Chronic Liver Disease Detection and Quantification

Paul Martin Trembling

UCL

Thesis submitted for the degree of Doctor of Philosophy

2019

DECLARATION

I, Paul Martin Trembling confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

This work was funded by the Roddick Foundation.

ABSTRACT

Chronic liver disease (CLD) is a major cause of death in the UK. The major contributors are alcohol, fat and viral hepatitis. The common pathway towards cirrhosis is progressive liver fibrosis. The utility of the traditional method of evaluating fibrosis, liver biopsy, is limited by procedural risk, sampling error and variability in histological analysis. This has driven exploration of non-invasive markers of liver fibrosis.

I evaluated the performance of the Enhanced Liver Fibrosis (ELF) test to detect liver fibrosis in chronic hepatitis B and compared it to an alternative modality, transient elastography (TE), demonstrating good diagnostic performance in fibrosis assessment, with comparable performance to TE.

Further, liver biopsy is not feasible in community settings, and although the role of non-invasive markers of fibrosis is expanding, they have not been widely evaluated in community settings. I estimated the incidence of CLD in a large cohort of community-based postmenopausal women and investigated the contribution of alcohol and overweight / obesity to risk of CLD, observing more clinical events attributable to cirrhosis among those who were overweight or obese, with the highest risk in those who were overweight or obese consuming the most alcohol. I estimated the association between skirt size, as a surrogate for overweight / obesity, and CLD, finding significantly increased risk in those with larger or increasing skirt size. I demonstrated that the ELF test predicts CLD in women with risk factors comprising alcohol excess and / or overweight or obesity.

In addition to contributing to the epidemiological data in postmenopausal women, an important but under-evaluated group in terms of liver disease, I have provided data that could be used to design pathways for the early detection and stratification of CLD in the community.

IMPACT STATEMENT

This thesis has investigated the performance of non-invasive markers of liver fibrosis and provided the first evidence supporting a role for a non-invasive marker of fibrosis in predicting chronic liver disease (CLD) in a community based population. CLD is now the fifth commonest cause of death in the UK and the only major disease area to demonstrate increasing mortality. The findings of this work will be of interest to both clinicians and scientists with an interest in public health and those responsible for the strategic planning and delivery of healthcare for CLD, a major global health problem.

This work answers several important questions;

"What are the risks from alcohol and obesity in the development of CLD, and is there an interaction between these two risk factors?"

I have demonstrated that in women aged 50-74 there is an increased risk of morbidity and mortality attributed to liver disease with increasing body mass index (BMI). A "J-shaped" relationship is seen with alcohol consumption and liver disease, similar to that previously reported for alcohol and mortality with the risk higher in abstinent women and heavy drinkers than in women who drink small quantities of alcohol. The risk of chronic liver disease is highest in women who are overweight or obese and who consume high levels of alcohol, although no interaction was demonstrated between the two risk factors in this study.

"How can the risk of liver disease due to overweight / obesity be communicated to the public?"

I have proposed a potential simple public health measure to empower individuals with the knowledge of risk to 'self-stratify' their risk of liver disease. Women with higher skirt size in their 20s are at higher risk of CLD, and this risk remains in middle age.

"Can non-invasive tests for liver fibrosis be used to predict chronic liver disease in a community based population?"

I have demonstrated that a blood test, the Enhanced Liver Fibrosis (ELF) test, a marker of liver fibrosis used in clinical practice, is able to predict liver disease in a population of postmenopausal women.

From an academic perspective, my work has established a framework in which to study the epidemiological aspects of liver disease – incidence, assessment of relative contributions of risk factors, and potential confounders. I designed and performed a prospective cohort study, and although this is not the first study of its kind in liver disease in a general population, it is one of the few to use a data definition for liver disease that not only covers liver disease itself, but the consequences of liver disease, thereby increasing the ability to identify cases of CLD during follow up. This approach can be adopted and applied more widely in studies of liver disease.

From a clinical perspective, my work will be of immediate interest to a wide range of clinicians in primary care and secondary care centres, and those working in specialist alcohol services, where the conclusions could be applied to modify risks and behaviours related to alcohol consumption and obesity. My findings offer the possibility to intervene in the development of CLD at an early stage, when liver disease is preventable. The use of skirt size as an easy to

understand indicator of risk and a blood test that predicts the risk of liver disease could be incorporated into public health strategies in community settings to intervene with harm-reduction measures ultimately to impact on the ever-rising burden of liver disease.

To date, my work has resulted in three peer-reviewed publications, and several presentations at international conferences.

TABLE OF CONTENTS

Declaration 2
Abstract 3
Impact statement 5
Table of contents 8
Supervisors 14
Acknowledgements 15
Relevant publications and awards 16
List of abbreviations 20
List of figures 24
List of tables 26
Chapter 1. Aims and Objectives 29
1.1 Introduction 29
1.2 Overall aims and objectives 29
1.3 Research questions 30
1.4 Aims and objectives 31
1.5 Outline of thesis 34
Chapter 2. Background of Chronic Liver Disease and Non-invasive Markers of Liver Fibrosis 35
2.1 The liver and its function 35
2.2 Chronic liver disease 35
2.2.1 Aetiology of chronic liver disease 35
2.2.2 Pathophysiology of liver fibrosis 36
2.2.3 Complications of chronic liver disease 38
2.3 Ageing and the liver 43
2.4 Prevalence and classification 43
2.4.1 ICD classification 44
2.5 Diagnosis of liver fibrosis and cirrhosis 45
2.5.1 Liver biopsy and histological classification of liver fibrosis 45
2.6 Liver chemistry and markers of liver function 49
2.7 Introduction to non-invasive assessment of liver fibrosis 51
2.7.1 Serum-based markers 52
2.7.2 The Enhanced Liver Fibrosis (ELF) test 58
2.7.3 Validation of disease biomarkers with a focus on stability 71
2.7.4 Transient elastography 78

2.8 Conclusion 79
Chapter 3. Critical Appraisal of Systematic Review and Structured Literature Review 82
3.1 Overview 82
3.2 Introduction 82
3.3 Critical appraisal of systematic review 83
3.3.1 Method 83
3.3.2 Results 84
3.3.3 Conclusion 88
3.4 Structured literature review 89
3.4.1 Method 89
3.4.2 Search results 90
3.5 Conclusion 107
Chapter 4. Diagnostic Performance of the Enhanced Liver Fibrosis Test in Patients with Chronic Hepatitis B 109
4.1 Introduction 109
4.2 Aims of study 110
4.3 Principles of statistical methods used in this chapter – sensitivity, specificity, positive predictive value, negative predictive value and receiver operator characteristic curves 110
4.4 Methods 114
4.4.1 Ethical approval 114
4.4.2 Study population 114
4.4.3 Liver biopsy 115
4.4.4 Transient elastography 116
4.4.5 Blood markers 116
4.4.6 ELF testing 117
4.4.7 Statistical analysis 117
4.5 Results 121
4.5.1 Baseline characteristics 121
4.5.2 Discrimination between fibrosis stages 123
4.5.3 Performance of the ELF test and TE to diagnose liver fibrosis 124
4.5.4 Diagnostic performance in e-antigen negative patients 128
4.5.5 Influence of transaminases on diagnostic performance 128
4.5.6 Spectrum bias 130
4.5.7 Clinical utility models for the ELF test and transient elastography 131

4.5.8 A model combining both modalities 136
4.5.9 Further exploration of combined models – a serial algorithm 137
4.6 Discussion 145
4.6.1 Other studies 146
4.6.2 Generalisability of the findings 155
4.7 Strengths and limitations 156
4.7.1 Strengths 156
4.7.2 Limitations 156
4.8 Conclusion 157
Chapter 5. Investigation of Chronic Liver Disease in a General Population. 158
5.1 Introduction 158
5.2 Aims of study 158
5.3 Principles of statistical methods used in this chapter – Survival analysis 159
5.4 Methods 162
5.4.1 Background to UKCTOCS 162
5.4.2 Ethical approval 163
5.4.3 Study population 164
5.4.4 Exposures 164
5.4.5 Covariates 165
5.4.6 Follow up 165
5.4.7 Outcomes 166
5.4.8 Database creation 167
5.4.9 Statistical analysis 167
5.5 Results 170
5.5.1 Incidence of chronic liver disease 173
5.5.2 BMI and risk of liver-related events 174
5.5.3 Alcohol consumption and risk of liver-related events 175
5.5.4 Risk of liver-related events due to other covariates 175
5.5.5 Interaction between alcohol and BMI 182
5.5.6 Kaplan-Meier estimator results 183
5.5.7 Risk of liver-related events in participants grouped in to combinations of BMI and alcohol use 186
5.6 Discussion 190
5.6.1 Main findings 190
5.6.2 Strengths and limitations 192

5.6.3 Other studies 195
5.6.4 Implications of this study 202
5.6.5 Conclusion 202
Chapter 6. Association Between Skirt Size and Chronic Liver Disease 203
6.1 Introduction 203
6.2 Aims of this study 203
6.3 Methods 203
6.3.1 Ethical approval 203
6.3.2 Study population 204
6.3.3 Exposures 204
6.3.4 Categorisation of exposure variables 205
6.3.5 Covariates 205
6.3.6 Follow up 206
6.3.7 Outcome 206
6.3.8 Statistical analysis 206
6.4 Results 208
6.4.1 Sample characteristics 208
6.4.2 Frequencies and distributions of BMI and skirt size 211
6.4.3 Crude event rates 213
6.4.4 Survival analysis 217
6.5 Discussion 231
6.5.1 Main findings 231
6.5.2 Menopause and the role of oestrogens 232
6.5.3 Strengths and limitations 234
6.5.4 Other studies 235
6.5.5 Body mass index as a ‘gold’ standard 238
6.6 Conclusion 240
Chapter 7. Performance of the ELF Test in Predicting Clinical Outcomes of CLD in a General Population with Risk Factors 241
7.1 Introduction 241
7.2 Aims and objectives of study 242
7.3 Methods 243
7.3.1 Ethical approval 243
7.3.2 Study design 243
7.3.3 Study population and exposures 243

7.3.4 Covariates 245
7.3.5 Missing data 245
7.3.6 Follow up 246
7.3.7 Selection of cases and controls 247
7.3.8 Sample collection and sample selection 247
7.3.9 Serum marker testing 250
7.3.10 Statistical analysis 251
7.3.11 Stability of the ELF assay 252
7.4 Results 252
7.5 Discussion 265
7.5.1 Strengths and limitations 267
7.5.2 Other studies investigating the performance of biomarkers to predict clinical outcomes 269
7.6 Conclusion 273
Chapter 8. Discussion 275
8.1 Statement of main findings 275
8.1.1 Validation of the ELF test in chronic hepatitis B and comparison with transient elastography 275
8.1.2 Defining the incidence of chronic liver disease in a general population 276
8.1.3 Evaluating the risks of alcohol and above normal body mass index on the risk of chronic liver disease 276
8.1.4 Exploring the interaction between alcohol and body mass index and the risk of chronic liver disease 277
8.1.5 Evaluating the association between skirt size and chronic liver disease 278
8.1.6 Determining the performance of the ELF test to predict liver-related clinical outcomes in postmenopausal women with risk factors 279
8.1.7 Overall statement of findings 279
8.2 Strengths and limitations 280
8.2.1 Strengths 280
8.2.2 Limitations 281
8.2.3 Missing data 282
8.3 Implications of research findings for clinical practice 284
8.3.1 Hepatology services for NAFLD and alcohol and pathways of care 286
8.4 Application of this work to the general population – Health Survey for England 295

8.4.1 Alcohol consumption 296
8.4.2 Body mass index and waist circumference 297
8.4.3 Smoking 298
8.4.4 Hypertension and diabetes 298
8.4.5 Summary 299
8.5 Future research 299
8.5.1 Screening for liver disease 300
8.6 Overall summary 304
8.7 Overall conclusions 305
References 307
Appendices 346

SUPERVISORS

Primary supervisor	Professor William Rosenberg
	Peter Scheuer Professor of Liver Diseases
	Institute for Liver and Digestive Health
	Division of Medicine
	University College London
Secondary supervisor	Professor Usha Menon
	Professor of Gynaecological Cancer
	MRC Clinical Trials Unit at UCL
	Institute of Clinical Trials and Methodology
	Faculty of Population and Health Sciences
	University College London

ACKNOWLEDGEMENTS

I am grateful to Professor William Rosenberg for his supervision, and many discussions where we have developed ideas, and his friendship. The research methods that he has taught me have not only been central to this work, but have also made me a better doctor.

I am grateful to Professor Usha Menon for inviting me in to her group. Her encouragement and support has been invaluable.

Professor Julie Parkes, Professor of Public Health and head of Primary Care and Population Sciences, University of Southampton, has been a constant source of support and guidance.

Team Rosenberg, comprising Dr Sudeep Tanwar, Carmen Velazquez, Dr Simran Singh and Dr Brian Hogan have been great companions during this project and the source of constructive advice.

The UKCTOCS team, in particular Dr Sophia Apostolidou, Dr Alex Gentry-Maharaj and Dr Andy Ryan have been a great support.

Thanks to Dr Scott Harris, Southampton Clinical Research Institute, University of Southampton, for his guidance in statistics.

Finally, thanks to family, friends and colleagues for all their patience and support.

RELEVANT PUBLICATIONS AND AWARDS

(refer to appendix A)

Publications

Related to this thesis

Trembling PM, Lampertico P, Parkes J, Tanwar S, Viganò M, Facchetti F, Colombo M, Rosenberg WM. Performance of Enhanced Liver Fibrosis test and comparison with transient elastography in the identification of liver fibrosis in patients with chronic hepatitis B infection. J Viral Hepat. 2014;21:430-8

Trembling PM, Apostolidou S, Gentry-Maharaj A, Parkes J, Ryan A, Tanwar S, Burnell M, Jacobs I, Menon U, Rosenberg WM. Risk of chronic liver disease in post-menopausal women due to body mass index, alcohol and their interaction: a prospective nested cohort study within the United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS). BMC Public Health. 2017;17:603

Trembling PM, Apostolidou S, Gentry-Maharaj A, Parkes J, Ryan A, Tanwar S, Burnell M, Menon U, Rosenberg WM. Association between skirt size and chronic liver disease in post-menopausal women: a prospective cohort study within the United Kingdom Trial of Ovarian Cancer Screening (UKCTOCS). BMC Public Health. 2018;18:409

Not related to this thesis

Kennedy OJ, Parkes J, Tanwar S, **Trembling PM**, Rosenberg WM. The Enhanced Liver Fibrosis (ELF) panel: analyte stability under common sample storage conditions used in clinical practice. J Appl Lab Med. 2017;1(6):720-8

Presentations

Trembling P, Apostolidou S, Gentry-Maharaj A, Parkes J, Ryan A, Tanwar S, Burnell M, Harris S, Menon U, Rosenberg W. Enhanced Liver Fibrosis (ELF) Test predicts liver-related outcomes in postmenopausal women with risk factors in the community [Abstract]. J Hepatol 2018;68(Suppl 1):S641, Poster presentation, European Association for the Study of the Liver International Liver Congress, Paris 2018

Trembling PM, Apostolidou S, Parkes J, Ryan A, Gentry-Maharaj A, Tanwar S, Menon U, Rosenberg WM. Influence of BMI and Alcohol on Liver-Related Morbidity and Mortality in a Cohort of 108,000 Women from the General Population from UKCTOCS [Abstract]. J Hepatol 2013;58(Suppl 1):S51. Oral presentation, European Association for the Study of the Liver International Liver Congress, Amsterdam 2013

Trembling PM, Lampertico P, Parkes J, Tanwar S, Viganò M, Facchetti F, Colombo M, Rosenberg WM. Comparison of enhanced liver fibrosis test and transient elastography for the non-invasive assessment of liver fibrosis in chronic hepatitis B [Abstract]. Hepatology 2011;54(4)(Suppl):1223A. Poster presentation, American Association for the Study of Liver Diseases The Liver Meeting, San Francisco 2011

Trembling PM, Cheung M, Tanwar S, Rosenberg WM. Concordance of non-invasive markers of liver fibrosis in a mixed population of liver diseases [Abstract]. Gut 2012;61(Suppl):A408. Poster presentation, Digestive Disorders Federation Meeting, Liverpool 2012

Trembling PM, Parkes J, Tanwar S, Burt AD, Rosenberg WM. Enhanced liver fibrosis test accurately identifies liver fibrosis and predicts clinical outcomes in alcoholic liver disease [Abstract]. J Hepatol 2012;56(Suppl 2):S424. Poster presentation, European Association for the Study of the Liver International Liver Congress, Barcelona 2012

Hogan BJ, **Trembling PM**, Tanwar S, Yu D, O'Beirne JP, Rosenberg WM. Do all arterialised nodules become hepatocellular carcinoma? The outcome of 4 years magnetic resonance imaging surveillance [Abstract]. J Hepatol 2012;56(Suppl 2):S282. Poster presentation, European Association for the Study of the Liver International Liver Congress, Barcelona 2012

Awards

Shire Innovation Fund for Specialist Registrars in Gastroenterology

- Awarded for collaboration with the UKCTOCS team. 2011

British Association for the Study of the Liver Travel Award

- Awarded for abstract at the British Association for the Study of the Liver annual meeting. 2011

LIST OF ABBREVIATIONS

AFP	Alpha-fetoprotein
AIH	Autoimmune hepatitis
ALFI	Algorithm for Liver Function Investigations
ALT	Alanine aminotransferase
ANOVA	Analysis of variance
APRI	AST to platelet ratio index
ARFI	Acoustic radiation force impulse
AST	Aspartate transaminase
AUDIT	Alcohol use disorders identification test
AUROC	Area under the receiver operator characteristic curve
BAFLD	Both alcoholic and non-alcoholic fatty liver disease
BMI	Body mass index
CA125	Cancer Antigen 125
CASP	Critical Appraisal Skills Programme
CE	Conformité Européene
CHB	Chronic hepatitis B
CLD	Chronic liver disease
DANA	Difference between advanced and non-advanced fibrosis stages

DOR	Diagnostic odds ratio
ELF	Enhanced Liver Fibrosis
Fib-4	Fibrosis-4 index
FN	False negative
FP	False positive
GGT	Gamma-glutyl transpeptidase
HA	Hyaluronic acid
HOMA-IR	Homeostatic model assessment for insulin resistance
HBV	Hepatitis B virus
HCC	Hepatocellular carcinoma
HES	Hospital Episode Statistics
HR	Hazard ratio
HSC	Hepatic stellate cell
HVPG	Hepatic venous pressure gradient
INR	International normalised ratio
ICD-10	International classification of disease 10th revision
IMD	Index of multiple deprivation
LFT	Liver function test
LR	Likelihood ratio

LRE	Liver-related event
MRI	Magnetic resonance imaging
NAFLD	Non-alcoholic fatty liver disease
NASH	Non-alcoholic steatohepatitis
NHANES	National Health and Nutrition Examination Survey
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NPV	Negative predictive value
NRES	National Research Ethics Service
OELF	Original European Liver Fibrosis (test)
P3NP	Aminoterminal propeptide of procollagen type III
PPV	Positive predictive value
PRoBE	Prospective-specimen-collection, retrospective-blinded-evaluation
PSC	Primary sclerosing cholangitis
RT-E	Real-time elastography
SE	Standard error
SR	Success rate
SS	Skirt size

SWE	Shear-wave elastography
TE	Transient elastography
TIMP-1	Tissue inhibitor of matrix metalloproteinase-1
TIPS	Transjugular intrahepatic portosystemic shunt
TN	True negative
TP	True positive
UKCTOCS	United Kingdom Trial of Ovarian Cancer Screening
ULN	Upper limit of normal
WC	Waist circumference
WHO	World Health Organization

LIST OF FIGURES

Figure 2.1. Liver fibrosis staging according to Metavir classification.....48

Figure 2.2. Significance and source of abnormalities in commonly used tests of liver disease 50

Figure 2.3. Typical ranges of aminotransferase levels for various disease....51

Figure 3.1. Studies screened and assessed for structured literature review..92

Figure 4.1. 2x2 contingency table comparing two variables; the true disease state and the diagnostic test result.....111

Figure 4.2. Algorithm for using two non-invasive markers in series 114

Figure 4.3. Box plots showing median and quartiles for a) TE and b) ELF scores for diagnosing Metavir fibrosis stages 125

Figure 4.4. Receiver operator characteristic curves of ELF and TE predicting liver fibrosis stages.....127

Figure 4.5. Clinical utility models for ELF and TE predicting any fibrosis and severe fibrosis at thresholds with sensitivity and specificity of 80%, 85% and 90% 134

Figure 5.1. The composition of the final study cohort and its derivation from the UKCTOCS cohort.....171

Figure 5.2. Crude rates of first liver-related events (per 1000 person years) according to a) BMI category and b) alcohol consumption category over mean follow-up of 5.1 years 177

Figure 5.3. Forest plot showing hazard ratio (HR) estimates and 95% confidence intervals for univariate associations between baseline characteristics and liver-related event.....181

Figure 5.4. Kaplan-Meier curves of hazard of liver-related event to eight years for groups of UKCTOCS participants based on thresholds of self-reported alcohol consumption and body mass index 185

Figure 5.5. Matrices showing hazard ratio estimates of first liver-related event for each BMI / alcohol combination 189

Figure 6.1. Composition of the final study cohort and its derivation from the UKCTOCS cohort.....209

Figure 6.2. Distributions of a) BMI, b) skirt size in 20s, c) skirt size at questionnaire completion, and d) change in skirt size per year 213

Figure 6.3 Crude rates of LRE per 1000 person years, for a) BMI, b) skirt size in 20s, c) skirt size at questionnaire completion and d) change in skirt size.....216

Figure 6.4. Forest plot showing hazard ratio (HR) estimates and 95% confidence intervals for univariate associations between baseline characteristics and liver-related event 222

Figure 6.5. Forest plots of hazard ratio (HR) estimates for liver-related events, for univariate models and adjusted models 230

Figure 6.6. BMI and risk of non-alcoholic fatty liver disease 237

Figure 7.1. Study design for evaluating performance of the ELF test in predicting liver-related events in participants with risk factors 245

Figure 7.2. Sample time collection points in UKCTOCS ELF test study 248

Figure 7.3. Derivation of the study cohort for ELF testing 255

Figure 7.4. Venn diagrams showing contributions from BMI and alcohol to the study group 257

Figure 7.5. Box plots showing changes and distributions of change in mean ELF scores from recruitment to a) second sample and b) third sample, in cases and controls 260

Figure 7.6. Cumulative hazards for LRE using ELF threshold of 9.8 262

Figure 7.7. Cumulative hazards for LRE using ELF threshold of 10.51 263

Figure 8.1. Self-reported alcohol consumption in women in the Health Survey for England 2015 297

Figure 8.2. Study design to evaluate performance of the ELF test in low risk participants, to evaluate its role as a screening tool 302

Figure 8.3. Proposed two-stage liver disease risk stratification pathway for middle-aged women in the community 304

LIST OF TABLES

Table 2.1. Causes of chronic liver disease....36

Table 2.2. Summary of drugs currently being evaluated in phase 3 clinical trials....55

Table 2.3. Median ELF test scores and performance of the ELF test in predicting liver fibrosis in subjects with alcoholic liver disease....65

Table 3.1. Overview of studies included in structured literature review of prevalence of liver disease in community-based populations using non-invasive markers....93

Table 3.2. Baseline characteristics of participants in studies of non-invasive markers in the community in individuals with type 2 diabetes....101

Table 4.1. Baseline characteristics of subjects who underwent TE and ELF testing....123

Table 4.2. Median scores and diagnostic performance of the ELF test and TE according to Metavir fibrosis stage....126

Table 4.3. Diagnostic performance of the ELF test and TE according to categories of serum ALT levels....129

Table 4.4 Diagnostic accuracy of TE and ELF for predicting any, significant and severe fibrosis....130

Table 4.5. Obuchowski measures for the ELF test and TE for each liver fibrosis stage pair....131

Table 4.6. Diagnostic performance indices for ELF and TE in the identification of severe fibrosis (F3,4) and cirrhosis (F4) at a range of thresholds....133

Table 4.7. Sensitivities and specificities of the ELF test and TE identifying severe fibrosis (Metavir F≥3) and any fibrosis (F≥1) using thresholds with sensitivity and specificity of 80%, 85% and 90%....135

Table 4.8 Odds ratios calculated by logistic regression for prediction of liver fibrosis using models comprising the ELF test and TE individually, and the ELF test and TE combined....137

Table 4.9. Calculated Youden index for each liver fibrosis stage....139

Table 4.10. Diagnostic performance characteristics of models comprising serial non-invasive markers....141

Table 4.11. Summary of studies of the performance of the ELF test in subjects with chronic hepatitis B....152

Table 5.1. Baseline characteristics of the study cohort according to BMI category and in all participants....172

Table 5.2. ICD-10 codes and / or death certificate text of first LREs 174

Table 5.3. Crude rates of first liver-related events (per 1000 person years) according to BMI category and alcohol consumption category over mean follow-up of 5.1 years 176

Table 5.4. Hazard ratio estimates of first liver-related events, according to BMI category 178

Table 5.5. Hazard ratio estimates of first liver-related events, according to alcohol category 179

Table 5.6. Event rates of LRE for each covariate and hazard ratios for univariate associations between each covariate and liver-related event 180

Table 5.7. Event rate of first liver-related event for each BMI / alcohol combination 187

Table 5.8. Hazard ratio estimates of first liver-related event for each BMI / alcohol combination 188

Table 5.9. Summary of studies reporting incidence of chronic liver disease using various data definitions 200

Table 6.1. Baseline characteristics according to BMI category, and for all participants 210

Table 6.2 Skirt size data for all participants and according to BMI group 212

Table 6.3. Event rate of first LRE for each BMI category, and for skirt size in 20s categories, change in skirt size categories and skirt size at time of questionnaire categories 214

Table 6.4. ICD-10 codes and / or death certificate text of first LREs 215

Table 6.5. Hazard ratio estimates for first liver-related events for covariates 218

Table 6.6. Hazard ratios of first liver-related events for skirt size in 20s, skirt size at questionnaire completion, BMI and change in skirt size, for three models 224

Table 7.1. Baseline characteristics of study participants and comparisons between cases and controls 256

Table 7.2 Assay results for individual components of the ELF test and calculated ELF test score for cases and controls 258

Table 7.3. Mean ELF scores for cases and controls in recruitment samples, subsequent samples and in the combined subsequent samples 259

Table 7.4. Hazard ratio estimates for liver-related event at ELF thresholds of 9.8 and 10.51 261

Table 7.5. ICD-10 codes and / or death certificate text of first LREs for the cases 265

Table 7.6. Benefits and limitations of non-invasive markers of liver fibrosis to be considered during evaluation as a prognostic marker 273

Chapter 1. AIMS AND OBJECTIVES

1.1 Introduction

There is a need for better tests for liver disease and methods of stratification to identify those with chronic liver disease (CLD) in the community. Although risk factors for CLD are well described, their relative contribution to CLD is not. Both heavy alcohol use and high body mass index (BMI) are prevalent in the population and more clarity is needed to understand their roles and interaction in CLD. Further, the burden of CLD in middle-aged women is not well understood.

Compared to the traditional role of evaluating liver fibrosis, liver biopsy, the role of non-invasive markers of liver fibrosis in diagnosing CLD is in its infancy.

1.2 Overall aims and objectives

The aim of this thesis was to investigate the epidemiology of CLD in a general population; to determine the incidence of CLD and to understand the association between disease and common risk factors. Secondly, this thesis aimed to evaluate the use of a non-invasive marker of liver fibrosis to predict CLD.

The objectives of the thesis were to utilise data from two sources; a cohort of patients with chronic hepatitis B, and a large cohort of women participating in the United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS).

1.3 Research questions

The precise interaction between the two common causes of CLD, obesity and alcohol use, is not well understood. Although middle-aged women are considered to be a relatively high-risk group for developing CLD, the prevalence of CLD in this population is not well characterised. Further, CLD is often detected late in its natural history, and there is a need to develop strategies for earlier detection of CLD. There are a number of non-invasive techniques emerging for the evaluation of liver disease, however these have mainly been confined to secondary care populations. This thesis aims to combine these areas of unanswered questions;

1. Both serum markers of liver fibrosis and transient elastography (TE) are used to stratify individuals in terms of liver disease. How do the performances of these two diagnostic tests compare in the detection of fibrosis in a population of patients with chronic hepatitis B? Although transient elastography is being considered for use in community settings, blood test markers have the advantage that they can be incorporated in to routine investigations; could this comparison study provide confidence that one such blood test marker, the Enhanced Liver Fibrosis (ELF) test, may also have a role as an easier to use alternative to TE?

2. What is the incidence of CLD in a population of postmenopausal women in the general population?

3. What are the relative contributions of obesity and alcohol use to the development of CLD in this population?

There is an urgent need to prevent the development of liver disease in individuals and to recognise risk factors for CLD. Public health measures need to provide the public with simple, easy to understand ways to assess individual risk.

4. Could skirt size be used as a simple public health measure to inform postmenopausal women about their individual risk of CLD, thus empowering them to make lifestyle changes to reduce risk?

The risk factors for CLD are common (and will be elucidated in this population) and individuals with risk factors for CLD require the most urgent attention. However, not all individuals with risk factors will develop CLD, and there is a need to stratify the population, which may be possible by employing non-invasive markers of liver fibrosis.

5. Does the ELF test predict the development of CLD in postmenopausal women with risk factors in the form of high BMI, high alcohol use, or both?

1.4 Aims and objectives

The aims and objectives were;

1. To compare the performance of the ELF test to that of transient elastography.

In a population of patients with chronic hepatitis B;

i) to evaluate the ability of the ELF test and transient elastography to distinguish between fibrosis stages using liver biopsy as the reference standard.

ii) to evaluate the diagnostic performance of the ELF test and transient elastography in predicting fibrosis stages, using liver histology as the reference.

iii) to compare the performance of the ELF test and transient elastography in determining fibrosis stage.

iv) to model and compare the clinical utility of the ELF test and transient elastography in predicting any fibrosis and severe fibrosis, and to determine the proportion of cases where biopsy was avoided incorrectly, biopsy was avoided correctly, and where an indeterminate result required biopsy for clarification.

2. To investigate the epidemiology of CLD in a population of postmenopausal women in a general population.

By extracting data from UKCTOCS, to perform a prospective cohort study nested in the UKCTOCS trial;

i) using relevant International Classification of Diseases tenth revision (ICD-10) codes, to estimate the incidence of CLD using a number of data sources.

ii) to estimate the prevalence of risk factors for CLD, namely BMI and alcohol use.

iii) to examine the association between a number of risk factors and risk of CLD using time-to-event analysis.

3. To investigate the association between skirt size and CLD in a population of postmenopausal women in a general population.

By extracting data from UKCTOCS, to perform a prospective cohort study nested in the UKCTOCS trial to;

i) describe the distribution of skirt size of participants at the time of study participation and in their 20s.

ii) examine the association between skirt size in women aged 20-30, skirt size in middle age (at the time of trial participation) and change in skirt size, and risk of CLD while controlling for potential confounders.

4. To evaluate the performance of a non-invasive marker of liver fibrosis to predict CLD in a population of postmenopausal women with risk factors in a general population.

By extracting data from UKCTOCS and using the PRoBE (prospective-specimen-collection, retrospective-blinded-evaluation) design to perform a case-control study nested in the UKCTOCS trial;

i) to measure the ELF test score in participants with risk factors for CLD, comprising self-reported high alcohol use and/or BMI ≥25 kg/m^2.

ii) using time-to-event analysis, determine the ability of the ELF test to predict CLD at various ELF score thresholds.

1.5 Outline of thesis

This thesis begins with an overview of CLD and its risk factors, followed by a critical appraisal of a systematic review, and a structured literature search to identify previous work relating to the early detection of CLD in the community using non-invasive tests of liver fibrosis.

Chapter 4 compares the performance of the ELF test with transient elastography.

Chapters 5, 6 and 7 are based on data from the UKCTOCS trial.

In chapter 8, the main findings of the thesis are presented, with an overall discussion and conclusion. Future work is outlined.

Chapter 2. Background of Chronic Liver Disease and Non-invasive Markers of Liver Fibrosis

2.1 The liver and its function

The liver has a multitude of functions including carbohydrate, lipid and protein metabolism, notably the synthesis of albumin, glycogen, cholesterol and triglycerides. It provides a key element in digestion, through the synthesis of bile salts. It has a major role in drug metabolism and hormone breakdown and production. In addition to glycogen, the liver stores vitamins (including A, B12, D, E and K) and minerals and has an important role in iron metabolism by release and storage of iron in the form of ferritin.[1] It has a major role in production of clotting factors.

2.2 Chronic liver disease

Chronic liver disease (CLD), by definition, describes disease of the liver that persists for longer than six months. In the clinical context, it represents a process of damage and regeneration of the liver that has ultimately resulted in progressive liver fibrosis to cirrhosis. Cirrhosis is defined as the histological development of regenerative nodules in the liver, surrounded by fibrous bands in response to chronic injury.[2]

2.2.1 Aetiology of chronic liver disease

The main causes of CLD are shown in table 2.1. Globally, chronic hepatitis B and C virus infection represent the most common causes of CLD.[3] However, there are wide geographical variations. In the UK, the most common causes

of CLD are alcohol misuse, non-alcoholic fatty liver disease and chronic viral hepatitis.

Table 2.1. Causes of chronic liver disease

Category	Cause
Viral	Hepatitis B virus Hepatitis C virus
Drugs	Alcohol Other drugs including methotrexate, amiodarone
Metabolic	Non-alcoholic fatty liver disease Haemochromatosis Wilson's disease
Autoimmune	Primary biliary cholangitis Autoimmune hepatitis Primary sclerosing cholangitis
Genetic	Alpha-1-antitrypsin deficiency Glycogen storage disorders Cystic fibrosis Porphyria
Other	Budd-Chiari Heart failure Biliary obstruction

2.2.2 Pathophysiology of liver fibrosis

The key feature of progressive liver damage is liver fibrosis. The functional unit of the liver is the lobule, comprised of hepatocytes arranged in branching plates around a central vein connecting to the portal tract. The portal tract comprises an arteriole branch of the hepatic artery, a venule branch of the portal vein and

bile ductules. Blood enters the lobule through the branches of the portal vein and hepatic artery, flowing through sinusoids between the hepatocyte plates. Hepatocyes are separated from the sinusoids by the space of Disse which is filled with blood plasma and contains hepatic stellate cells (HSC). On one side, fenestrated endothelia lie on a sheet of connective tissue. On the other, microvilli extending in to the space of Disse allow absorption of proteins and other plasma components in to hepatocytes.

