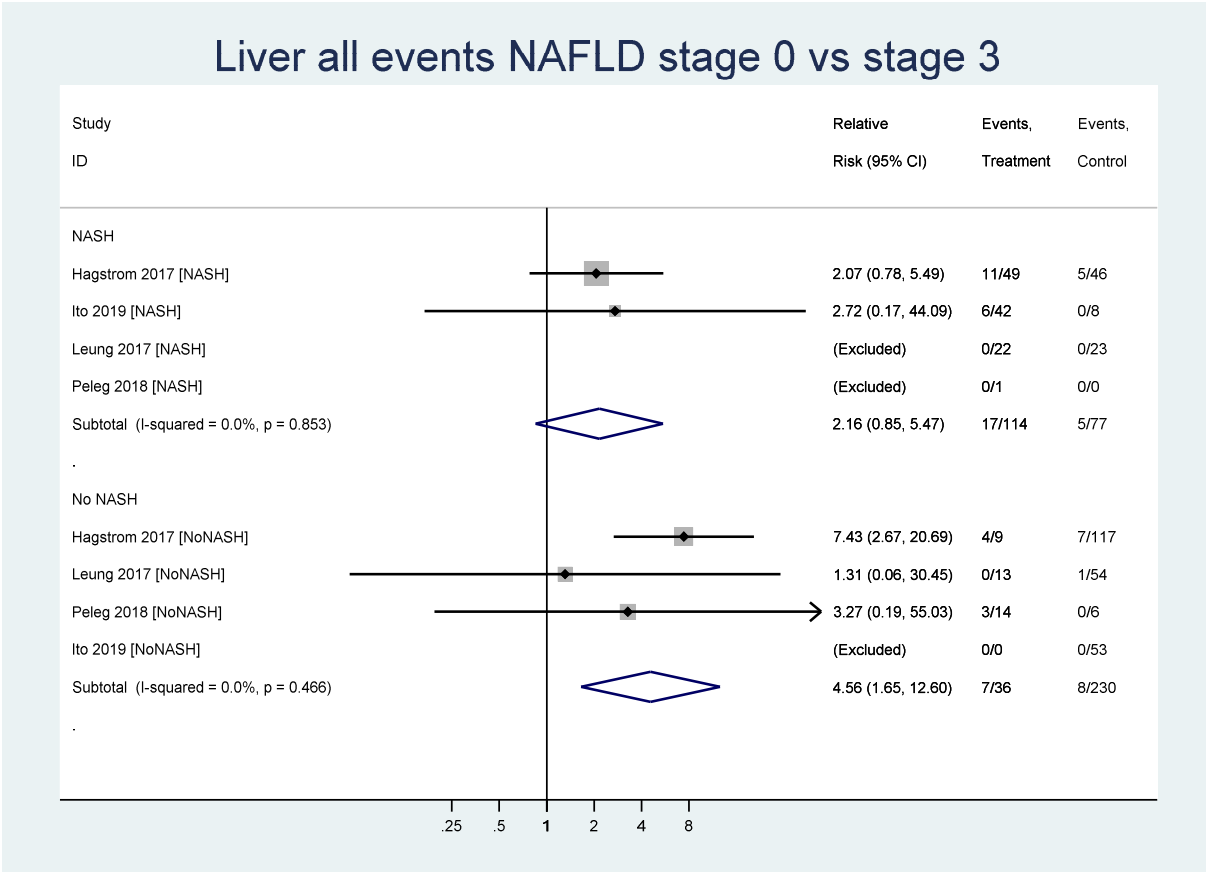
Stage 0 vs 1



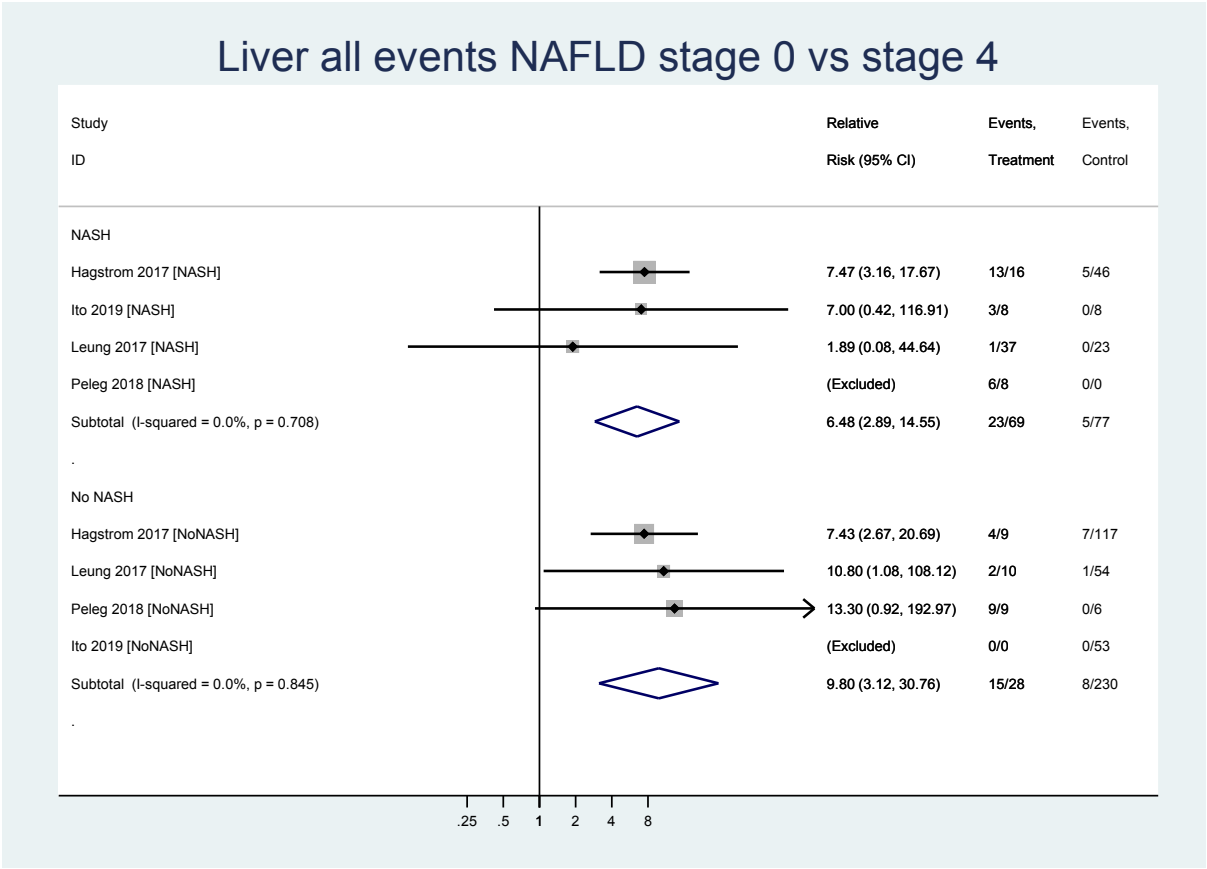
Stage 0 vs 2


Liver all events NAFLD stage 0 vs stage 2
Study
ID
Relative
Risk (95% CI)
Events,
Treatment
Events,
Control
NASH
Hagstrom 2017 [NASH]
0.94 (0.35, 2.53)
12/117
5/46
Ito 2019 [NASH]
2.13 (0.13, 36.12)
4/37
0/8
Leung 2017 [NASH]
4.00 (0.17, 92.57)
1/17
0/23
Peleg 2018 [NASH]
(Excluded)
2/5
0/0
Subtotal (I-squared = 0.0%, p = 0.620)
1.21 (0.51, 2.91)
19/176
5/77
.
No NASH
Hagstrom 2017 [NoNASH]
3.13 (1.13, 8.67)
6/32
7/117
Leung 2017 [NoNASH]
1.15 (0.05, 26.79)
0/15
1/54
Peleg 2018 [NoNASH]
3.21 (0.20, 51.20)
5/23
0/6
Ito 2019 [NoNASH]
(Excluded)
0/1
0/53
Subtotal (I-squared = 0.0%, p = 0.835)
2.85 (1.12, 7.24)
11/71
8/230
.
.25
.5
1
2
4
8

Stage 0 vs 3


Liver all events NAFLD stage 0 vs stage 3
Study
ID
Relative
Risk (95% CI)
Events,
Treatment
Events,
Control
NASH
Hagstrom 2017 [NASH] 2.07 (0.78, 5.49) 11/49 5/46
Ito 2019 [NASH] 2.72 (0.17, 44.09) 6/42 0/8
Leung 2017 [NASH] (Excluded) 0/22 0/23
Peleg 2018 [NASH] (Excluded) 0/1 0/0
Subtotal (I-squared = 0.0%, p = 0.853) 2.16 (0.85, 5.47) 17/114 5/77
.
No NASH
Hagstrom 2017 [NoNASH] 7.43 (2.67, 20.69) 4/9 7/117
Leung 2017 [NoNASH] 1.31 (0.06, 30.45) 0/13 1/54
Peleg 2018 [NoNASH] 3.27 (0.19, 55.03) 3/14 0/6
Ito 2019 [NoNASH] (Excluded) 0/0 0/53
Subtotal (I-squared = 0.0%, p = 0.466) 4.56 (1.65, 12.60) 7/36 8/230
.
.25
.5
1
2
4
8

