

Systematic Review

Parker, Richard; Hodson, James; Rowe, Ian A C

DOI:
[10.1111/jgh.13625](https://doi.org/10.1111/jgh.13625)

Document Version
Peer reviewed version

Citation for published version (Harvard):
Parker, R, Hodson, J & Rowe, IAC 2016, 'Systematic Review: Current Evidence in Non-Alcoholic Fatty Liver Disease Lacks Relevance to Patients with Advanced Fibrosis', *Journal of Gastroenterology and Hepatology*.
<https://doi.org/10.1111/jgh.13625>

[Link to publication on Research at Birmingham portal](#)

Publisher Rights Statement:
Checked for eligibility: 04/11/2016.

This is the peer reviewed version of the following article: Parker, R., Hodson, J., and Rowe, I. A. C. (2016) Systematic Review: Current Evidence in Non-Alcoholic Fatty Liver Disease Lacks Relevance to Patients with Advanced Fibrosis. *Journal of Gastroenterology and Hepatology*, which has been published in final form at [10.1111/jgh.13625](https://doi.org/10.1111/jgh.13625). This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Systematic Review: Current Evidence in Non-Alcoholic Fatty Liver Disease Lacks Relevance to Patients with Advanced Fibrosis

Richard Parker¹

James Hodson²

Ian AC Rowe³

1. Centre for Liver Research, University of Birmingham United Kingdom
2. Statistics Department, University Hospitals Birmingham NHS Foundation Trust, Birmingham UK
3. Leeds Institute for Data Analysis, University of Leeds, UK

Corresponding author:

Dr Richard Parker

Centre for Liver Research

5th Floor IBR

University of Birmingham

Birmingham B15 2TT

United Kingdom

richardparker@nhs.net

(+44) (0) 7971 118036

Disclosure

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/jgh.13625

None of the authors have any conflicts of interest to declare

This work was supported by the European Association for the Study of the Liver (EASL) through a Sheila Sherlock Physician Scientist award to Dr Parker. There is no specific grant number attached to this award.

Word count:

Abstract: 238

Manuscript: 2588

Tables: 1

Figures: 4

Supplementary material: 5 tables, 4 figures

Author contributions

RP conceived the idea for this study, performed literature searches, data extraction, analysis and wrote the draft manuscript.

JH assisted with data analysis and reviewed the draft manuscript.

IAR conceived the idea for this study, data extraction and reviewed the draft manuscript.

All authors approved the final manuscript.

Abstract

Objectives

Epidemiological data have shown that individuals with advanced fibrosis are at greatest risk of premature morbidity in NAFLD. Individuals included in clinical trials are often highly selected to remove confounding factors but selection can introduce bias and limit external validity. We examined the external validity of trials in non-alcoholic fatty liver disease (NAFLD) by examining characteristics of participants in observational studies (OS) and randomised controlled trials (RCT) in NAFLD.

Design

A systematic review was performed with structured literature searches for relevant OS and RCT using PubMed and Ovid Embase (1948 - 2016). Identified studies were screened for inclusion by the authors and data extracted. Study populations were compared using t-tests to compare means and variances, in each case weighted by the size of individual studies. Dichotomous data were compared by Chi-squared test.

Results

In total 148 studies were included: 67 RCT and 81 OS including data from 44,860 individuals. Fifteen RCT participants differed from individuals in OS with regard to age, BMI, prevalence of DM, and gender ($p < 0.001$ in each case). The most pronounced differences were seen between RCT participants and patients with advanced fibrosis. Co-morbid conditions prevalent amongst individuals with NAFLD were frequent exclusion criteria in RCT.

Conclusions

The characteristics of participants in randomised controlled trials differ to those of the wider population of individuals with NAFLD. These differences may reduce the utility of trial data to individuals with NAFLD at greatest risk of death.

Key words: Non-alcoholic fatty liver disease, Non-alcoholic steatohepatitis, Systematic review

Non-alcoholic fatty liver disease (NAFLD) is a spectrum of disease, ranging from simple hepatic steatosis through non-alcoholic steatohepatitis (NASH), to cirrhosis (1). NAFLD is common: hepatic steatosis is present in up to one-third of individuals (2) and NASH is seen in approximately 10% (3). Whilst the association of NAFLD *per se* with mortality is debated (4), large studies have shown that it is the subset of patients with advanced fibrosis who have an increased risk of liver-related morbidity as well as cardiovascular and neoplastic disease (5, 6). Given the high prevalence of NAFLD there is a need for effective therapies to prevent progression of disease and to treat established fibrosis. However, whilst many treatments have been trialed, few have shown categorical benefit in NASH (7).

The external validity of a trial describes its relevance to a population outside of the trial's participants (8). Systematic analyses of trials in cardiology (9) and respiratory medicine (10, 11) have demonstrated have poor external validity, where commonly occurring medical conditions and age often exclude individuals from participation in trials (12). Poor external validity may promote ineffective treatments or, conversely, limit the acceptance of effective treatments (8). We used systematic review methods to assess the characteristics of participants included in randomised controlled trials (RCT) for treatment of NAFLD and the characteristics of individuals described in observational studies (OS) of NAFLD. This allowed comparison of the two groups and therefore an estimate of the external validity of RCTs performed in NAFLD to date.