In response to toxins or other drivers, HSC become activated and proliferate. They phenotypically become myofibroblast-like, producing extracellular matrix in the space of Disse. The endothelial fenestrations are lost (sinusoidal capillarisation) and hepatic microvilli are damaged, resulting in impaired absorption of nutrients in to hepatocytes. The accumulation of matrix results in architectural distortion of the liver parenchyma and resistance to sinusoidal blood flow. This is the basis of portal hypertension.

It is important to note that liver fibrosis is dynamic, with accumulation of matrix accompanied by matrix degradation and remodelling, with the process being more dominant in one direction or the other. Even when cirrhosis is reached, remodelling and, to some degree, regression can still occur.[4] This balance is regulated by multiple factors including age and gender. Further, there is a genetic influence, and future understanding of liver fibrosis may focus on identifying genes that are dysregulated in the liver. The role of genes regulating the role of oestrogens in liver fibrosis is discussed in chapter six. In addition, the process of chronic liver disease not only involves fibrosis, but also comprises chronic inflammation and neo-angiogenesis (driven in large part by hypoxia due to extracellular matrix accumulation).[5]

2.2.3 Complications of chronic liver disease

The progression of liver fibrosis to cirrhosis may be clinically silent, with no impairment of liver synthetic function and no evidence of portal hypertension. This state may persist even when cirrhosis has been reached. Individuals at this point are described as having compensated cirrhosis.

The clinical features of decompensated cirrhosis are related to relative impairment of the functions of the liver and the effects of portal hypertension.

2.2.3.1 Liver failure

This form of decompensation is due to a reduction in the functional performance of the liver. In clinical practice patients may present with jaundice and elevated serum bilirubin. Other markers of synthetic dysfunction include coagulopathy, with prolonged prothrombin time, and decreased serum albumin levels.

2.2.3.2 Portal hypertension

There are a number of clinical consequences of hypertension in the portal vein and its branches.

2.2.3.2.1 Ascites

Ascites in the context of CLD has been generally thought to be due to elevated capillary hydrostatic pressure in the splanchnic bed, resulting in transudation of fluid in to the peritoneal space. A feature of cirrhosis is systemic vasodilatation, causing reduced effective arterial blood volume and compensatory hyperdynamic circulation. In order to maintain blood pressure in the presence of vasodilatation, activation of the renin-angiotensin system results in renal vasoconstriction which causes reduced sodium delivery to the kidney and

increased sodium, and therefore water, reabsorption, which promotes transudation to the peritoneal cavity. In addition, hyperactivation of the sympathetic nervous system results in increased absorption of sodium and water in the proximal tubules.

There is some controversy associated with this mechanism, with other theories now being considered. For example, the role of pathological translocation of bacteria or bacterial products from the gut to the systemic circulation may have a role, possibly by stimulating release of pro-inflammatory cytokines that cause inflammation and further release of vasodilators, increasing splanchnic arterial vasodilation.[6]

Ascites is associated with development of spontaneous bacterial peritonitis and hepatorenal syndrome, and confers a poor prognosis, associated with 85% and 57% survival at 1 and 5 years, respectively.[6,7]

Management of ascites includes dietary salt restriction, diuretics (aldosterone antagonists and loop diuretics), large volume paracentesis and transjugular intrahepatic portosystemic shunt (TIPS).[8]

2.2.3.2.2 Variceal haemorrhage

Normal portal venous pressure is 5-10 mmHg, and a hepatic venous pressure gradient (HVPG) of ≥5 mmHg is consistent with portal hypertension. Resistance of blood flow through the liver in portal hypertension promotes the spontaneous development of shunts to the systemic circulation, and commonly occur at the cardia through the intrinsic and extrinsic gastro-oesophageal veins (which drain in to the portal vein) resulting in oesophageal or gastric varices, dilated superficial veins which are susceptible to necrosis and ulceration with resultant

haemorrhage. HVPG of >10 mmHg is associated with the development of varices, and HVPG above 12 mmHg is associated with variceal haemorrhage.

Primary prophylaxis for variceal haemorrhage includes non-selective beta-blockers (or endoscopic band ligation of oesophageal varices). Management of oesophageal or gastric variceal haemorrhage centres on endoscopic intervention, with secondary prophylaxis with beta-blockers, or TIPS.[9]

2.2.3.2.3 Hepatic encephalopathy

Hepatic encephalopathy is brain dysfunction as a result of liver dysfunction and portal hypertension, manifesting as a spectrum of neuropsychiatric abnormalities. The pathogenesis is thought to be related to the inability of damaged hepatocytes to metabolise nitrogen-containing compounds, most notably ammonia, from the gut and / or due to direct delivery of these compounds to the systemic circulation via collaterals described above. Ammonia easily crosses the blood-brain barrier and is absorbed in to astrocytes, increasing the synthesis of glutamine to glutamate. The resultant increase in osmotic pressure causes cell swelling and damage. In addition, there is an accumulation of natural benzodiazepines seen in patients with hepatic encephalopathy which act on the GABA receptors in the brain, causing neuroinhibition.

Hepatic encephalopathy may be sub-clinical, termed minimal or covert encephalopathy, but when symptomatic is described as overt hepatic encephalopathy. Diagnosis of minimal hepatic encephalopathy requires neuropsychometric testing to elucidate cognitive dysfunction in the absence of the clinical features seen in overt hepatic encephalopathy. Neurophysiological

investigations including electroencephalogram testing may be useful. Overt hepatic encephalopathy is usually classified using the West Haven criteria; grade 0, no abnormality; grade 1, short attention span; grade 2, impaired performance of addition or subtraction; grade 3, confusion, disorientation, somnolence; grade 4, coma. Prevalence of overt encephalopathy in patients with cirrhosis is 30-35%.[10]

There are little data to support treatment of minimal hepatic encephalopathy. Treatment of overt hepatic encephalopathy centres on targeting the areas of pathogenesis. Lactulose, a non-absorbable disaccharide, is generally used as first line therapy, its laxative effect reducing the gut nitrogenous load. Rifaximin, a non-absorbable antibiotic, has been shown to be effective in reducing urease-producing bacteria in the gut. Ultimately, liver transplantation may be indicated.[11]

2.2.3.3 Hepatocellular carcinoma

Although most cases of hepatocellular carcinoma (HCC) occur in patients with cirrhosis, incidence is associated with CLD aetiology, with the highest proportion seen in patients with chronic viral hepatitis. 10-20% of cases of HCC are seen in patients with non-alcoholic fatty liver disease (NAFLD).[12] Five year cumulative risk of HCC in alcoholic liver disease has been reported to be 8%,[13] whilst recent data indicate annual cumulative incidence of HCC in NAFLD is 2-12%.[14]

Pathogenesis of HCC is probably related to development of regenerative nodules with small cell dysplasia through to invasive HCC. Screening for HCC is an important component of managing patients with CLD and comprises six-

monthly imaging and serum alpha-fetoprotein (AFP) level, although the clinical utility of AFP is questionable due to its low sensitivity and specificity.

There are little data on the longitudinal follow-up of hepatic nodules in cirrhosis. In a study by myself and colleagues (see list of publications) at the Royal Free Hospital, reports of all magnetic resonance imaging (MRI) scans between 2006 and 2011 were searched for the term 'nodule', identifying 630 such scans in 369 patients. Patients were excluded if there was less than one year follow up, if HCC was diagnosed on index scan, if an alternative diagnosis was made or if no significant arterialised lesion was reported despite previous suspicion on ultrasound scan. This yielded for analysis the scans of 29 patients with a diagnosis of regenerative, indeterminate or dysplastic arterialised nodules at baseline and more than one year follow-up with MRI and AFP surveillance. We found that 31% of lesions described as arterialised nodules on index scan developed in to HCC within two years.

The Barcelona staging classification system has been widely adopted for prognostication.[15] Management of HCC depends on tumour size, tumour number and spread of disease. Treatment may be with curative intent, for example liver resection or transplantation, or radiofrequency ablation. "Bridging" therapies to potential transplant include transarterial chemoembolisation. Sorafenib, an oral multikinase inhibitor is employed as a systemic therapy for advanced HCC, with palliative intent.

Other clinical complications of advanced liver disease include malnutrition and hepatorenal syndrome.

2.3 Ageing and the liver

As will be described later, this thesis will focus on liver disease in a cohort of middle-aged individuals. Although there are changes in the structure and function of the liver with age, progressive liver damage is not considered to be a predominant feature of the ageing process.[16] There is, however, evidence that risk factors for chronic liver disease increase with age, with increased vulnerability of the liver to injury with age.[17]

Age has clearly been shown to be a risk factor for non-alcoholic fatty liver disease.[18,19] Further, evaluation of liver fibrosis in patients with NAFLD has demonstrated more advanced fibrosis in the elderly.[20,21]

Ageing is associated with down-regulation of genes associated with fibrosis degradation, and reduced expression of antioxidants with resultant increase in reactive oxygen species-mediated tissue damage, inflammation and fibrosis. Compared to males, fibrosis progression is slower in females until menopause.[5] The role of oestrogen in fibrogenesis is further discussed in chapter six.

These observations raise an important clinical question – what is the risk of chronic liver disease in older populations?

2.4 Prevalence and classification

In contrast to other major disease areas, standardised mortality rates for liver disease are rising, with an increase of 400% since 1970, with an estimated prevalence of CLD in England and Wales of 60,000.[22] Accurate data on incidence and prevalence of CLD, however, are problematic. In contrast to cancers, there is no system of registration for CLD. Death registry data may not

capture all cases of CLD, and due to the large number of causes of liver disease and variation in terminology, there is no widely accepted definition for CLD.

2.4.1 ICD classification

The International Classification of Diseases (ICD), which is maintained by the World Health Organization (WHO) is a healthcare classification system which lists codes for classifying diseases.[23] As will be outlined later, this thesis presents work based on utilisation of the tenth revision, ICD-10, which was introduced in 1994.

One strategy to improve accuracy of data collection on incidence and prevalence of liver disease is to use multiple sources. A recent study has suggested that there has been a general underestimation of incidence of cirrhosis and that variation in the coding employed may result in misleading estimates.[24] A previous study by this group had interrogated two data sources, the Hospital Episode Statistics (HES) database (described in chapter 5) for secondary care data and the Clinical Practice Research Datalink for primary care data.[25] A wide definition of liver disease was adopted, searching for codes, including (as with my study) codes for clinical events related to decompensated liver disease. Between 1998 and 2009, incidence of liver disease was estimated to be 30.7 per 100,000 participant years. In women, incidence was 25.4 per 100,000 participant years. In this group's subsequent study, mortality due to cirrhosis was estimated using data from the Office for National Statistics death registry (as with my study). To demonstrate differences due to data definitions, estimates were made using three different sets of codes; a broad code commonly used in studies of CLD epidemiology developed by Leon and

McCambridge,[26] a restrictive definition used in a study of alcoholic liver disease by Jepson, K70.3,[27] and the set of codes used in the authors' initial study. Standardised mortality rates were 8.8, 5.1 and 5.4 per 100,000 participant years for the three definitions, respectively. Translating their previous incidence data to the new study, the authors note low mortality rates compared to incidence, and comment that this apparent underestimation may be related to the choice of data definition. They highlight the value of interrogating routinely collected data, which has both cost-effective and methodological strengths. This concept is a central strand of my thesis.

An alternative explanation for this discrepancy could be the failure to code for liver disease in death certification and this is discussed in chapter five.

2.5 Diagnosis of liver fibrosis and cirrhosis

2.5.1 Liver biopsy and histological classification of liver fibrosis

2.5.1.1 Liver biopsy

The traditional method for assessing liver fibrosis, and the current reference standard, is liver biopsy. This is an invasive procedure that can be performed either trans-abdominally or via the transjugular route (and less often, surgically). In addition to assessment of fibrosis, benefits of histological examination of liver tissue include assessment of liver inflammation and the potential to provide information on the aetiology of the liver disease. However, transabdominal liver biopsy is often a painful experience for the patient and is potentially hazardous. A single centre study reported an overall complication rate of 6.4%. The most common complication was major bleeding, requiring

blood transfusion and / or surgical intervention, in 3.5% of the cohort. The overall death rate was 1.6%.[28]

A UK-wide study of 1,500 procedures reported a complication rate due to bleeding of 1.7%, requiring transfusion in less than half the cases. Overall mortality was 0.13-0.33%.[29] More recently, a Canadian study investigated over 4,000 liver biopsy procedures over nine years and reported a rate of significant complications of 0.75%. Pain was the most common complication, responsible for 69% of complications, followed by bleeding (47%).[30]

An additional problem with liver biopsy is sampling variability, with only 1/50,000 of the liver being sampled, and variation in interpretation. This has been well described in studies of patients with chronic hepatitis C infection and NAFLD. One study indicated that sampling error may have led to under diagnosis of cirrhosis by 15%,[31] whilst another reported good inter-observer agreement, which was influenced by level of experience and not specimen length.[32] In a study of patients with NALFD, where evaluation of hepatic steatosis and distinction between simple steatosis steatohepatitis is also important, two samples were taken from each of fifty-one patients.[33] There was significant sample variability in all histological features, including bridging fibrosis.

2.5.1.2 Classification of liver fibrosis

A number of scoring systems have been developed for both the staging of liver fibrosis and grading of inflammation in biopsies, and have most often been used in the context of viral hepatitis. Histological assessment of fibrosis has traditionally used a Masson trichrome stain. Fibrosis progresses from no

fibrosis, to fibrous portal expansion to bridging fibrosis, then to incomplete cirrhosis and established cirrhosis[34] (figure 2.1). The Ishak fibrosis score has been commonly used in cases of viral hepatitis, and has the most stages of the scoring systems, with seven stages.[35] Others include the Scheuer[36] and Metavir[37] classifications. A number of scoring systems have been developed for NAFLD including Brunt's[38] and Kleiner's[39] systems. Both of these, in addition to including a fibrosis score, include evaluation of inflammation and steatosis.

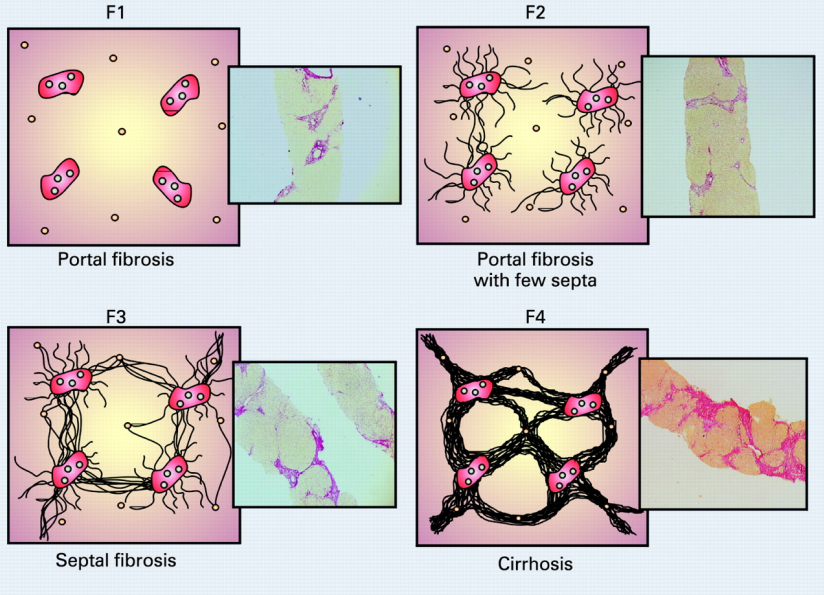


Figure 2.1. Liver fibrosis staging according to Metavir classification
From Asselah, T.*et al*.[40] Reproduced with permission

An important factor to consider when utilising liver fibrosis scoring systems is that histological staging represents a categorical variable, and progression from one stage to the next does not necessarily represent an ordinal progression in matrix accumulation.[4]

A large population-based study in Canada reported an annual liver biopsy rate of 54.8 per 100,000 of the general population.[30] During the eight year study period, between 1994 and 2002, the background population increased by 23%, but the number of biopsies performed increased by 41%. The authors postulate that this trend is due to rising prevalence of liver disease and limited availability of non-invasive markers of fibrosis. This study further highlights the need for alternative safer and cheaper tools for evaluating liver fibrosis in the general population. Later in this chapter, the role of liver biopsy as the reference standard for validation of non-invasive markers of liver fibrosis will be

discussed. As the thesis progresses, there will be a focus on investigation of liver fibrosis in the community and evaluation of alternatives to liver biopsy.

2.6 Liver chemistry and markers of liver function

Standard liver chemistry tests ("liver function tests", LFTs) are commonly employed both in secondary care and primary care to assess liver disease. This panel usually includes the aminotransferases alanine aminotransferase (ALT) and aspartate aminotransferase (AST), alkaline phosphatase (ALP), gamma-glutamyl transpeptidase (GGT) and bilirubin. Rather than a measure of function, increased levels of aminotransferases in serum represent hepatocellular damage. As discussed earlier, serum albumin and prothrombin time (or international normalised ratio, INR) are more accurate measures of function. Platelet count is a sensitive marker of portal hypertension due in part to the resultant hypersplenism (figure 2.2).

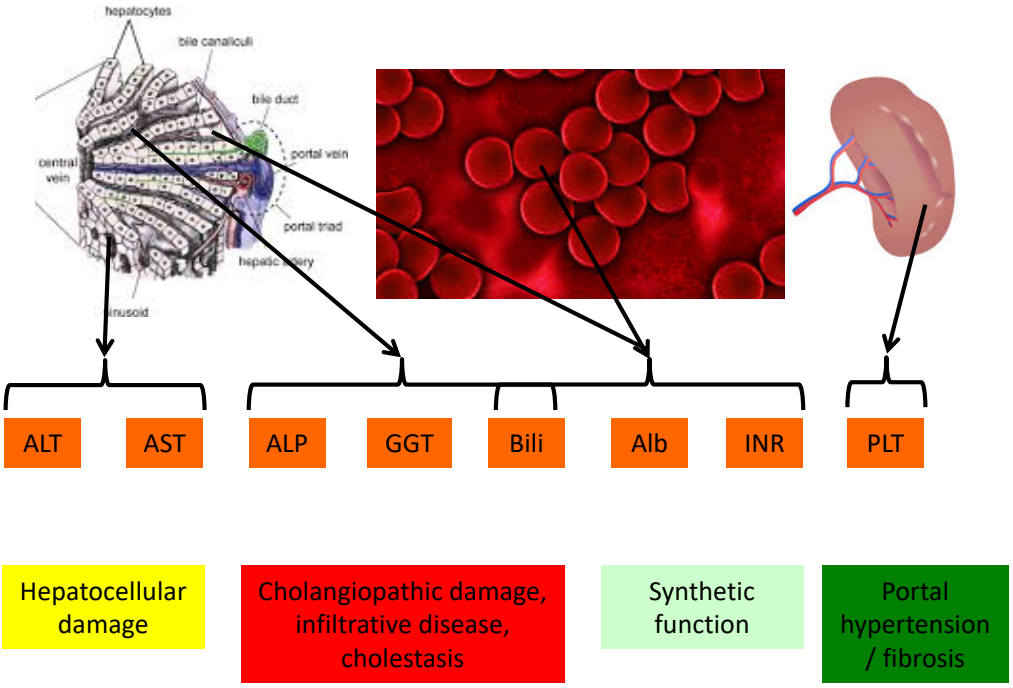


Figure 2.2. Significance and source of abnormalities in commonly used tests of liver disease

ALT, alanine aminotransferase; AST, aspartate transaminase; ALP, alkaline phosphatase; GGT, gamma-glutyl transpeptidase; Bili, bilirubin; Alb, albumin; INR, international normalised ratio; PLT, platelet count.

Liver chemistry in isolation, however, is not a good predictor of liver fibrosis. A large population-based retrospective study investigated patients with no obvious liver disease presenting with abnormal liver chemistry.[41] Over a follow up period of 4 years, just over 1% developed liver disease. Severely and mildly elevated transaminases were associated with liver disease with hazard ratios (HR) of 13 and 4, respectively compared to normal levels. This low sensitivity with high specificity may in part be due to the relatively short follow up period.

Therefore, relying on abnormal liver chemistry to detect liver fibrosis or cirrhosis may be falsely reassuring (figure 2.3) and a number of studies have supported this.[42,43]

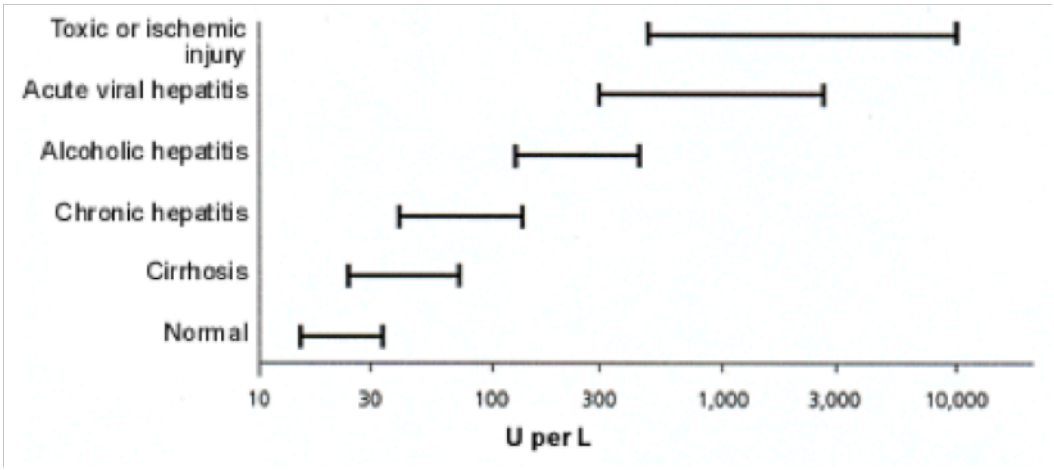


Figure 2.3. Typical ranges of aminotransferase levels for various disease
From Johnston DE.[44] Reproduced with permission

2.7 Introduction to non-invasive assessment of liver fibrosis

The asymptomatic nature of liver fibrosis progression, leading to cirrhosis, results in individuals often presenting with life-threatening features of decompensation in the form of ascites, variceal haemorrhage, hepatic encephalopathy, liver failure or hepatocellular carcinoma.[45] Non-invasive evaluation of liver fibrosis is now established in clinical practice, but remains largely confined to secondary and tertiary care environments, in those with known or suspected liver disease. In view of the rising incidence of CLD,[22,25,26,46] there is an urgent need to identify liver disease and the risk of progressive fibrosis in primary care, not least in those with risk factors, where,

as discussed above, the reliance on measurement of serum liver enzymes may be falsely reassuring.

The next section discusses non-invasive markers of liver fibrosis in the context of two major causes of liver disease, NAFLD and alcoholic liver disease, ahead of comparison of two common markers and exploration of their use in community settings later in the thesis.

2.7.1 Serum-based markers

Serum (non-invasive) biomarkers of liver fibrosis can be divided in to indirect (simple) markers and direct markers.

Indirect markers are those related to hepatic function and inflammation rather than direct involvement in hepatic fibrosis, including animotransferases, bilirubin, GGT, INR and platelet count. These markers are routinely available but serum levels of these markers can be affected by a wide range of factors outside of the liver. Indirect markers can be combined in algorithms with demographic indices, for example age.

Direct markers of fibrosis include the products of matrix synthesis or degradation. These markers can be divided in to enzymatic markers, collagen-related markers, glycoproteins and matrix-metalloproteinase markers and glycosaminoglycan markers. Of these, glycosaminoglycans (hyaluronic acid, HA) and collagen-related markers are the most widely used. As with indirect markers, direct markers can be combined in to algorithms ('marker panels').[47]

2.7.1.1 Serum markers for liver fibrosis in NAFLD

Non-alcoholic fatty liver disease

NAFLD is the hepatic manifestation of the metabolic syndrome. The central pathological feature is hepatic fat accumulation (steatosis), and the prevalence of steatosis is 20-30%. NAFLD is strongly associated with obesity, type 2 diabetes, hyperlipidaemia, insulin resistance and hypertension. Hepatic fat accumulation may lead to inflammation (non-alcoholic steatohepatitis, NASH) resulting in progressive fibrosis and cirrhosis.[48]

At least 90% of individuals with NAFLD will not progress from simple steatosis to NASH, and simple hepatic steatosis may be, in itself, a benign condition. A study of 129 subjects with biopsy-proven NAFLD followed for 14 years found no increase in mortality with simple steatosis on index biopsy, but significantly lower survival in subjects with NASH, due to both liver and cardiovascular diseases.[49] Progression from steatosis to NASH has previously been explained by the 'two hit hypothesis'.[50] This describes the initial 'hit' as fat accumulation which sensitises the liver to a second 'hit' that activates inflammatory cascades leading to fibrogenesis. The situation is likely to be more complicated, with a multiple hit hypothesis being suggested,[51] comprising multiple parallel processes contributing to development of steatosis and steatohepatitis. A combination of dietary, genetic and environmental factors promotes insulin resistance (causing increased hepatic lipogenesis and impaired lipolysis, resulting in fatty acids entering the liver) and obesity (the resulting elevated cholesterol and free fatty acids entering the liver). The consequence is mitochondrial dysfunction with oxidative stress leading to inflammation and

fibrosis. Changes in the gut flora may contribute by increased production and absorption of free fatty acids.

Current management of NAFLD centres on the identification and optimisation of metabolic risk factors, in particular weight loss. In a prospective study of 293 overweight or obese patients with biopsy-proven NASH who were given weight loss advice and reassessed at one year, weight loss was associated with improvement in histological features of NASH.[52] In patients who lost ≥10% of their weight, 90% had resolution of NASH. A study of one hundred and ninety-one morbidly obese patients with biopsy-proven NASH who underwent bariatric surgery demonstrated a significant reduction in mean BMI from 49.3 to 37.4 kg/m^2 one year after surgery.[53] NASH disappeared in 85% of patients, and fibrosis was reduced in 34%. Gastric bypass surgery was more effective than laparoscopic gastric banding. Although there is no liver-specific treatment, there is growing evidence for the roles of statins, vitamin E and insulin sensitisers for example pioglitazone. It is postulated that vitamin E may down-regulate the oxidative stress which is thought to be involved in the 'second hit', where inflammatory cytokines are activated in steatosis, inducing NASH. The PIVENS trial investigated subjects with biopsy-proven NASH and no diabetes, and demonstrated an improvement in histological features of NASH after 96 weeks of treatment with vitamin E, but not with pioglitazone, compared to placebo.[54] Subsequent meta-analyses have shown, however, that doses of vitamin E of ≥400 IU/day may increase all-cause mortality.[55] Antifibrotics are currently being evaluated in clinical trials, with four drugs currently being evaluated in phase three clinical trials (table 2.2). These drugs interfere with processes involved in fatty acid transport and beta-oxidation (elafibranor),

insulin sensitivity and hepatic gluconeogenesis and circulating triglyceride levels (obeticholic acid), migration of pro-inflammatory cytokines to the liver (cenicriviroc), and stress-response pathways leading to apoptosis in the liver (selonsertib).

Table 2.2. Summary of drugs currently being evaluated in phase 3 clinical trials
Mechanism of action, trial name and identifier registered with *clinicaltrials.gov* are presented for each drug

Drug	Mechanism of action	Clinical trial name (reference)	ClinicalTrials.gov Identifier
Elafibranor	PPARα/δ receptor agonist	GOLDEN-505[56]	NCT01694849
Obeticholic acid	FXR receptor agonist	Farnesoid X Receptor Ligand Obeticholic Acid in NASH Treatment trial (FLINT)[57]	NCT01265498
Cenicriviroc	CCR2/5 receptor antagonist	CENTAUR[58]	NCT02217475
Selonsertib	ASK1 inhibitor	STELLAR 3 & 4[59]	NCT03053050 / NCT03053063

PPAR, peroxisome proliferator-activated receptor; FXR, farnesoid X nuclear receptor agonist; CCR, C-C motif chemokine receptor; ASK1, apoptosis signal-regulating kinase 1

Serum markers for NAFLD

Previous studies have evaluated the performance of individual serum markers.[60] In addition to direct markers, Fibrotest-Fibrosure and Hepascore panels include other parameters including age, sex, bilirubin, GGT.[61,62] The NAFLD fibrosis score combines age, BMI, glucose, platelets, albumin and

AST/ALT ratio and was specifically designed for NAFLD, reporting a diagnostic performance based on area under the receiver operator characteristic curve (AUROC) 0.82 for advanced fibrosis.[63] A large French cross-sectional study in 452 subjects with biopsy-proven NAFLD assessed the performance of TE and 8 serum-based markers in diagnosing F3 fibrosis and found that TE and FibroMeter were the most accurate, with AUROC values of 0.83 and 0.82, respectively.[64] Obuchowski indices were 0.83 and 0.80, respectively (see chapter 4 for an explanation of the Obuchowski index).

Comparisons of simple marker panels in patients with NAFLD, and of combination and simple panels have not been extensively studied. In diagnosis of advanced fibrosis, Fib-4 has demonstrated superior performance over the AST to platelet ratio index (APRI) and AST/ALT ratio.[65] A large study comparing APRI and BARD with Hepascore, Fibrotest and Fib-4 found that the combination panels were more accurate than the simple panels, with Hepascore the most accurate and BARD the least accurate for diagnosis of significant, advanced and severe fibrosis.[66] More recently, in 741 patients with a histological diagnosis of NAFLD, a number of non-invasive tests (APRI, AST/ALT, BARD, Fib-4, NAFLD fibrosis score and GGT/platelet ratio) were applied, in addition to transient elastography.[67] Using thresholds for F3 or above from published literature, performance was assessed using AUROC values. The best performing tests were TE, NAFLD fibrosis score and Fib-4 (AUROC 0.86, 0.77, 0.79, respectively). Combining TE with NAFLD fibrosis score or Fib-4 reduced the likelihood of wrongly classifying patients, but increased the area of uncertainty. However, using two tests in series, i.e. using

one test first, then another in those with a result in the indeterminate area, increased diagnostic accuracy and reduced the indeterminate range.

As discussed above, the development of NASH is a key step in the risk of fibrosis. Only 3-5% of the NAFD population will have NASH and 1-2% will have progressive fibrosis.[68] Therefore the diagnosis of NASH at an early stage of fibrosis is clinically useful. A recent study from our group of 172 individuals with biopsy-proven NAFLD found that in patients without advanced fibrosis serum levels of P3NP were associated with a histological diagnosis of NASH, with AUROC 0.77-0.82 in patients with F0-2 fibrosis and 0.82-0.84 in those with F0-3 fibrosis.[69]

The performance of the ELF test in NAFLD is discussed below.

2.7.1.2 Serum markers for liver fibrosis in alcoholic liver disease

Alcohol is a major cause of chronic liver disease, and alcoholic liver disease in common with other CLD aetiologies may remain clinically silent until an event associated with hepatic decompensation occurs. Further, only a minority of those who drink heavily will develop CLD and identifying these individuals at risk is problematic.[70] Detection of fibrosis at an early stage offers the opportunity for intervention to change behaviour and avoid progression to advanced liver disease. Screening tools such as the Alcohol Use Disorders Identification Test (AUDIT) have been shown to be effective,[71] and these could be used in conjunction with serum markers to allow more appropriate targeting of people at risk.

A systematic review has highlighted the paucity of studies examining the performance of serum markers of fibrosis in alcoholic liver disease.[72] Hyaluronic acid (HA) is the most extensively evaluated single marker, and is a component of several marker panels; Hepascore, Fibrometer and ELF. The performance of HA as a single marker is variable, but the panels containing HA consistently perform well, particularly in the diagnosis of cirrhosis.

2.7.2 The Enhanced Liver Fibrosis (ELF) test

The Enhanced Liver Fibrosis (ELF) test comprises a panel of direct serum markers that has now become established in clinical practice as an accurate, reproducible and repeatable non-invasive test of liver fibrosis. In this section I will discuss the development of the ELF test and the liver diseases in which it has been validated.

2.7.2.1 Development of the ELF test

The development of the ELF test was driven by the deficiencies of liver biopsy, which is associated with procedural discomfort and risk, cost, skill to interpret, and misrepresentation of liver disease due to the patchy nature of liver disease and inter- and intraobserver variation of histological assessment. Of note, there was acknowledgement of the ethical and acceptability issues of repeated liver biopsies that limited its use to monitor responses to medications in development.

A number of individual assays for the products of matrix synthesis and degradation that had demonstrated some ability to diagnose fibrosis levels in specific liver diseases were investigated in a cross-sectional study to determine

if a panel of markers may improve performance.[73] Over two years, 1,021 liver biopsy specimens were prospectively collected from subjects under investigation for chronic liver disease who had abnormal liver chemistry for at least six months (the 'Bayer' cohort). The aetiology in this cohort comprised chronic hepatitis C (496), alcoholic liver disease (64), fatty liver disease (61), primary biliary cholangitis or primary sclerosing cholangitis (53), recurrent disease after liver transplantation (48), autoimmune hepatitis (45), heamochromatosis (32), cryptogenic cirrhosis (19), hepatitis B and C (4) and other (138). Biopsies were assessed using the Scheuer classification (stage 0, 24.4%; stage 1, 35.5%; stage 2, 13.4%; stage 3, 14.9%; stage 4, 11.8%), (the Ishak classification was also assessed with similar results) and serum samples taken on the same day as biopsy. Nine different circulating serum markers were measured.

Histology, serum samples and clinical information were available for 921 patients, comprising the analysis cohort. Logistic regression was applied to a sub-set (the test set) to generate an algorithm predicting significant fibrosis, and then applied to the validation cohort.

The final algorithm comprised three markers; hyaluronic acid (HA), amino-terminal propeptide of type III collagen (P3NP) and tissue inhibitor of matrix metalloproteinase 1 (TIMP-1), in addition to age. This demonstrated good discrimination between fibrosis stages and good performance for detection of significant fibrosis in the validation group, with AUROC of 0.804, and for detection of severe fibrosis / cirrhosis, 0.887. Performance to detect significant fibrosis in specific liver diseases within the cohort were; chronic hepatitis C, AUROC 0.773; NAFLD, 0.870; alcoholic liver disease, 0.994.

HA is a high molecular weight polysaccharide and is a major component of extracellular matrix in almost every tissue of the body. In the liver, it is synthesised by hepatic stellate cells and degraded by sinusoidal endothelial cells.[74] Type III collagen is synthesised from procollagen III by the removal of the N- and C-terminal extensions of procollagen III by proteinases during the final stages of collagen synthesis. P3NP is a product of this cleavage.[75] TIMP-1 is also synthesised by hepatic stellate cells, and as with other members of the TIMP family, is involved with inhibition of matrix metalloproteinases. Increased expression of TIMP-1 therefore is associated with reduced degradation of extracellular matrix.[76] The ELF test utilises enzyme-linked immunosorbent assays. The TIMP-1 and P3NP assays use 2 monoclonal antibodies that bind to independent binding sites on their antigens. HA uses HA-binding protein, isolated from cow nasal septum.