Stage 0 vs 4


Liver all events NAFLD stage 0 vs stage 4
Study
ID
Relative
Risk (95% CI)
Events,
Treatment
Events,
Control
NASH
Hagstrom 2017 [NASH]
7.47 (3.16, 17.67)
13/16
5/46
Ito 2019 [NASH]
7.00 (0.42, 116.91)
3/8
0/8
Leung 2017 [NASH]
1.89 (0.08, 44.64)
1/37
0/23
Peleg 2018 [NASH]
(Excluded)
6/8
0/0
Subtotal (I-squared = 0.0%, p = 0.708)
6.48 (2.89, 14.55)
23/69
5/77
No NASH
Hagstrom 2017 [NoNASH]
7.43 (2.67, 20.69)
4/9
7/117
Leung 2017 [NoNASH]
10.80 (1.08, 108.12)
2/10
1/54
Peleg 2018 [NoNASH]
13.30 (0.92, 192.97)
9/9
0/6
Ito 2019 [NoNASH]
(Excluded)
0/0
0/53
Subtotal (I-squared = 0.0%, p = 0.845)
9.80 (3.12, 30.76)
15/28
8/230
.25
.5
1
2
4
8

a-Appendix 6. HRQoL by fibrosis stage across individual studies

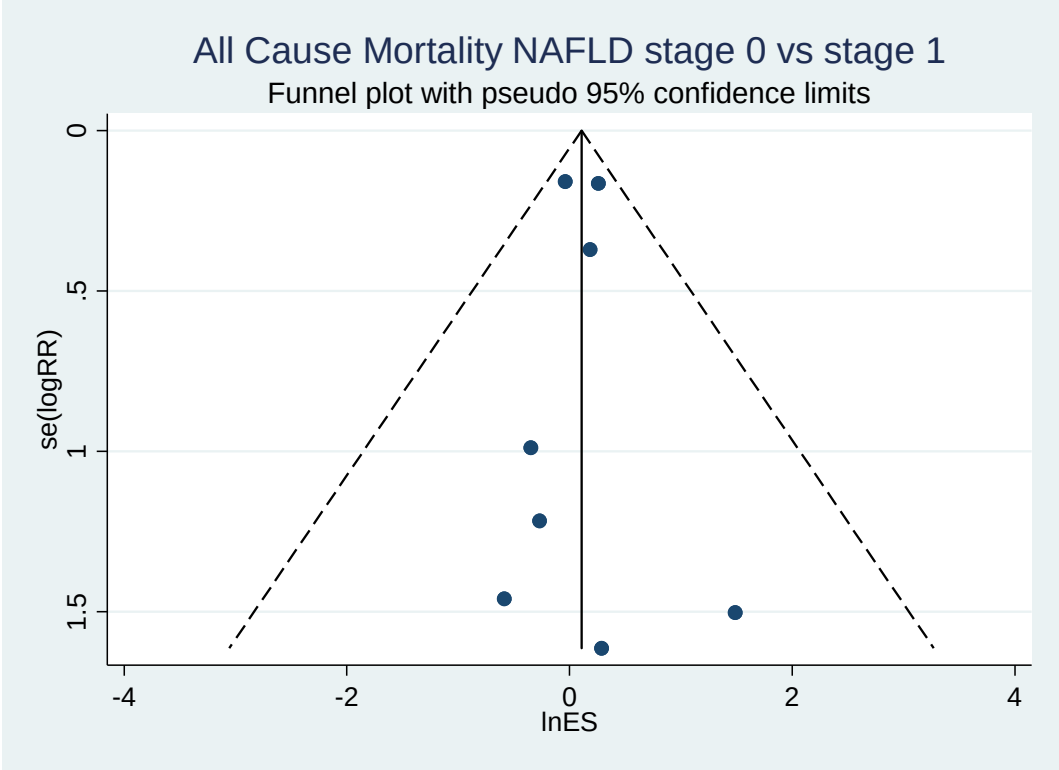
	Stage 0	**Stage 1**	**Stage 2**	**Stage 3**	**Stage 4**
David (2009)	N, median (IQR)	N, median (IQR)	N, median (IQR)	N, median (IQR)	N, median (IQR)
SF-36 PCS	167, 50 (42.5, 56)	211, 50 (39, 54)	138, 47 (36, 54)	131, 48 (37, 53)	66, 37 (31, 48)
SF-36 MCS	NR	NR	NR	NR	NR
Huber (2018)	N, mean (SD)	N, mean (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)	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)	-