Methodology

Literature search

Literature searching was undertaken using three search strategies, to identify OS for NAFLD, OS reporting advanced fibrosis, and to identify randomised controlled trials in NAFLD. PubMed/MedLine and Ovid Embase were searched (1948 - 2016), using the search terms: (((randomised controlled trial) AND non alcoholic steatohepatitis) OR

NASH) OR NAFLD) OR non-alcoholic fatty liver disease) for RCT, and (((Prevalence[Title] OR Natural history[Title])) AND (((non-alcoholic fatty liver disease) OR NASH) OR NAFLD) OR non-alcoholic steatohepatitis) and (fibrosis OR histology) AND (NASH OR non-alcoholic steatohepatitis) for OS. Results were limited to human studies and those published in English. The literature search was performed 1/7/2014 and updated on 11/8/2016. To examine 'grey literature' for suitable studies the reference lists of included studies were searched for other suitable studies, and papers citing included studies were also reviewed. The title and abstract of papers found by the literature search were reviewed and unsuitable manuscripts or duplicate results excluded. Remaining papers were reviewed independently by two authors (RP, IAC) and disagreements resolved by consensus. This report is the only published account of this protocol.

Inclusion and exclusion criteria

Papers were included if they were full papers describing RCTs of interventions (lifestyle or pharmacological) in adult patients with any stage of NAFLD, or full papers reporting OS of prevalence or natural history of NAFLD in adults. Instances where the same cohort was described in more than one study were identified and included only once, with the study containing the most data included or the most recent report if descriptions were similar. Studies that had been published as abstracts, and those published in languages other than English were excluded. OS that only included particular groups, for example, studies reporting the prevalence of NAFLD amongst patients undergoing bariatric surgery, were also excluded.

Data extraction

Papers were reviewed and data extracted into a pre-prepared spreadsheet. Mean values for characteristics of trial participants in RCTs and individuals in OS were noted, specifically age, BMI, gender (% of male participants), prevalence of Diabetes mellitus (DM). For RCTs, type of intervention, primary outcome, secondary outcome and exclusion criteria were

recorded, as well as whether the trial reported a positive or negative finding. In the first instance, all OS were included in analysis of NAFLD as a whole, encompassing all stages of disease. Subsequent analyses were undertaken for biopsy-proven NASH and advanced fibrosis, including only OS that used and reported biopsy findings. For measurement of fibrosis in studies including biopsies, the Kleiner/Brunt classification was often used where F3 or F4 was taken to indicate advanced fibrosis, as is usual in the field. In studies that did not use the Kleiner/Brunt classification, advanced fibrosis was considered as bridging fibrosis or cirrhosis. The quality of included studies was assessed using the CONSORT guidelines for RCT (13) and the STROBE guidelines for observational studies (14).

Data synthesis and analysis

Many papers reported mean values for each arm of a trial but not an overall mean. Where this occurred, overall weighted means and variances were produced for the study as a whole, using the formulae $(\sum_1^i n_i \bar{x}_i) / (\sum_1^i n_i)$ and $(\sum_1^i [n_i - 1] s_i^2) / (\sum_1^i [n_i - 1])$ respectively. Once every study was represented by a single mean and variance, they were then split by study type (RCT or OS), and overall pooled means and variances calculated using the same approach.

Comparisons were initially made between the study types with Kruskal-Wallis tests to produce an unweighted comparison of reported means or prevalences between the study types. As this does not take the size of studies into account, comparisons were then made using t-tests on the pooled means and variances. Variances were tested with F test, with Welch's correction used when they differed significantly by study type. Sub-group analyses were performed by stratifying data by geography, and by quality of RCTs.

For the dichotomous outcomes (diabetes mellitus, gender), the overall rates were calculated by based on the total number of individuals with diabetes, or total number of

males in each type of study and calculating a percentage based on the total available data. This again removes the bias of larger studies dominating the data when only reported percentages were considered. Comparisons between study types were then performed using Chi-square tests. A p value of less than 0.05 was considered significant in all analyses. Prism v5.0 (Carlsbad, California USA) was used for statistical analysis.

Results

Literature search

The search strategies resulted in a total of 1165 studies being identified. After removal of irrelevant studies and duplicated data, a total of 143 studies were included (**figure 1**). The quality of included studies, assessed with reference to the CONSORT and STROBE guidelines for RCT and OS respectively, showed most studies to be of high quality (**supplementary figure 1**).

Characteristics of included studies

In total eighty-one observational studies were included, including data on a total of 40,014 individuals with NAFLD. Twenty-four studies were population-based studies that described individuals with NAFLD in a general population, and fifty-two described characteristics of patients with NAFLD in secondary or tertiary centres. Seventy-three studies described all stages of NAFLD, including 34,147 individuals (**supplementary table 1**). Eighteen studies including 2,780 individuals described patients with biopsy-proven NASH (**supplementary table 2**) and 28 studies with 1938 individuals described characteristics of patients with advanced fibrosis (AF) (**supplementary table 3**). Sixty-seven RCTs were included. These studies included a total of 4,846 individuals (**supplementary table 4**).