The ELF test requires no more than 300 µl serum (to allow for repeat testing and pipetting dead volume). It is a Conformité Européene (CE) marked diagnostic test and is manufactured by Siemens Healthineers Inc., Tarrytown, NY, USA. The manufacturer's thresholds for interpreting the ELF test scores are; <7.7 (none to mild fibrosis), 7.7 - <9.8 (moderate fibrosis), ≥9.8 (severe fibrosis).[77]

The original assays were performed using a Bayer IMMUNO 1 autoanalyser. The algorithm, known as the Original European Liver Fibrosis (OELF) panel, which incorporates age as a variable and is analyser-specific, is:

Discriminate (ELF) score = -6.38 – (ln(age)*0.14) + (ln(HA)*0.616) + (ln(P3NP)*0.586) + (ln(TIMP-1)*0.472)

2.7.2.2 Subsequent validation studies

2.7.2.2.1 Non-alcoholic fatty liver disease

The algorithm was evaluated in an external cohort of 196 patients with a diagnosis of NAFLD on liver biopsy, with serum samples taken within 3 months of biopsy.[78] A five-stage fibrosis scoring system was used, part of the histological scoring system for NAFLD devised by Kleiner and colleagues.[39] This is a modified version of the system devised by Brunt and colleagues, with descriptive subdivisions of stage 1 (although these subdivisions were not used in the ELF study).[38] The OELF panel discriminated between fibrosis scores with good accuracy, as did an algorithm without age. Therefore, diagnostic performance was assessed using this new algorithm, the Enhanced Liver Fibrosis (ELF) score. AUROC values for any fibrosis, moderate fibrosis and severe fibrosis were 0.82, 0.90 and 0.93, respectively. A clinical utility model showed that, by using the ELF test, 82% of biopsies could be potentially avoided for the diagnosis of severe fibrosis.

The ELF panel algorithm, which again is specific for the IMMUNO 1 autoanalyser and reagents, that was generated by regression analysis, is:

Discriminant (ELF) score = -7.142 + (ln(HA)*0.681) + (ln(P3NP)*0.775) + (ln(TIMP-1)*0.494)

2.7.2.2.2 Primary biliary cholangitis

The ELF test was evaluated in an external cohort of 161 subjects with primary biliary cholangitis, in whom serial liver biopsies and serum samples were collected two-yearly for a median of 7 years as part of a study evaluating the

effect of adding methotrexate to ursodeoxycholic acid therapy on disease progression.[79] Performance of the ELF test to predict both significant fibrosis and cirrhosis was good, AUROC 0.75 and 0.76, respectively. In addition, ELF predicted survival, which was significantly lower in those in the highest ELF tertile.

2.7.2.2.3 Paediatric NAFLD

Prevalence of NAFLD in children and adolescents is between 3 and 10%. The ELF test was evaluated in 112 children with a diagnosis of NAFLD in whom liver biopsy was performed.[80] Again, the ELF test discriminated well between fibrosis stages and had excellent performance in diagnosis of fibrosis, AUROC 0.99 for advanced fibrosis / cirrhosis.

2.7.2.2.4 Chronic hepatitis C

A study investigating the ELF tests in patients with chronic hepatitis C first compared the OELF algorithm to the ELF algorithm in subjects with chronic hepatitis C in the original cohort described above.[81] As with the external NAFLD cohort, the performances of OELF and ELF were similar. Using three external cohorts of patients with chronic hepatitis C, similar performance of the two algorithms was confirmed. Performance, in terms of AUROC, of the ELF test to predict advanced fibrosis / cirrhosis was 0.87-0.90. There was also good performance in predicting moderate fibrosis. Algorithms were generated with reference to both Metavir and Ishak systems, and as the assays were performed on the IMMUO 1 analyser, are specific to this platform;

Ishak:

$$ELF = -8.468 + (\ln(HA)*0.892) + (\ln(P3NP)*0.759) + (\ln(TIMP\text{-}1)*0.410) + 10$$

Metavir:

ELF = -7.412 + (ln(HA)*0.681) + (ln(P3NP)*0.775) + (ln(TIMP-1)*0.494) + 10

2.7.2.2.5 Primary sclerosing cholangitis

Primary sclerosing cholangitis (PSC) is a progressive disease and attempts to define predictors of disease have not proved very reliable. Serum samples were used from 167 (derivation group) and 138 (validation group) subjects with PSC, 100 healthy controls and from 96 subjects with active ulcerative colitis and 47 patients with colitis in remission.[82]

At the time of this study, the IMMUNO 1 platform had been superseded by the Siemens ADVIA Centaur autoanalyser, with corresponding new algorithms depending on the Centaur model. These are the currently used algorithms built in to the analysers that generate the ELF scores and published by Siemens;[77]

ADVIA Centaur XP (used in this study):

ELF = 2.278 + (0.851*ln(HA)) + (0.751*ln(P3NP)) + (0.394*ln(TIMP-1))

ADVIA Centaur CP:

ELF = 2.494 + (0.846*ln(HA)) + (0.735*ln(P3NP)) + (0.391*ln(TIMP-1))

Median ELF scores were higher in PSC patients compared to healthy controls and participants with ulcerative colitis (11.1, 10.2, 9.7, respectively). In subjects with PSC, when ELF score was divided in to tertiles, transplant-free survival was significantly longer in the lowest ELF tertile. The ELF test distinguished between mild and severe disease (defined by transplantation or death), with AUROC 0.81.

This study compared the Siemens-manufactured ELF assay with other commercially available assay kits. Overall, performance was similar, although the Siemens assay was superior in separating patients with particularly low risk.

2.7.2.2.6 Alcoholic liver disease

With colleagues, I performed a study of ELF in patients with alcoholic liver disease using the original ELF cohort (see list of publications).

2.7.2.2.6.1 Aims

The aims of the study were to evaluate the performance of the ELF test in diagnosing fibrosis and in predicting clinical outcomes in the sub-set of participants in the original Bayer study with alcoholic liver disease.

2.7.2.2.6.2 Methods

The ELF test was performed on 81 subjects from the original cohort and diagnostic performance assessed by calculating AUROC values for each Scheuer fibrosis stage.

Clinical outcomes were assessed 7 years after liver biopsy by reviewing clinical notes, routine data sources and contacting primary care physicians.

2.7.2.2.6.3 Results

Median age was 47 years. Biopsy stages were; F0, 17%; F1, 21%; F2, 6%; F3, 21%; F4, 35%. The ELF test demonstrated good performance in predicting all stages of fibrosis (table 2.3). ELF predicted liver-related events at 7 years, AUROC 0.81 (95% CI 0.71-0.90). An ELF score $\geq$9.5 was better at predicting

outcome than cirrhosis on biopsy (p = 0.002), correctly predicting outcomes in 84% compared to biopsy predicting 55%.

Table 2.3. Median ELF test scores and performance of the ELF test in predicting liver fibrosis in subjects with alcoholic liver disease

Fibrosis stage	Median ELF score (IQR)	AUROC (95% CI)
0 vs 1-4	9.26 (1.05) vs 11.06 (2.66)	0.89 (0.82-0.96)
0,1 vs 2-4	9.47 (1.12) vs 11.47 (2.05)	0.91 (0.84-0.98)
0-2 vs 3,4	9.49 (0.86) vs 11.74 (2.02)	0.89 (0.82-0.97)
1-3 vs 4	9.85 (1.62) vs 11.75 (1.96)	0.82 (0.73-0.92)

AUROC, area under the received operator characteristic curve; CI, confidence interval; IQR, interquartile range

2.7.2.2.6.4 Conclusion

In this cohort of subjects with alcoholic liver disease, the ELF test correlates closely with histological staging conducted by an expert pathologist at all stages of fibrosis, and performs better than biopsy staging at predicting outcomes.

2.7.2.3 Other ELF studies

2.7.2.3.1 Performance of the ELF test to predict clinical outcomes

Utilising the original ELF cohort, the ELF test has been shown to predict liver-related outcomes with good accuracy.[83] This is discussed in chapter 7.

The ELF test has been shown to predict transplant-free survival in subjects with primary sclerosing cholangitis; AUROC for transplant or death compared to milder outcomes was 0.81.[82]

Using cryopreserved samples from subjects with chronic hepatitis C up to twenty-five years old, the ELF test predicted clinical outcomes; 74% of those

with ELF scores ≥11.3 developed events compared to 3.2% in those with ELF scores <9.7. Further, the bio-stability of the ELF test appeared to be preserved in samples up to 25 years old; samples were compared to paired liver biopsies and the median values of the ELF components and the ELF score remained stable.[84]

ELF was measured two yearly in a cohort of HCV/HIV co-infected subjects for a median of eight years to assess its ability to predict all-cause mortality, demonstrating an AUROC at year of death or last visit of 0.85.[85]

An Australian study aimed to determine the accuracy of the ELF test threshold of 9.8 to identify advanced fibrosis in a secondary care cohort of mixed aetiology liver disease.[86] The study confirmed the reliability of the threshold of 9.8, but suggested that the ELF score may be influenced by age, with an increased false positive rate seen in subjects ≥45 years. This will require further study given the convincing data in previous studies indicating that age can be safely excluded from the algorithm. Interestingly, the ELF test was less likely to misclassify fibrosis in the presence of steatosis. A follow up study of this cohort evaluated the ability of the ELF test to predict liver-related events, showed that 19% of those with ELF ≥9.8 and <1% with ELF <9.8 experienced events, and a unit increase in ELF was associated with a 2.5-fold increase in risk of event.[87]

A study in the general population assessed the ability of ELF to predict development of hepatocellular carcinoma. After adjusting for age, sex, BMI, smoking status, diabetes, coffee and alcohol consumption, an ELF score of ≥9.89 had an odds ratio of 25 for predicting event.[88]

A number of studies have evaluated the performance of the ELF test to identify clinically significant portal hypertension, using the hepatic venous pressure gradient as the reference, with AUROC values of 0.68,[89] 0.88[90] and 0.88[91] demonstrated. The ELF test has been evaluated in a group of obese patients undergoing bariatric surgery and who had suspected NAFLD, with a significantly higher ELF score seen in those with NASH and / or fibrosis on biopsy compared to those with normal histology or steatosis.[92]

As outlined in chapter one, my thesis will further explore the ability of the ELF test to predict liver-related outcomes in a general population

2.7.2.3.2 Evaluation of the ELF test compared to other fibrosis tests

In a prospective study of patients with a mixed aetiology of liver disease, performance of the ELF test was compared with transient elastography (TE) and another imaging-based test, acoustic radiation force impulse (ARFI).[93] For detection of F2 fibrosis and above, analysis of the performance of the three modalities yielded AUROC values of 0.764, 0.861 and 0.879, for ELF, TE and ARFI, respectively. For detection of cirrhosis, AUROC values were 0.841, 0.918 and 0.936. When the ELF test was combined with ARFI or TE, the positive and negative predictive values for both levels of fibrosis were increased, suggesting a role for combinations of tests. Using thresholds of 9.4 for the ELF test and 8.3 kPa for TE, there was disconcordance in 34%. Failure rate of TE was 11%.

The failure rate of TE seen in this study is similar to a study I performed at the Royal Free Hospital (see publication list). 99 patients with mixed liver disease aetiology who underwent ELF test and TE within one year were studied. TE thresholds were based on a study of subjects with mixed viral aetiology (mild

fibrosis, <5.2 kPa; moderate-severe fibrosis, ≥5.2-12.8 kPa; cirrhosis, >12.8 kPa).[94] ELF thresholds were; mild fibrosis, <7.7; moderate-severe fibrosis, 7.7-9.8; cirrhosis, >9.8. Invalid TE results (success rate <60%, interquartile range of median stiffness <30%) were recorded in 66% and excluded from analysis. The Pearson correlation between the ELF scores and median stiffness values was 0.6. The kappa statistic was used as the measure of agreement between the two modalities. For predicting moderate fibrosis, kappa was 0.14 (34% discrepancy) and for cirrhosis was 0.49 (17% discrepancy). Where subjects had also undergone liver biopsy within two years, kappa values for agreement with histology for moderate fibrosis were 0.19 (ELF test) and 0.13 (TE), and for cirrhosis 0.71 (ELF test) and 0.79 (TE). These results indicate poor agreement between TE and ELF based on these thresholds. Both modalities perform well in agreement with cirrhosis on biopsy. A failure rate of TE of 11% was seen.

Using histology as the reference standard, performance of the ELF test, FibroTest, elastography, and other simple serum marker panels (including APRI and Fib-4) were compared in a primary care cohort with a history of excess alcohol use. Using a cut off value of 10.5, the ELF test diagnosed advanced fibrosis with high accuracy (AUROC = 0.89), with similar performance to FibroTest and elastography, and more accurate than the simple marker panels.[95]

A recent study from our group measured the ELF test in patients on second-line therapy for chronic hepatitis C (pegylated interferon with or without silymarin) and found that change in ELF score at one year from baseline predicted the two-year ELF score with high accuracy, raising the potential utility

of the ELF test to be used to monitor antifibrotic therapy and determine early those who will benefit.[96]

2.7.2.3.2.1 Performance of the ELF test compared to simple liver fibrosis markers

A number of studies have compared the performance of the ELF test to simple markers.

A prospective Danish study investigated performance of the ELF test in patients with a history of alcohol excess but no clinical signs of CLD recruited from primary (128) and secondary (161) care.[95] Liver biopsy (staged using the Kleiner classification), ELF testing and TE were performed, and blood tests performed to calculate APRI, age-platelet index, Fib-4, Forns index, AST:ALT ratio and GGT-to-platelet ratio. ELF scores of 10.5 and 9.8 were evaluated as cut-off values for advanced fibrosis (F≥3). Median stiffness of 15 kPa was used for diagnosing advanced fibrosis using TE. Prevalence of advanced fibrosis on biopsy was 23% (6% in primary care, 36% in secondary care). The ELF test accurately diagnosed F≥2, F≥3 and F4 fibrosis: AUROC 0.84, 0.92 and 0.94, respectively. Performance of TE was similar; 0.85, 0.89 and 0.87, respectively. Both ELF and TE were more accurate than the indirect panels. The ELF test performed equally well in both the primary and secondary care cohorts. In the primary care cohort, an ELF cut off of 10.5 demonstrated a negative predictive value (NPV) of 98%, indicating good performance in ruling out advanced fibrosis, but a positive predictive value (PPV) of 60%. All markers had high NPVs in the primary care cohort, but PPVs were low, highlighting a general deficiency in performance of these tests in a low prevalence setting. Reducing

the ELF threshold from 10.5 to 9.8 reduced specificity from 97% to 89% (i.e. reduced ability to rule in advanced fibrosis). The authors proposed an algorithm for primary care using an (inexpensive) indirect maker panel initially to rule out advanced fibrosis, followed by the ELF test, using a threshold of 10.5, in those with high probability of advanced fibrosis based on the first test. ELF and indirect serum marker testing was successful in all subjects. TE was unreliable in 2% and failed in 2.4% yielding reliable results in 95.5% of cases. Although use of the higher ELF threshold improved the PPV in this study, the risk in this strategy is of missing cases of advanced fibrosis. This two stage serial approach, using an initial (cheap) test with (relatively) high sensitivity to rule out disease followed by a test with high specificity to rule in disease has been proposed in the context of NAFLD,[97,98] and is further discussed in the next chapter.

A study of 109 subjects with chronic hepatitis C compared performance of the ELF test, TE, APRI and Fib-4 in diagnosing F≥2, F≥3 and F4 fibrosis using liver biopsy as the reference.[99] AUROC values increased for all modalities at increasing fibrosis stages. AUROC values for diagnosing F≥3 fibrosis were 0.7 (ELF), 0.83 (TE), 0.69 (APRI) and 0.76 (Fib-4). AUROC values for diagnosing F4 fibrosis were 0.94, 0.99, 1 and 1, respectively. However, the ROC curves at this level appear very unstable, presumably due to the small number (2) of subjects with cirrhosis on biopsy, which is not commented on by the authors.

The ELF test was compared to APRI and Fib-4 in 119 females with chronic hepatitis B.[100] TE was used as the reference, based on the premise that TE demonstrates high accuracy to detect fibrosis on liver biopsy in chronic hepatitis

B. AUROC values for detecting F≥3 fibrosis were 0.65, 0.66 and 0.66 for ELF, APRI and Fib-4, respectively. The use of TE as the reference is controversial; although biopsy may not be considered the 'gold' standard, it is generally accepted as the reference standard for studies of non-invasive markers of fibrosis.

2.7.2.3.3 Studies reporting normal values of the ELF test

Several studies have defined the normal ELF score in healthy populations. In a South Korean population where heart disease, diabetes, metabolic syndrome, hepatitis B, hepatitis C and liver dysfunction was excluded, Yoo *et al* reported that the ELF test score was between 5.69 and 8.67 in females, and 6.72 and 8.93 in males, with the highest median score in those over 60 years.[101]

Lichtinghagen *et al* measured ELF scores using serum samples from four hundred blood donors reporting ELF scores of 6.6 - 9.3 in females and 7.0 - 9.9 in males, with a statistically significant difference between genders.[102]

2.7.3 Validation of disease biomarkers with a focus on stability

A review by Duffy *et al* outlined key areas that must be considered when developing a disease biomarker for use in clinical practice.[103] Some of these are relevant to both the ELF test and my study and will be discussed.

Sample collection

The development of a biomarker should include attention to patient-related factors that may influence performance (for example age, sex, fasting status) and specimen-related factors. These include centrifugation parameters and other processing factors, stability during transport and sample storage, which

all require evaluation to ensure samples are handled appropriately when the test is used in clinical practice.

Analytical validation

Analytes comprising the biomarker need to be validated in any assay that will be used in clinical practice to ensure accuracy, repeatability (measurements made under the same conditions), reproducibility (measurements made under different conditions), limit of detection and robustness. The ELF assay has been validated in these key areas (see below).

Clinical validation

This process ensures that the results of the biomarker test stratifies subjects in to different groups, for example disease and no disease. These studies are usually reported as diagnostic accuracy, including sensitivity, specificity, positive predictive value, negative predictive value, likelihood ratio and AUROC. Thresholds (which may be disease-specific) should be determined. These terms will be explained in this thesis as needed.

The context in which the biomarker is being evaluated is important, i.e. for screening, diagnosis, prognosis or prediction. Duffy *et al* warn about the risk of bias during validation and define bias as;

> "A systematic erroneous association of a characteristic in a group that distorts comparison with another group".

Bias may result in positive findings that are unrelated to the clinical reality. This can occur if groups are not appropriately matched (e.g. for sex or age) resulting in any observed differences in marker score possibly being due to these

differences rather than the presence or absence of disease. Another potential source of bias is differences in transportation, storage or processing of samples between the two groups.

Pepe *et al* suggest that bias could be eliminated by using a nested case-control study in which samples are collected prospectively before a diagnosis is established and are then evaluated in a blinded fashion retrospectively (prospective-specimen-collection, retrospective-blinded-evaluation, PRoBE).[104] Only after data are available are random samples from cases and controls selected for the study. This retrospective and random approach minimises the problem of baseline non-equivalence because samples are collected without knowledge of outcome, so both selection of study participants and sample collection should be completely objective. Systematic bias is eliminated by similar handling of samples from cases and controls. This design can be used in the evaluation of screening, diagnostic and prognostic studies and forms the basis of my study and is outlined in chapter seven.

Demonstration of clinical value

In addition to clinical validation the clinical utility of a biomarker needs to be demonstrated. This will be explored in the context of chronic hepatitis B in chapter four. Ultimately, systematic reviews followed by meta-analyses of studies can be undertaken. Duffy *et al* highlight that inclusion of data from unpublished studies is important as this can reduce publication bias, given that positive studies are more likely to be published than negative ones.

Regulatory approval and post-marketing evaluation

In Europe, biomarkers for clinical use require the CE mark, which represents the manufacturer's declaration that the test meets the European Union requirements. As mentioned above, the ELF test is CE marked.

Once the biomarker is approved, rigorous quality control procedures are mandatory, including participation in external quality assessments assessed against national or international standards.

2.7.3.1 Stability of the ELF test assay

One particular aspect of analytical validation of a biomarker that is pertinent to my study is that of robustness. Duffy *et al* define this as;

> "the precision of an assay following changes in assay conditions".

Related to this is the stability of the ELF assay when samples are stored or frozen. There are limited studies evaluating the ELF test in this context.

Our group evaluated the stability of the ELF assay (see list of publications).[105]

Three experiments were conducted. In the first experiment, sample stability under medium to long term storage at -80°C was evaluated. After concentrations of each ELF test component were measured immediately after collection from 10 individuals, samples were frozen for a mean of 220 days. After thawing, the ELF component concentrations were measured, and the average values compared to the baseline measurements. In addition, the ELF score was calculated in baseline and frozen samples.

In the second experiment, samples from 5 subjects were subjected to three cycles over three days of freezing at -80°C for two hours, thawing at room temperature for three hours and freezing for at least two hours, with ELF constituent concentrations measured after each cycle.

In the third experiment, samples were refrigerated at 4°C for four days after being frozen at -80°C, then measured on the day of thawing and at days 1, 2 and 4.

Results from experiment 1 showed that the level of P3NP did not change, but TIMP-1 increased and HA decreased, with no significant change in ELF score. Results from experiment 2 showed that although there was some variation in P3NP, again ELF score did not significantly change. Results from experiment 3 showed some variation in TIMP-1 and P3NP, but no significant variation in the ELF score.

Although this study highlighted some instabilities of the ELF components, the calculated ELF score remained stable in what can be considered common storage conditions of samples used for ELF testing, including refrigeration storage.

As described above, Puigvehi *et al* evaluated the long-term biological stability of the ELF test using cryopreserved samples, noting stability of the assay components in samples cryopreserved for over twenty years.[84]

The ELF score was measured in 949 healthy blood donors in Brazil and (after exclusion of nine donors with scores considered as outliers) ELF ranges were 6.52 – 10.85, 6.52 – 10.6 and 6.54 – 10.85 in the whole cohort, females and

males, respectively.[106] The difference between genders was statistically significant. There was a positive correlation between ELF and BMI (which was seen in the study by Loo *et al* but not Lichtinghagen *et al*). Analyte stability was verified in three samples that were split in to several aliquots and subjected to increasing numbers of freeze / thaw cycles. Two additional aliquots from each sample were kept at room temperature for 24 hours. The analytes remained stable in all sets of conditions.

Finally, a recent study further evaluated the biological variation and analytical performance of the ELF test. Initially, the reference ranges for the ELF assay components and ELF score were determined in forty healthy volunteers aged 20-50 years.[107] Blood samples were then obtained from a subgroup of 20 subjects weekly for seven weeks. Blood sampling and sample preparation was standardised; all subjects were in a sitting position for 5-10 minutes before phlebotomy using 21-guage needles, and samples were centrifuged within 60 minutes before aliquoting. Serum aliquots were stored at -80ºC before ELF testing. The ELF assay was performed using the ADVIA Centaur CP analyser and Siemens reagents. The ELF score was calculated using the appropriate algorithm. The ELF scores ranged from 7.14 to 9.55. The within-subject and between-subject variations were highest in the HA assay, however these effects were diminished in the ELF equation because of the use of the natural logarithm. The authors conclude that the ELF test has suitable analytical and acceptable biological performance characteristic for clinical practice.

In summary, the values of the ELF score in the normal populations studies appear to lie in the range of moderate fibrosis in the liver disease cohorts. This may in part be due to the ELF test being evaluated using liver biopsy in liver

disease, which is a categorical score. Establishing the range of ELF scores in the general population may be of use in the context of screening. The ELF assay appears stable in a range of storage conditions.

2.7.3.2 National Institute for Health and Care Excellence guidelines for NAFLD

The National Institute for Health and Care Excellence (NICE) published guidelines for assessment and management of NAFLD in 2016.[108] The document is primarily aimed at general practitioners, with guidance for the assessment of advanced liver fibrosis in people with NAFLD. It suggests that people with NAFLD should be offered testing for advanced fibrosis, and that the ELF test should be used in this context. The authors made this recommendation based on effectiveness and cost effectiveness following comparison with other methods, despite using an inflated cost for ELF in their modelling (£92 compared to the NHS price of £42 per test). No alternative tests are suggested. An ELF test threshold of ≥10.51 is stated to diagnose advanced fibrosis / cirrhosis. This differs from the data derived threshold of 9.8 that is recommended by Siemens but is based on analysis of data contained in the study by Guha *et al* in adults and Nobili *et al* in children. In those with an ELF test score below 10.51, it is recommended that retesting occurs every 3 years.

Alternative approaches to stratification of patients with NAFLD in primary care have been proposed, and some of these are discussed in chapter 3. In chapter 7, the two ELF thresholds described above are explored in detail.

2.7.3.3 ELUCIDATE study

The ELUCIDATE study (Enhanced Liver Fibrosis (ELF) Test to Uncover Cirrhosis as an Indication for Diagnosis and Action for Treatable Events) is a randomised controlled clinical trial with the primary aim of evaluating the ELF test in early detection of CLD, and to assess the ability of this strategy to implement prophylaxis for oesophageal varices, ascites and hepatic encephalopathy, to prevent variceal haemorrhage and to facilitate earlier detection of HCC.[109] The study comprises two arms; the ELF arm, where patients undergo follow-up screening for CLD with the ELF test, and the standard care arm, where patients undergo standard follow-up screening for cirrhosis. The study protocol includes an economic evaluation to demonstrate cost-effectiveness of early intervention. Recruitment is now complete. Early analysis revealed that the process of care was executed in accordance with the protocol. The planned long-term analysis powered to evaluate clinical outcomes will be performed in 2020.

2.7.4 Transient elastography

Transient elastrography (TE) is a technique for evaluating the stiffness of soft tissue. The FibroScan machine measures shear wave velocity through the liver. A transducer on the end of an ultrasound probe sends a 50 MHz wave in to the liver. The probe also has a transducer that measures the velocity of the shear wave. This velocity is converted to stiffness of the liver in kilopascals, with the assumption that increased stiffness is associated with increased fibrosis. The median stiffness of ten readings is used as the result, and the test takes around five minutes. The manufacturer (Echosens) considers a test successful when

at least 60% of attempts to measure velocity are successful and the interquartile range of the results obtained is less than 30% of the final median stiffness. A volume equivalent to 1/400 of the liver is analysed using this technology.

Contraindications to TE include pregnancy, presence of ascites, and patients with cardiac pacemakers. Further technical aspects of FibroScan are discussed in chapter 4.

Since the development of FibroScan further elastography techniques including acoustic radiation force impulse (ARFI), real-time elastography (RT-E), magnetic resonance elastography (MRE), and shear wave elastography (SWE) have been developed.

2.7.4.1 National Institute for Health and Care Excellence guidelines

NICE guidelines for assessment and management of cirrhosis, 2016, recommend offering TE to diagnose cirrhosis in people with chronic hepatitis C, those consuming excess alcohol or those with alcoholic liver disease.[110] It is also advised for those in whom an ELF test score of 10.51 has been found via the NAFLD guidelines.

2.8 Conclusion

CLD is often clinically silent until life-threatening and usually irreversible complications have developed. Earlier detection of CLD offers the opportunity to intervene with measures to reduce the underlying factors driving the liver disease and therefore prevent progression to advanced fibrosis and cirrhosis. Most cases of chronic liver disease are preventable.

Furthermore, detection of cirrhosis prior to the development of its most serious complications, namely variceal haemorrhage and hepatocellular cancer permits interventions that have been shown to reduce morbidity and mortality including prescription of beta-adrenoreceptor blockers, variceal band ligation and resection, ablation or transplantation, respectively.

Non-invasive tests for liver fibrosis have distinct advantages over liver biopsy, however their evaluation is limited by their reliance on liver histology as their reference standard. In an attempt to circumvent the subjectivity inherent in histological staging of liver biopsies attempts have been made to use collagen staining, and automated measurement of the proportion of a biopsy stained for matrix but even this approach requires field sampling to eliminate large connecting biliary triads and liver capsule and it still necessitates liver biopsy with all the associated problems.

Instead of seeking more objective measures of liver fibrosis as a reference standard some investigators have acknowledged that liver fibrosis is often investigated as a surrogate marker for clinical outcomes. Acknowledging that fibrosis assessment is often evaluated to prognosticate morbidity and mortality some studies have chosen clinical outcomes as the objective measure against which tests for liver fibrosis are evaluated. The challenges for these studies are the scarcity of events and the long duration of follow-up required to generate meaningful and informative data. This thesis will move on to comparison of two commonly used fibrosis markers, the ELF test and TE, and then focus on investigation of liver disease in a community setting. It will expand on the concept of clinical outcomes having more ‘real world’ value than liver fibrosis

and evaluate the performance of a non-invasive marker to generate outcome data that could be utilised in practice.

Chapter 3. CRITICAL APPRAISAL OF SYSTEMATIC REVIEW AND STRUCTURED LITERATURE REVIEW

3.1 Overview

This section comprises a critical appraisal of a recent systematic review of non-invasive markers of liver fibrosis in community settings, and a structured literature review of studies published since this systematic review was performed.

3.2 Introduction

The vast majority of assessments of liver fibrosis occur within the secondary care environment, comprising of either histological evaluation using liver biopsy or using non-invasive markers of fibrosis. The aims of this thesis include the estimation of incidence of CLD in a community-based population and the evaluation of a non-invasive marker of liver fibrosis, the ELF test, in this setting.

There is an urgent need to identify both established CLD and early CLD, particularly in those at highest risk of progressive liver fibrosis, in the community. As discussed earlier, the use of standard liver chemistry to diagnosis liver disease is not reliable in the detection of CLD, and LFTs can often be normal in established liver disease.[44] The clinical significance of this suboptimal strategy is highlighted by the finding that nearly 50% of individuals only receive a diagnosis of CLD when they present to hospital with a decompensating event.[45]

Barriers to improving early diagnosis of CLD include lack of data on prevalence of CLD in the community, an absence of a screening strategy to identify and stratify individuals, reliance on unreliable tests for liver disease and paucity of studies evaluating the clinical utility of non-invasive markers of liver fibrosis in the primary care setting.

3.3 Critical appraisal of systematic review

The aim of this critical appraisal of a systematic review is to establish its methodological quality to determine validity of the results and conclusions. The use of a similar search strategy employed in this systematic review to conduct a structured literature review of studies published after this systematic review would provide a valuable update on the topic, but only if this systematic review was conducted with methodological rigour.

The systematic review by Harris *et al*[111] was published in 2017 and comprised a systematic review of non-invasive markers used to stratify patients at risk of CLD in the general population and using these data, to estimate the prevalence of CLD in this setting. The primary aim of the review was to evaluate the proportion of the populations studied who had CLD as defined by the non-invasive markers. The secondary aims included identifying the proportion of individuals with CLD diagnosed by non-invasive markers who had normal transaminases.

3.3.1 Method

Although systematic reviews are considered to provide the highest level of evidence for evidence-based medicine, there are few tools available to assess

the quality of a systematic review, and no consensus for a gold standard. Examples of tools include the critical appraisal of systematic reviews checklist from the Centre for Evidence-Based medicine (CEBM), 'A Measurement Tool to Assess Systematic Reviews' (AMSTAR), 'Critical Appraisal Skills Programme' (CASP),[112] the National Institute for Health and Care Excellence Methodology Checklist' (NICE) and the tool from the Scottish Intercollegiate Guidelines Network (SIGN). In considering which tool to use, I compared a number of tools in particular CEBM, AMSTAR, CASP and SIGN. A particular strength of the CASP and SIGN tools are whether the conclusions can be trusted based on assessment of methodological quality. SIGN does not have open-ended questions. AMSTAR does not question whether the study is relevant to my clinical question, and furthermore it focuses on evaluation of randomised controlled trials and meta-analyses. The most relevant and comprehensive checklist found was that from the Centre for Evidence Based Management (Critical Appraisal of a Met-analysis or Systematic Review), based on the principles outlined by Crombie[113] combining elements of the CEBM, Cochrane Centre and BMJ Editor's checklists[114] which was used in this study.

3.3.2 Results

Evaluation of the validity of the systematic review

1. *Did the review address a clearly focused question?*

   Yes. This review aimed to explore the use of non-invasive markers of liver fibrosis in the general population, with clear and specific questions; the primary aim was to report the prevalence of the populations studied

found to have clinically significant liver disease defined by the non-invasive tests used.

2. *Was a comprehensive literature search conducted using relevant research databases?*

   Yes. A number of databases were searched and search dates reported (Embase, 1 January 1980 to 21 January 2015; MEDLINE, 1 January 1946 to 21 January 2015; Web of Science, 1 January 1900 to 21 January 2015). In addition, reference lists of original studies were searched, as were conference proceedings. Both MeSH terms and text words were used. However, the search was limited to English language only (appendix B).

3. *Is the search systematic and reproducible?*

   Yes. The supplementary material included the search algorithms employed for the electronic databases. A flow diagram was presented with numbers of records identified through each database search and numbers of records excluded.

4. *Has publication bias been prevented as far as possible?*

   To some extent. This review reported disease prevalence using non-invasive tests rather than performance, therefore publishing bias may be less relevant in this context. The omission of a negative study related to performance may result in bias. Studies were restricted to English. In addition, there was no inclusion of unpublished data which may result in

publishing bias and may compromise the comprehensiveness of the review.

5. *Are the inclusion and exclusion criteria clearly defined?*

   Yes. Inclusion criteria were defined and included adults, studies performed in non-hospital settings, use of validated non-invasive markers, and recruitment of participants from an unselected population or on the basis of defined risk factors. Exclusion criteria were defined; if the study setting or non-invasive marker threshold were not reported, and if the study was not written in English.

Interpretation of the results of the review

6. *Was the methodological quality of each study assessed using predetermined quality criteria?*

   Yes. To have been included, studies must have evaluated a validated non-invasive marker and presented the prevalence of liver disease based on a defined threshold for that marker.

7. *Are the key features of the included studies described?*

   Yes. The review presents the risk factor prevalence for each study, the outcome measure (any liver fibrosis, significant fibrosis, advanced fibrosis), the non-invasive test threshold and the disease prevalence.

8. *How are the results presented and is this appropriate to the data?*

   The review included 19 relevant studies. Results were presented in table form, which was an appropriate format because it allows easy cross-

referencing of studies, and included relevant measures for example prevalence of risk factors in the population, threshold used for the non-invasive test and estimated disease prevalence.

9. *Were the results similar from study to study?*

   Of the 19 studies, transient elastography was evaluated in 5. The serum-based non-invasive markers were the NAFLD fibrosis score, FibroTest, BARD score, AST/ALT, APRI, Fib-4, BAAT, hyaluronic acid, ELF Test and the Southampton Traffic Light test. Five studies investigated an unselected population, the remainder stratified the population (e.g. age, diagnosis of NAFLD, diagnosis of type 2 diabetes, hazardous alcohol use). All studies reported liver fibrosis prevalence as defined by the threshold for the non-invasive test used. The studies of an unselected population all used transient elastography, with a large range in fibrosis prevalence (2-19%), but used liver stiffness thresholds from 6.8 to 9.6 kPa. Similarly, large ranges in fibrosis prevalence were seen in studies that stratified patients with risk factors. Prevalence of cirrhosis was reported in seven studies, with higher prevalence in studies comprising individuals with risks.