CLDQ: Chronic Liver Disease Questionnaire; MCS: Mental Component Score NR: not reported; PCS: Physical Component Score; SD: standard deviation; SF-36: Short Form-36

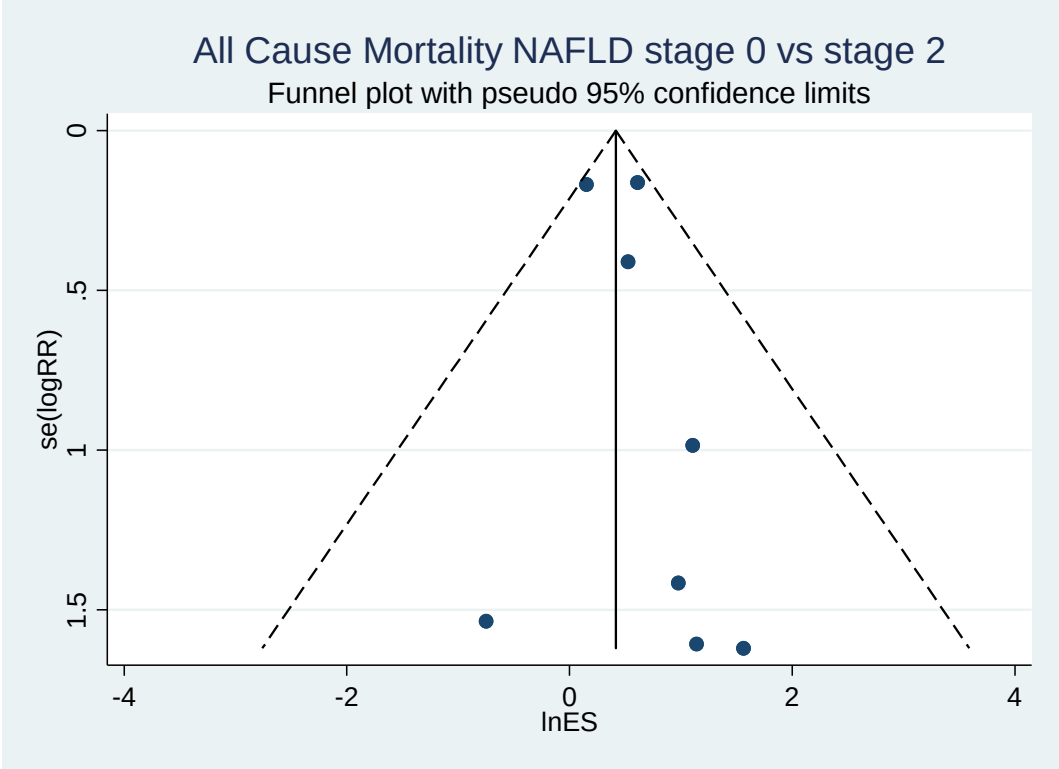
e-Appendix 7. Univariate meta-regression analysis

	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	P = 0.774	P = 0.47	P = 0.48	P = 0.20
High vs. moderate risk of bias	P = 0.92	P = 0.67	P = 0.67	P = 0.46
Duration of follow up (years)	P = 0.36	P = 0.11	P = 0.46	P = 0.12
All-liver related events				
Retrospective vs. prospective study design	P = 0.41	P = 0.97	P = 0.28	P = 0.47
High vs. moderate risk of bias	P = 0.670	P = 0.80	P = 0.73	P = 0.98
Duration of follow up (years)	P = 0.61	P = 0.13	P = 0.28	P = 0.51