The methods used to diagnose NAFLD varied. Most population-based studies used imaging techniques, predominantly ultrasound, to define NAFLD whilst the majority of

secondary-care based studies used histology. NASH and AF were usually defined using the system proposed by Kleiner and Brunt (15), although some earlier studies used descriptive terms. In these cases, bridging fibrosis and/or cirrhosis were regarded as advanced fibrosis. Several studies did not report the variables of interest, or did so in such a manner that we were unable to extract data for use in the present study. Available data are summarised in **supplementary table 5**.

Comparison of Randomised Controlled Trials and Observational Studies in NAFLD

Analysis of unadjusted study-reported means from included studies showed significant variance in age, BMI, prevalence of diabetes and distribution of gender across all study cohorts (**figure 2**). To compare age and BMI between types of study weighted means were calculated to reflect the relative size of each study and compared with student's t-test, and absolute prevalence of diabetes and male gender compared with Chi-squared test (**table 1**).

RCT and OS showed statistically significant differences with respect to age (mean age RCT 50.0 years (SEM 0.09) vs. OS 49.4 years (SEM 0.06) student's t-test $p < 0.001$) and BMI (32.1 kg/m^2 (0.05) vs. 29.3 (0.03), $p < 0.001$) (**table 1, figure 3**). Prevalence of diabetes and gender were compared by Chi-squared test (**table 1**). RCT and OS showed statistically significant differences with regard to the prevalence of diabetes (8%, 337 of 4186 participants, vs. 24%, 5561 of 23162 individuals, $p < 0.001$) and gender of participants (55% male, 2617 of 4733 participants, vs. 50%, 12,405 of 24,648 individuals, $p < 0.001$) (**table 1, figure 4**).

Subgroup analyses were performed to compare individuals with NASH or advanced fibrosis in OS to RCT participants. Age did not differ between RCT and individuals with biopsy-proven NASH (mean age RCT 50.0 years (0.09) vs. 49.9 (0.18), $p = 0.520$) but did differ between RCT and individuals with advanced fibrosis (mean age 54.0 years (0.19), $p < 0.001$) (**table 1, figure 3**). BMI differed significantly between RCT participants and

individuals with NASH (mean BMI RCT 32.1 (0.04) vs. 32.4 kg/m² (0.09) p=0.009), and also between RCT participants and individuals with advanced fibrosis (33.8 kg/m² (0.08), p<0.001) (table 1, figure 3).

The prevalence of DM and gender also differed significantly between RCT participants and individuals with NASH or advanced fibrosis. In RCTs, 8%, (337 participants) had DM. In observational studies, 39% of individuals with NASH (829 of 2145 participants) and 45% of individuals with advanced fibrosis (782 of 1726 participants) had DM (Chi squared test p<0.001 in each case)(table 1, figure 4). In RCTs, 55% of participants were male compared to 47% of individuals with NASH (1170 of 2476 individuals) and 38% of individuals with advanced fibrosis (712 of 1869 individuals) (Chi squared test p<0.001 in each case) (table 1, figure 4).

Analysis by geography

Ethnicity is associated with marked phenotypic differences in NAFLD (16, 17). In view of this observational studies were stratified based on geographic location (supplementary figure 2). BMI and gender distribution showed significant differences by geographic location (Kruskal-Wallis test p<0.0001 and <0.01 respectively). Accordingly, BMI and gender distribution in observational studies and RCTs were compared by geographical location. Significant differences remained between observational studies and RCTs (t-test of weighted means p<0.001). Again, these differences were most pronounced between RCTs and individuals with advanced fibrosis, with the exception of European data (supplementary figure 3). When considering results for gender, European data showed no differences between populations. In Asia and North America, RCTs contained significantly more men than were observed in epidemiological studies of advanced fibrosis (supplementary figure 4).

Exclusion criteria of RCT in NAFLD

The exclusion criteria of all identified RCT were reviewed. Eighteen trials including 2008 participants (41% of all participants) excluded patients with cirrhosis. Participants with DM were excluded in 17 studies (1144 participants, 24% of total) and exclusion criteria based on medications to treat DM were reported in a further 22 studies (1643 participants, 34% of total). Thus, the presence of DM represented an absolute or relative exclusion criterion in 39 trials including 2,787 participants (58% of all RCT participants). Individuals using drugs to treat dyslipidaemia were excluded in 7 trials (375 participants, 8%).

Discussion

These data, derived from a robust systematic review, show that characteristics of individuals with NAFLD in observational studies differ from those included in RCT. These differences are statistically significant but are often small and may not be clinically significant. However, marked differences exist between RCT cohorts and individuals with advanced fibrosis who are more likely to progress to liver-related morbidity. This is compounded by the frequent exclusion of patients with cirrhosis or diabetes from trials. These differences may limit the application of RCT trial data to high-risk patients with NAFLD.