Applicability of the results of the review in clinical practice

10. *Are the results clinically relevant?*

    Yes. This study aimed to generate data to inform strategies to stratify individuals in community settings at risk of CLD. By evaluating studies that investigated the use of non-invasive markers in these populations it

has indicated that prevalence of CLD is likely to be higher than previous estimates. However, the lack of consensus on thresholds, contributing to the large range of liver disease prevalence between studies highlights the uncertainty around which is the most appropriate test. The authors suggest that the differing thresholds may be due to the heterogeneous populations used to derive thresholds.

In the studies of general populations, 40-74.6% of individuals with liver fibrosis according to the non-invasive test had normal serum ALT levels. In studies of those with risk factors, this range was 26.5-87.5%.

11. *Does this review answer my clinical question?*

Yes. This study highlights several points relevant to this thesis. It reinforces that serum transaminases are poor predictors of liver disease. The small number of studies in general or community settings compared to secondary care settings highlights the need for more work in this area, strengthened by the finding of a higher prevalence of liver disease in these populations.

3.3.3 Conclusion

Critical appraisal of this systematic review finds that it employed a systematic and explicit methodology, with appropriate interpretation of the data, providing valid conclusions that can be applied to the population of interest. It highlights the need for further work to determine incidence and prevalence of CLD in community settings, and to evaluate non-invasive markers of liver fibrosis particularly in those with risk factors.

The positive critical appraisal of this review will allow me to perform a structured literature review of subsequent studies, using a similar search strategy to permit both integration and comparison of any subsequent data.

3.4 Structured literature review

I undertook a literature review to build on the systematic review of Harris *et al* and will focus on the same aims; to report the prevalence of liver disease estimated in each study defined by the non-invasive test being used.

In contrast to the evaluation of non-invasive markers of liver fibrosis in secondary care populations, their use in primary care and community populations has not been extensively explored. Further, no reference standard for fibrosis assessment exists in community studies, which renders assessment of performance of non-invasive markers difficult in these lower prevalence settings. It could be postulated that, in the diagnosis of liver disease, there are more likely to be a higher proportion of false positives resulting in a lower positive predictive value for non-invasive markers.

3.4.1 Method

3.4.1.1 Search strategy

The search strategy employed was that outlined by Harris *et al*;[111] the search algorithms can be found in appendix B. Using these algorithms, the on-line databases Embase and MEDLINE were searched. Harris *et al* searched database records between 1 January 1990 and 21 January 2015 (Embase) and 1 January 1946 and 21 January 2015 (MEDLINE), therefore I searched records between 30 January 2015 and 17 March 2018.

Web of Science was searched for ‘non-invasive markers’ and within results, for ‘liver’ and ‘community’.

3.4.1.2 Inclusion and exclusion criteria

Inclusion criteria comprised;

- Study written in English
- Adults over 18 years
- Community, primary care or outreach setting
- Study was of a validated non-invasive marker of fibrosis
- Prevalence of liver disease, defined by the threshold of the non-invasive marker was reported
- Participants were either unselected or had risk factors of alcoholic or non-alcoholic liver disease

Exclusion criteria comprised;

- Threshold for non-invasive marker not reported
- The study population was investigated for liver disease that was not either alcoholic or non-alcoholic liver disease

3.4.1.3 Data extraction

Data extracted included details of the non-invasive marker used and the prevalence of liver disease based on the threshold employed.

3.4.2 Search results

The searches identified 94 (MEDLINE), 62 (Embase), and 10 (Web of Science) citations. After accounting for duplicate citations, 141 citations remained for

screening. Review of titles and abstracts identified 28 citations that potentially met the inclusion criteria. These studies were reviewed in full, resulting in 11 that met the inclusion criteria. Two additional studies were identified after reviewing reference lists. The results of the search, with reasons for exclusions, are shown in figure 3.1. There were 4 abstracts and 9 papers.

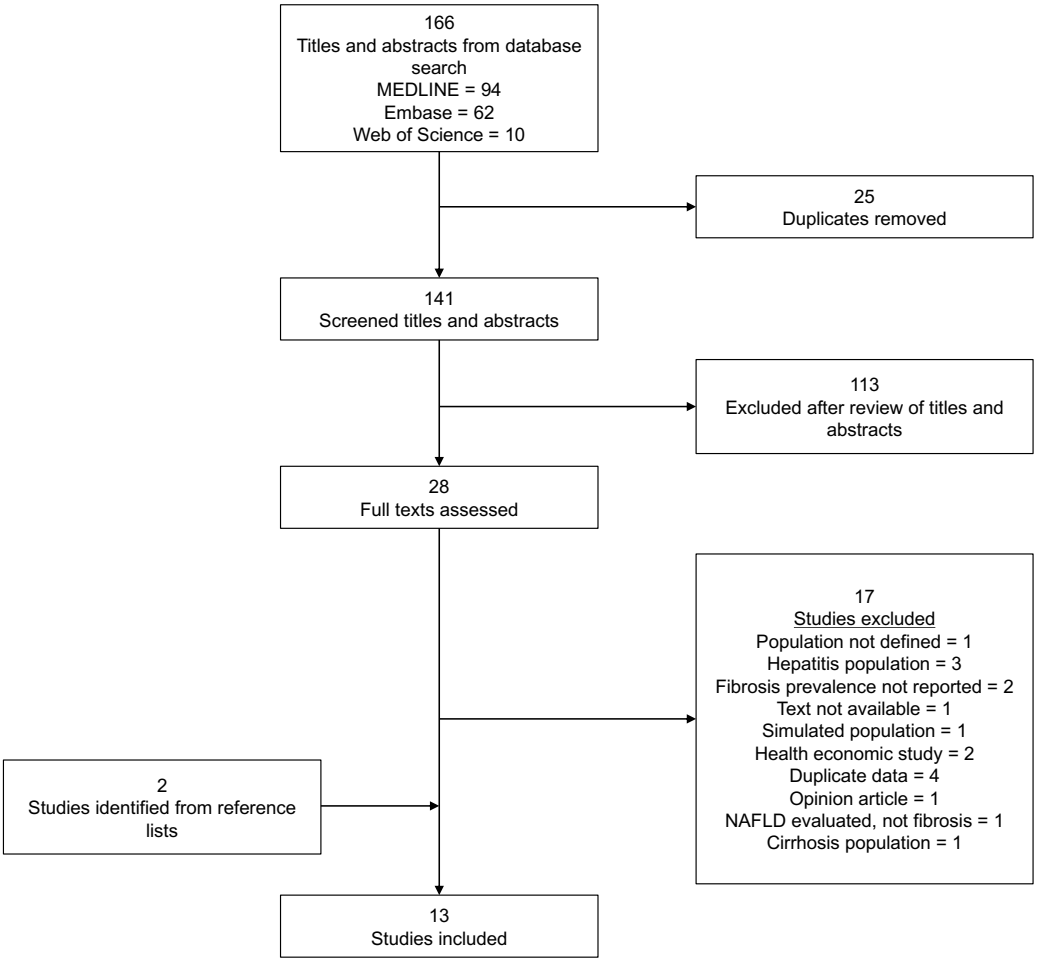


Figure 3.1. Studies screened and assessed for structured literature review
Three databases were searched, MEDLINE, Embase and Web of Science. After removing publications for reasons described in the figure, thirteen studies were included in the analysis

Table 3.1. Overview of studies included in structured literature review of prevalence of liver disease in community-based populations using non-invasive markers

Author & year	Population (Location)	Sample size	Mean age	Male %	Risk factor (prevalence)	Non-invasive modality	Threshold	Corresponding fibrosis severity	Prevalence %
Srivastava *et al.* [115] 2015 (abstract)	Primary care (UK)	112	Not reported	Not reported	NAFLD (100%)	Fib-4	>3.25	Cirrhosis	1.8
Roulot *et al.* [116] 2017	Primary care (France)	705	58	56	Type 2 diabetes (100%)	TE	>8 kPa	Significant fibrosis	12.7
							>9.5	Advanced fibrosis	7.3
							>13	Cirrhosis	2.1
Harman *et al.* [117] 2015	Primary care (UK)	504	Not reported (screening group)	70 (screening group)	One or more of hazardous alcohol (>14 units/week), type 2 diabetes, raised ALT	Blood-based biomarker screen:		To rule out significant fibrosis.	88% had high blood-based biomarker)
						AST:ALT for hazardous alcohol use	≥0.8		
						BARD for other risk	≥2		

Author & year	Population (Location)	Sample size	Mean age	Male %	Risk factor (prevalence)	Non-invasive modality	Threshold	Corresponding fibrosis severity	Prevalence %
						factors			
			62 (TE group)			TE if high blood-based biomarker result	>8	Significant fibrosis	26.8 (of 378) in those undergoing TE after high blood-based marker
Caballeria *et al.* [118] 2016 (abstract)	Primary care (Spain)	3076	54	43	Unselected	TE	>6.8		9.3
							>7.6		7.1
							>8.0		6.0
Patel *et al.* [119] 2017 (abstract)	Primary and secondary care (Australia)	197	58 (LSM <8.2) 58 (LSM ≥8.2) 56 (ELF <9.8) 66 (ELF ≥9.8)	Not reported	Tertiary care diabetes clinic or primary care patients with suspicion of NAFLD	Blood-based biomarker screen, with Fib-4 and NAFLD fibrosis score then referred to tertiary centre for ELF test and TE	Pre-screen: Indeterminate or high Fib-4 or NAFLD fibrosis score		4% (diabetes clinic) 58% (primary care)
						TE	≥8.2	Advanced	28.9

Author & year	Population (Location)	Sample size	Mean age	Male %	Risk factor (prevalence)	Non-invasive modality	Threshold	Corresponding fibrosis severity	Prevalence %
								fibrosis	
						ELF test	≥9.8	Advanced fibrosis	30.6
Doycheva *et al.* [120] 2016	General population and primary care (USA)	100	60	53	Type 2 diabetes	Magnetic resonance elastography	≥3.6 kPa	Advanced fibrosis	7.1
Harman *et al.* [121] 2018	Primary care (UK)	919	60 (LSM <8) 63 (LSM ≥8)	74 (LSM <8) 70 (LSM ≥8)	Hazardous alcohol use and/or type 2 diabetes	TE	≥8 kPa	Advanced fibrosis	25.6
Srivastava *et al.* [97] 2016 (abstract)	Primary care (UK)	452	Not reported	Not reported	NAFLD and abnormal transaminases	Fib-4 ELF test	Fib-4 ≥3.25 or (Fib-4 1.30-3.25 & ELF >9.5)	Advanced fibrosis	25.2
Cheng *et al.* [122] 2016	General population screening for liver disease	559	56	38	Unselected (those with risk factors including HBV, HCV, alcohol misuse were	TE	≥8 kPa	Significant fibrosis	4.7
							≥13 kPa	Advanced fibrosis/cirrhosis	1.3

Author & year	Population (Location)	Sample size	Mean age	Male %	Risk factor (prevalence)	Non-invasive modality	Threshold	Corresponding fibrosis severity	Prevalence %
	(Taiwan)				excluded)				
Morling *et al.* [123] 2016	Diabetes register (UK)	923	69	52	Older patients with type 2 diabetes	Range of biomarkers including: AST:ALT ELF test Fib-4 Hyaluronic acid NAFLD fibrosis score		Cirrhosis	2.2
Chen *et al.* [124] 2015	General population from a population based cross-sectional survey (China)	2550	55 (low NFS) 62 (indeterminate NFS) 71 (high NFS)	33 (low NFS) 31 (indeterminate NFS) 29 (high NFS)	NAFLD	NAFLD fibrosis score	Low (<-1.455)	No advanced fibrosis	48
							Indeterminate (-1.455-0.676)	Indeterminate	48
							High (>0.676)	Advanced fibrosis	4
Koehler *et*	General population	3342	66	45	Unselected	TE	≥8 kPa	Clinically significant	5.6

Author & year	Population (Location)	Sample size	Mean age	Male %	Risk factor (prevalence)	Non-invasive modality	Threshold	Corresponding fibrosis severity	Prevalence %
al. [125] 2016	from a population-based cohort study (The Netherlands)							fibrosis	
							≥13 kPa	Cirrhosis	0.6
Conti *et al.* [126] 2016	General population from a cross-sectional community-based survey (Italy)	331	47	31	Healthy cohort	TE	≥8 kPa	Significant fibrosis	4.5

ALT, alanine aminotransferase; AST, aspartate transaminase; ELF, Enhanced Liver Fibrosis; Fib-4, Fibrosis-4 index; HBV, hepatitis B virus; HCV, hepatitis C virus; kPa, kilopascal; LSM, liver stiffness measurement; NAFLD, non-alcoholic fatty liver disease; NFS, NAFLD fibrosis score; TE, transient elastography

3.4.2.1 Comparing and contrasting of studies

3.4.2.1.1 Non-invasive markers used

Transient elastography was the most commonly used test,[116-119,121,122,125,126] employed in eight studies. The only other imaging-based test used was magnetic resonance elastography.[120]

The serum-based tests using simple markers that were studied were Fib-4, AST:ALT, BARD and the NAFLD fibrosis score. Hyaluronic acid was used in one study as part of a range of tests,[123] and the ELF test was evaluated in three studies.[97,119,123]

Several studies used more than one non-invasive test, with some employing multiple-stage algorithms to stratify participants.

3.4.2.1.2 Populations studied

Although all studies were based in the community, the characteristics of study groups varied, with some selecting cohorts on the basis of one or a number of risk factors being present. Five studies analysed data from a general population or a population considered to be healthy (using participants in population-based cohort or cross-sectional studies).[120,122,124-126] Four studies included participants with, or with suspicion of, NAFLD.[97,115,119,124] Six studies included individuals with type 2 diabetes,[116,117,119-121,123] and two studies selected individuals reporting hazardous alcohol use.[117,121] Two studies included participants with abnormal liver chemistry.[97,117]

3.4.2.1.3 Fibrosis prevalence

General populations

In the studies of a general or healthy population, prevalence of fibrosis was assessed using TE in all studies, and liver stiffness >8 kPa ranged between 4.5 and 6%. Harris *et al* reported a wider range of fibrosis prevalence, of 2-19%, however prevalence estimates were based on a number of liver stiffness thresholds depending on the study. Only two studies evaluating the general population in the review by Harris *et al* reported prevalence of advanced fibrosis, with prevalence of 0.9% using FibroTest and 2% using TE. As four of the five studies in my search reported results using the same test and the same threshold, a more direct comparison was possible.

NAFLD populations

Comparison of studies of NAFLD populations is more difficult as there was more variation in the non-invasive test used and fibrosis level being assessed. In addition, some cohorts comprised additional risk factors for liver disease. In the two studies of NAFLD-only participants, prevalence of cirrhosis was 1.8% in the study using Fib-4 and 4% in the study using NAFLD fibrosis score. The study using either Fib-4 or NAFLD fibrosis score reported an indeterminate or high score in 74%. Prevalence of advanced fibrosis was 25% in the population with NAFLD and abnormal liver chemistry.

3.4.2.1.4 Hazardous alcohol use

The association with alcohol was evaluated in two studies, both from the same centre, and both reporting similar prevalence of significant fibrosis (18% and

19%) using TE with a threshold of ≥8 kPa.[117,121] Hazardous alcohol use was defined as consumption of ≥14 units / week (women) or ≥21 units / week (men) or an AUDIT questionnaire score of ≥8 or Read code related to hazardous, harmful or dependent alcohol use. Fibrosis prevalence increased considerably if alcohol risk was combined with other risks. In the review by Harris *et al*, four studies reported prevalence of any fibrosis of 80%-100% (only one study reported prevalence of significant fibrosis which was 14.4%. This study used TE with a threshold of ≥5.9 kPa).

3.4.2.1.5 Type 2 diabetes

As with the NAFLD cohorts, studies of patients with type 2 diabetes used a range of non-invasive markers and participants sometimes had additional risk factors. In addition, there appears to be a greater spread of prevalence compared to other risk groups. In the studies by Harman *et al*,[117,121] using TE, prevalence of significant fibrosis was 32%, and 34% in the study where participants were first screened with a simple marker panel. Patel *et al*[119] used either TE with a threshold of 8 kPa or the ELF test with a threshold of 9.8, reporting advanced fibrosis in 29% and 31%, respectively. However, in a study using magnetic resonance elastography, prevalence of advanced fibrosis was 7%.[120] This is more consistent with results from Roulot *et al*, who reported in a population of individuals with type 2 diabetes using TE, prevalence of fibrosis of 13% and 7% using thresholds of 8 and 9.5 kPa, respectively.[116] A low prevalence was also seen in a study using the ELF test or simple serum marker panels which found a prevalence of cirrhosis of 2.2% in a cohort of older people with type 2 diabetes.[123] Table 3.2 shows baseline characteristics of the

populations in these studies. Apart from a higher BMI in the study by Patel *et al*, characteristics are similar, suggesting that differences in prevalence may be related to test modality and that direct comparisons between MRE and other tests may need further evaluation.

Table 3.2. Baseline characteristics of participants in studies of non-invasive markers in the community in individuals with type 2 diabetes

Study	Mean age (years)	Male %	Mean BMI	Hypertension %
Harman *et al.*[121]	59.7	53	30.8	66
Patel *et al.*[119]	58	N/A	32.5 (TE <8.2 kPa) 38.5 (TE >8.2 kPa)	N/A
Roulot *et al.*[116]	58	56	29.6	N/A
Morling *et al.*[123]	68	52	31	Mean SBP 138 mmHg

kPa, kilopascal; SBP, systolic blood pressure; TE, transient elastography

3.4.2.1.6 Elevated serum ALT

Harman *et al* reported fibrosis prevalence in participants with a range of risk factors including elevated serum ALT levels. After a positive screen using a simple marker panel, in those participants where raised ALT was the only risk factor (i.e. absence of type 2 diabetes or hazardous alcohol use) prevalence of significant fibrosis measured using TE was 31%.

3.4.2.2 Review of studies

All thirteen studies reported prevalence of fibrosis, however there was variation in markers used, thresholds selected and fibrosis levels considered of interest. Several studies employed a multi-step algorithm. A retrospective study based in an inner-city primary care centre reviewed all referrals from general practitioners to a secondary care hepatology service, and retrospectively applied the Fib-4 test.[115] 66% of 112 patients had a low Fib-4 score, consistent with low risk of advanced fibrosis. Fib-4 was indeterminate in 31% and high in 3%. These data suggest that two thirds of referrals were potentially avoidable if it is accepted that patients without significant liver disease can be managed in primary care. These data, which suggested a large indeterminate score in this population, informed a prospective study comprising a two-step approach.[97] Patients with NAFLD and abnormal transaminases were entered into the pathway. The initial stratification stage was use of Fib-4. Those with a Fib-4 score predicting low risk of significant fibrosis (Fib-4 <1.30) remained in primary care for management. Those with a high risk of significant fibrosis (Fib-4 ≥3.25) were referred to secondary care, and those with an indeterminate score underwent an ELF test. Of the 452 patients entering the pathway in the first year, 25% had an indeterminate Fib-4 score and overall 75% had a low risk of significant fibrosis and 25% had high risk. In those seen in secondary care, 53% had a subsequent diagnosis of cirrhosis compared to 4% of those referred outside of the pathway.

Other studies have also reported two-stage algorithms to screen individuals. A UK study evaluated prevalence of significant fibrosis in participants with risk factors, both individually and in combination.[117] A simple blood marker panel

was used as a screening tool, followed by TE in those with a high screening result. A high screening blood test marker was seen in 88%. In those that proceeded to TE, 27% were found to have significant fibrosis, but there was considerable variation depending on risk factor. Prevalence was 34% in those with type 2 diabetes and 18% in those with hazardous alcohol use. In those with elevated ALT, prevalence was 31%. In those with more than one risk factor, prevalence was 49%.

An Australian study also used a two-stage stratification approach, employing a simple marker panel, Fib-4 or NAFLD fibrosis score in patients in primary care or patients attending a tertiary hospital diabetes clinic.[119] These patients then underwent TE and ELF testing, with 29% and 31% being diagnosed with advanced fibrosis using thresholds of 8.2 kPa and 9.8, respectively. In patients who then underwent liver biopsy, 81% and 82%, respectively, had advanced fibrosis or cirrhosis histologically. Interestingly age appeared to be independently associated with ELF score.

The ELF test was also evaluated as part of a study involving 931 participants in the Edinburgh Type 2 Diabetes Study, with the aim of determining the prevalence and incidence of cirrhosis in a group of older people with type 2 diabetes.[123] Participants were followed up over 6 years. Fibrosis was assessed using a range of markers including AST:ALT, Fib-4, hyaluronic acid, ELF test and NAFLD fibrosis score and a diagnosis of cirrhosis was made by combining these markers with radiological and clinical assessment. Prevalence of CLD at baseline was 2.2%, with 15 subsequent incident cases (1.4%; 2.9/1000 person years) over the follow up period. Almost 70% of CLD was attributable to NAFLD in this population.

One further community-based study used simple marker panels, measuring the NAFLD fibrosis score in 2550 individuals with a diagnosis of NAFLD on ultrasound scan.[124] A high NAFLD fibrosis score was seen in 4%, and was indeterminate in 48%. Those with advanced fibrosis were more likely to have higher BMI, waist circumference and HOMA-IR, with increased prevalence of type 2 diabetes, hypertension and cardiovascular disease. The aim of this study was, by employing a non-invasive marker of liver fibrosis, to determine whether advanced fibrosis was associated with subclinical atherosclerosis. A positive association between advanced liver fibrosis and measures of atherosclerosis was seen and is a reminder of the link between NAFLD and cardiovascular disease.

A number of TE-based studies were identified, most assessing general populations. However, two studies evaluated individuals with risk factors. One study comprised patients with type 2 diabetes, and reported prevalence of fibrosis using TE thresholds of 8 kPa, 9.6 kPa and 13 kPa, with prevalence of 12.7%, 7.3% and 2.1%, respectively.[116] The authors used multivariate analysis of a range of covariates to derive a predictive model for significant fibrosis based on age, BMI and serum GGT. The authors suggest that this could be a useful screening tool for patients with diabetes where AST is not routinely available. Area under the receiver operator characteristic curve (AUROC) was 0.712 indicating relatively good test performance. Liver biopsy was offered to those with liver stiffness ≥8 kPa, and advanced fibrosis was seen histologically in 51%. In those with liver stiffness ≥13 kPa, 57% had cirrhosis on biopsy and 43% had advanced fibrosis. No cirrhosis was seen in patients with liver stiffness <13 kPa.

A large study selected patients from both suburban and inner-city general practices, using electronic records to identify patients with hazardous alcohol use (>14 units/week, an AUDIT score ≥8 or a Read code for hazardous, harmful or dependent alcohol use) and / or type 2 diabetes.[121] Using a threshold of 8 kPa, 26% were found to have advanced fibrosis (19% in hazardous drinkers, 32% in patients with type 2 diabetes and 38% in those with both risk factors). In those with elevated liver stiffness, 3% were found to have cirrhosis. This study adds weight to the risk of synergism of risk factors, and to the value of screening in high risk groups within the general population.

All studies of the general population used TE. In a study of patients attending primary care physicians for any reason, 52% agreed to undergo liver biopsy.[118] The percentage of patients with significant fibrosis (F≥2) was related to the TE threshold, being 31%, 38% and 44% for thresholds of 6.8, 7.6, and 8, respectively. The threshold with the greatest accuracy for diagnosing significant fibrosis was 9.2 kPa, with 92% sensitivity and 80% specificity and AUROC 0.87.

In a Taiwanese study of a general population of 559, those with excess alcohol use were excluded, and TE performed.[122] In addition, blood tests for liver chemistry were taken. Waist and hip circumference was measured and abdominal ultrasonography performed. 5% had liver stiffness ≥8 kPa, and 1% had liver stiffness ≥13 kPa consistent with advanced fibrosis / cirrhosis. Liver stiffness increased with increasing BMI or waist circumference, and was higher in those with type 2 diabetes or in the presence of hepatic steatosis on ultrasound scan, or with raised transaminases.

In a study nested in the Rotterdam Study (a prospective cohort study investigating cardiovascular, endocrine, hepatological, endocrine and other disease areas) consecutive participants visiting the research centre were included over a two-and-a-half-year period and data including medical, smoking and alcohol histories gathered in addition to anthropometric measurements, blood sampling and liver ultrasound scan.[125] TE was performed, using thresholds of ≥8.0 kPa and ≥13.0 kPa for diagnosing clinically relevant liver fibrosis and cirrhosis, respectively. Failure rate of TE was 5%. In this general population over 51 years old, 36% had steatosis on ultrasound scan. Liver stiffness was ≥8.0 kPa in 5.6% and ≥13.0 kPa in 0.6%. In those with NAFLD, liver stiffness was ≥8.0 kPa in 8.4%.

A study, nested in an Italian community-based survey investigating liver disease, studied participants with NAFLD, and a healthy cohort by excluding participants with any systemic disease.[126] Median liver stiffness was 5.1 kPa in those with NAFLD compared to 4.4 kPa in the healthy group. The authors used the 95th percentile of the liver stiffness range as the threshold (6.8 kPa) for diagnosing significant fibrosis. Using this threshold, 5% of the healthy group and 18% of the NAFLD group had significant fibrosis.

Finally, one study aimed to assess feasibility of liver fibrosis assessment using magnetic resonance elastography in a primary care setting.[120] In this cross-sectional study, 100 consecutive patients with type 2 diabetes were recruited from primary care and newspaper advertisements. After clinical history, anthropometric measurements and blood testing, participants underwent magnetic resonance elastography which has been shown to accurately

distinguish between non-advanced and advanced fibrosis. Prevalence of advanced fibrosis was 7%.

Summary

These studies show that there are a number of non-invasive markers of liver fibrosis that are able to detect liver fibrosis in community settings and stratify individuals. The studies raise a number of general limitations however. Prevalence estimates are likely to vary depending on the modality used and the thresholds selected. The lack of clear thresholds in these populations is a limitation. The different cut-off values seen in different studies may be related to the different prevalences of liver fibrosis in the study populations, so different thresholds may be related to spectrum bias (discussed in chapter four) rather than truly different thresholds related to mechanistic differences in fibrosis biology between aetiology. Validation of these markers in secondary care settings has been achieved by using liver biopsy as the reference standard. Use of liver biopsy in community settings is unlikely to be feasible, and the translation of data from secondary to primary care settings for this purpose is likely to be unreliable due to the difference in prevalence of liver disease. Longitudinal cohort studies in community settings are required.

3.5 Conclusion

The overall theme of studies was the potential clinical utility of non-invasive markers of fibrosis in community settings and their value as a screening tool. Many studies revealed the increased rate of advanced fibrosis in participants with risk factors, highlighting the role of using non-invasive markers for screening in high risk groups. These studies will help to define thresholds for

use in community settings. No studies, however, reported the performance of non-invasive markers to predict CLD. I explore this potential clinical utility using a serum-based marker, the ELF test later in the thesis.

Chapter 4. DIAGNOSTIC PERFORMANCE OF THE ENHANCED LIVER FIBROSIS TEST IN PATIENTS WITH CHRONIC HEPATITIS B

4.1 Introduction

Chronic hepatitis B (CHB) caused by infection with the hepatitis B virus (HBV) is characterised by periods of continuous or fluctuating inflammation of the liver, leading to fibrosis, which may remain occult, with no signs or symptoms at the time of diagnosis of CHB. Morbidity and mortality in patients with CHB is related to persistence of viral replication and the development of liver fibrosis that, as has been described earlier, may progress to cirrhosis and its complications, particularly portal hypertension and liver cancers including hepatocellular cancer, and an increased risk of intra- and extrahepatic biliary cancer.[127,128]

In common with other liver diseases, the assessment of liver fibrosis is, therefore, an essential component in the initial evaluation of patients with CHB and informs the decision to commence antiviral therapy. Liver fibrosis assessment using invasive or non-invasive tests is a key feature of international guidelines.[129,130] Continued monitoring of fibrosis is critical in order to determine changes in fibrosis over time and to assess the efficacy of therapy and the necessity for interventions to manage portal hypertension and screen for liver cancer and progression to cirrhosis.

As discussed, although the traditional method for assessing liver fibrosis has been needle biopsy of the liver, this is expensive, frequently painful and potentially hazardous for the patient, and subject to sampling error and variation

in interpretation.[28,31] As with other forms of liver disease, transient elastography (TE) and serum markers are now being evaluated in patients with CHB.[131-133]

Previous studies comparing the performance of non-invasive markers of liver fibrosis in CHB have reported contradictory results. Performance defined by the area under the receiver-operator curve (AUROC) of TE to identify F≥2 has been reported in several studies to range from 0.61 to 0.87.[134-139]

4.2 Aims of study

The aim of this study was to evaluate and validate the performance of the ELF test in a cohort of patients with CHB and to compare the ELF test with a different non-invasive modality, TE, in the assessment of liver fibrosis defined by histological staging of liver biopsies.

4.3 Principles of statistical methods used in this chapter – sensitivity, specificity, positive predictive value, negative predictive value and receiver operator characteristic curves

The performance of diagnostic tests can be described by their sensitivity and specificity. Sensitivity is the true positive rate, that is the proportion of the sample with the 'disease' who test positive. A highly sensitive test will correctly identify a larger proportion of those with the disease, and by inference a negative test result accurately rules-out the disease.

Specificity is the true negative rate, describing the proportion of the sample without the disease who test negative. A highly specific test will correctly identify a larger proportion of those without the disease who test negative, and a positive test result will accurately rule-in the disease.

The positive predictive value (PPV) is the proportion of the sample with a positive test result that is truly positive. Using a test with a high PPV, it is more likely that the individual who tests positive will have the disease. The negative predictive value (NPV) is the proportion of the sample who test negative who do not have the disease. As discussed in chapter 3, the PPV and NPV (unlike sensitivity and specificity) depend on the pre-test probability of the disease (i.e. the prevalence) and also the performance of the test (sensitivity and specificity). In a low prevalence population, the observed PPV of even a good test will be lower than that observed in a high prevalence population.

These characteristics of a diagnostic test can be calculated after construction of a 2x2 contingency table. If the data is non-binary it will need to be dichotomized, and the relationship between two binary variables can be evaluated to allow the comparison between the 'diagnostic test' and the 'reference standard' or 'true status'. The structure of the 2x2 table is shown in figure 4.1.

		Disease state		
		Positive	Negative	TOTAL
Test result	Positive	True positive (TP)	False positive (FP)	TP+FP
	Negative	False negative (FN)	True negative (TN)	FN+TN
	TOTAL	TP+FN	FP+TN	TP+FP+FN+TN

Figure 4.1. 2x2 contingency table comparing two variables; the true disease state and the diagnostic test result

The sensitivity, specificity, PPV and NPV as defined above, can now be calculated;

- Sensitivity = TP / (TP+FN)
- Specificity = TN / (FP+TN)
- PPV = TP / (TP+FP)
- NPV = TN / (TN+FN)

The diagnostic accuracy of the test, the proportion of true positive and true negative tests is (TP+TN) / (TP+FP+FN+TN).

To increase the PPV and the NPV of a test, two thresholds can be employed, a low threshold with a good sensitivity to rule out the disease, and a high threshold with good specificity to rule in the disease. This strategy creates a range between the thresholds where the results are indeterminate as to whether the disease exists. Resolution of these results may be achieved using an alternative test.

In order to select two thresholds for a test, the sensitivity and specificity values are calculated for a range of thresholds which is then plotted (sensitivity v 1-specificity or true positive rate v false positive rate), producing a receiver operating characteristic (ROC) curve. The area under the ROC (AUROC) curve is a measure of the overall diagnostic accuracy of the test. In the context of liver fibrosis, the AUROC represents the probability that the non-invasive test will correctly rank two randomly chosen individuals, one with a liver biopsy showing advanced fibrosis (for example) and the other with a biopsy showing no advanced fibrosis. The diagnostic accuracies of two non-invasive tests are compared by comparing two AUROC values (and an appropriate statistical

test).[140] An AUROC value above 0.9 is considered an excellent test. A value between 0.8 and 0.9 is considered good. A value of less than 0.5 indicates that the performance of the test is worse than chance. The Youden index is the point on the curve that maximises both the sensitivity and specificity of the test. This 'optimal cut off' is however a trade-off between the sensitivity and specificity.

The clinical utility of the test can then be evaluated by producing models that have an upper threshold with an acceptably high specificity and therefore high PPV to rule in the disease, and a lower threshold with an acceptably high sensitivity and therefore high NPV to rule out the disease.

One strategy that may have beneficial resource implications in clinical practice is the use of two thresholds for the first test, with individuals who have an indeterminate result progressing to the next modality. The cross-sectional evaluation of liver fibrosis requires conversion of a continuous test score (ELF or TE) in to a categorical variable, i.e. to produce a cut off value for the presence of 'disease', and to maximise sensitivity and specificity, two cut off values can be used, but this results in some indeterminate scores, where a second test is needed. By combining the contingency tables of the two tests, the overall diagnostic accuracy of the algorithm can be calculated. This strategy is shown in figure 4.2, in the context of the ELF test and TE, and will be explored in this chapter.

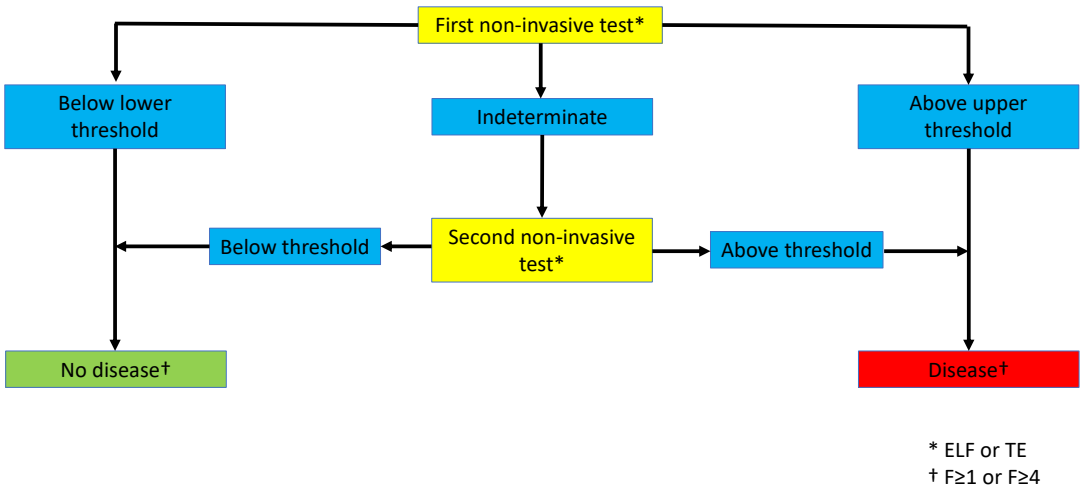


Figure 4.2. Algorithm for using two non-invasive markers in series
The first non-invasive marker, either ELF or TE, results in cases classified as having no disease, disease or indeterminate. The second non-invasive marker classifies the indeterminate cases as disease or no disease

4.4 Methods

4.4.1 Ethical approval

Ethical approval was sought by the custodians of the dataset, under the lead investigator, Professor Pietro Lampertico, and was approved by the local ethics committee in Milan. All participants gave their written consent to the study.