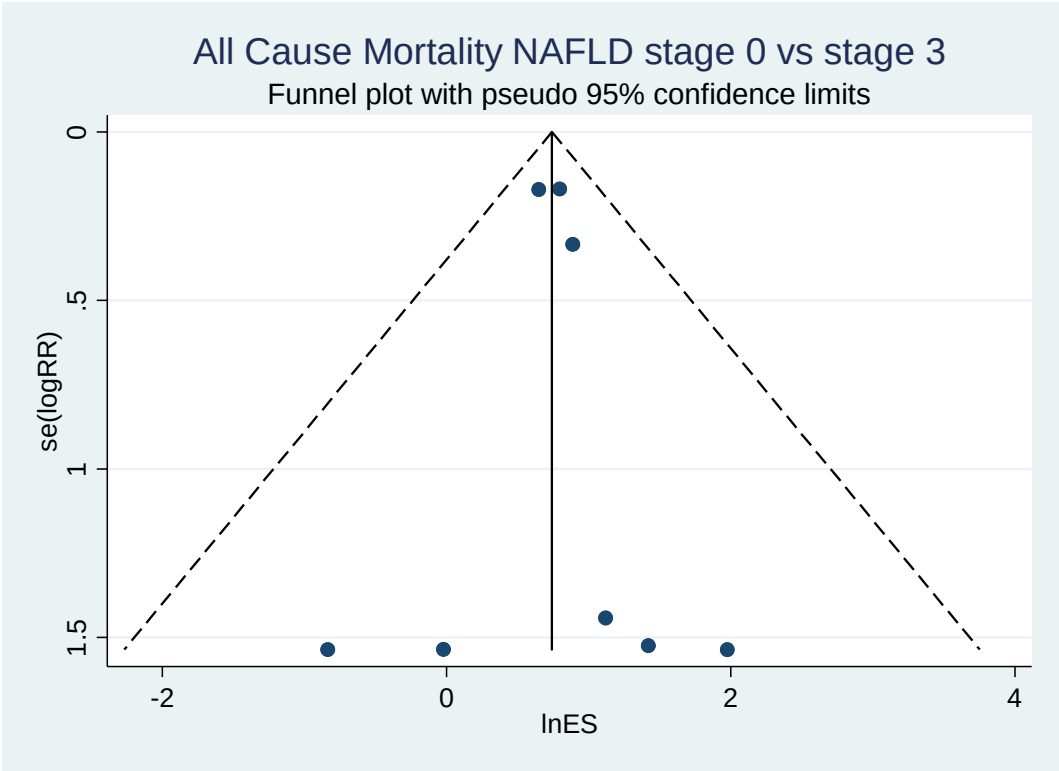
eAppendix 8. - Assessment of small study bias