There are important differences in susceptibility to insulin resistance and fatty liver between individuals of differing ethnicity (2, 16, 18). This was evident when studies were stratified by geography. Importantly, significant differences remained between OS and RCT populations.

An explanation for the differences seen between OS and RCT may lie in the current paradigm that steatohepatitis is required for the development of liver fibrosis. Since this process takes many years to develop, RCTs in patients with NASH have often relied on histological criteria to recruit patients and assess efficacy. For instance the accepted endpoint of an ≤ 2 point improvement in the NAFLD activity score (NAS) without

worsening of fibrosis (19), specifically excludes patients with cirrhosis since there is no way to evaluate whether fibrosis has worsened in this group. Whilst RCT are necessarily different to real life clinical practice, use of this endpoint skews trial populations away from the groups at greatest risk of liver related morbidity and mortality, and towards younger patients with earlier disease. This questions the value of current surrogate endpoints in NAFLD trials and raises important issues regarding the definition of such outcome measures in early phase studies for patients with NASH. In other liver diseases, such as hepatitis B virus infection, when liver disease is treated in patients with cirrhosis there is evidence of a reduction in fibrosis progression (20) and a concurrent reduction in the risk of liver related events. It is important to discover whether this can also be achieved in patients with NASH and advanced fibrosis.

An additional cause for the differences in observational cohorts and trial participants is the stringent exclusion criteria applied in RCT. In particular, diabetes *per se* or use of medications for diabetes is a frequent cause of exclusion of patients from trials, while nearly a quarter of patients (24%) have DM in observational studies of NAFLD and nearly half of patients (45%) with advanced fibrosis have DM. In some trials, for example trials of metformin or thiazolidinediones, limitations on diabetic patients or diabetic medications may be justified but the applicability of these findings to patients with advanced fibrosis is then limited. The design of trials that exclude patients with both diabetes and cirrhosis is thus a major barrier to external validity since many of those patients at the greatest risk of liver related death are not represented in these studies. Recent notable trials in NAFLD have suggested that studies are becoming more inclusive. For example, the recent trial of Elafibranor included around reported that over 30% of participants had diabetes (21).

The American Association for the Study of Liver Disease (AASLD) published a consensus statement regarding trial design in NAFLD (22). This was added to in 2015 by a report of a meeting between representatives of the AASLD the US Food and Drug Administration (FDA) (19). These documents provide guidance that aims to achieve greater consistency in

trial design and to define outcomes relevant to patients. They do not provide recommendations on measures that might strengthen external validity *per se* but do recommend that trials in NAFLD should target specific groups, in particular those at risk of progression to cirrhosis, those with cirrhosis, and post-transplant patients. Advice from regulatory bodies regarding measures to improve or review external validity of trials is either lacking or non-binding.

These findings are similar to studies in other disease areas. In cardiovascular disease advanced patient age and the presence of co-morbidity were identified as important factors limiting external validity (9). The issue of co-morbidity is also important in patients with NAFLD. The association of NAFLD with the metabolic syndrome raises the risk of co-existing cardiovascular disease, prior extra-hepatic cancer, and other obesity related complications (23) that may also impact on recruitment into clinical trials. It is important that these co-morbidities are considered in future trial design. It is likely that inclusion of patients with significant co-morbidities would increase the risk of competing mortality such that in large scale, long duration licensing studies the number of patients required for sufficient statistical power would also be increased. Thus there is a short-term disincentive to include such patients in registration studies. In the long-term however it is preferable that physicians and patients understand the likely benefits and harms of treatment and this can only be achieved through the inclusion of such patients in licensing studies of novel therapeutics for patients with NASH.

This study has several limitations. It is limited by missing data, which also reflects data being presented in a manner that was not suitable for analysis in this study. Nevertheless, the comprehensive systematic review and consequent large numbers studies included goes some way to limiting the impact of these missing data. There is also a risk of bias in the selection of patients for inclusion in observational studies of NASH and advanced fibrosis, through selection for biopsy, which may only be performed in patients who appear at high risk to clinicians. This is also the case in observational studies based in

secondary care where a degree of selection bias may be present. The larger community based studies using non-invasive methods to diagnose NAFLD are less at risk of these biases.

In conclusion, we have identified differences in the characteristics of patients identified in observational studies of NASH, particularly those with advanced fibrosis, and those patients enrolled in RCTs of new therapeutic approaches. These findings, whilst partially explained by the therapies that have been trialed and by the choice of surrogate endpoint, highlight a risk that future studies will have limited external validity.