4.4.2 Study population

This study utilised serum samples and digital data from a cohort of participants with CHB and was conducted in collaboration with the cohort's guardian Professor Pietro Lampertico, AM and A Migliavacca Center for Liver Disease, 1st Division of Gastroenterology, Department of Medicine, Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico, Università degli Studi di Milano, Milan, Italy.

The cohort has been described elsewhere. [132] Subjects were recruited at a single Italian centre. Among 224 treatment-naive patients with CHB who were consecutively referred for a liver biopsy and TE evaluation to the Liver Center, Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico, Milan, those with a stored serum sample available for ELF testing were included. Patients with hepatitis C virus, hepatitis delta virus and human immunodeficiency virus coinfections, other concomitant liver diseases, current or previous hepatic decompensation, current or previous antiviral treatment and / or an absolute contra-indication to liver biopsy (platelet count <60 x 10^9/l, INR >1.35) were excluded. In all patients, serum sampling, liver biopsy and TE were performed on the same day.

4.4.3 Liver biopsy

Under the care of Professor Lampertico, all patients underwent an ultrasound-guided liver biopsy with a semiautomatic modified Menghini system (16G, Bio-Mol, Hospital Service, Pomezia, Italy, Philips iU22, Bothell, WA, USA) to stage severity of hepatitis. All the procedures were carried out by two highly experienced hepatologists. Liver specimens were considered of adequate size if longer than 2 cm. Patients with a smaller specimen underwent a repeat procedure during the same session. Five-micron thick sections of formalin-fixed, paraffin-embedded liver tissue were stained with haematoxylin–eosin and Masson tri-chrome, and read by a single liver pathologist blind to TE and ELF results and clinical data. Grading and staging were evaluated according to Metavir system including staging (F0 = fibrosis absent, F1 = portal fibrosis without septa, F2 = portal fibrosis with few septa, F3 = severe fibrosis, F4 = cirrhosis).[141]

4.4.4 Transient elastography

Transient elastography was performed in Milan, under the direction of Professor Lampertico. After an overnight fast, patients underwent a FibroScan® (Echosens, Paris, France) utilising a 5-MHz ultrasound transducer probe mounted on the axis of a vibrator that was operated by three experienced hepatologists who were blind to clinical, biochemical and histological data. [142,143] Mild amplitude and low-frequency vibrations (50 Hz) are transmitted to the liver, thus inducing an elastic shear wave propagating through the underlying liver tissue. Velocity of the wave is directly related to tissue stiffness. The tip of the transducer was covered with a drop of gel and placed perpendicularly in the intercostal space with the patient lying in dorsal decubitus with the right arm in maximal abduction. Under control time motion and A-mode, the operator chose a liver portion within the right liver lobe at least 6 cm thick, free of large vascular structures and gallbladder. Ten successful acquisitions were performed on each patient. The success rate (SR) was calculated as the ratio of the number of successful acquisitions over the total number of acquisitions. The median value, expressed in kPa, was kept as representative of the liver stiffness. The manufacturer recommends that liver stiffness measurements are considered reliable using the following criteria: (i) number of valid acquisitions at least 10, (ii) SR at least 60% and an interquartile range of the median of 30% or less.

4.4.5 Blood markers

In Milan, quantitative polymerase chain reaction amplification for HBV DNA was performed using Amplicor HBV Monitor® (Roche Diagnostics, Branchburg, NJ,

USA), and serology for hepatitis B e-antigen (HBeAg) status was assessed with standard assays, and serum alanine aminotransferase (ALT) and aspartate transaminase (AST) levels were measured using standard enzymatic immunoassays.

4.4.6 ELF testing

Serum samples were shipped to the central ELF laboratory at UCL and analysed for levels of HA, TIMP-1 and P3NP using the proprietary assays developed for the ELF test by Siemens Healthineers Inc. Samples were analysed on an ADVIA Centaur® immunoassay system (Siemens Healthineers Inc.). Results were entered into the manufacturer's published algorithm to derive an ELF score (see algorithm in chapter two).

4.4.7 Statistical analysis

Following ELF testing of the serum samples I used data from the central dataset in Milan for the statistical analysis. I created a database which combined the ELF results with anonymised data for corresponding participants. I performed one-way analysis of variance (ANOVA) to evaluate the null hypothesis that there is no significant difference between liver fibrosis groups for baseline characteristics. I then performed post hoc comparisons to evaluate pairwise differences among group means using the Tukey test. I determined median values and interquartile ranges for each diagnostic test for each liver fibrosis stage. I then assessed the diagnostic performances of ELF and TE by deriving the area under receiver operator characteristic (AUROC) curves. AUROC and 95% confidence intervals of AUROC were calculated. Comparisons of AUROC values for ELF and TE were determined for each stage of fibrosis using the

DeLong method to calculate the chi-squared value for the comparison and expressed as the significance of difference (*p* value).[140]

I determined the optimal cut-off values for discriminating positive and negative cases at each fibrosis stage for ELF and TE by identifying the point of maximum sensitivity and specificity on the ROC curve, and calculated sensitivity, specificity, PPVs and NPVs, and positive and negative likelihood ratios.

I evaluated the clinical utility of each test by analysing performance, by selecting an upper threshold with high specificity, therefore high PPV to 'rule in' fibrosis and a low threshold with high sensitivity and therefore high NPV to 'rule out' fibrosis.

Logistic regression analysis was conducted to further investigate the relationship both between individual non-invasive modalities and fibrosis, and within a model combining both ELF and TE.

4.4.7.1 Spectrum bias

Several methodological issues have been raised in relation to the application of ROC curve analysis to compare non-invasive tests with liver biopsy. The spectrum effect (the differences in the distributions of fibrosis stages in the sample and reference populations) may result in the performance of a non-invasive test varying between the populations giving rise to apparent differences in performance of tests between different sample populations, called spectrum bias. In addition, ROC analysis assumes the reference standard to be binary, whereas the Metavir scoring system employs a five-stage ordinal scale. To overcome these potential flaws and allow more accurate

comparison of studies, a number of statistical methods have been developed to allow a correction to be applied to data.

4.4.7.1.1 The difference between advanced and non-advanced (DANA) fibrosis

The difference between advanced and non-advanced (DANA) fibrosis stages[144] is defined as the difference between the mean fibrosis stage of advanced fibrosis minus the mean fibrosis stage of non-advanced fibrosis and was devised to standardise AUROC values for tests identifying advanced fibrosis.

Mean fibrosis stage for advanced fibrosis:

- ((F2 prevalence x 2) + (F3 prevalence x 3) + (F4 prevalence x 4)) / (F2 prevalence + F3 prevalence + F4 prevalence)

Mean fibrosis stage for non-advanced fibrosis:

- ((F0 fibrosis x 0) + (F1 fibrosis x 1)) / (F0 prevalence + F1 prevalence)

DANA ranges from 1 to 4. A DANA of 1 would be obtained if advanced fibrosis cases were all F2 and non-advanced cases were all F1 (i.e. a central clustering, with no F0 or F3/4). A DANA of 4 would be seen if advanced fibrosis cases were all F4 and the non-advanced fibrosis cases were all F0 (i.e. extremes of the ranges). When prevalences of each fibrosis stage are equal, DANA is 2.5.

The standardised AUROC is calculated using a regression formula which includes the standard DANA value of 2.5 to give an adjusted AUROC for the diagnosis of advanced fibrosis which is independent of the prevalence of fibrosis stages defining advanced and non-advanced fibrosis.

DANA was initially derived using a population of patients with chronic hepatitis C, but was subsequently modified in a group of patients with chronic hepatitis B, generating the following equation which I will use in my analysis;[145]

- Adjusted AUROC = observed AUROC + 0.0482(2.5 - DANA)

4.4.7.1.2 The Obuchowski measure

ROC analysis is based on the use of a reference standard test with a binary outcome, against which the diagnostic test under evaluation is compared. The accuracy of the test is measured by estimating sensitivity and specificity at various thresholds of the diagnostic test. However, as discussed in chapter 2 (Classification of liver fibrosis), the usual reference standard for evaluating non-invasive markers of liver fibrosis, histological assessment is based on an ordinal scale (and also depends on the distribution of fibrosis stages in the study group). To create a binary outcome, the fibrosis stages have to be divided in to two groups (e.g. no advanced fibrosis and advanced fibrosis) and this may lead to bias. As outlined above, the AUC can also be biased if the fibrosis distribution in the sample population differs from the reference population. The DANA correction aims to overcome this by standardising the AUROCs, but is not yet well validated. The Obuchowski measure summarises all pair-wise comparisons of fibrosis stages defined by biopsy, with a weighting scheme and penalty function based on a reference distribution.[146] This reduces the spectrum bias related to distribution of fibrosis stages. If the same Obuchowski measure (using a standard weighting scheme) is used as standard, results from different studies could be compared. I applied the Obuchowski measure using previously described penalty functions[147] to correct for the degree of difference

between the histological stages ascribed by pathological staging and conversion of ELF test scores.

4.4.7.2 Exploring a serial algorithm model

I applied the serial model described above (figure 4.2) to this cohort. Data-derived thresholds from the clinical utility modelling (to produce dual thresholds), and the Youden index (to produce a single threshold) were used as thresholds for tests one and two, respectively, in a serial algorithm model (shown in figure 4.2). The diagnostic characteristics, calculated by using 2x2 tables for each test were combined to calculate the overall diagnostic accuracy of the algorithm for detecting any fibrosis and severe fibrosis.

4.5 Results

Of the 224 subjects consecutively recruited, 188 had a stored serum sample. TE acquisition was unsuccessful in 6 of these subjects (3%); therefore, paired ELF and TE data were available for 182 subjects. Replacing values for missing TE results by both imputation of simple mean and expectation maximisation methods did not change the significance of difference between ELF and TE in ROC analysis, therefore only subjects with paired results were used in the analysis.

4.5.1 Baseline characteristics

Baseline characteristics of subjects are shown in table 4.1. All subjects had a diagnosis of CHB and were treatment-naïve. Median age was 46 years, 71% were male, and 71% were HBV e-antigen negative. Seventy-nine (43%) were overweight (BMI ≥25 kg/m^2). Biopsies reported any fibrosis (Metavir F≥1) in

90.1%, moderate fibrosis (Metavir F2) in 25.8% and severe fibrosis / cirrhosis (Metavir F≥3, equivalent to Ishak stage 4-6) in 36.8%. Median age increased with increasing fibrosis stage from F0 to F3, but was lower in the F4 group compared to the F3 group. These differences were significant. AST (upper limit of normal = 38 IU/l) level increased with fibrosis stage and again was significant, but there was no overall significant difference in ALT (upper limit of normal = 40 IU/l) level or HBV DNA level between the groups.

There was a significant correlation between TE and ALT values (Pearson correlation 0.213, p = 0.004) and between ELF scores and ALT values (Pearson correlation 0.260, p <0.001).

Table 4.1. Baseline characteristics of subjects who underwent TE and ELF testing

Data are presented for all subjects. Liver biopsy reported any liver fibrosis in over 90% of subjects. There were significant differences in means between fibrosis groups for age and AST, but not AST or HBV DNA level, at the 5% level. Post hoc comparisons to evaluate pairwise differences among group means using the Tukey test, showed significant pairwise differences between mean groups for age as follows; between F0 and all other groups, between F1 and F0/F3/F4, between F2 and F0, between F3 and F0/F1, and between F4 and F0. For AST, significant differences were seen between F0 and F4, between F1 and F3/F4, between F3 and F1, and between F4 and F0/F1. The full results of pairwise comparisons is shown in appendix C

Characteristic		All subjects	By Metavir stage					Between groups *p* values
			F0	F1	F2	F3	F4	
Number of subjects		182	18 (9.9)	50 (27.5)	47 (25.8)	31 (17.0)	36 (19.8)	N/A
Age, median (range)		46 (18-67)	32.5 (21-54)	44.0 (18-65)	46 (20-67)	55 (27-65)	50 (29-65)	<0.001
AST (IU/l), mean (SD)		69.7 (64.1)	47.3 (31.9)	49.2 (31.6)	66.4 (38.5)	86.6 (71.1)	97.1 (105.5)	0.002
ALT (IU/l), mean (SD)		110.3 (103.4)	86.4 (78.1)	86.7 (72.0)	110.2 (68.0)	148.1 (167.0)	122.6 (112.4)	0.082
e-antigen status	**Positive, *n* (%)**	53	7 (13.2)	12 (22.6)	10 (18.9)	12 (22.6)	12 (22.6)	N/A
	Negative, *n* (%)	129	11 (8.5)	38 (29.5)	37 (28.7)	19 (14.7)	24 (18.6)	N/A
HBV DNA (log_{10} copies/ml) mean		7.96	7.97	7.82	8.07	7.93	7.98	0.885

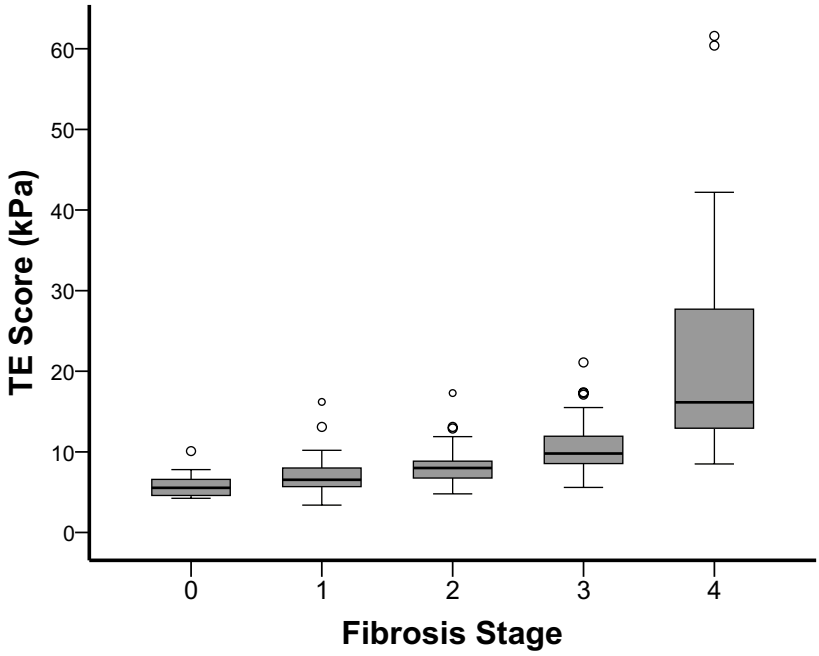
4.5.2 Discrimination between fibrosis stages

Both ELF and TE discriminated different fibrosis stages well, with linear progression, (figure 4.3) and both modalities performed well in predicting fibrosis stage.

4.5.3 Performance of the ELF test and TE to diagnose liver fibrosis

The AUROC for the diagnosis of each stage of fibrosis for ELF and TE are shown in table 4.2 and figure 4.4. The AUROCs for the diagnosis of any fibrosis for ELF and TE were 0.77 and 0.86, respectively (p = 0.09). The AUROCs for the diagnosis of severe fibrosis / cirrhosis for ELF and TE were 0.80 and 0.90, respectively (p <0.01). The AUROCs for the diagnosis of cirrhosis (Metavir F4) were 0.83 and 0.95, respectively (p <0.01).

a)



b)

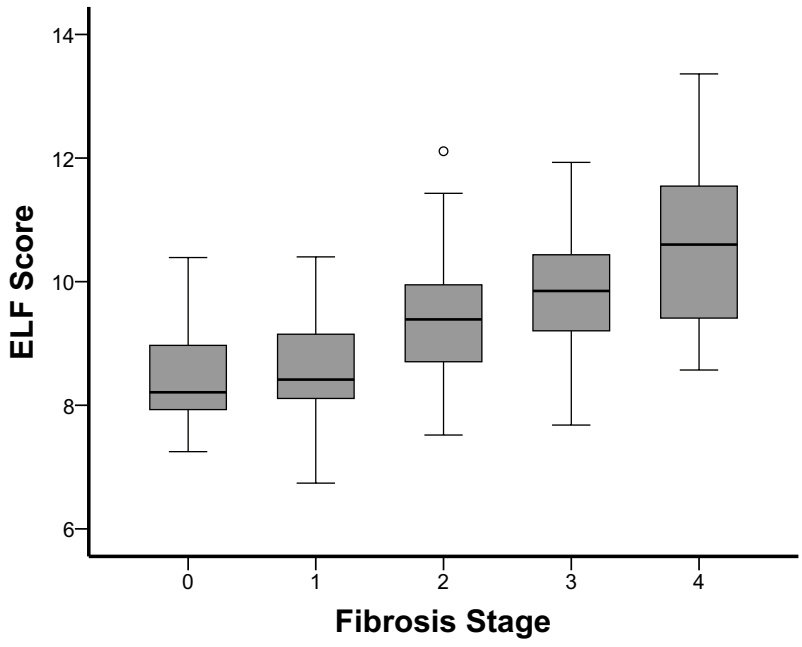


Figure 4.3. Box plots showing median and quartiles for a) TE and b) ELF scores for diagnosing Metavir fibrosis stages

Table 4.2. Median scores and diagnostic performance of the ELF test and TE according to Metavir fibrosis stage
Data are presented describing the performance of ELF and TE in the differentiation of histological stages of liver fibrosis

Fibrosis stage	ELF score (*n* = 182)			TE (kPa) (*n* = 182)			*p* value[†]
	Median (IQR)	AUROC (95% CI)	Adjusted AUROC	Median (IQR)	AUROC (95% CI)	Adjusted AUROC	
0 *vs* 1-4	8.21 (1.08) *vs* 9.39 (1.81)	0.77 (0.67-0.87)	0.81	5.55 (2.08) *vs* 8.50 (5.93)	0.86 (0.78-0.94)	0.90	0.09
0,1 *vs* 2-4	8.35 (1.13) *vs* 9.82 (1.53)	0.82 (0.763-0.88)	0.86	6.30 (2.47) *vs* 9.80 (6.43)	0.86 (0.80-0.91)	0.89	0.34
0-2 *vs* 3,4	8.75 (1.35) *vs* 10.06 (1.83)	0.80 (0.73-0.87)	0.83	6.90 (2.60) *vs* 13.00 (11.10)	0.90 (0.85-0.95)	0.94	<0.01
0-3 *vs* 4	9.01 (1.61) *vs* 10.60 (2.16)	0.83 (0.76-0.90)	0.86	7.60 (2.93) *vs* 16.15 (14.77)	0.95 (0.91-0.98)	0.96	<0.01

IQR, interquartile range; AUROC, area under receiver operator characteristic curve; CI, confidence interval
[†] Significance of comparison of observed ELF and TE AUROC values

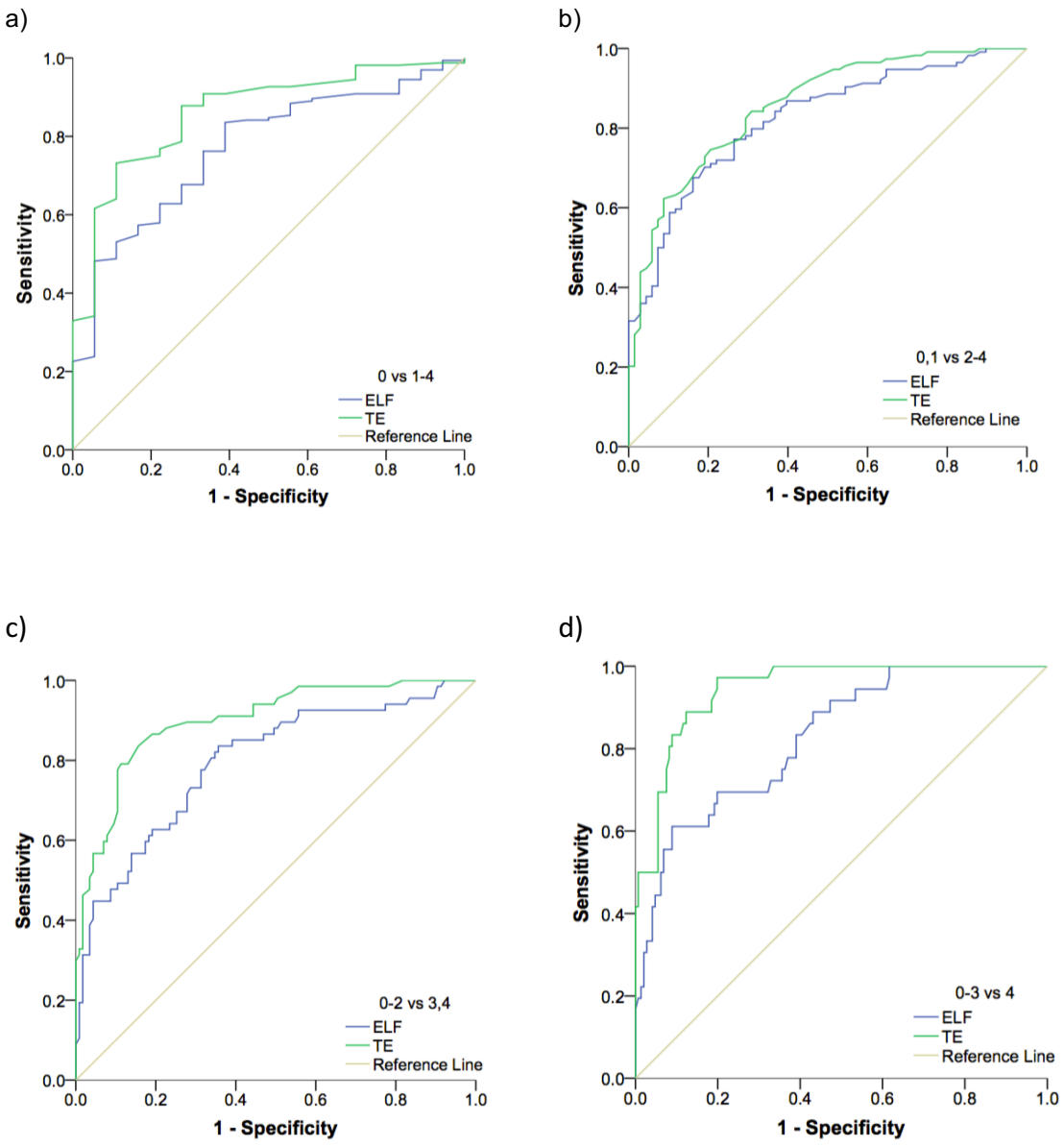


Figure 4.4. Receiver operator characteristic curves of ELF and TE predicting liver fibrosis stages

Both modalities demonstrated at least good performance (AUROC ≥0.8) at all liver fibrosis stages. ROC curves are shown for predicting the following liver fibrosis stages; a) 0 vs 1-4, b) 0,1 vs 2-4, c) 0-2 vs 3,4 and d) 0-3 vs 4

4.5.4 Diagnostic performance in e-antigen negative patients

A sub-analysis of the performance in e-antigen negative patients showed similar performance of ELF and TE to that for the whole cohort. AUROC values for ELF and TE for F≥1, F≥2, F≥3 and F4 stages were 0.71, 0.80, 0.79, 0.81 and 0.81, 0.83, 0.90, 0.95, respectively, with a significant difference in performance at F≥3 and F4.

4.5.5 Influence of transaminases on diagnostic performance

Effect of serum ALT level on test performance was assessed. Diagnostic performance appears to be maintained with both modalities when ALT is 3 or 5 times above the upper limit of normal (ULN) (120 and 200 IU/l, respectively) (table 4.3). In the diagnosis of severe fibrosis both modalities maintained their performance in all categories of ALT. The AUROC values indicate that in the diagnosis of any fibrosis, ELF is less accurate when ALT is below ULN compared to when ALT is above ULN, and accuracy of TE improves when ALT is below ULN. The 95% confidence interval for ELF in diagnosing any fibrosis is very large in this small cohort.

When the diagnostic accuracy of each modality is calculated, using the threshold with maximal sensitivity and specificity, diagnostic accuracy is reduced when ALT values are normal compared to elevated ALT, in the diagnosis of any fibrosis or severe fibrosis (table 4.4). For the diagnosis of significant fibrosis, diagnostic accuracy of ELF is higher in subjects with elevated ALT, and the reverse is seen with TE.

Table 4.3. Diagnostic performance of the ELF test and TE according to categories of serum ALT levels

ALT category	Modality	AUROC (95% CI)			
		Fibrosis stage			
		0 *vs* 1-4	0,1 *vs* 2-4	0-2 *vs* 3,4	0-3 *vs* 4
≤ULN (≤40 IU/l) *n* = 24	ELF	0.51 (0.06-0.96)	0.86 (0.66-1.00)	0.81 (0.62-1.00)	0.89 (0.76-1.00)
	TE	0.94 (0.85-1.00)	0.95 (0.87-1.00)	0.92 (0.82-1.00)	1.00 (1.00-1.00)
>ULN (>40 IU/l) *n* = 158	ELF	0.81 (0.76-0.94)	0.82 (0.76-0.89)	0.80 (0.72-0.87)	0.82 (0.74-0.90)
	TE	0.85 (0.76-0.94)	0.84 (0.78-0.90)	0.90 (0.85-0.95)	0.94 (0.90-0.98)
>3xULN (>120 IU/l) *n* = 55	ELF	0.85 (0.70-1.00)	0.83 (0.72-0.94)	0.81 (0.69-0.92)	0.87 (0.77-0.97)
	TE	0.99 (0.97-1.00)	0.79 (0.65-0.93)	0.90 (0.81-0.98)	0.95 (0.89-1.00)
>5xULN (>200 IU/l) *n* = 30	ELF	0.82 (0.63-1.00)	0.83 (0.68-0.98)	0.78 (0.61-0.95)	0.88 (0.75-1.00)
	TE	1.00 (1.00-1.00)	0.83 (0.64-1.00)	0.87 (0.74-1.00)	0.97 (0.89-1.00)

ALT, alanine transaminase; AUROC, area under receiver operator characteristic curve; CI, confidence interval; ULN, upper limit of normal range

Table 4.4 Diagnostic accuracy of TE and ELF for predicting any, significant and severe fibrosis

Data are presented for all subjects, and then divided in to subsets with ALT above and below the upper limit of normal. The threshold for each modality is where the sensitivity and specificity of the test is maximal (Youden index)

Threshold	Modality	Diagnostic accuracy (%)		
		All subjects *n* = 182	ALT ≤ULN *n* = 28	ALT >ULN *n* = 154
F≥1	TE	73.6	67.9	74.7
	ELF	81.3	78.6	81.8
F≥2	TE	76.4	82.1	75.3
	ELF	73.6	71.4	74.0
F≥3	TE	84.1	85.7	83.8
	ELF	71.4	78.6	70.1

ALT, alanine transaminase; ULN, upper limit of normal

4.5.6 Spectrum bias

The DANA value in this cohort is 2.169, therefore the adjusted AUROC values for diagnosing F2-4 are 0.839 and 0.872 for the ELF test and TE, respectively. For F3,4, adjusted AUROC values are 0.812 and 0.916 for the ELF test and TE, respectively. Therefore, using the DANA method to calculate the adjusted AUROC, diagnostic performance increased at all fibrosis stages with both modalities. Adjustment using the Obuchowski method showed that the overall mean accuracy (unweighted Obuchowski measure) was 0.91 for ELF and 0.95 for TE. For distinguishing between F3 and F4, performance was 0.59 for ELF, and 0.73 for TE (table 4.5).

Table 4.5. Obuchowski measures for the ELF test and TE for each liver fibrosis stage pair

Fibrosis Stage Pair	ELF		TE	
	Estimate	Standard Error	Estimate	Standard Error
1 *vs* 2	0.58	0.08	0.71	0.07
1 *vs* 3	0.80	0.06	0.87	0.06
1 *vs* 4	0.82	0.06	0.94	0.04
1 *vs* 5	0.94	0.03	0.99	0.01
2 *vs* 3	0.74	0.05	0.70	0.05
2 *vs* 4	0.79	0.06	0.86	0.04
2 *vs* 5	0.92	0.03	0.98	0.01
3 *vs* 4	0.59	0.07	0.73	0.06
3 *vs* 5	0.78	0.05	0.96	0.02
4 *vs* 5	0.69	0.06	0.85	0.05
Overall	0.91	0.01	0.95	0.01

4.5.7 Clinical utility models for the ELF test and transient elastography

Table 4.6 shows the sensitivities, specificities, predictive values and diagnostic odds ratios of ELF and TE predicting severe fibrosis / cirrhosis, and cirrhosis. If two thresholds with high sensitivity and specificity are used to “rule in” fibrosis (upper threshold with high specificity, therefore high positive predictive value) or “rule out” fibrosis (lower threshold with high sensitivity, therefore high negative predictive value), the clinical utility of each modality can be evaluated. Figure 4.5 and table 4.7 show the models for identifying any fibrosis and severe fibrosis using thresholds with a sensitivity and specificity of 80%, 85% and 90%. For example, using ELF to identify severe fibrosis at data-derived thresholds of 9.08 and 9.94 (sensitivity and specificity of 85%, respectively), 60% of patients

would have correctly avoided liver biopsy and 15% would have incorrectly avoided biopsy. 25% would have had an indeterminate result – a value between the thresholds.

Using TE to identify severe fibrosis with thresholds of 8.75 and 8.95 (sensitivity and specificity of 85%, respectively) would have resulted in biopsy correctly being avoided in 82% and incorrectly avoided in 15%, with an indeterminate result in 3%. At higher sensitivity and specificity, the proportion avoiding biopsy decreases. For example, if sensitivity and specificity thresholds are increased to 90%, the proportion of incorrectly classified cases (i.e. the false positive and false negative rates) substantially decreases to around 10% for both modalities for diagnosis of both severe and any fibrosis. However, this is at the cost of increased proportions of indeterminate cases.