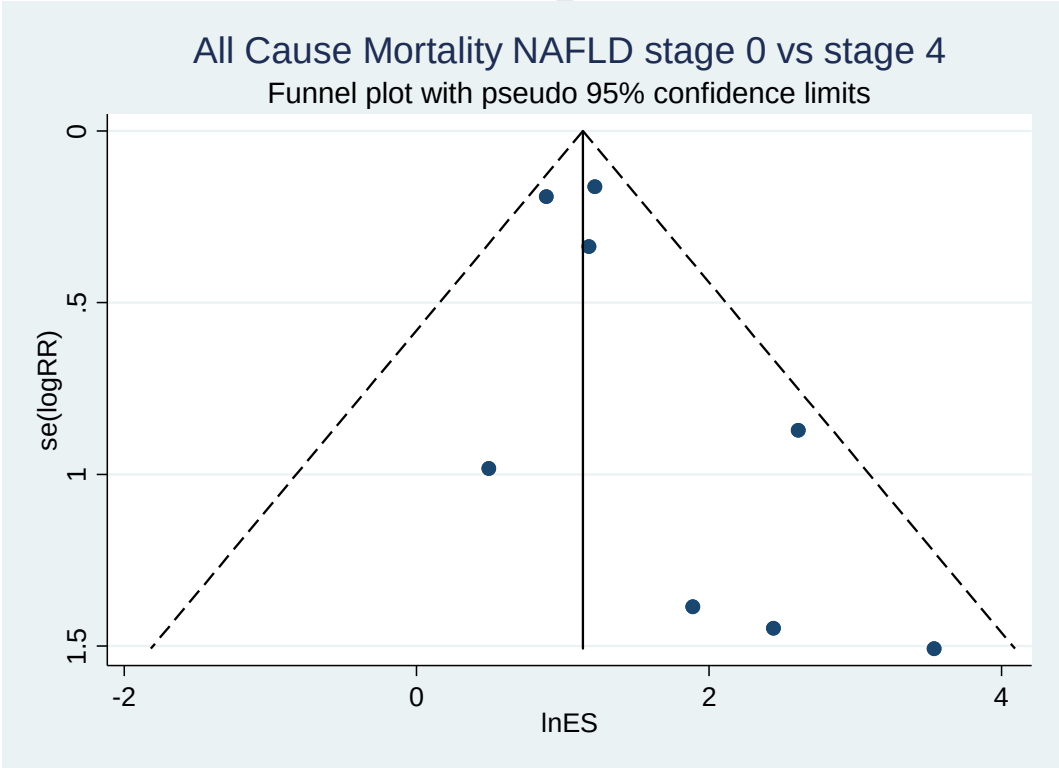
Egger test p-value=0.996



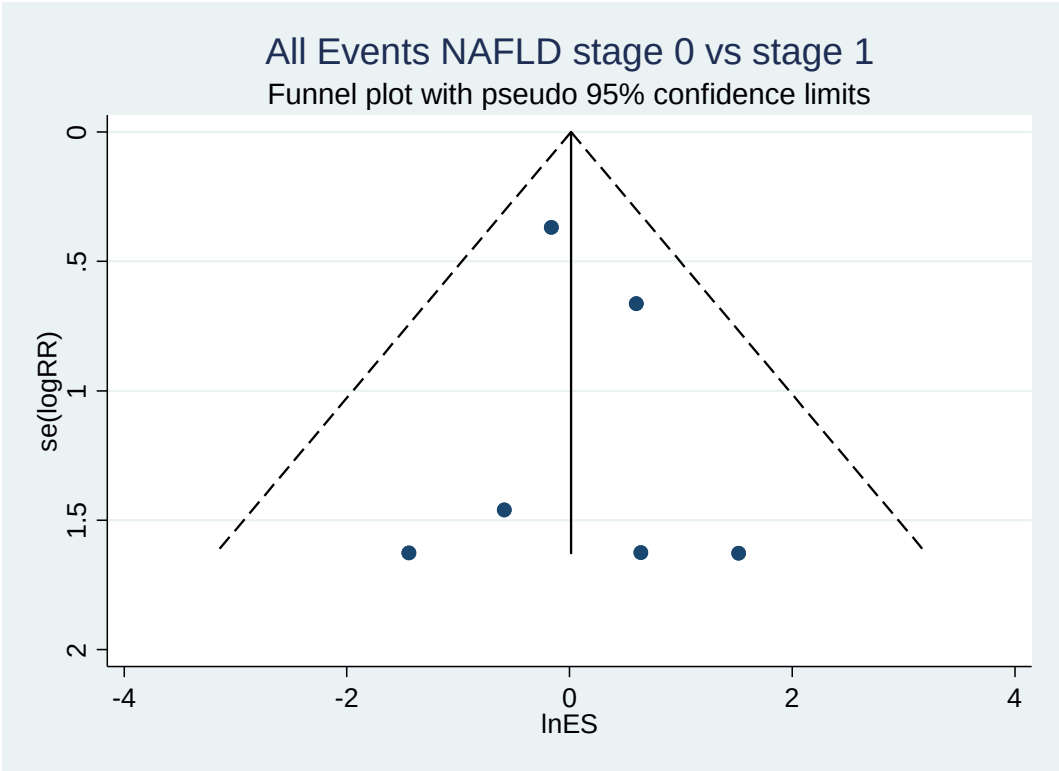
Egger test P-value=0.485



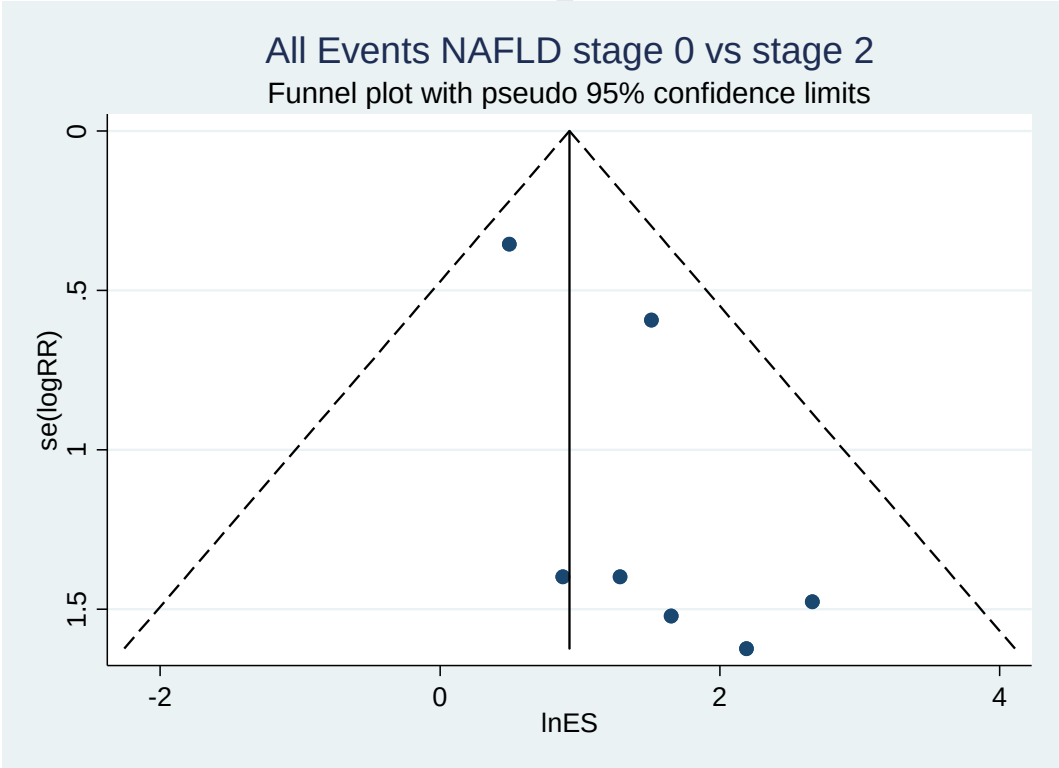
Egger test P-value=0.89



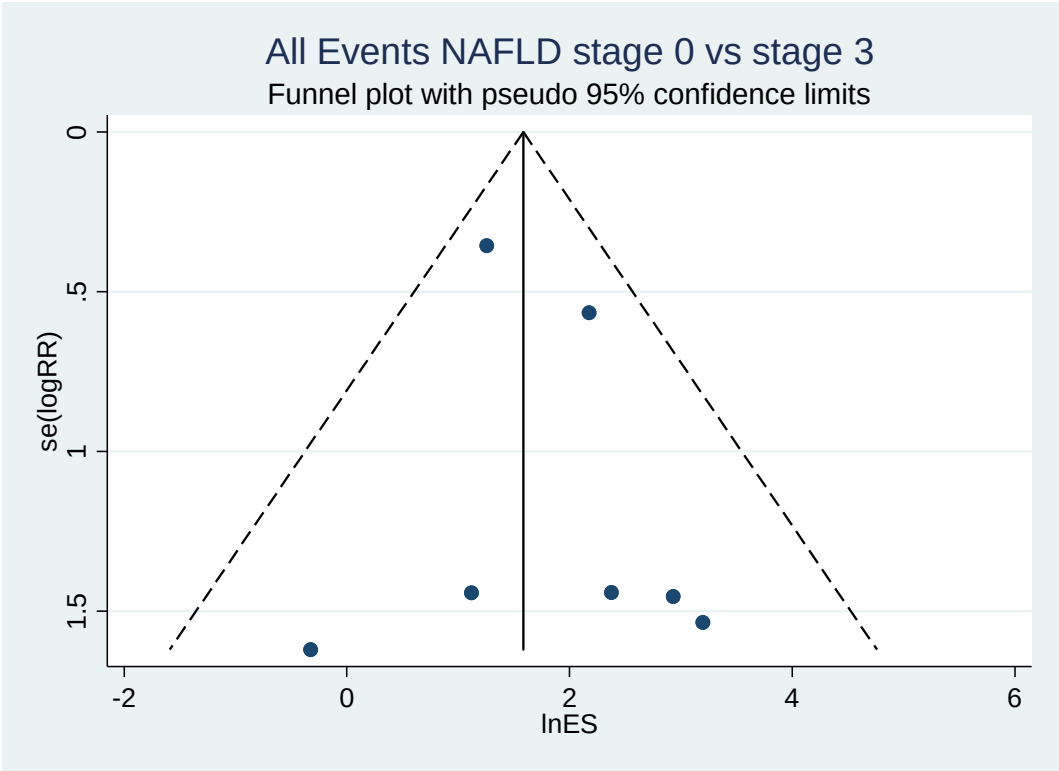
Egger test P-value=0.11



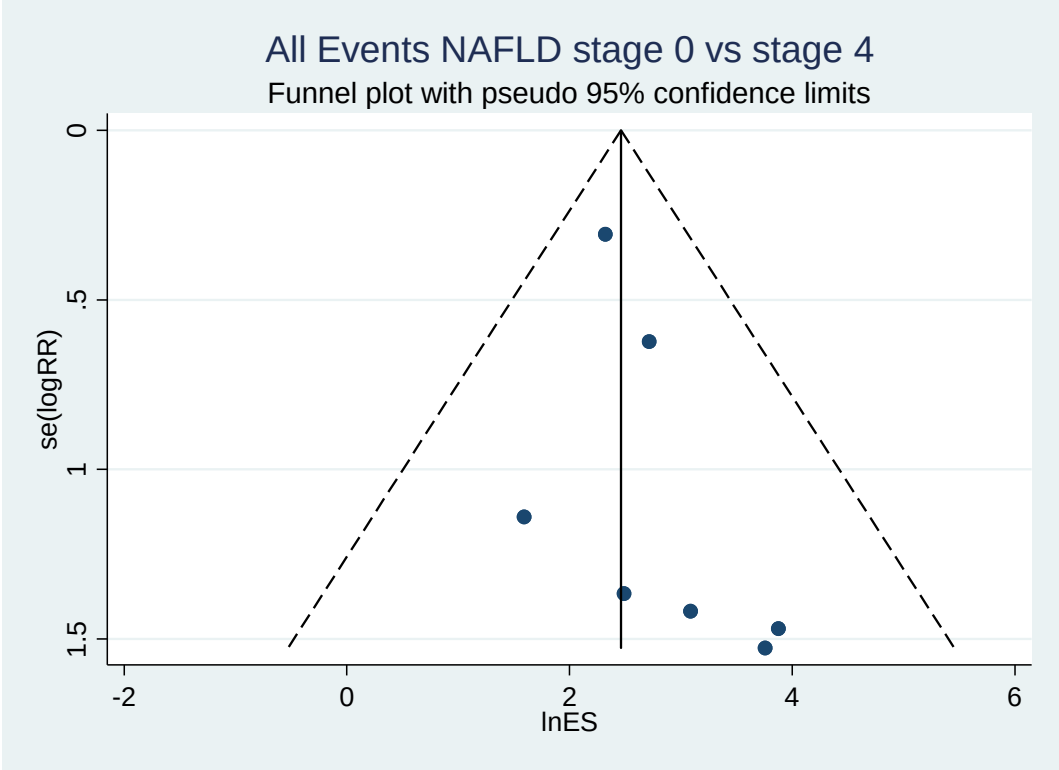
Egger test P-value=0.75



Egger test P-value=0.05



Egger test P-value=0.49



Egger test P-value=0.20

e-Appendix 9. – GRADE assessment of quality of evidence for liver fibrosis as prognostic marker for NAFLD

	Study design	Risk of bias	Indirectness	Imprecision	Additional considerations[5]	Quality
All-cause mortality	Observational	Serious[1]	Not serious[2]	Not serious[3]	large effect; no publication bias; confounder adjusted	**High**