References

1. Chalasani N, Younossi Z, Lavine JE et al. The Diagnosis and Management of Non-Alcoholic Fatty Liver Disease: Practice Guideline By the American Gastroenterological Association, American Association for the Study of Liver Diseases, and American College of Gastroenterology. *Gastroenterology*. 2012;142(7):1592-1609.
2. Browning JD, Szczepaniak LS, Dobbins R et al. Prevalence of Hepatic Steatosis in an Urban Population in the United States: Impact of Ethnicity. *Hepatology*. 2004;40(6):1387-1395.
3. Williams CD, Stengel J, Asike MI et al. Prevalence of Nonalcoholic Fatty Liver Disease and Nonalcoholic Steatohepatitis Among a Largely Middle-Aged Population Utilizing Ultrasound and Liver Biopsy: A Prospective Study. *Gastroenterology*. 2011;140(1):124-131.
4. Lazo M, Hernaes R, Bonekamp S et al. Non-Alcoholic Fatty Liver Disease and Mortality Among Us Adults: Prospective Cohort Study. *Bmj*. 2011;343
5. Adams LA, Lymp JF, St Sauver J et al. The Natural History of Nonalcoholic Fatty Liver Disease: A Population-Based Cohort Study. *Gastroenterology*. 2005;129(1):113-121.
6. Ekstedt M, Hagström H, Nasr P et al. Fibrosis Stage is the Strongest Predictor for Disease-specific Mortality in Nafld After Up to 33 Years of Follow-up. *Hepatology*. 2015;61 (5):1547 - 54.
7. Musso G, Gambino R, Cassader M et al. A Meta-Analysis of Randomized Trials for the Treatment of Nonalcoholic Fatty Liver Disease. *Hepatology*. 2010;52(1):79-104.
8. Rothwell PM. External Validity of Randomised Controlled Trials:“to Whom Do the Results of This Trial Apply?”. *The Lancet*. 2005;365(9453):82-93.
9. Steg PG, López-Sendón J, De Sa EL et al. External Validity of Clinical Trials in Acute Myocardial Infarction. *Archives of internal medicine*. 2007;167(1):68-73.
10. Travers J, Marsh S, Williams M et al. External Validity of Randomised Controlled Trials in Asthma: To Whom Do the Results of the Trials Apply? *Thorax*.

- 2007;62(3):219-223.
11. Travers J, Marsh S, Caldwell B et al. External Validity of Randomized Controlled Trials in Copd. *Respiratory medicine*. 2007;101(6):1313-1320.
 12. Van Spall HGC, Toren A, Kiss A et al. Eligibility Criteria of Randomized Controlled Trials Published in High-Impact General Medical Journals: A Systematic Sampling Review. *JAMA*. 2007;297(11):1233-1240.
 13. Schulz KF, Altman DG, Moher D. Consort 2010 Statement: Updated Guidelines for Reporting Parallel Group Randomised Trials. *BMC medicine*. 2010;8(1):18.
 14. Von Elm E, Altman DG, Egger M et al. The Strengthening the Reporting of Observational Studies in Epidemiology (Strobe) Statement: Guidelines for Reporting Observational Studies. *International Journal of Surgery*. 2014;12:1495 - 1499.
 15. Kleiner DE, Brunt EM, Van Natta M et al. Design and Validation of a Histological Scoring System for Nonalcoholic Fatty Liver Disease. *Hepatology*. 2005;41(6):1313-1321.
 16. Mohanty SR, Troy TN, Huo D et al. Influence of Ethnicity on Histological Differences in Non-Alcoholic Fatty Liver Disease. *Journal of Hepatology*. 2009;50:797-804.
 17. Bambha K, Belt P, Abraham M et al. Ethnicity and Nonalcoholic Fatty Liver Disease. *Hepatology*. 2012;55:769-780.
 18. Raji A, Seely EW, Arky RA et al. Body Fat Distribution and Insulin Resistance in Healthy Asian Indians and Caucasians. *The Journal of Clinical Endocrinology & Metabolism*. 2001;86(11):5366-5371.
 19. Sanyal AJ, Friedman SL, McCullough AJ et al. Challenges and Opportunities in Drug and Biomarker Development for Nonalcoholic Steatohepatitis: Findings and Recommendations From an American Association for the Study of Liver Diseases - U.s. Food and Drug Administration Joint Workshop. *Hepatology*. 2015;61 (4):1392 - 1405.
 20. Marcellin P, Gane E, Buti M et al. Regression of Cirrhosis During Treatment With Tenofovir Disoproxil Fumarate for Chronic Hepatitis B: A 5-Year Open-Label Follow-Up Study. *The Lancet*. 2013;381(9865):468-475.
 21. Ratziu V, Harrison SA, Francque S et al. Elafibranor, an Agonist of the Peroxisome Proliferator-Activated Receptor α , Induces Resolution of Nonalcoholic Steatohepatitis Without Fibrosis Worsening. *Gastroenterology*. 2016;150(5):1147-1159. e5.
 22. Sanyal AJ, Brunt EM, Kleiner DE et al. Endpoints and Clinical Trial Design for Nonalcoholic Steatohepatitis. *Hepatology*. 2011;54(1):344-353.
 23. Armstrong MJ, Adams LA, Canbay A et al. Extrahepatic Complications of Nonalcoholic Fatty Liver Disease. *Hepatology*. 2014;59(3):1174-1197.