Table 4.6. Diagnostic performance indices for ELF and TE in the identification of severe fibrosis (F3,4) and cirrhosis (F4) at a range of thresholds

Modality	Threshold	Sensitivity %	Specificity %	PPV	NPV	LR +	LR -	DOR
Severe fibrosis (prevalence = 37%)								
ELF	8.02	96	17	40	86	1.10	0.24	4.58
	8.45	93	41	48	90	1.58	0.17	9.29
	8.96	85	56	53	86	1.93	0.27	7.15
	9.39	73	70	58	82	2.43	0.39	6.23
	9.88	60	83	67	78	3.53	0.48	7.35
	10.41	45	95	83	75	9.00	0.58	15.52
TE	6.85	96	50	52	95	1.92	0.08	24.00
	7.70	91	60	57	92	2.28	0.15	15.20
	8.45	88	77	69	92	3.83	0.16	23.94
	9.35	79	87	78	88	6.08	0.24	25.33
	10.15	64	90	80	81	6.40	0.40	16.00
	11.95	57	96	88	79	14.25	0.45	31.67
Cirrhosis (prevalence = 20%)								
ELF	8.61	94	39	28	97	1.54	0.15	10.27
	9.43	72	64	34	90	2.00	0.44	4.55
	9.66	69	72	38	90	2.46	0.43	5.72
	9.99	67	81	47	91	3.53	0.41	8.61
	10.34	61	87	54	90	4.69	0.45	10.42
	10.68	44	95	70	87	8.80	0.59	14.92
TE	9.70	94	80	54	98	4.70	0.08	58.75
	10.30	89	86	62	97	6.36	0.13	48.92
	11.85	83	90	67	96	8.30	0.19	43.68
	12.95	75	92	71	94	9.38	0.27	34.74
	14.15	61	95	74	91	12.20	0.41	29.76
	15.45	50	95	72	88	10.00	0.53	18.87

PPV, positive predictive value; NPV, negative predictive value; LR +, positive likelihood ratio; LR -, negative likelihood ratio; DOR, diagnostic odds ratio

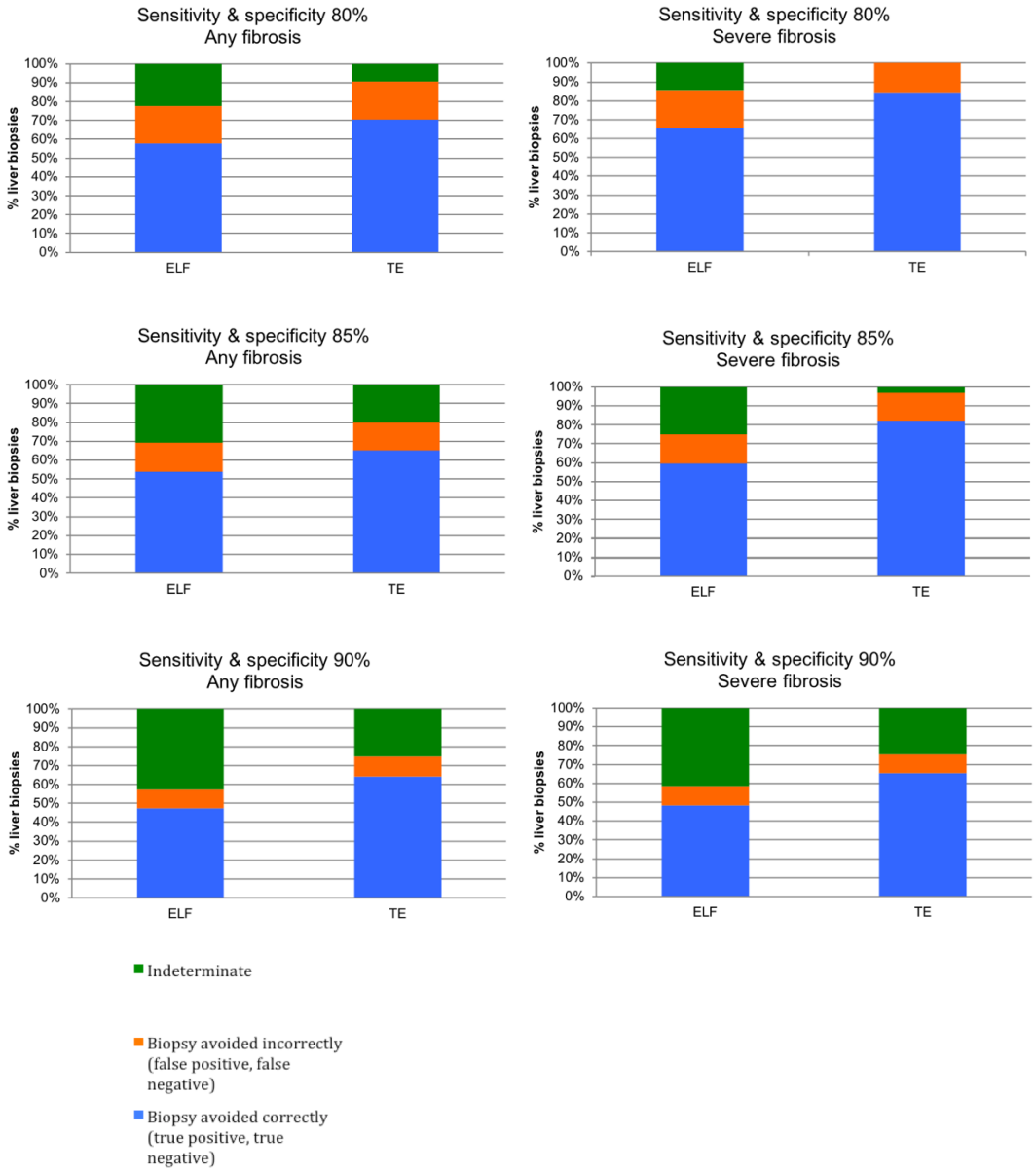


Figure 4.5. Clinical utility models for ELF and TE predicting any fibrosis and severe fibrosis at thresholds with sensitivity and specificity of 80%, 85% and 90%

Each model shows the percentage of individuals in whom biopsy would be avoided correctly or avoided incorrectly and in whom the result would be indeterminate, for data derived thresholds with sensitivity and specificity values of 80%, 85% and 90%

Table 4.7. Sensitivities and specificities of the ELF test and TE identifying severe fibrosis (Metavir F≥3) and any fibrosis (F≥1) using thresholds with sensitivity and specificity of 80%, 85% and 90%

Fibrosis stage	Modality	Thresholds	Sensitivity (%)	Specificity (%)	Correctly avoided (%)	Incorrectly avoided (%)	Indeterminate (%)
Sensitivity and specificity 80%							
Severe fibrosis	ELF	9.24	80.0	66.1	65.4	20.2	14.4
		9.66	62.9	79.7			
	TE	8.85	83.6	84.3	84.0	16.0	0
Any fibrosis	ELF	8.30	80.6	61.1	57.9	19.7	22.4
		9.15	57.6	77.8			
	TE	6.40	79.9	27.8	70.3	20.3	9.4
		6.75	75.0	77.8			
Sensitivity and specificity 85%							
Severe fibrosis	ELF	9.08	84.4	61.0	59.5	15.5	25.0
		9.94	57.1	84.7			
	TE	8.75	85.0	86.0	82.0	15.0	3.0
		8.95	82.0	85.0			
Any fibrosis	ELF	8.22	84.2	50.0	53.7	15.4	30.9
		9.24	54.1	83.3			
	TE	6.15	85.0	65.0	65.0	15.0	20.0
		7.70	64.0	89.0			
Sensitivity and specificity 90%							
Severe fibrosis	ELF	8.74	90.0	48.3	48.4	10.1	41.5
		10.2	48.6	89.8			
	TE	7.95	89.6	66.1	65.4	9.9	24.7
		10.15	64.2	90.4			
Any fibrosis	ELF	8.03	90.0	38.9	47.3	10.1	42.6
		9.40	48.2	88.9			
	TE	5.75	89.6	66.7	64.3	10.5	25.2
		7.70	64.0	88.9			

4.5.8 A model combining both modalities

Logistic regression analysis found that in a model combining both modalities, in the prediction of Metavir F≥1 and F4, ELF was a non-significant predictor. In the prediction of F≥2 and F≥3 ELF significantly improved the prediction of fibrosis when combined with TE. Respective ELF and TE odds ratios in the combined models were: 1.45 (95% CI 0.75-2.83) and 1.99 (1.31-3.02), 2.47 (1.55-3.94) and 1.54 (1.25-1.90), 1.61 (1.03-2.51) and 1.55 (1.31-1.83), 1.32 (0.75-2.32) and 1.44 (1.23-1.68) for F≥1, F≥2, F≥3 and F4, respectively (table 4.8). Combining the two tests results in AUROC values of 0.87, 0.88, 0.90 and 0.95 for diagnosis of F≥1, F≥2, F≥3 and F4 stages, respectively.

Table 4.8 Odds ratios calculated by logistic regression for prediction of liver fibrosis using models comprising the ELF test and TE individually, and the ELF test and TE combined

Fibrosis stage		Model			
		ELF	TE	Combined	
				ELF	TE
0 *vs* 1-4	**OR**	2.58	2.14	1.45	1.99
	95% CI	1.45-4.60	1.44-3.19	0.75-2.83	1.31-3.02
	***p* value**	0.001	<0.001	0.27	0.001
	R^2	0.16	0.32	0.33	
0,1 *vs* 2-4	**OR**	3.75	1.75	2.47	1.54
	95% CI	2.45-5.75	1.43-2.15	1.55-3.94	1.25-1.90
	***p* value**	<0.001	<0.001	<0.001	<0.001
	R^2	0.39	0.44	0.53	
0-2 *vs* 3,4	**OR**	2.85	1.64	1.61	1.55
	95% CI	2.02-4.04	1.40-1.93	1.03-2.51	1.31-1.83
	***p* value**	<0.001	<0.001	0.04	<0.001
	R^2	0.34	0.56	0.58	
0-3 *vs* 4	**OR**	3.00	1.49	1.32	1.44
	95% CI	2.04-4.43	1.30-1.72	0.75-2.32	1.23-1.68
	***p* value**	<0.001	<0.001	0.34	<0.001
	R^2	0.36	0.62	0.62	

OR, odds ratio; CI, confidence intervals; R^2, Nagelkerke pseudo R-square values

4.5.9 Further exploration of combined models – a serial algorithm

Serial models using ELF then TE or TE then ELF to diagnose any fibrosis or severe fibrosis were constructed, as outlined in figure 4.2.

Identification of cut off scores for each modality

- For single cut-off scores, the data-derived threshold with the maximal sensitivity and specificity (Youden index) was used, for each modality for any fibrosis and for severe fibrosis.

Identification of dual cut off scores for each modality

- Using the upper and lower thresholds derived from the clinical utility modelling thresholds, presented in table 4.7, for identifying any fibrosis

and severe fibrosis the number of subjects in this dataset who would have an indeterminate ELF score can be determined. These thresholds have sensitivity and specificity of 85% and were used in this model.

Using this “fully assigned” approach, the first non-invasive test was applied using dual thresholds, resulting in subjects being diagnosed with ‘disease’, ‘no disease’ or ‘indeterminate for disease’. The second test was then applied to the indeterminate cases using a single cut off, resulting in these individuals being diagnosed with ‘no disease’ or ‘disease’. Using this approach, no individuals are unassigned.

The sensitivity, specificity, PPV, NPV and diagnostic accuracy for the first test was calculated for those individuals who were assigned. The second test was applied to the unassigned (indeterminate) subjects, again calculating the sensitivity, specificity, PPV, NPV and diagnostic accuracy for this group. The overall performance of the algorithm was calculated by combining the two 2x2 tables to calculate overall sensitivity, specificity, PPV, NPV and diagnostic accuracy.

The algorithm was applied using thresholds for diagnosis of any fibrosis and for severe fibrosis.

Table 4.9 shows the calculated Youden indices for each fibrosis stage, and the associated specificity and sensitivity values associated for each threshold. The Youden indices for fibrosis stages 0 vs 1-4 and 0-2 vs 3-4 were used as the thresholds for the second test in the serial algorithm.

Table 4.9. Calculated Youden index for each liver fibrosis stage
The table presents the corresponding threshold for each non-invasive modality and respective sensitivity and specificity values

Fibrosis stage	ELF				TE			
	Youden index	Threshold	Sensitivity %	Specificity %	Youden index	Threshold	Sensitivity %	Specificity %
0 *vs* 1-4	44.6	8.255	83.5	61.1	60.9	6.95	72.0	88.9
0,1 *vs* 2-4	51.3	9.285	67.5	83.8	54.0	7.95	74.6	79.4
0-2 *vs* 3,4	47.9	9.155	83.6	64.3	67.9	8.85	83.6	84.3
0-3 *vs* 4	52.2	10.44	61.1	91.1	77.3	9.53	97.2	80.1

The performance characteristics for each of the two stages in the models and the overall performance characteristics of the algorithm are shown in table 4.10. Using either ELF or TE alone, employing the single threshold derived from the Youden index above, the diagnostic accuracy values for ELF and TE for diagnosing any fibrosis were 81.3% and 73.6%, respectively. For the diagnosis of severe fibrosis, the values were 71.4% and 84.1%, respectively. The optimal models are those where TE is the first test.

The overall diagnostic accuracy values in the serial modality models are similar to the values when a single threshold / single test strategy is used. In the serial models for the diagnosis of any fibrosis the diagnostic performance of the TE then ELF model is higher than the TE alone model, but lower than the ELF alone model. In the serial models for the diagnosis of severe fibrosis, the TE alone model has the highest diagnostic performance (84.1%). The advantage of the models is that all cases are assigned, with no overall indeterminate cases. This modelling demonstrates that this is achieved with minimal loss of overall diagnostic performance. In clinical practice, the order of any serial test will depend on the resources available to clinicians, in terms of access to TE and ELF, and the prevailing costs of each test.

Table 4.10. Diagnostic performance characteristics of models comprising serial non-invasive markers
Each model comprises an initial test with two thresholds. The second test is applied to the indeterminate cases from the first test. The overall performance characteristics of the algorithm are presented. Models are presented for ELF then TE predicting any fibrosis and severe fibrosis, and for TE then ELF predicting any fibrosis and severe fibrosis

a) ELF then TE, predicting any fibrosis

Modality	Threshold		Indeterminate, n (%)	PPV %	NPV %	Sensitivity %	Specificity %	Diagnostic performance %
Modality 1 (ELF)	Lower	8.22 *n* = 35 (19.2%)	54 (29.7)	96.8	25.7	77.6	75.0	77.3
	Upper	9.24 *n* = 93 (51.1%)						
Modality 2 (TE)	Threshold	6.95	N/A	96.7	20.8	60.4	83.3	63.0
	Below, *n* (% of indeterminates)	24 (44.4)						
	Above, *n* (% of indeterminates)	30 (44.6)						
Overall algorithm	N/A		N/A	96.7	23.7	72.6	77.8	73.1

b) TE then ELF, predicting any fibrosis

Modality	Threshold		Indeterminate, n (%)	PPV %	NPV %	Sensitivity %	Specificity %	Diagnostic performance %
Modality 1 (TE)	Lower	6.15 *n* = 38 (20.9%)	37 (20.3)	98.1	34.2	80.8	86.7	81.4
	Upper	7.70 *n* = 107 (58.8%)						
Modality 2 (ELF)	Threshold	8.255	N/A	92.6	10.0	73.5	33.3	70.3
	Below, *n* (% of indeterminates)	10 (27.0)						
	Above, *n* (% of indeterminates)	27 (73.0)						
Overall algorithm	N/A		N/A	97.0	29.2	79.3	77.8	79.1

c) ELF then TE, predicting severe fibrosis

Modality	Threshold		Indeterminate, n (%)	PPV %	NPV %	Sensitivity %	Specificity %	Diagnostic performance %
Modality 1 (ELF)	Lower	9.08 *n* = 80 (44.0%)	46 (25.3)	67.9	87.5	79.2	79.5	79.4
	Upper	9.94 *n* = 56 (30.8%)						
Modality 2 (TE)	Threshold	8.85	N/A	66.7	86.4	84.2	70.4	76.1
	Below, *n* (% of indeterminates)	22 (47.8)						
	Above, *n* (% of indeterminates)	24 (52.2)						
Overall algorithm	N/A		N/A	67.5	87.3	80.6	77.4	78.6

d) TE then ELF, predicting severe fibrosis

Modality	Threshold		Indeterminate, n (%)	PPV %	NPV %	Sensitivity %	Specificity %	Diagnostic performance %
Modality 1 (TE)	Lower	8.75 *n* = 105 (57.7%)	5 (2.75)	84.6	84.8	76.4	90.5	84.7
	Upper	8.95 *n* = 72 (39.6%)						
Modality 2 (ELF)	Threshold	9.155	N/A	40.0	N/A	100.0	0.00	40.0
	Below, *n* (% of indeterminates)	0 (0)						
	Above, *n* (% of indeterminates)	5 (100)						
Overall algorithm	N/A		N/A	81.4	84.8	77.0	88.0	83.5

4.6 Discussion

This study has set out to investigate the performance of the ELF test in the diagnosis of liver fibrosis as assessed by histology, and compared it to TE in a cohort of subjects with CHB. The ELF test accurately assesses liver fibrosis severity in subjects with CHB. TE is superior to the ELF test and logistic regression shows that the combination of modalities is superior to TE. Although using two thresholds improves sensitivity and specificity, a category of indeterminate cases is generated. This study has shown that use of a serial algorithm can overcome this problem without substantial loss of diagnostic accuracy.

Although there was a significant correlation between TE and ELF scores and ALT, diagnostic performance was maintained in subjects with elevated ALT. The diagnostic accuracy was generally higher in subjects with above normal ALT. As this study was cross-sectional rather than longitudinal, the effect of elevated ALT or 'flares' of hepatitis on future performance in these individuals of these non-invasive modalities cannot be evaluated. These data, however, do provide some reassurance that even in subjects with ALT values of 1, 3 and 5 times the upper limit of normal (40, 120 and 200 IU/l) both modalities accurately evaluate liver fibrosis.

As discussed in chapter 2, ELF has been validated in external disease-specific cohorts of patients with non-alcoholic fatty liver disease, primary biliary cholangitis, primary sclerosing cholangitis and chronic hepatitis C.[78-82,148,149] It predicts liver-related outcomes at 7 years at least as well as biopsy, with a unit change in ELF associated with a doubling of risk. [83] Of the 25 patients with CHB

followed up for over 7 years in that study, 2 died of a liver-related cause and one experienced a non-fatal liver-related outcome by 7 years (median for the whole cohort) after biopsy and ELF test. In all 3 cases the incident ELF score exceeded 7.8. The median ELF score was 8.63 for the whole cohort of CHB patients that were followed up.

This study reports the external validation of the ELF test in subjects with CHB. Performance in patients with CHB in the original OELF cohort (*n* = 44) was good at all fibrosis stages and was maintained in this validation cohort. Logistic regression, which included age and simple biochemical parameters (AST, ALT), did not improve performance. These data suggest a role for ELF in the assessment of patients with CHB and in informing the decision-making process when antiviral therapy is being considered.

4.6.1 Other studies

A study[150] reporting the performance of ELF in 58 patients with CHB used the original ELF algorithm and that for the IMMUNO 1 autoanalyser for Metavir staging[78] but not the immune-assays that have been specifically developed for the ELF test. Further, the study assessed liver biopsies using the Ishak system. AUROC values for predicting Ishak fibrosis stages 1-6 and 2-6 (equivalent to Metavir F≥1) were 0.66 and 0.59, respectively, lower than the values I have found. AUROC for predicting Ishak stages 3-6 was 0.83, similar to my results. The inferior performance of the test in this cohort may be attributable to the use of assays that were not specifically developed for the ELF test and failure to use the appropriate ELF algorithm.

In my study TE performed as well or better than in other studies in patients with CHB. For example in the detection of F4 fibrosis, AUROC values in other studies range from 0.8 in a study of fifty nine subjects (with a TE failure rate of 1.7%)[134] to 0.94.[151] A meta-analysis of non-invasive tests for liver disease severity in NAFLD[152] found that the collective performance of TE in detecting F≥2 and F≥3 fibrosis was 0.84 (95% CI 0.79-0.90) and 0.94 (95% CI 0.86-0.99), respectively. Regression analysis found that success was unaffected by the severity of inflammation or steatosis. The authors do not comment on how included studies handled TE failures, but state that TE failure rate was 5-13% and most failures were in obese subjects.

Compared to the failure / invalid TE rate of 3% in my study, a major review of clinical performance of over 13,000 TE examinations found a failure rate (defined as no TE value obtained after at least ten attempts) of 3.1% and unreliable results (defined as less than ten valid measurements, success rate less than 60%, or IQR more than 30%) in 15.8% of cases.[153] The authors cite obesity, operator experience of fewer than 500 examinations and subject age over 52 years as predictors of failure or unreliable results. Studies investigating TE in patients with CHB report success rates for acquiring valid TE results ranging between 79% and 99.6%.[134,139,153-157] TE reproducibility has been shown to be excellent for both inter- and intra-observer agreement, but this is reduced at lesser stages of fibrosis and in patients with steatosis, high body mass index and in particular waist circumference.[158,159] All 6 patients in my study excluded due to TE failure were overweight (*n* = 2) or obese (*n* = 4).

Several studies have compared TE to the ELF test in subjects with CHB, for example in a cohort of 170 Asian subjects with CHB.[160] The appropriate ELF

algorithm was used. Reliable TE results were defined (at least ten valid measurements, success rate ≥60% and interquartile range/median value ratio <30%). Only subjects with successful TE results were included, and the authors do not state how many of the original 253 screened patients were not included due to TE (or ELF) failure. AUROC values for predicting F≥2, F≥3 and F4 were 0.90, 0.86 and 0.86 for ELF and 0.94, 0.96 and 0.96 for TE, respectively. TE was significantly better than ELF for predicting F≥3 and F4.

TE was compared to the ELF test using the Ishak staging system in 102 subjects with CLD of which 55% were known to have viral hepatitis.[161] Although the proprietary ELF test reagents were used, the original ELF algorithm was used rather than the ELF algorithm. TE was successful in all subjects. AUROC values for TE and ELF identifying Ishak stages ≥2 (0.92 vs 0.87, respectively) and ≥5 (0.95 vs 0.93, respectively) were similar. Data-derived thresholds with the best compromise between sensitivity and specificity (Youden index) were 8.5 and 8.99 for predicting F≥2 for TE and ELF, respectively. For predicting F≥5, values were 17.45 and 9.39, respectively. The ELF test was less discriminative in the low / moderate fibrosis stages with overlap between F0-1 and F2-4. Both TE and ELF correlated with ALT / AST levels, although regression analysis found that neither modality was influenced by inflammatory marker levels in predicting F≥2 or F≥3 fibrosis stages.

An Australian study set out to investigate the accuracy of the ELF threshold of 9.8 to identify advanced fibrosis, in a cohort of subjects with liver disease of mixed aetiology.[86] ELF was reliable, with sensitivity of 74.4% and specificity of 92.4% to detect advanced fibrosis using liver biopsy as reference. In terms of

aetiology, subjects with CHB were significantly more likely than subjects with NAFLD to have a false negative ELF score. The authors postulate that a lower ELF threshold may be needed to detect advanced fibrosis in patients with CHB. As the majority of the false negatives in the CHB cohort were in those with less inflammation on biopsy, and as indicated by my data, the influence of ALT flares requires more work.

In a study of participants with CHB in Hong Kong, 238 subjects underwent liver biopsy, followed by TE and the proprietary ELF test using the appropriate ELF algorithm. In the training cohort, AUROC values for TE and ELF for identifying F≥2 and F≥3 fibrosis were 0.82 vs 0.59 and 0.83 vs 0.69, respectively.[162] In the validation cohort, values were 0.74 vs 0.76 and 0.73 vs 0.68. An ELF score ≤8.4 had a sensitivity of 95% to exclude advanced fibrosis and ELF >10.8 had a specificity of 92% to confirm advanced fibrosis. Using the recommended threshold of 9.8, specificity of ELF to detect F≥2 was 66.3%. An algorithm incorporating TE and ELF was generated. As both modalities had high sensitivity to exclude advanced fibrosis, two models were proposed; TE then ELF and ELF then TE. Both tests used dual thresholds therefore there were unassigned cases following application of the second test. Diagnostic accuracy was improved using a combined model. This study also calculated AUROC for each modality to predict liver-related events, of which 4% of participants experienced over a 4 year follow up period, which were 0.6 for ELF and 0.71 for TE.

A recent study has evaluated TE and the ELF test in a cohort of 222 subjects with CHB.[163] Using the Batts and Ludwig criteria, ≥F3 or F4 fibrosis was seen

in 64% and 53% of liver biopsies, respectively. Of the original 265 subjects who underwent liver biopsy, 15 were excluded due to failure of TE or ELF; although the authors do not clarify, it is more likely that failures are due to TE than ELF. TE performed better than the ELF test in identifying ≥F2, ≥F3 and F4 with AUROC values of 0.857 vs 0.802, 0.887 vs 0.703 and 0.853 vs 0.706, respectively. The previously established sequential TE-ELF algorithm described in the study above was applied and compared to a concurrent TE-ELF algorithm generated by regression analysis. The sequential algorithm had a higher probability of preventing liver biopsy to diagnose ≥F3 and F4 fibrosis than the concurrent algorithm, preventing unnecessary biopsy in 69 to 73% of subjects in these fibrosis groups.

TE and ELF were compared in a study of 119 women with perinatally acquired CHB with ALT levels below twice the upper limit of normal, using Metavir-scored liver biopsy as reference.[100] The correct ELF algorithm was used and TE was used as the reference standard, using the following thresholds; Metavir F01 <7.2 kPa, F2 7.2-8.0 kPa, F3 8.1-11.0 kPa, F4 >11.0 kPa. AUROC for the detection of F3 was 0.65. The authors argue that TE is an acceptable reference standard due to its published performance in detecting severe fibrosis in several liver diseases. This is, however, not established practice.

The efficacy of the ELF test to predict clinical outcomes in patients with CHB has been investigated.[164] 170 subjects who had undergone liver biopsy, TE and ELF test were followed up for a median of 41 months for event (decompensation, hepatocellular carcinoma and / or liver-related death), with events recorded in 23%. AUROC values for predicting outcome were 0.81 for

ELF, 0.73 for TE and 0.71 for liver biopsy. These data suggest a role for the ELF test in predicting liver-related events in CHB and support the outcome data from the original ELF cohort[83] and subsequent studies.[82,85,87-90,96,165-167]

More recently, a review of non-invasive testing of liver fibrosis in patients with chronic hepatitis B reports meta-analyses of the serum marker panels APRI (8,855 subjects) and Fib-4 (6,455 subjects) and reports the diagnostic performance of each to identify F≥2 fibrosis as AUROC 0.74 and 0.73, respectively. Performance in diagnosing cirrhosis was 0.78 and 0.84, respectively. The authors reported a meta-analysis of TE, comprising 4,386 subjects, with performance of diagnosing F≥2 fibrosis AUROC 0.88 and cirrhosis 0.93. The conclusion of reviews of studies of serum biomarkers, including APRI, Fib-4, FibroTest, Fibrometer, Hepascore and ELF, and of TE found that these markers are generally less accurate in detecting significant fibrosis than cirrhosis, and that these tests should be used as first line tests of liver fibrosis in CHB. Further, the potential for these markers to be used as screening tests in populations at risk of CHB or in the general population was highlighted.[168]

A summary of published studies of the performance of the ELF test is shown in table 4.11.

Table 4.11. Summary of studies of the performance of the ELF test in subjects with chronic hepatitis B
Studies comprising mixed liver aetiologies were not included if the cohort of subjects with CHB was small

Author (Year)	Cohort	ELF assay platform	Correct ELF algorithm	Reference		Outcome measure / fibrosis thresholds	Performance of ELF test (AUROC (95% CI) unless otherwise stated)
				Liver biopsy fibrosis staging system	Other		
Kim *et al.*[160] (2012)	CHB (170)	ADVIA Centaur XP	Yes	Batts and Ludwig (5 stages)		F≥2 F≥3 F4	0.901 (0.849-0.953) 0.860 (0.805-0.915) 0.862 (0.809-0.915)
Wahl *et al.*[161] (2012)	Mixed (55, Viral hepatitis; 7, AIH; 4, Wilson's; 22, NAFLD; 14, unknown aetiology)	IMMUNO 1	Yes (OELF algorithm)	Ishak		2 -6 5 -6	0.87 (0.78-0.96) 0.93 (0.88-0.99)
Gümüsay *et al.*[150] (2013)	CHB (58) and healthy controls (30)	Platform not stated. Non-Siemens reagents	OELF and IMMUNO 1 algorithms	Ishak		1 – 6 2 – 6 3 - 6	0.651 (OELF) / 0.663 (ELF) 0.571 (OELF) / 0.588 (ELF) 0.833 (OELF) / 0.830 (ELF) (CIs not reported)
Harkisoen *et al.*[100] (2014)	CHB	ADVIA Centaur XP	Yes		TE	F≥3	0.65 (0.51–0.80)

Author (Year)	Cohort	ELF assay platform	Correct ELF algorithm	Reference		Outcome measure / fibrosis thresholds	Performance of ELF test (AUROC (95% CI) unless otherwise stated)
				Liver biopsy fibrosis staging system	Other		
Wong *et al*.[162] (2014)	HBV (238, training set; 85, validation set)	ADVIA Centaur	Yes (for Centaur XP)	Metavir		Training F≥2 F≥3 F4 Validation F≥2 F≥3 F4	 0.59 (0.51-0.67) 0.69 (0.63-0.75) 0.68 (0.61-0.75) 0.76 (0.65-0.86) 0.69 (0.63-0.75) 0.68 (0.61-0.75)
Trembling *et al*.[169] (2014)	CHB (182)	ADVIA Centaur	Yes (not stated)	Metavir		F≥1 F≥2 F≥3 F4	0.77 (0.67-0.87) 0.82 (0.76-0.88) 0.80 (0.73-0.87) 0.83 (0.76-0.90)
Kim *et al.*[164] (2014)	CHB (170)	ADVIA Centaur XP	Yes	Batts and Ludwig		F≥2 F≥3 F4 Ability of ELF to predict outcome	0.901 (0.849-0.953) 0.860 (0.805-0.915) 0.862 (0.809-0.915) AUROC to predict outcome: 0.808 (ELF), 0.713 (histology)

Author (Year)	Cohort	ELF assay platform	Correct ELF algorithm	Reference		Outcome measure / fibrosis thresholds	Performance of ELF test (AUROC (95% CI) unless otherwise stated)
				Liver biopsy fibrosis staging system	Other		
Heo *et al.*[163] (2018)	CHB (222)	ADVIA Centaur XP	Yes	Batts and Ludwig		F≥2 F≥3 F4	0.802 0.703 0.706 (CIs not reported)

AIH, autoimmune hepatitis; CHB, chronic hepatitis B; CI, confidence interval; ELF, Enhanced Liver Fibrosis; NAFLD, non-alcoholic fatty liver disease; OELF, Original European Liver Fibrosis

4.6.2 Generalisability of the findings

Both ELF and TE represent alternative and potentially complimentary approaches to assessing liver fibrosis, and are associated with minimal discomfort and hazard to the patient when compared to biopsy. Logistic regression analysis suggests that the performance of ELF is improved with the addition of TE, although TE does not improve with the addition of ELF.

Both modalities track fibrosis stage linearly, with TE having superior discrimination and closer correlation with histological staging, particularly at higher fibrosis stages. The performance of TE predicting F≥2 fibrosis in this study was superior to most of the previous studies assessing TE in CHB. The diagnostic performance of each modality was evaluated at various sensitivities and specificities; the median diagnostic odds ratio for ELF for detecting severe fibrosis between sensitivity and specificity of 95% was 7.3 and for TE 24.0. Clinical utility modelling supports a role for these modalities in the assessment of patients and in treatment decisions.

Applying previously published thresholds to these results allows for some generalisability of the model. Recent studies investigating ELF and TE both in a heterogeneous population[93] and in CHB[160] did not report dual thresholds, making comparison difficult. However, using thresholds reported in separate studies allows some comparisons to be drawn. A study of TE in CHB[132] reported that cut-off values of 9.4 and 6.2 which had sensitivity and specificity of >90% ruled in and ruled out F≥2 in 56% of cases, with 90% accuracy. Applying these thresholds to my results, 57% of patients would have F≥2 ruled in or ruled out, with 91% accuracy. Data from patients with chronic hepatitis C[81]

found that using ELF cut-off values of 9.59 and 10.22, with sensitivity and specificity of 85%, 81% of patients could avoid biopsy by having severe fibrosis (F≥3) ruled in or ruled out, with 81% accuracy. Applying these thresholds to my results, 77% of patients would avoid biopsy, with 86% accuracy.

Application of the DANA method to calculate the adjusted AUROC, increased diagnostic performance at all fibrosis stages with both modalities. This method assumes equal prevalence in all fibrosis stages, which may not be reflective of true prevalence and may overestimate prevalence at the extremes of fibrosis stage. Further validation of this method is required.

4.7 Strengths and limitations

4.7.1 Strengths

Strengths of this study include the method of data collection. Liver biopsy, TE and serum sampling were all performed on the same day. ELF tests were performed in one central laboratory, ensuring quality control and consistency. It is important to note that, in contrast to some studies, my study used the proprietary ELF assays in accordance with the manufacturer's instructions, and the ELF algorithm specific for the assay platform and histology staging system (table 4.11).

4.7.2 Limitations

There are several potential limitations to my study. The low failure rate of TE in this study was at odds with much larger reports of clinical practice. The relatively high prevalence of fibrosis in this cohort means that the findings may

not be reliably applied to lower prevalence populations such as the primary care setting, where the positive predictive value of the test will be lower.

4.8 Conclusion

This study has demonstrated that the performance of ELF in detection of liver fibrosis in subjects with CHB is good and is reproducible. Both ELF and TE perform well in the prediction of fibrosis at all stages, with TE superior at detecting severe fibrosis and cirrhosis in this cohort that contained a high prevalence of severe fibrosis. Further analyses in cohorts of subjects with CHB are required.

This study provides reassurance that the ELF test performs well when compared to TE. The two modalities may complement each other as demonstrated by modelling of a serial algorithm. When considering transferring non-invasive fibrosis assessment to community settings, there are a number of potential problems with TE. The equipment remains expensive and training is needed to perform TE. Results are known to be affected by liver inflammation and steatosis.[153] The ELF test has been shown to perform well in patients with steatosis and obesity.[86] More data exploring ELF in primary care settings where prevalence of advanced fibrosis is lower is required. This thesis will next focus on liver disease in a community population and ultimately explore the performance of the ELF test in this setting.

Chapter 5. INVESTIGATION OF CHRONIC LIVER DISEASE IN A GENERAL POPULATION

5.1 Introduction

As described in chapter 2, CLD is a major cause of death in the UK and overweight and alcohol consumption are major contributors to CLD. NAFLD can be considered the pathological manifestation in the liver of the metabolic syndrome, of which BMI is a key feature.

The precise influence of BMI on the risk of liver disease in women is not conclusive and previous studies using small subsets of ICD-10 codes to identify CLD may have underestimated the impact of BMI and alcohol.[170,171] Further, interaction between alcohol and BMI and risk of CLD is not well understood.

5.2 Aims of study

The aim of my study was to estimate the incidence of CLD in a large community population by extracting data from the UKCTOCS trial to perform a prospective cohort study nested in the UKCTOCS trial cohort. Additional aims were to define the prevalence of risk factors in the UKCTOCS participants, comprising overweight and obesity, and alcohol use, and to evaluate the contribution of the risk factors and to examine the interaction between risk factors.

5.3 Principles of statistical methods used in this chapter – Survival analysis

As will be shown, my data comprise censored participants and participants who did experience an event, from a fixed point (return of UKCTOCS questionnaire), therefore time to event analysis is appropriate.

I considered alternative approaches including Poisson regression, which would generate incident rate ratios that may be an easily understandable way to describe risk to patients. However, Poisson regression considers multiple presentations of event. From a clinical hepatology perspective, *first* presentation with liver disease is key (rather than the number of times an individual presents), and my aim is to investigate strategies for earlier first presentation of liver disease.

Kaplan-Meier analysis allows estimation of survival (or hazard) over time, even when participants are lost from follow up or are followed up for different amounts of time. For each interval, survival or hazard probability is calculated (participants surviving divided by participants at risk). Censored participants are not included in the denominator. Probability of surviving to a time point is estimated from the cumulative probability of surviving each of the previous intervals (calculated as the product of previous probabilities). Although the probability calculated at any time point may not be particularly accurate because of the small number of events, the overall probability of surviving is more accurate. The vertical axis represents the estimated probability of survival (or hazard) from a hypothetical cohort, not the actual percentage surviving. The log rank test tests whether there is a statistically significant difference between

groups (i.e. curves on the plot). This method, however, does not take into account adjusting for other covariates / confounding variables. In order to adjust for confounders, Cox proportional hazards regression analysis can be used.

Proportional hazards models are survival models. Survival models provide estimates of the effect that a factor (risk factor, covariate, exposure) has on the time to event. At the end of the observation period, the event may not have occurred in an individual. Another possibility is that an individual has left the study early or been lost to follow up. The survival times in these situations are censored.

Cox proportional hazards regression allows the effect of multiple risks / exposures on survival time to be considered simultaneously. The resulting coefficients are interpreted in the same way as in standard multiple linear regression.

The hazard ratio (HR) is the ratio of the hazard (the chance of event) in one group (e.g. presence of risk factor) divided by the hazard in another group (e.g. absence of risk factor) in a particular time period.

The survival model consists:

- The baseline hazard function (how the risk of events per unit time changes over time at the baseline level of covariate / exposure)

  And:

- Effect parameters (how the hazard change due to the covariate / exposure)

The proportional hazards condition states that the covariate is multiplicatively related to the hazard, e.g. the hazard may double with a risk factor or a unit change in exposure (while the baseline hazard may vary).

The partial likelihood is a key component of Cox regression and describes the situation when only the parameters of interest occur in the likelihood. The likelihood does not contain the shape of the hazard over time.

The Cox partial likelihood is obtained by using Breslow's estimate of the baseline hazard function, putting this in to the full likelihood, then observing that the result is the product of two factors:

- The partial likelihood – in which the baseline hazard has been 'cancelled out'
- A factor free of the regression coefficient which depends on the data only through the censoring pattern

Therefore, the effects of the covariates estimated by any proportional hazards model can be reported as HRs.

Cox regression is based on assuming proportional hazards, so that the effect parameter can be estimated without any consideration of the baseline hazard function. The HR assumes that, apart from the exposure, everything else is constant between the two (or more) groups, i.e. it assumes proportionality of the hazard functions.