Liver-related mortality	Observational	Serious[1]	Not serious[2]	Serious[4]	large effect; no publication bias; confounder adjusted	**Moderate**
Liver transplantation	Observational	Serious[1]	Not serious[2]	Serious[4]	Large effect; no publication bias; confounder adjusted	**Low**
All-liver events	Observational	Serious[1]	Serious[6]	Not serious[3]	Large effect; no publication bias; confounder adjusted	**Low**
HRQoL	Observational	Serious[1]	Not serious[2]	Serious[7]	Large effect; no publication bias; confounder adjusted	**Low**

1. risk of bias for individual studies judged to be moderate or high (see Table 2); 2. appropriate population and outcomes; 3. sufficient number of events; 4. insufficient number of events (especially for stage 0 fibrosis); 5. positive considerations can allow upgrading of GRADE rating; 6. Inconsistent definition of composite outcome of all liver events across studies; 7. Small number of studies; 8: From Iorio et al (2014): High: very confident that the true prognosis (probability of future events) lies close to that of the estimate; Moderate: 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: confidence in the estimate is limited: the true prognosis (probability of

BACKGROUND AND CONTEXT: The stage (or extent) of liver fibrosis, confirmed by biopsy, is believed to be 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.

Lay Summary: In an analysis of data from 13 previously published studies, this study found stage of fibrosis, determined by biopsy analysis, to be associated with mortality and morbidities in patients with NAFLD.