Weighted means of age and BMI														
	RCT cohorts		Observational studies											
			All NAFLD				NASH				Adv. fibrosis			
	Mean	SEM	Mean	SEM	Difference between means	p t-test	Mean	SEM	Difference between means	p t-test	Mean	SEM	Difference between means	p t-test
Age (years)	50.0	0.09	49.4	0.04	-0.64	<0.0001	49.8	0.18	-0.16	0.424	54.0	0.19	3.96	<0.0001
BMI (kg/m ²)	32.1	0.05	29.3	0.01	-2.83	<0.0001	32.3	0.09	0.20	0.043	33.8	0.08	1.67	<0.0001
Calculated prevalence of diabetes, distribution of gender														
	RCT cohorts		Observational studies											
			All NAFLD				NASH				Advanced fibrosis			
	prevalence	prevalence	prevalence	p Chi-squared	prevalence	p Chi-squared	prevalence	p Chi-squared						
Prevalence of Diabetes (%)	8	24	<0.001	39	<0.001	46	<0.001							
Male gender (%)	56	51	<0.001	48	<0.001	39	<0.001							

Table 1: Characteristics of RCT and observational study cohorts. p-value refers to difference of observational studies to RCT cohort.

Figure Legends

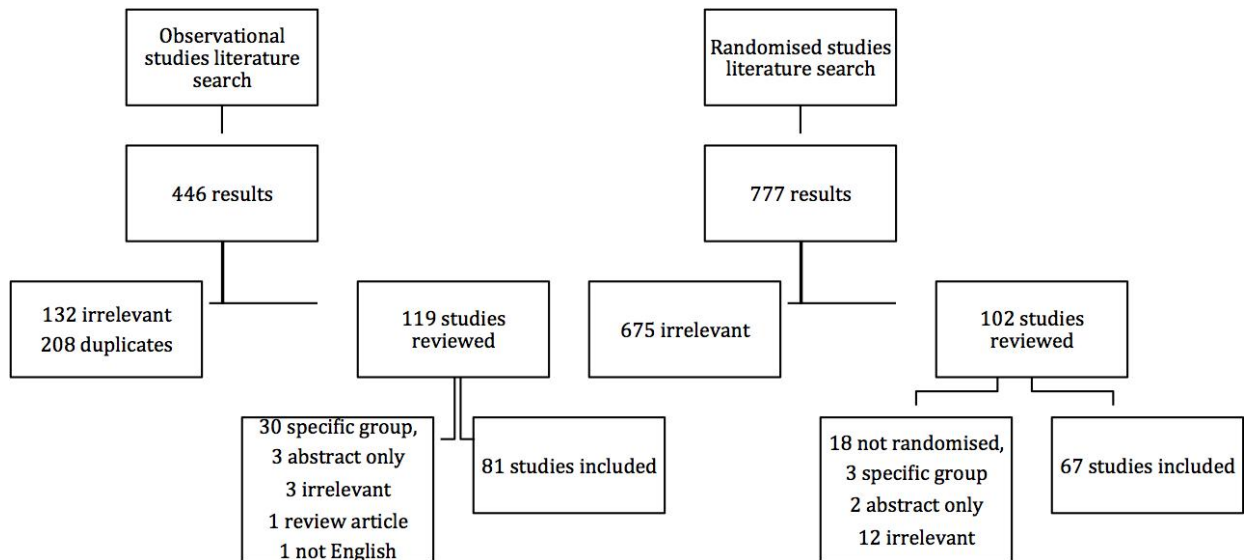


Figure 1. Flow chart of literature search outcomes.