5.4 Methods

5.4.1 Background to UKCTOCS

The United Kingdom Trial of Ovarian Cancer Screening (UKCTOCS) is a multi-centre UK-based randomised controlled trial designed to define the effect of ovarian cancer screening on mortality. Between April 2001 and October 2005, 202,638 women aged between 50 and 74 were recruited in England, Wales and Northern Ireland. The study is coordinated by the Gynaecological Cancer Research Centre at University College London, under the direction of Professor Usha Menon. The study design is comprehensively explained elsewhere.[172-175] Briefly, women were randomly selected from 27 local authority registers. Exclusion criteria included bilateral oophorectomy, increased risk of familial ovarian cancer, previous ovarian cancer and active non-ovarian cancer. Those who accepted the invitation were given written and verbal information about the trial, and watched an information video at one of 13 regional trial centres. Participants provided written consent and a baseline serum sample. Participants completed a baseline questionnaire (appendix D). Questions sought information on participants' cancer history and asked participants to record their height and weight.

Participants were randomly allocated to one of three arms; no screening, annual serum Cancer Antigen 125 (CA125) measurement with transvaginal ultrasound as a second line test (multimodal arm, MMS), or transvaginal ultrasound only. Recent data from the trial have demonstrated the predictive value of changes in CA125 levels to predict ovarian cancer, and potentially reduce mortality in the multimodal arm.

Follow up questionnaire

All participants in UKCTOCS were sent a follow up questionnaire approximately 3.5 years after randomisation (appendix E). This covered areas including education, current weekly alcohol consumption, smoking status, skirt size and medical history.

Longitudinal follow up

Participants in the MMS arm were screened annually with blood tests for at least 7 years. In addition, via their National Health Service (NHS) number in England and Wales participants were linked to NHS Digital (formerly the Health and Social Care Information Centre and prior to this the Office for National Statistics) for cancer and death registrations. For participants in England, data was also obtained from the National Cancer Intelligence Network, and from the Hospital Episodes Statistics (HES) records. Data for participants in Northern Ireland were obtained from the Central Services Agency and Northern Ireland Cancer Registry.[175]

5.4.2 Ethical approval

UKCTOCS was approved by the UK North West Multicentre Research Ethics Committee (North West MREC 00/8/34), with site-specific approval from the local regional ethics committees and the Caldicott guardians (data controllers) of the primary care trusts. At recruitment, written informed consent was obtained from all volunteers for use of data and samples in future ethically approved secondary studies.

Ethics approval for the work relating to this thesis was approved by the National Research Ethics Service (NRES) Committee London - Bentham (Ref: 05/Q0505/57) on 10th August 2011.

5.4.3 Study population

My study was nested within UKCTOCS. As HES data was available for those participants in England, this study was restricted to women recruited via the recruiting centres in England.

5.4.4 Exposures

My exposures of interest were BMI and current weekly alcohol consumption. I used self-reported height and weight data from the baseline questionnaire to calculate BMI (BMI (kg/m^2) = weight (kg) / (height (m))^2) and categorised BMI according to the World Health Organization's definitions; normal (<25 kg/m^2), overweight (25 - <30 kg/m^2) and obese (≥30 kg/m^2). As there are no existing population estimates for the range of BMI, I adopted a pragmatic approach to selecting participants with plausible BMI values. Participants with a height outside the range 140-210 cm, or a weight outside the range 25-200 kg, or a calculated BMI outside the range 16-65 kg/m^2 were excluded.

The UKCTOCS follow up questionnaire asked participants to estimate their alcohol consumption as the number of alcoholic drinks consumed per week (none, less than 1, 1-3, 4-6, 7-10, 11-15, 16-20 or ≥21 drinks), assuming one drink is a glass of wine, half a pint of beer or cider, or a measure of spirits. I calculated alcohol units using the convention that one drink is the equivalent of

one UK unit (10 ml or 8 g of pure alcohol).[176] Participants with no alcohol response were excluded.

5.4.5 Covariates

I used data from the UKCTOCS follow-up questionnaire to identify possible covariates. The follow-up questionnaire asked participants to report on known comorbidities including heart disease, hypercholesterolaemia, hypertension and type 2 diabetes mellitus, and whether they currently smoked (all categorised as yes / no). Socioeconomic status was estimated using the Index of Multiple Deprivation 2007 (IMD) (continuous variable).[177] This ascribes a deprivation score to participants based on their postcode, with a higher score indicating higher deprivation.

5.4.6 Follow up

I extracted follow-up data from the UKCTOCS database for participants included in my study. As described above, participants in UKCTOCS are 'flagged' for clinical events via a number of sources. The NHS Information Centre for Health and Social Care in England and Wales provides data on cancer registrations and deaths, with the diagnosis / cause of death coded according to the International Classification of Diseases, version 10 (ICD-10). 99.98% of UKCTOCS participants were successfully flagged. The HES database provided hospital inpatient and outpatient episode data for 2001-2010. Each HES record reports a main diagnosis and up to 19 (inpatient admissions) and 11 (outpatient admissions) further diagnoses. Each death record reports the primary cause of death and additional contributory causes recorded on the death certificate.

I searched the UKCTOCS follow-up databases for ICD-10 codes of interest (see below) by linking my Microsoft Access file to the UKCTOCS follow-up Access files to import the date and code for each event for each participant who experienced an event from the start date of my study. The starting point for participants entering my study was the date that they returned the follow-up questionnaire to UKCTOCS and therefore I prospectively followed up participants using data extraction from this point. Women with known pre-existing liver disease were not included, by excluding those where a code of interest had been registered between recruitment to UKCTOCS and return of follow-up questionnaire.

5.4.7 Outcomes

The main outcome measure was first liver-related event (LRE), defined as first presentation of either a hospital admission, outpatient appointment, cancer registration with, or death from, an ICD-10 code of interest. I searched for the following codes for liver disease: K70 (alcoholic liver disease), K73 (chronic hepatitis) and K74 (fibrosis and cirrhosis). These codes are consistent with other UK studies of cirrhosis.[26,170] K76 (other diseases of liver, including fat) was also included in order to widen the search for liver disease beyond cirrhosis to include fatty liver disease.

In addition, I searched for codes relating to sequelae of decompensated liver disease; I85 (oesophageal varices), Z94.4 (liver transplant) and C22.0 (hepatocellular carcinoma). In addition to ICD-10 codes, death certificates were also searched for any mention of alcoholic liver disease or non-alcoholic fatty liver disease.

A table listing all ICD-10 codes relevant to this thesis (including those in other studies that are discussed in the thesis) is presented in appendix F.

5.4.8 Database creation

I created a Microsoft Access database for this study which was held securely on the UKCTOCS hard drive. All UKCTOCS participants were anonymised for this study by assigning a unique identification number ("hepatology ID") to create an autonomous database. Any additional baseline or new data could be added to my database through a separate, secure Access file that linked the UKCTOCS participant ID to the hepatology ID. Initially I imported baseline data for participants in my study from the central UKCTOCS Access databases, and subsequently added follow-up data of LRE as described above. I exported data from my master Access database in to statistical software packages for data analysis.

5.4.9 Statistical analysis

Prior to individual level (survival) analysis, I calculated crude incidence rates of first LRE using person-years of follow-up as the denominators, for each BMI group, each alcohol group and each BMI / alcohol combination, to appreciate the absolute rates of event in my population. For each participant, person-years of follow-up were accrued from date that the follow-up questionnaire was returned (as this was the date that current alcohol use was ascertained), to the censorship date (February 1, 2013), date of first presentation with LRE, or death from any other cause, whichever was first. Participants who experienced a LRE at any time from randomisation to return of questionnaire were excluded.

5.4.9.1 Separate influences of BMI and of alcohol on incident liver disease

I used Cox proportional hazards models so that specification of an underlying hazard was not required. Hazard ratios (HRs) of first LRE in three categories of BMI were calculated using normal BMI as the reference. Similar analysis was performed for alcohol, with no alcohol consumption as the reference. The proportional hazards models were adjusted for BMI or alcohol, respectively.

All potential confounding risk factors (these were IMD, and via the UKCTOCS questionnaires self-reported smoking, hypertension, heart disease, hypercholesterolaemia and type 2 diabetes) were included individually in a Cox regression model to estimate their univariate associations with LREs, to guide their utility in the models evaluating risk associated with BMI and alcohol.

5.4.9.2 Interaction (and background to interaction)

It is important to appreciate the difference between confounding and interaction. A confounder variable is associated with the exposure *and* associated with the outcome. It can 'confuse' the association between a variable of interest and outcome / incidences. It can be dealt with by 'adjusting' for it in multivariate analyses.

Interaction indicates that the effect of one variable on outcome is different at different values of the other variable. It is tested by adding a term to the model in which the two variables are multiplied. The two variables combine to produce an effect that is not simply additive, i.e. they are not acting independently and

produce a greater or lesser effect than the sum of the effects of each variable acting on its own.

An interaction term was calculated between alcohol and BMI groups.

5.4.9.3 Influences of combinations of BMI and alcohol

Survival analyses were initially performed to examine the cumulative hazard of LRE using Kaplan-Meier curves, censoring at death or 'last known alive'. Participants were divided in to groups based on alcohol and BMI thresholds to examine the effects of 'high alcohol use', 'high BMI' and combinations. High alcohol was considered ≥21 units / week and high BMI was considered as overweight in the first analysis, and obese in the second analysis.

To further investigate the effects of combinations of alcohol and BMI, using Cox proportional hazards regression analysis, HRs were calculated for twelve BMI and alcohol combinations using the normal BMI / no alcohol consumption category as the reference, adjusted for potential confounders with significant HRs for LRE at the 5% level, and then adjusted only for factors associated with the metabolic syndrome (hypertension, hypercholesterolemia, heart disease and type 2 diabetes). The proportional hazards assumption was checked by examining the log minus log plot.

All analyses were performed using SPSS (version 22, SPSS Inc, Chicago, IL, USA), STATA statistical software (StataCorp 2007. Release 10. College Station, TX, USA: StataCorp LP) and R (R Foundation for Statistical Computing, Vienna, Austria).

5.5 Results

Derivation of study cohort

Of the 157,996 UKCTOCS participants resident in England, 62,870 were excluded. Participants who did not return the follow-up questionnaire to UKCTOCS could not be included. The other excluded participants were those who experienced an LRE or died between recruitment and return of questionnaire, those who did not provide a response to the alcohol or smoking question and those where there was no valid BMI. The ages and other characteristics appear similar between the study group and the remainder of the participants in England who returned the follow-up questionnaire, although differences were statistically different at the 5% level (appendix G). The composition of the final study cohort of 95,126 participants and its derivation is shown in figure 5.1.

Baseline characteristics

Baseline characteristics are shown in table 5.1. 97.1% of the participants were white. 36% were smokers. 55% were either overweight (37%) or obese (19%). 23.4% reported consuming no alcohol and 1.5% reported drinking more than 21 units / week. The most common comorbidity reported by participants was hypertension (32%), followed by hypercholesterolaemia (24%), heart disease (6%) and type 2 diabetes (5%). Increasing BMI correlated with increased reporting of hypertension, heart disease, hypercholesterolaemia and type 2 diabetes. The mean Index of Multiple Deprivation score in the study cohort was 18.5, and this too increased with increasing BMI.

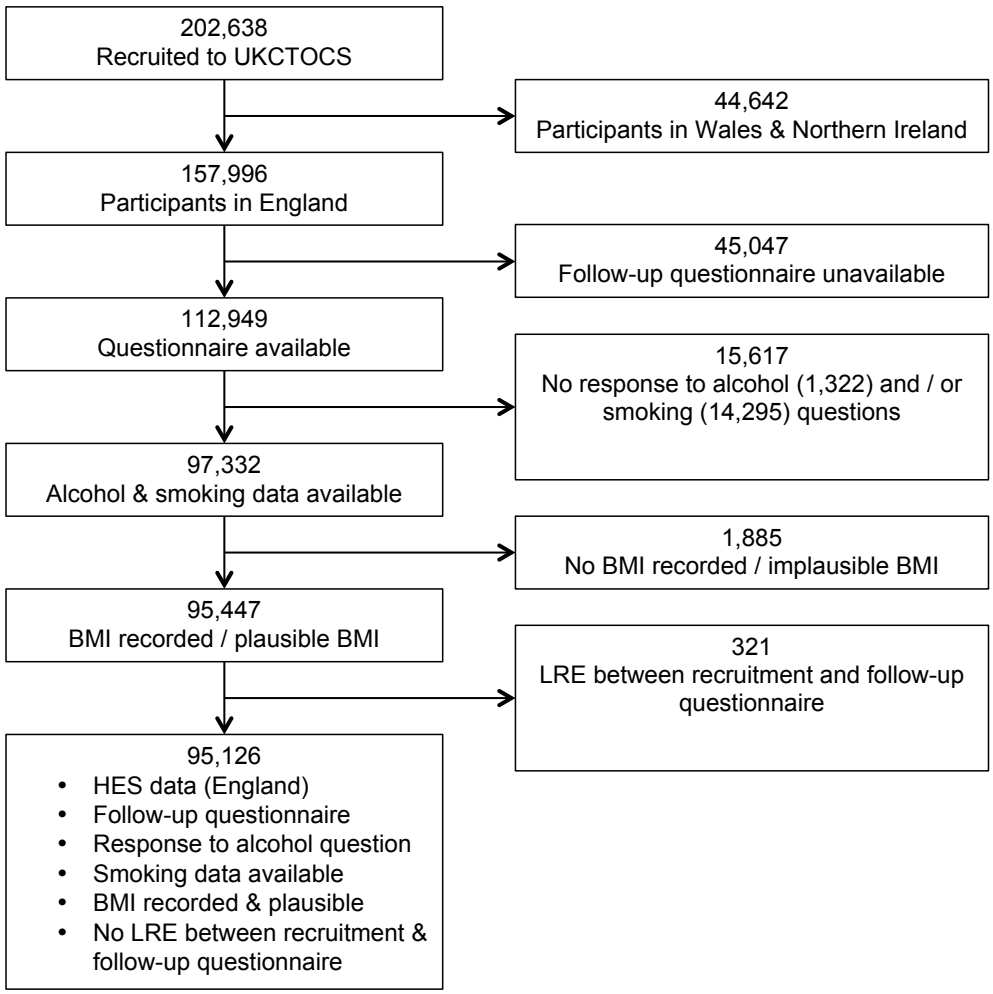


Figure 5.1. The composition of the final study cohort and its derivation from the UKCTOCS cohort

The final cohort comprised 95,126 participants. Participants in Wales and Northern Ireland were excluded due to lack of HES data. Participants who did not return the UKCTOCS questionnaire or did not respond to the alcohol or smoking status questions were removed, as were those who did not record BMI or if the recorded BMI was implausible. Participants who experienced a liver-related event before the start of the study were excluded

BMI, body mass index; HES, Hospital Episode Statistics; LRE, liver-related event; UKCTOCS, United Kingdom Collaborative Trial of Ovarian Cancer Screening

Table 5.1. Baseline characteristics of the study cohort according to BMI category and in all participants

Data are presented for the entire cohort, comprising 95,126 participants, and categorised according to WHO BMI category. Self-reported comorbidities, smoking and alcohol status are shown

Characteristic	BMI category (kg/m^2)			All participants
	<25	25 - <30	≥30	
Total, *n* (%)	42,452 (44.6)	35,073 (36.9)	17,601 (18.5)	95,126
Recruitment age in years, median (IQR)	60.0 (50-74)	61.0 (50-74)	60.0 (50-74)	60.0 (50-74)
Smoker, *n* (%)	14,740 (34.7)	12,616 (36.0)	6621 (37.6)	33,977 (35.7)
Hypertension, *n* (%)	9477 (22.3)	12,116 (34.5)	8440 (48.0)	30,033 (31.6)
Heart disease, *n* (%)	1721 (4.1)	2086 (5.9)	1416 (8.0)	5223 (5.5)
Hypercholesterolemia, *n* (%)	8001 (18.8)	9148 (26.1)	5440 (30.9)	22,589 (23.7)
Type 2 diabetes, *n* (%)	836 (2.0)	1689 (2.6)	2263 (12.9)	4788 (5.0)
IMD, mean	17.0	18.7	21.3	18.5
Alcohol consumption (units/week)				
None	8479 (20.0)	8189 (23.3)	5547 (31.5)	22,215 (23.4)
<1-15	31,811 (74.9)	25,324 (72.2)	11,473 (65.2)	68,608 (72.1)
16-20	1448 (3.4)	1067 (3.0)	366 (2.1)	2881 (3.0)
≥21	714 (1.7)	493 (1.4)	215 (1.2)	1422 (1.5)

BMI, body mass index; IMD, Index of Multiple Deprivation; LRE, liver-related event; WHO, World Health Organization

5.5.1 Incidence of chronic liver disease

325 (0.34%) women experienced a first LRE over a total of 509,561 person-years of follow-up (mean 5.1 years), equivalent to 0.64 first events per 1000 person-years (3.3 per 1000 women over 5 years). The most common ICD-10 code for the study definition of LRE was K76, ‘other diseases of liver’ (table 5.2).

Table 5.2. ICD-10 codes and / or death certificate text of first LREs
The number of codes / death certificate text results is higher than the number of LREs (325) as some participants had more than one code when presenting with first LRE. Numbers of participants with codes of interest are divided by source of the code (hospital admission (HES), outpatient appointment (HES), cancer registration (ONS) and death certification)

Source	Code or text	Number of participants (% of those with LRE)
Hospital admission	K70	15 (4.6)
	K73	9 (2.8)
	K74	45 (13.8)
	K76	183 (56.3)
	C22.0	7 (2.2)
	I85	12 (3.7)
	Z94.4	33 (10.2)
Outpatient appointment	K74	1 (0.3)
	Z94.4	11 (3.4)
Cancer registration	C22.0	12 (3.7)
Death certificate	K70	6 (1.8)
	K74	7 (2.2)
	K76	10 (3.1)
	C22.0	2 (0.6)
	Mention of alcoholic liver disease	8 (2.5)
	Mention of non-alcoholic fatty liver disease	8 (2.5)

HES, Hospital Episode Statistics; ICD-10, International Classification of Diseases, Version 10; LRE, liver-related event; ONS, Office for National Statistics

One thousand two hundred and thirty-seven (7% of the obese group) women could be classified as morbidly obese (BMI ≥40 kg/m^2) and in this group, the event rate was highest (1.98 events per 1000 person years (95% CI; 1.05-3.38)). There were 2713 (2.9%) deaths from any cause.

5.5.2 BMI and risk of liver-related events

Crude rates of LREs increased with rising BMI (figure 5.2, table 5.3). HRs for LRE were significantly higher in both overweight (1.44, 95% CI; 1.10-1.87) and

obese (2.25, 95% CI; 1.70-2.97) categories compared to the normal BMI group. A partially adjusted model (adjusted for components of the metabolic syndrome) and a fully adjusted model are presented incorporating adjustment for all confounders with significant HRs (table 5.4).

5.5.3 Alcohol consumption and risk of liver-related events

The rate of LRE was lowest in the group consuming <1-15 units weekly and increased with abstinence and increasing alcohol use (figure 5.2, table 5.3). This tendency towards a “J-shaped” relationship between LRE and alcohol consumption was seen in the unadjusted HR estimates, and preserved after adjustment for BMI, with lowest HRs in the <1-15 units/week group, although the there was no statistically significant difference between the HRs for this group and the reference group. A partially adjusted model (adjusted for components of the metabolic syndrome) and a fully adjusted model are shown, adjusted for variables with significant HRs for LRE (table 5.5).

In the group reporting no current alcohol consumption the proportion of LREs that were alcohol-related was 3.96% compared to 11.16% in those consuming any alcohol.

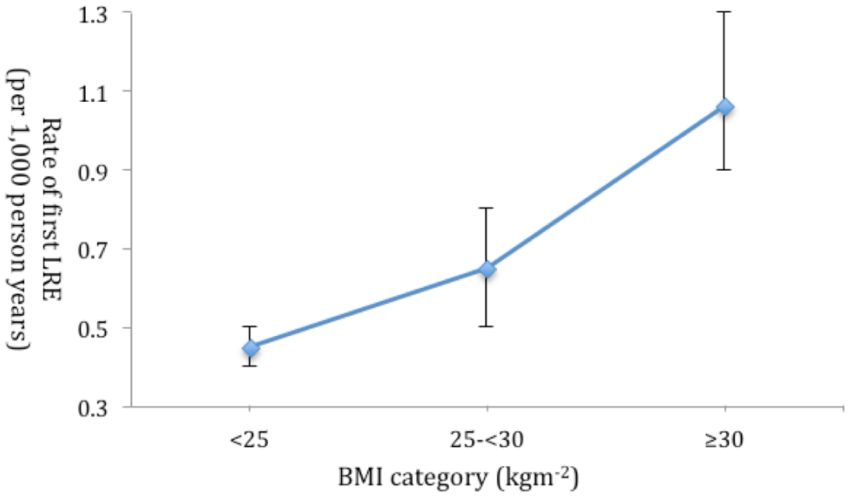
5.5.4 Risk of liver-related events due to other covariates

Other covariates also demonstrated independent association with liver-related events (table 5.6 and figure 5.3). Significant HRs were seen with smoking, hypertension, heart disease, hypercholesterolaemia, type 2 diabetes and IMD.

Table 5.3. Crude rates of first liver-related events (per 1000 person years) according to BMI category and alcohol consumption category over mean follow-up of 5.1 years

BMI or alcohol categories	Number of events Person year follow up Incidence per 1000 person years (95% confidence intervals)
BMI category	
<25	102 227211.7 0.45 (0.37-0.54)
25 - <30	123 188186.8 0.65 (0.55-0.78)
≥30	100 94162.6 1.06 (0.87-1.29)
Alcohol category	
None	101 117198.8 0.86 (0.71-1.04)
<1-15	202 368293.6 0.55 (0.48-0.63)
16-20	11 16097.7 0.68 (0.35-1.19)
≥21	11 7971.2 1.38 (0.73-2.40)

a)



b)

Figure 5.2. Crude rates of first liver-related events (per 1000 person years) according to a) BMI category and b) alcohol consumption category over mean follow-up of 5.1 years

The change in crude rates for first liver related events associated with increasing BMI and alcohol consumption are depicted. The rate of LRE rises with increasing BMI but follows a "J" or "U" shaped relationship with alcohol consumption in which abstinence is associated with a greater risk than moderate consumption (<1-15 units / week)

Table 5.4. Hazard ratio estimates of first liver-related events, according to BMI category

Hazard ratio estimates are presented, adjusted for alcohol category, and the following models; unadjusted, partially adjusted (alcohol category, hypertension, heart disease, hypercholesterolemia and type 2 diabetes) and fully adjusted (alcohol category, smoking, hypertension, heart disease, hypercholesterolaemia, type 2 diabetes and IMD)

BMI category (kg/m^2)	Hazard ratio (95% confidence intervals)			
	Adjusted for alcohol category	Unadjusted	Partially adjusted	Fully adjusted
Continuous variable	1.062 (1.043-1.081)	1.065 (1.05-1.08)	1.052 (1.032-1.072)	1.048 (1.028-1.069)
<25	1 (reference)	1 (reference)	1 (reference)	1 (reference)
25 - <30	1.44 (1.10-1.87)	1.457 (1.12-1.90)	1.37 (1.05-1.79)	1.33 (1.01-1.73)
≥30	2.25 (1.70-2.97)	2.367 (1.80-3.12)	1.99 (1.49-2.66)	1.87 (1.41-2.53)

BMI, body mass index; IMD, Index of Multiple Deprivation

Table 5.5. Hazard ratio estimates of first liver-related events, according to alcohol category

Hazard ratio estimates are presented, adjusted for BMI (continuous variable), and the following models; unadjusted, partially adjusted (BMI, hypertension, heart disease, hypercholesterolemia and type 2 diabetes) and fully adjusted (BMI, smoking, hypertension, heart disease, hypercholesterolaemia, type 2 diabetes and IMD)

Alcohol category	Hazard ratio (95% confidence intervals)			
	Adjusted for BMI	Unadjusted	Partially adjusted	Fully adjusted
None	1 (reference)	1 (reference)	1 (reference)	1 (reference)
<1-15	0.70 (0.55-0.88)	0.64 (0.51-0.82)	0.75 (0.59-0.95)	0.78 (0.61-1.00)
16-20	0.93 (0.50-1.73)	0.82 (0.44-1.53)	1.02 (0.55-1.91)	0.97 (0.52-1.82)
≥21	1.82 (0.97-3.39)	1.66 (0.89-3.09)	1.99 (1.06-3.71)	1.83 (0.97-3.44)

BMI, body mass index; IMD, Index of Multiple Deprivation

Table 5.6. Event rates of LRE for each covariate and hazard ratios for univariate associations between each covariate and liver-related event

Characteristic		Number of events Person year follow up Incidence per 1000 person years (95% confidence intervals)	Hazard ratio (95% confidence intervals)
Recruitment age			1.01* (0.99-1.02)
Smoker		170 197397.8 0.86 (0.74-0.99)	1.89** (1.52-2.35)
Hypertension		126 158768.2 0.79 (0.66-0.94)	1.38** (1.11-1.73)
Heart disease		36 27616.1 1.30 (0.93-1.79)	2.17** (1.53-3.06)
Hypercholesterolemia		111 119504.9 0.93 (0.77-1.11)	1.68** (1.33-2.11)
Type 2 diabetes		41 25066.1 1.64 (1.19-0.20)	2.76** (1.99-3.83)
IMD			1.019* (1.01-1.03)
IMD tertile	1	76 169816.9 0.45 (0.36-0.56)	1** (reference)
	2	93 165934.5 0.56 (0.45-0.68)	1.248** (0.92-1.69)
	3	151 171152.8 0.88 (0.75-1.03)	2.013** (1.53-2.65)

* continuous variable; ** categorical variables
BMI, body mass index; IMD, Index of Multiple Deprivation

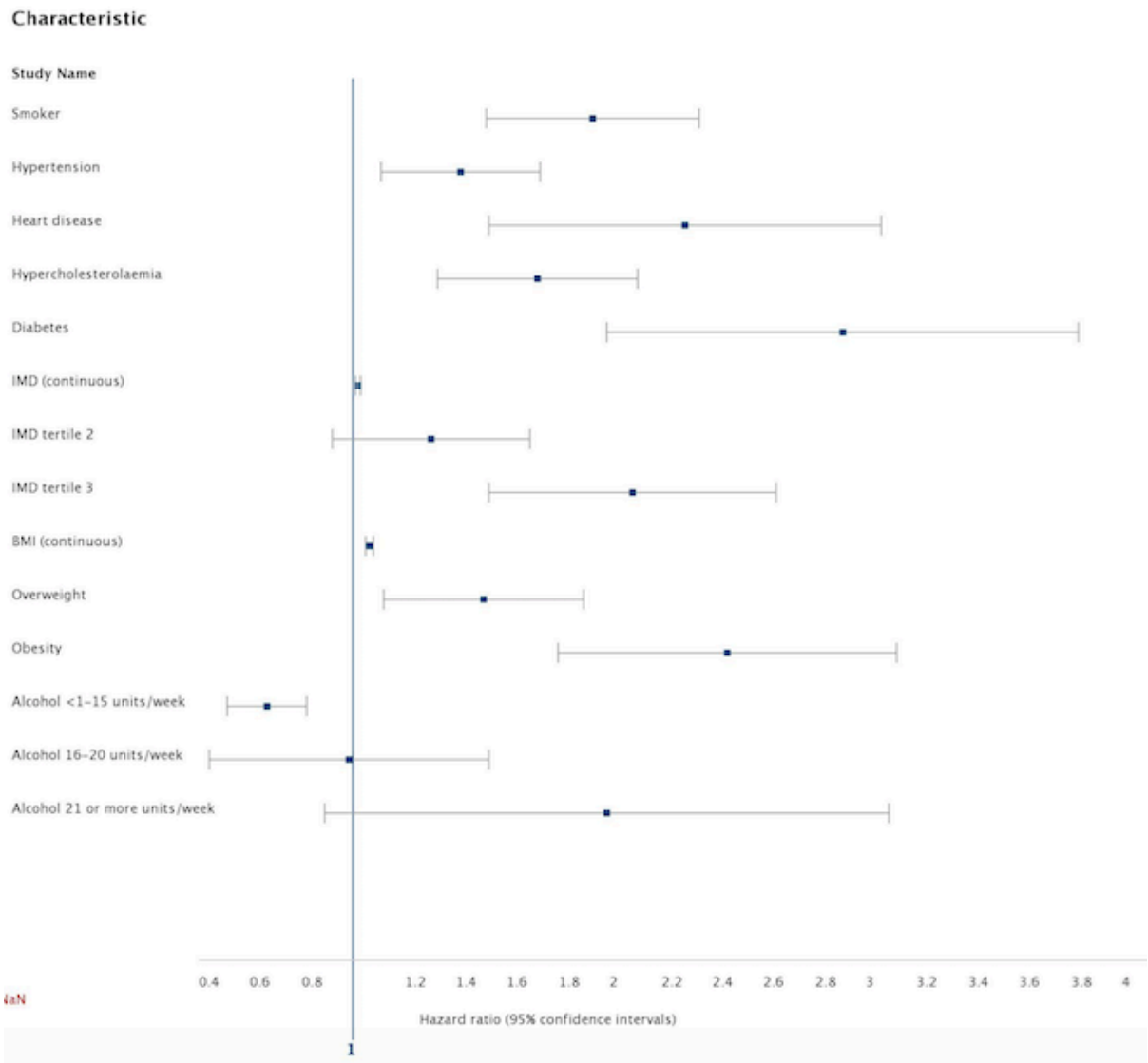


Figure 5.3. Forest plot showing hazard ratio (HR) estimates and 95% confidence intervals for univariate associations between baseline characteristics and liver-related event

Reference HRs for smoking, hypertension, heart disease, hypercholesterolaemia and type 2 diabetes are absence of the covariate. Reference HR for IMD tertiles 2 and 3 is IMD tertile 1. Reference HR for overweight and obesity is normal BMI. Reference HR for the alcohol groups is abstinence

IMD, Index of Multiple Deprivation

Forest plot generated using DistillerSR Forest Plot Generator from Evidence Partners (https://www.evidencepartners.com/resources/forest-plot-generator/#forest_plot_5_graph_edit_linebyline)

5.5.5 Interaction between alcohol and BMI

Interaction between alcohol groups and BMI groups was calculated (SPSS output shown below). There is no significant interaction seen (p = 0.966).

Omnibus Tests of Model Coefficients[a]

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change ...
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square
7366.287	62.323	11	.000	52.940	11	.000	52.940

Omnibus Tests of Model Coefficients[a]

Change From Previous	
df	Sig.
11	.000

a. Beginning Block Number 1. Method = Enter

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI Lower
BMI_categories			16.396	2	.000		
BMI_categories(1)	.478	.267	3.210	1	.073	1.613	.956
BMI_categories(2)	1.025	.259	15.612	1	.000	2.787	1.676
Alcohol_categories			3.595	3	.309		
Alcohol_categories (1)	-.206	.240	.734	1	.392	.814	.509
Alcohol_categories (2)	-.009	.542	.000	1	.987	.992	.343
Alcohol_categories (3)	.701	.542	1.674	1	.196	2.016	.697
Alcohol_categories *BMI_categories			1.396	6	.966		
Alcohol_categories (1)*BMI_categories (1)	-.170	.314	.294	1	.587	.844	.456
Alcohol_categories (2)*BMI_categories (1)	.051	.722	.005	1	.943	1.053	.256
Alcohol_categories (3)*BMI_categories (1)	.111	.722	.024	1	.878	1.117	.271
Alcohol_categories (1)*BMI_categories (2)	-.279	.316	.777	1	.378	.757	.407
Alcohol_categories (2)*BMI_categories (2)	-.338	.904	.139	1	.709	.713	.121
Alcohol_categories (3)*BMI_categories (2)	-.521	.904	.332	1	.565	.594	.101

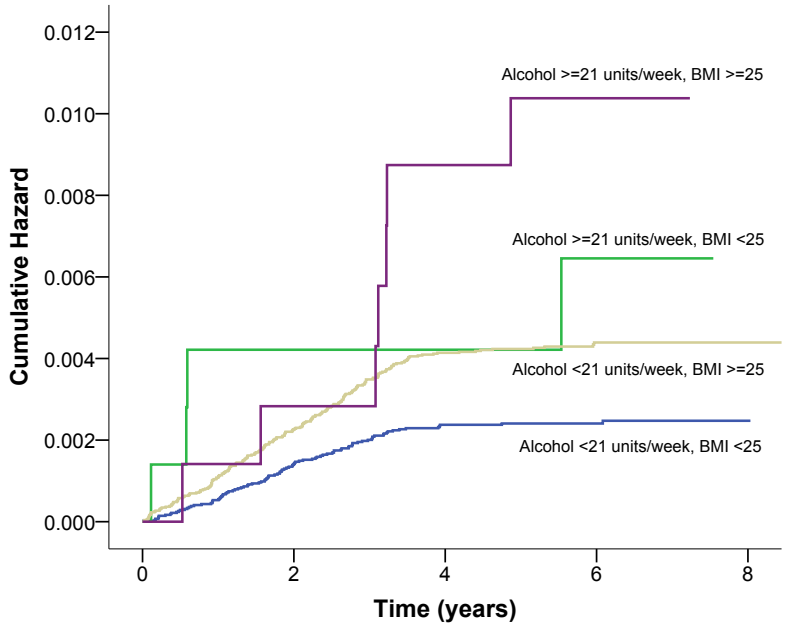
Because p ≠ 1 for interaction, there appears to be some impact from interaction, but as this is not significant, in this population there is no 'clinically significant' interaction. Therefore, in my study group, the HR estimates for each BMI group are not varying by level of alcohol consumption, and vice versa.

As there is no interaction, the HR estimates for each of the twelve alcohol / BMI combinations may be calculated by multiplying the HR for BMI by the HR for alcohol group to generate the HRs for the additional six groups. However, a recent meta-analysis (Parkes, unpublished data) shows that interaction has been seen in other populated based studies. Therefore, I created a twelve-level variable (i.e. twelve BMI / alcohol combinations), which assumes an interaction term. This model allows for the effect of one variable (i.e. BMI) to be modified by the other (i.e. alcohol), allowing them to interact and not be independent. The model is presented in the next section, and the resultant HR estimates are lower than the calculated (independent) HRs, producing more conservative estimates of association (data not shown).