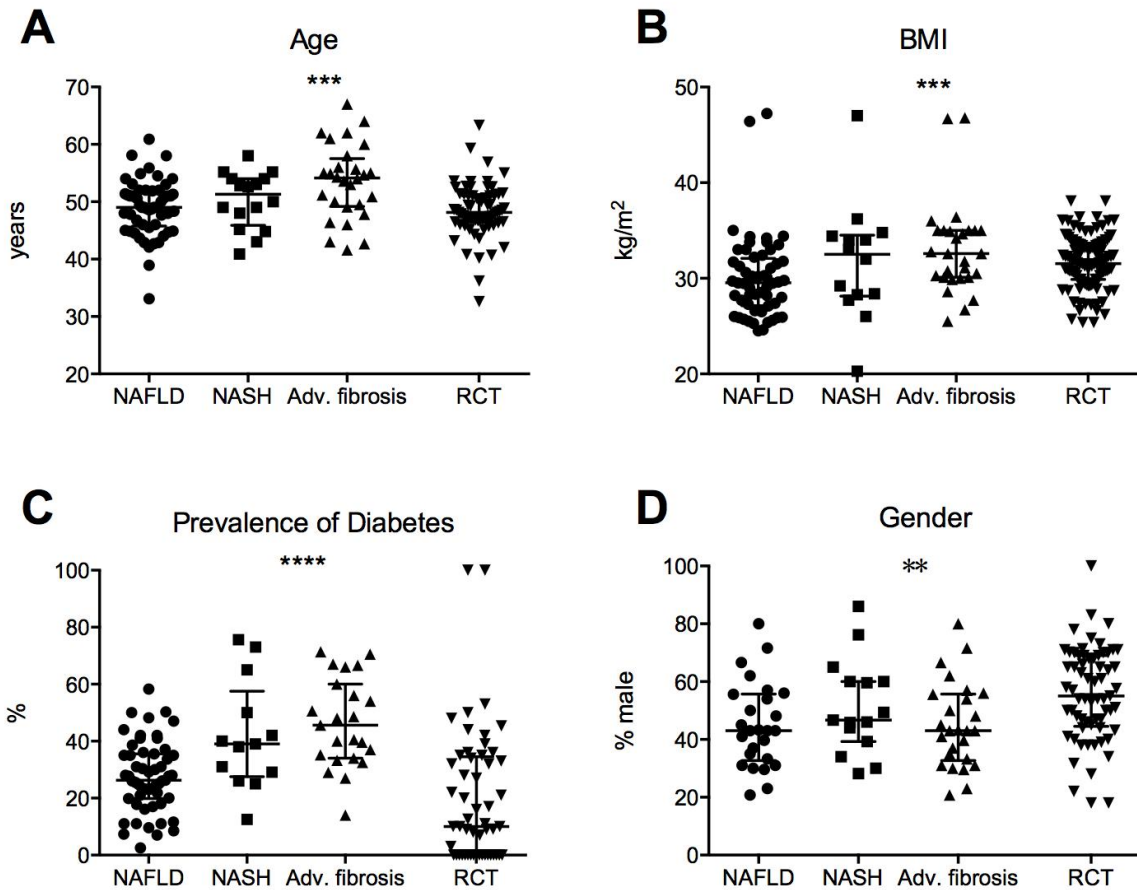


Figure 2: Distribution of variables between types of study: **A** age, **B** BMI, where each point represents the mean value from an individual trial. **C** prevalence of diabetes and **D** male gender, each data point represents prevalence within an individual study. Lines at median and interquartile range. * $p < 0.05$, ** $p < 0.001$, *** $p < 0.001$ by Kruskal-Wallis test.

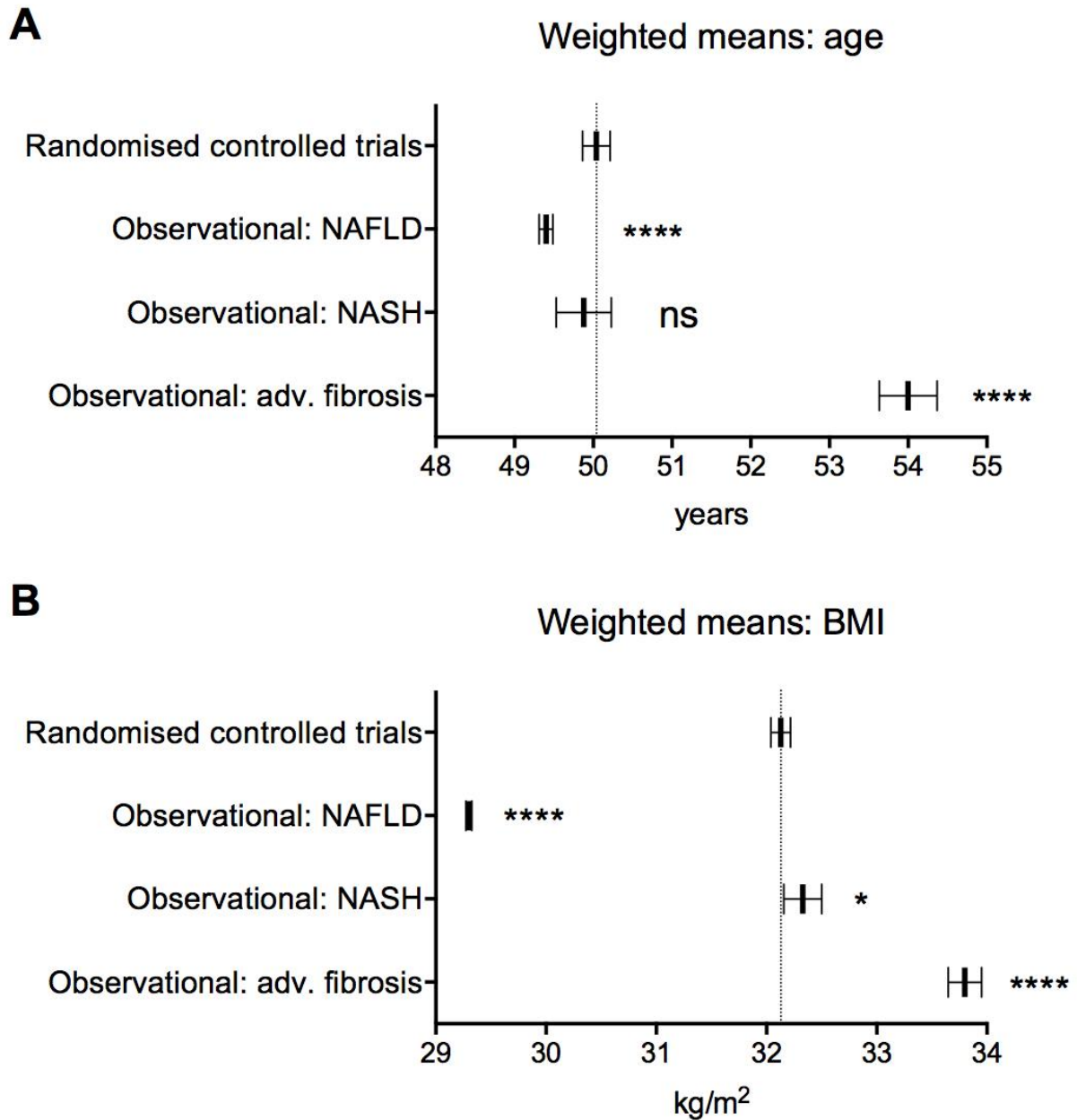


Figure 3: Age and BMI of participants in observational studies and RCTs. Data are shown as weighted mean, horizontal lines represent standard deviation. Dashed line at RCT mean. *** $p < 0.001$ by student's t-test

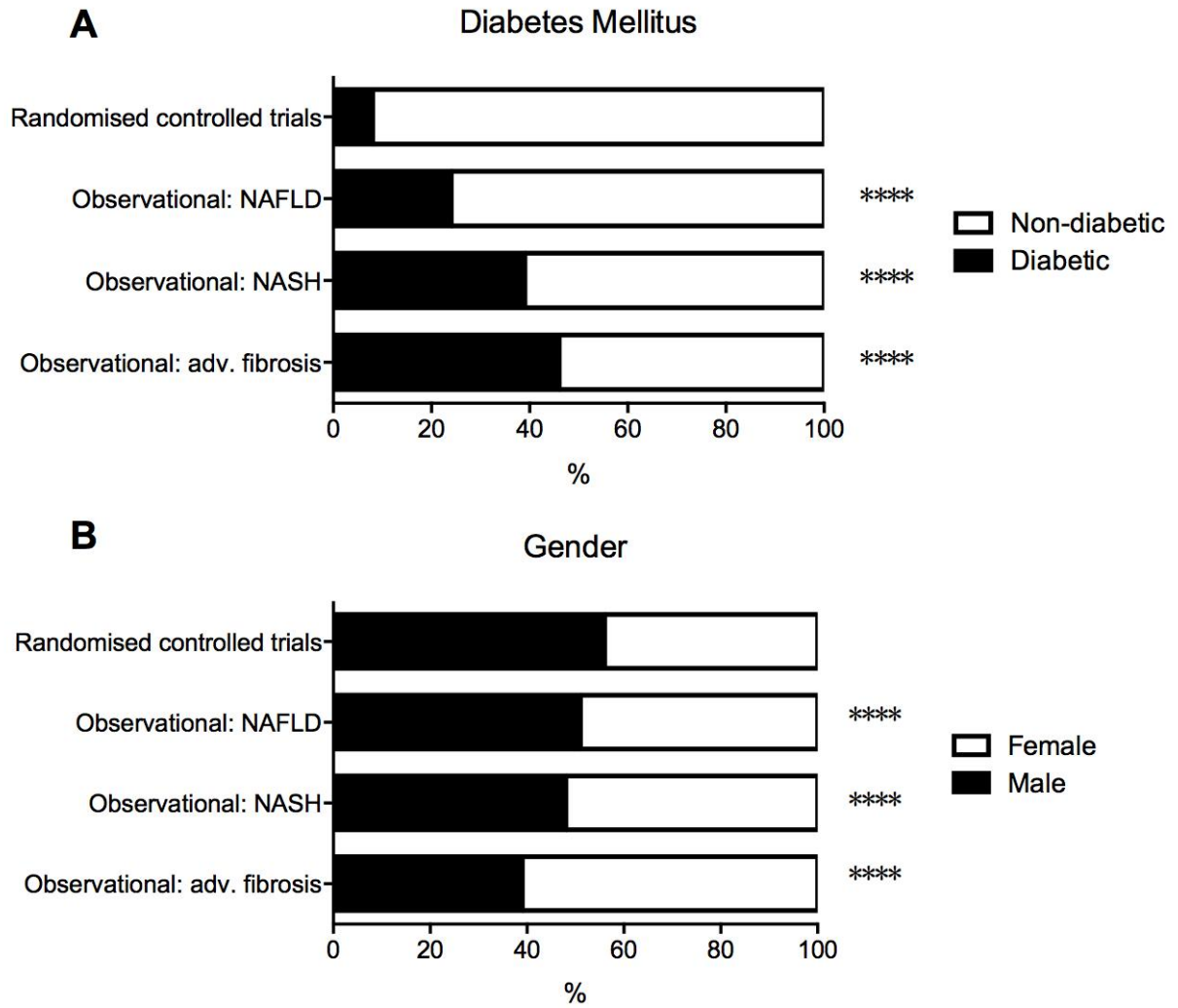


Figure 4: Prevalence of diabetes and distribution of gender in observational studies and RCTs. *** $p < 0.001$ by Chi-squared test