5.5.6 Kaplan-Meier estimator results

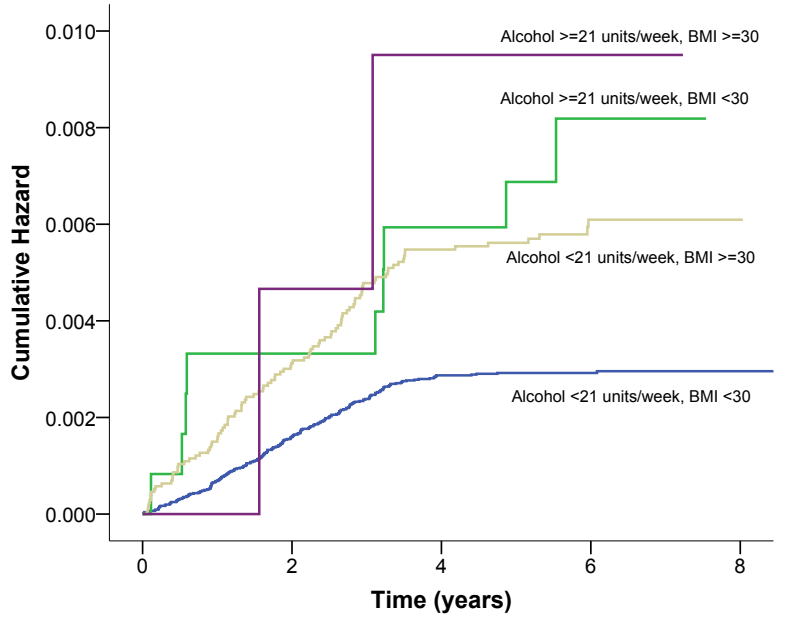
Crude unadjusted analyses by Kaplan-Meier plots to eight years showed that there was a graded relationship between alcohol / BMI and LRE. The highest cumulative hazard was seen in the group reporting alcohol consumption ≥21 units / week who were overweight. The lowest cumulative hazard was seen in the group reporting alcohol consumption <21 units / week who were not overweight. These estimates suggest that 'high alcohol' use may be riskier than 'high BMI' as of the two middle groups, the cumulative hazard was higher in the group consuming ≥21 units / week who were not overweight compared to the group consuming <21 units / week who were overweight. The same pattern was seen when obesity was used as the threshold for 'high BMI' (figure 5.4). Log rank tests showed that the groups in each analysis were significantly different (p <0.001 for both analyses).

a)



Time (years)	0	2	4	6
	Number of participants at risk (cumulative events)			
Alcohol ≥21 units/week, BMI ≥25	41738 (0)	41347 (59)	36097 (96)	16336 (97)
Alcohol ≥21 units/week, BMI <25	713 (1)	705 (3)	665 (3)	322 (4)
Alcohol <21 units/week, BMI ≥25	51966 (0)	51448 (117)	45143 (208)	19987 (216)
Alcohol <21 units/week, BMI <25	707 (1)	703 (2)	650 (6)	330 (7)

b)



Time (years)	0	2	4	6
	Number of participants at risk (cumulative events)			
Alcohol ≥21 units/week, BMI ≥30	76318 (0)	75620 (122)	66144 (212)	29740 (215)
Alcohol ≥21 units/week, BMI <30	1206 (1)	1195 (4)	1118 (7)	569 (19)
Alcohol <21 units/week, BMI ≥30	17386 (0)	17175 (55)	15096 (92)	6583 (98)
Alcohol <21 units/week, BMI <30	214 (1)	212 (1)	192 (2)	83 (2)

Figure 5.4. Kaplan-Meier curves of hazard of liver-related event to eight years for groups of UKCTOCS participants based on thresholds of self-reported alcohol consumption and body mass index

a) participants grouped using alcohol consumption threshold of 21 units/week and body mass index threshold of 25kg/m^2. b) participants grouped using alcohol consumption threshold of 21 units/week and body mass index threshold of 30kg/m^2. The plots show that cumulative hazard of liver-related event is highest in the group of participants who are overweight and consume ≥21 units/week of alcohol, and lowest in participants with normal body mass index who consume <21 units/week of alcohol. The same pattern is seen when obesity is used as the body mass index.

Numbers of participants 'at risk' for each group is shown with the cumulative number of liver-related events at each two-year time point

5.5.7 Risk of liver-related events in participants grouped in to combinations of BMI and alcohol use

Participants were grouped according to combinations of BMI and alcohol consumption. Table 5.7 shows the numbers of first LREs in each category and the corresponding rates of LRE. Table 5.8 and figure 5.5 show the hazard ratio estimates calculated using Cox proportional hazards regression for each category, using the group reporting abstinence from alcohol with normal BMI as the reference group. Three models are presented; a univariate model, a 'fully adjusted' model adjusted for smoking, hypertension, heart disease, hypercholesterolaemia, type 2 diabetes and IMD, and a 'partially adjusted' model adjusted for hypertension, heart disease, hypercholesterolemia and type 2 diabetes. In all models, the lowest risk is in those with normal BMI consuming <1-15 units / week. Within the normal BMI group, abstinence or consuming >16 units / week increases the risk of LRE, although there are wide confidence intervals. The highest HR estimates are seen in those with the highest BMI reporting the highest alcohol consumption.

Table 5.7. Event rate of first liver-related event for each BMI / alcohol combination

Data are presented for number of first LREs within each of the twelve BMI / alcohol categories, with corresponding person year follow up values and incidence rate per 1000 person years

Number of events Person year follow up Incidence per 1000 person years (95% confidence intervals)					
BMI category (kg/m^2)	**Alcohol category (units/week)**				**All participants**
	None	**<1 – 15**	**16 – 20**	**≥21**	
<25	23 44497.2 0.52 (0.3-0.8)	71 170623.6 0.42 (0.3-0.5)	4 8079.2 0.50 (0.1-1.3)	4 4011.7 1.00 (0.3-2.6)	102
25 - <30	36 43336.5 0.83 (0.6-1.2)	77 136109.4 0.57 (0.4-0.7)	5 5972.2 0.84 (0.3-2.0)	5 2768.8 1.81 (0.6-4.2)	123
≥30	42 29365.1 1.43 (1.0-1.9)	54 61560.5 0.88 (0.7-1.1)	2 2046.3 0.9 (0.1-3.5)	2 1190.7 1.68 (0.2-6.1)	100
Total	101	202	11	11	325

BMI, body mass index; LRE, liver-related event

Table 5.8. Hazard ratio estimates of first liver-related event for each BMI / alcohol combination

Data are presented for hazard ratio estimates for an unadjusted model, and models adjusted for hypertension, heart disease, hypercholesterolemia and type 2 diabetes and adjusted for smoking, hypertension, heart disease, hypercholesterolaemia, type 2 diabetes and IMD. For each model, the group of participants with BMI <25 kg/m^2 reporting no alcohol consumption were used as the reference group

BMI category (kg/m^2)	Alcohol category (units/week)			
	None	<1 – 15	16 – 20	≥21
	Hazard ratio (95% confidence intervals) Unadjusted model			
<25	1	0.81	0.99	2.02
	(reference)	(0.51-1.30)	(0.34-2.87)	(0.70-5.83)
25 - <30	1.61	1.11	1.68	3.63
	(0.96-2.72)	(0.70-1.77)	(0.64-4.43)	(1.38-9.55)
≥30	2.79	1.72	1.97	3.34
	(1.68-4.63)	(1.05-2.80)	(0.47-8.36)	(0.79-14.15)
	Hazard ratio (95% confidence intervals) Adjusted for hypertension, heart disease, hypercholesterolemia and type 2 diabetes			
<25	1	0.85	1.07	2.13
	(reference)	(0.53-1.37)	(0.37-3.09)	(0.74-6.17)
25 - <30	1.51	1.11	1.74	3.69
	(0.89-2.55)	(0.69-1.76)	(0.66-4.57)	(1.40-9.72)
≥30	2.35	1.59	1.89	3.16
	(1.40-3.95)	(0.97-2.60)	(0.44-8.01)	(0.74-13.41)
	Hazard ratio (95% confidence intervals) Adjusted for smoking, hypertension, heart disease, hypercholesterolaemia, type 2 diabetes and IMD			
<25	1	0.91	1.03	1.93
	(reference)	(0.56-1.47)	(0.35-2.99)	(0.66-5.62)
25 - <30	1.46	1.34	1.61	3.32
	(0.85-2.50)	(0.71-1.83)	(0.61-4.26)	(1.25-8.81)
≥30	2.28	1.58	1.67	2.86
	(1.35-3.86)	(0.96-2.61)	(0.39-7.15)	(0.67-12.21)

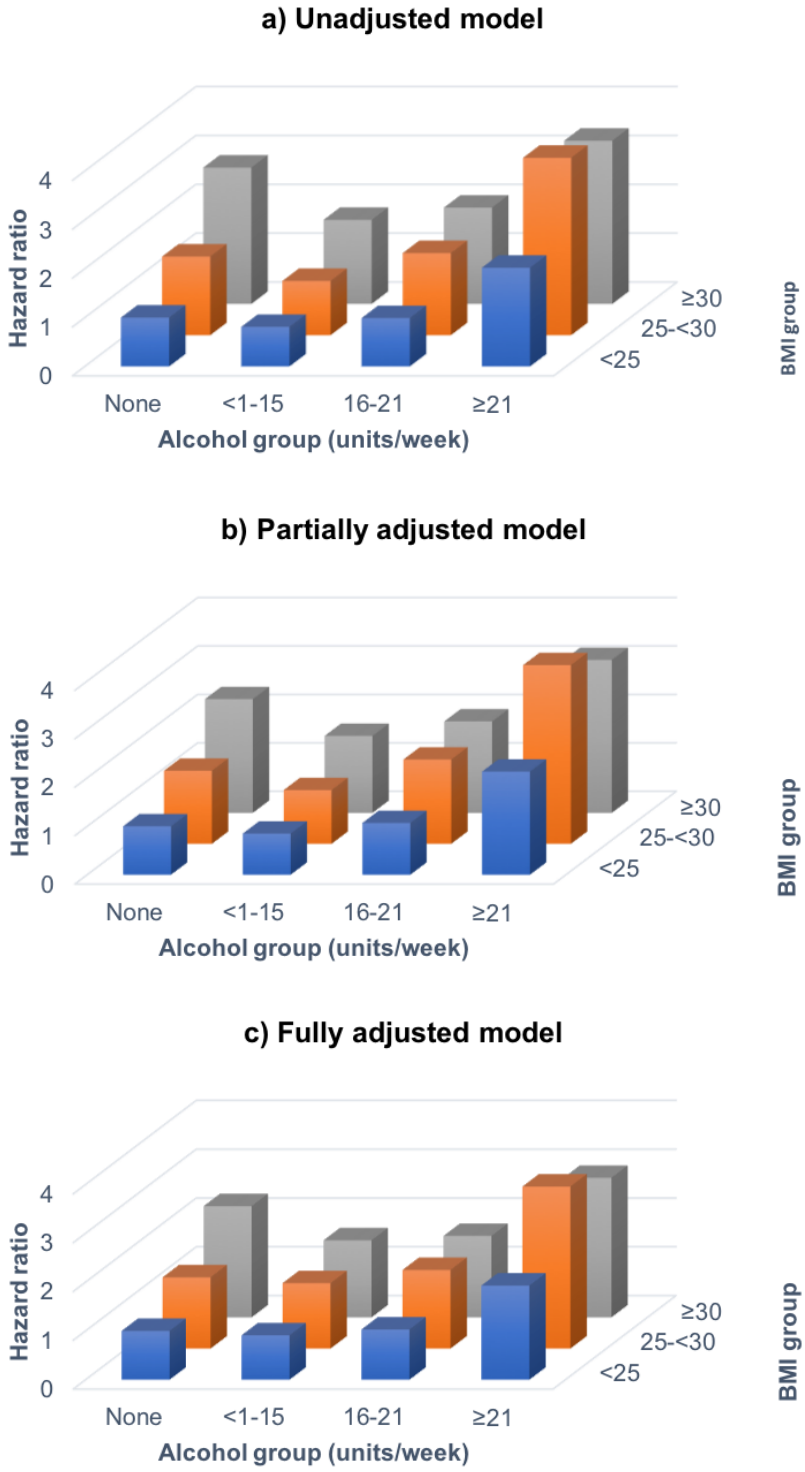


Figure 5.5. Matrices showing hazard ratio estimates of first liver-related event for each BMI / alcohol combination

Data are presented for hazard ratio estimates in the following models a) unadjusted b) adjusted for hypertension, heart disease, hypercholesterolemia and type 2 diabetes, and c) adjusted for smoking, hypertension, heart disease, hypercholesterolaemia, type 2 diabetes and IMD. For each model, the group of participants with BMI <25 kg/m^2 reporting no alcohol consumption were used as the reference group

Among overweight and obese women, the nadirs of risk were in the <1-15 units / week groups and, as in the normal BMI group, the risk was highest in the highest alcohol group (HR 3.32, 95% CI; 1.25-8.81; and HR 2.86, 95% CI; 0.67-12.21, respectively).

To estimate the effect of cardiovascular disease and type 2 diabetes on the morbidity associated with fatty liver disease, HRs were adjusted for confounding factors associated with the metabolic syndrome (partially adjusted model). When these elements of the metabolic syndrome were controlled for, risk of LRE attributable to heavier drinking increased. This suggests that the risk of liver disease attributable to BMI in patients with, or at risk of, metabolic syndrome is not entirely accounted for by hypertension, heart disease, hypercholesterolemia or type 2 diabetes, but may be partly attributable to steatosis itself.

When separated by BMI group, the trend to a “J-shaped” relationship of risk of LRE remains in all BMI groups, with risk highest in the abstainers and heavier drinkers, compared to those in the <1-15 units/week alcohol groups.

5.6 Discussion

5.6.1 Main findings

The most interesting finding of this study is the association of overweight / obesity with liver disease in postmenopausal women. These results add support to the data highlighting the adverse impact of heavy drinking compounding the effects of overweight and obesity.

This study suggests that in women aged 50-74, those consuming <1-15 units / week are at lowest risk of liver disease. Those consuming 16-20 units / week are only marginally more at risk. The UK Institute of Alcohol Studies defines hazardous drinking as more than 14 units / week and harmful drinking as >35 units / week which would be consistent with the observations in the UKCTOCS population.

Those that are overweight or obese have an increased risk of liver disease. Women of normal BMI who consume <1-15 units / week are at lowest risk, compared to those who consume more or who abstain. It is possible, however, that some abstainers had previously been heavy drinkers. This is supported by the finding that 4% of LREs in the abstainers were alcohol related.

When combinations of risk are considered, Cox proportional hazards estimates showed that, compared to a baseline of normal BMI and abstinence, higher BMI (≥30 kg/m^2) confers a greater risk than higher alcohol consumption (≥21 units / week) (this is in contrast to the Kaplan-Meier estimates and may be due to instability of the Kaplan-Meier estimate with time due to small numbers in some groups). The highest risk is in those who are overweight or obese and drink the most alcohol.

After adjustment for confounding due to metabolic risk factors, HRs in the two highest alcohol categories increased in all BMI groups, suggesting that these factors may contribute to the risk of CLD. It is biologically plausible that type 2 diabetes, hypercholesterolaemia and hypertension may contribute to liver disease over and above that caused by fatty liver disease and alcoholic liver toxicity. The corollary is that obesity can cause liver morbidity and mortality in

the absence of the metabolic syndrome, providing evidence that case ascertainment cannot be restricted to overweight or obese patients with features of the metabolic syndrome and challenging the “two hit” and “three hit” hypotheses discussed earlier.[50]

5.6.2 Strengths and limitations

5.6.2.1 Strengths

Strengths of this study include the size and duration of follow-up, the prospective design and the independence of the data capture for outcomes. This study was also able to adjust for confounding factors, which has not been possible in cross-sectional studies. In an effort to capture all morbidity and mortality attributable to liver disease, rather than just cirrhosis, ICD-10 codes were selected that encompass a clinically relevant group of diseases including codes for CLD and those relating to the consequences of decompensated liver disease. This was designed to maximise the ability to detect liver disease.

5.6.2.2 Limitations

Self-reporting

Limitations include reliance on self-reporting of alcohol consumption, co-morbidities, height and weight, which may be a factor in the wide confidence intervals seen for all HR estimates. However, good reliability of self-reporting height and weight,[178-182] and alcohol,[183-185] has been demonstrated in other studies.

Height and weight were reported at recruitment, and alcohol consumption reported, later, on the follow-up questionnaire. Participants were asked to

report current alcohol use, rather than lifetime patterns. Changes in drinking patterns would not have been identified, and this method of data collection may fail to identify episodic ("binge") drinkers. In this study, the convention that one drink is equivalent to one unit of alcohol was used. However, assumptions about alcohol content are difficult to make as measures of volume are likely to vary depending on where the alcohol is consumed, and the alcohol content of drinks continues to change. There is evidence that the number of units in alcoholic drinks in the UK have been undercounted.[186] In my study, the assumption that 1 drink = 1 unit was used as this remains a widely used convention, particularly in public health promotion.

Missing data and derivation of study group

My strategy for inclusion was for 'complete case analysis', for a number of reasons. Complete follow up data was only available for participants in England, and the absence of HES data in Wales and Northern Ireland represented a substantial lack of follow up data. As alcohol was a key exposure in my study, response to the alcohol question was essential.

Smoking was the only other covariate with a substantial amount of missing data. Although participants without smoking data could have been included, in that situation smoking should be removed from analyses. However, smoking is known to be an independent risk for developing liver disease, therefore it was important to include in models, hence those participants with no smoking data were excluded.

Participants were asked to indicate if they had certain comorbidities. The prevalence of comorbidities was derived from questions where participants only

indicated if they had the condition. Therefore, this could possibly underestimate the true prevalence, leading to a conservative estimate. Future design of a questionnaire might ask participants if they had or did not have a condition (yes / no), with non-responders representing missing data.

Miscoding

Reliance on ICD-10 to define events may result in errors due to miscoding. Additional codes to those used to define cirrhosis were used in order to maximize the capture of cases, but these may also be subject to miscoding. The risk of non-coding of events was reduced by using three independent sources. In addition, in the case of death certification, hand searching of key words in the text of death certificates is likely to have reduced the risk of missing events. The HES database may not capture some areas of healthcare, for example the private sector. The number of LREs that included ICD-10 Z94.4 is surprising (table 5.2). This may be because participants with liver transplants are engaged in hospital care and are easily identified and coded.

A possible reason for underestimation of liver disease is the failure to code for liver disease in death certification. One reason for this could be because the terminal event that leads to death in patients with CLD is often sepsis and so the underlying CLD is not recorded. Another reason may be the perception that ascribing liver disease as the cause of death is distressing to families and carers. Underreporting in the context of alcohol-related liver disease has been described. For example, in a study of over 18,000 male US army veterans, where medical records were retrospectively reviewed, six times the number of

deaths were found to be alcohol-related (133) compared to the original death certification (21).[187]

The variation in the choice of codes used in studies of liver disease and the resulting difficulty to compare studies is discussed in chapter two.

'Healthy volunteer effect'

Only postmenopausal women aged 50-74 were included with 97% being white. The loss to follow up rate in UKCTOCS was very small (0.02%). However, despite attempts to ensure that UKCTOCS was representative of the general population[173] there was a 'healthy volunteer effect'[188] on both overall and cause-specific mortality, which may have an effect on the generalisability of findings.[177] Although the health section of the follow-up questionnaire did not specifically ask about liver disease, those who had a code of interest recorded between recruitment to UKCTOCS and the start of this study were excluded. However, exclusion of all participants with known CLD could not be guaranteed.

As outlined in chapter 2, viral hepatitis is another major cause of liver disease. However, it is unlikely that viral hepatitis made a significant contribution to LRE based on low prevalence in the demographic of women in this study.[189] During the follow-up period in this study, only 21 (0.02%) of participants had a code for viral hepatitis recorded.

5.6.3 Other studies

A number of studies have demonstrated a reduced risk of liver disease in patients with NAFLD who consume low or moderate amounts of alcohol,[190-192] and it has been suggested that these levels of alcohol use may be associated

with beneficial effects of insulin sensitivity in postmenopausal women.[193] However, at higher extremes of BMI and alcohol use, data are not conclusive. Previous studies have attributed a lower incidence of CLD to BMI and alcohol, and as expected a lower incidence of CLD when only alcoholic cirrhosis is examined.[194] However these have relied on cirrhosis codes alone, ignoring complications characterising decompensated cirrhosis that are indicative of CLD and clearly associated with BMI and alcohol included in my study.

My study is in broad agreement with some other studies including the National Health and Nutrition Examination Survey (NHANES)[195] which found increasing risk with both increasing BMI and alcohol, but no excess risk in overweight or obese drinkers or in abstainers. A Scottish prospective study reported increasing risk with increasing BMI in men, but not in women.[171] A sub-analysis of men found the lowest risk of CLD in abstainers with normal BMI with a supra-additive interaction between BMI and alcohol.[196] The UK-based Million Women Study[170] used a limited range of ICD-10 codes to identify cirrhosis and reported a rate of hospital admission or death from liver disease less than half that found in this study. However, as in my study, highest risk was in overweight or obese women consuming the most alcohol.

Other studies reporting incidence are shown in table 5.9. As expected, there is a broad spread which is likely to reflect the variation in data extraction techniques and the high variation in codes used to define liver disease. The problems with comparing studies due to differences in data definition is discussed in chapter 8. As can be seen in table 5.9, some studies have included ICD-10 codes for conditions that do not always represent cirrhosis. For example, K70.1 codes for alcoholic hepatitis, a condition that can be found in

individuals without cirrhosis. In studies where incidence in women of similar age to my cohort was reported, incidence rates ranged between 0.24 and 0.50 per 1000 participant years, all slightly lower than in my cohort. A number of these studies were from low prevalence areas, for example Iceland. This does highlight the need for more data from higher prevalence populations.

In a study of patients with a history of alcohol excess who were admitted to hospital with an alcohol-related problem, risk of cirrhosis was twice as high among the overweight group as those with normal BMI.[197] A more recent prospective study of 107,735 middle-aged males used self-reported BMI and alcohol use to assess liver-related mortality ascertained from record linkage, using ICD-10 codes K70-K76, demonstrating a U-shaped relationship between alcohol and mortality and BMI and mortality. Although there was evidence of synergy between low BMI and high alcohol, as in my study there was no evidence of interaction between high BMI and high alcohol use.[198]

The finding in my study of increased risk in abstainers has precedent but is controversial. Previous studies have demonstrated the “J-shaped” relationship between alcohol and risk of mortality[199-202] or CLD.[203,204] Some prospective studies have found that men but not women abstainers were at increased risk,[170,204] in contrast to my study that provides a more comprehensive insight into the effects of weight and alcohol.

Using raised aminotransferase levels to diagnose suspected NAFLD in men and women in NHANES the highest risk was seen in non-drinkers compared to modest drinkers,[205] and in biopsy-proven NAFLD, moderate drinkers had lower risk of steatohepatitis compared to non-drinkers.[206] A prospective Danish study

investigating risk of alcohol-related cirrhosis in over 30,000 participants found a dose-dependent increase in risk of cirrhosis with increasing alcohol intake in women, rather than a “J-shaped” relation which they observed in males.[203] This contrasts to the trend towards a “J-shape” relationship seen in my study, which remains irrespective of BMI group.

The increased risk of alcoholic cirrhosis in abstainers compared to light drinkers may be due, in part, to this group containing previous drinkers who raise the overall risk in the abstainer group, rather than due to a true protective effect of alcohol in the light drinkers. One prospective study[194] demonstrated the loss of the “J-shaped” curve when lifetime abstainers were separated from current abstainers. In a small study of patients with biopsy-proven NAFLD, a comprehensive alcohol history was obtained and found to be higher than the original estimate at diagnosis in some patients, suggesting that some of these patients may have had alcohol-related liver disease rather than NAFLD.[207] My study found alcohol-related LREs in abstainers (although at less than half the rate seen in drinkers) which, although may partly be a function of miscoding, provides further evidence that this group comprises some ex-drinkers.

Interaction between higher levels of alcohol consumption and NAFLD may result in greater risk of liver disease. A study measuring aminotransferase activity found that increased BMI potentiates the harmful effect of alcohol on the liver.[208] Increased aminotransferase levels were associated with higher alcohol consumption and BMI. In those with normal BMI there was no association between alcohol and raised aminotransferase levels, but in the overweight and obese groups, alcohol increased risk of elevated aminotransferases. A study of an older population also found risk of elevated

aminotransferases with increased BMI and increased alcohol consumption (with lowest risk in abstainers), and a large synergistic effect in the obese group consuming more than three drinks / day.[209] This group also examined the risk of hepatocellular carcinoma in people with chronic hepatitis B, finding synergism between obesity and alcohol.[210,211]

Table 5.9. Summary of studies reporting incidence of chronic liver disease using various data definitions
The table presents the incidence estimations derived from the corresponding groups of ICD-10 codes or other sources of event data

Author	Location	Data definition / source	Population number	Demographic	Incidence (per 1000 person years)
Liu *et al.*[170]	UK (Million Women Study)	ICD-10: K70, K73, K74	1,230,662	Female Mean age 56	0.24
Ratib *et al.*[25]	UK GP and HES records	ICD-10: K70.3, K71.7, K72.1, K74.4, K74.5, K74.6, K76.6, I85.0, I85.9, I86.4, I98.2 And OPCS4 codes	5,118	Male and female Age 18-85+ Mean age 59	0.31
			2,153	Female Age 18-85+	0.25
			1,309	Female 55-64	0.50
			1,023	Female 65-74	0.54
Saunders *et al.*[212]	UK	Hospital admission & pathology records, death certificates, contacting GPs, 1959-1976 Denominator = population served by hospital	512	Not given	0.06 (1959) 0.15 (1974)
Ludviksdottir *et al.*[213]	Iceland	Death certificates 1951-1990, hospital, pathology and biopsy records, 1971-1990	Whole population of Iceland	Male and female	0.02 (alcoholic cirrhosis) 0.03 (non-alcoholic cirrhosis)
Gunnarsdottir *et al.*[214]	Sweden Iceland	Hospital inpatient and outpatient records (diagnosis based on clinical, laboratory, imaging, pathology data)	300,000 (whole population of Iceland)	Male and female	0.03

Author	Location	Data definition / source	Population number	Demographic	Incidence (per 1000 person years)
			600,000 (population of Gothenburg)	Male and female	0.15
Fleming *et al.*[215]	UK	General Practice Research Database, 1987-2002. Codes for cirrhosis, oesophageal varices, portal hypertension	UK-wide	Male and female	0.15
				Female Median age 61	0.18
				Female 55-64	0.24
				Female 65-74	0.25
Fialla *et al.*[216]	Denmark	Hospital admission data (from 1977) and outpatient data (from 1989) ICD-10: K70.1, K70.2, K70.3, K70.4, K73.2, K74.3, K74.4, K74.5, K83.0, B18.0, B18.1, B18.2, K71.7, K71.8, K75.8, K75.9, 73.9, 74.6, 72.1, 72.9, K76.1, E83.1, E88.0, Z94.4, I85.0, I85.9, I86.4, R18, K76.7, C22.0, B18.9	470,000	Male and female Mean age 56	0.33 (significantly higher in males in all age categories)
				Female	0.21
				Female 50-59	0.35
				Female 60-69	0.35
				Female 70+	0.2

GP, general practitioner; HES, Hospital Episode Statistics; ICD-10, International Classification of Diseases, Version 10; OPCS4, Office of Population Censuses and Surveys, Fourth Revision

5.6.4 Implications of this study

These results suggest a substantial influence of both elevated BMI and alcohol on risk of CLD. Although no significant interaction between BMI and alcohol was seen and this lack of synergy is reassuring, the compelling risk in the overweight and obese groups adds to the evidence that rising BMI and increasing alcohol use are risk factors for liver disease among women.

By considering the clinical consequences of liver disease beyond the diagnosis of cirrhosis this study revealed a greater burden of disease than previously recognised. Currently much CLD goes undiagnosed until complications of cirrhosis result in serious morbidity and mortality. Earlier identification of those at risk could avert illness and reduce costs by targeted interventions. While the risks associated with heavy alcohol consumption are frequently publicised these data emphasise the importance of disseminating awareness of the risks of liver disease associated with BMI, particularly in light of the growing prevalence of overweight and obesity throughout the world.[217]

5.6.5 Conclusion

This study of postmenopausal women suggests that elevated BMI and high alcohol intake are risk factors for liver disease. It strongly suggests that strategies for detecting liver disease and public health strategy should recognise the importance of BMI as well as alcohol when confronting the growing burden of liver disease. The next two chapters will investigate the potential clinical utility of two different tools to predict CLD in the UKCTOCS population, and this will be followed by a discussion of how such public health strategies may be informed by these data.

Chapter 6. Association Between Skirt Size and Chronic Liver Disease

6.1 Introduction

In the previous chapter, the association between increasing BMI and risk of CLD was demonstrated in the UKCTOCS population. Although BMI is a well-recognised tool for assessing overweight and obesity it is not easy to use on an individual level.

Skirt size (SS) could be an easily understood surrogate for BMI to communicate public health messages about the risks of obesity. As women will know their SS, this could provide a tool for women to self-stratify their risk. Increase in self-reported SS in participants in UKCTOCS has been shown to be associated with increased breast cancer risk. A unit increase in UK SS (e.g. 12 to 14) every 10 years between 25 and postmenopausal age is associated with postmenopausal breast cancer risk of 33%.[218] Validation of these results could provide women with a simple and easy to understand message, using SS.

6.2 Aims of this study

The aims of this study were to determine the association between SS and the incidence of CLD by extracting data from the UKCTOCS trial to perform a prospective cohort study nested in the UKCTOCS population.

6.3 Methods

6.3.1 Ethical approval

Ethical approval for this study is stated in chapter 5, section 5.3.2.

6.3.2 Study population

This study was nested within UKCTOCS. As HES data was available for those participants in England, this study was restricted to women recruited via a recruiting centre in England. The UKCTOCS trial is described in chapter 5, section 5.3.1.

6.3.3 Exposures

The exposures of interest were BMI and SS of participants. As outlined in chapter 5, at the time of recruitment, participants completed a questionnaire, which included self-reported height and weight, allowing me to calculate BMI. As previously discussed, there were some extreme values in self-reported data and as there are no existing population estimates for the range of BMI I adopted a pragmatic approach in order to include participants with plausible BMI values. Therefore, participants who reported a height outside the range 140-210 cm, or a weight outside the range 25-200 kg, or where the BMI was outside the range 16-65 kg/m^2 were excluded.

Via follow-up questionnaire 3-5 years post randomisation, participants were asked to estimate their UK SS when they were in their early twenties and to report their current SS (appendix E). I therefore extracted this SS data from UKCTOCS and used the two SS responses to calculate overall change in SS and change in SS per year. In the UK, SS range comprises of even numbers, therefore a one unit size increase in SS represents an increase in two nominal SS values (e.g. from 12 to 14).

6.3.4 Categorisation of exposure variables

BMI was categorised according to the World Health Organization's definitions; normal (<25 kg/m^2), overweight (25-<30 kg/m^2) or obese (≥30 kg/m^2). I categorised SS using UK dress sizes as ≤16 and ≥18; I selected this cut-off because of the association between SS ≥18 and an increased risk of cardiovascular morbidity.[219] The British Standards Institution defines UK size 16 as 100-104 cm, and size 18 as 105-109 cm, measured at the hips.[220] Change in SS was categorised as a decrease, no change or an increase in SS between when participants were in their early 20s and their current age.

6.3.5 Covariates

Participants reported, via the follow-up questionnaire, known comorbidities, comprising hypertension, heart disease, hypercholesterolaemia, stroke, type 2 diabetes, rheumatoid arthritis, osteoarthritis, osteoporosis and whether they currently smoked, all categorised as yes / no. As previously described, participants were asked "approximately how much alcohol on average do you drink each week, assuming one drink = a glass of wine, half a pint of lager or cider, a measure of spirits?" This was then categorised as none, <1-15 units / week, 16-20 units / week and ≥21 units / week, assuming one drink is equivalent to one UK unit (10 ml or 8 g of pure alcohol).[176]

Deprivation score was assigned to participants as previously described using the Index of Multiple Deprivation 2007 (IMD).

6.3.6 Follow up

As previously described in chapter 5, I searched the UKCTOCS follow-up databases for ICD-10 codes of interest by linking my Microsoft Access file to the UKCTOCS follow-up Access files to import the date and code for each event for each participant who experienced an event from the start date of my study. Follow-up data was extracted from the Hospital Episodes Statistics (HES) database, death records and cancer registrations. The starting point for participants entering my study was the date that they returned the follow-up questionnaire to UKCTOCS and therefore I prospectively followed up participants from this point. Participants were included in the study from the point of return of questionnaire. Participants with known pre-existing liver disease were not included, by excluding those where a code of interest had been registered between recruitment to UKCTOCS and return of questionnaire.

Only participants in England were included in my study, due to availability of their relevant HES data.

6.3.7 Outcome

The outcome measure was first liver-related event (LRE) which is defined in chapter 5.

6.3.8 Statistical analysis

For the incidence analyses I used person-years of follow-up as the denominator. Participants contributed person-years until the censorship date (February 1, 2013), date of first presentation with an LRE or death from any other cause.

I calculated crude incidence for each BMI group, each SS when aged in 20s group, each SS at questionnaire completion group, and change in SS group.

6.3.8.1 Survival analysis

All covariates listed above were included individually in a Cox regression model to estimate their univariate associations with LREs, to guide their utility in the models evaluating risk associated with SS.

I calculated hazard ratios (HRs) of first LRE, with 95% confidence intervals (CI) using Cox proportional hazards models. For each exposure described above, BMI, SS when aged in 20s and SS at questionnaire completion were analysed as continuous covariates, and then BMI, SS when aged in 20s, SS at questionnaire completion and overall change in SS as categorical covariates. For each outcome, univariate models were produced. Smoking and deprivation were then added (partially adjusted), and then all covariates listed above were added, with abstinence and alcohol consumption ≥21 units/week as individual indicator variables (fully adjusted).

All analyses were performed using SPSS (version 22, SPSS Inc, Chicago, IL, USA), STATA statistical software (StataCorp 2007. Release 10. College Station, TX, USA: StataCorp LP) and R (R Foundation for Statistical Computing, Vienna, Austria).

6.4 Results

6.4.1 Sample characteristics

Derivation of study cohort

Of the 157,996 UKCTOCS participants resident in England, 62,870 were excluded including 321 women who experienced an LRE or died between recruitment and return of questionnaire and 14,295 (9%) with no data on smoking. There was some missing SS data, and the resulting effective sample size for this study was 94,124 (figure 6.1).

Baseline characteristics

Baseline characteristics are shown in table 6.1. Forty five percent of the cohort had a normal BMI, thirty seven percent were overweight and eighteen percent were obese. Thirty six percent were smokers. The most common self-reported comorbidity was hypertension (32%), followed by hypercholesterolaemia (24%), osteoarthritis (16%), osteoporosis (7%), heart disease (6%), type 2 diabetes (5%), rheumatoid arthritis (5%) and stroke (1.5%). The mean Index of Multiple Deprivation score was 18.4. Prevalence of smoking, hypertension, heart disease, hypercholesterolaemia, stroke, type 2 diabetes, rheumatoid arthritis and osteoarthritis and the mean deprivation score all increased with increasing BMI.

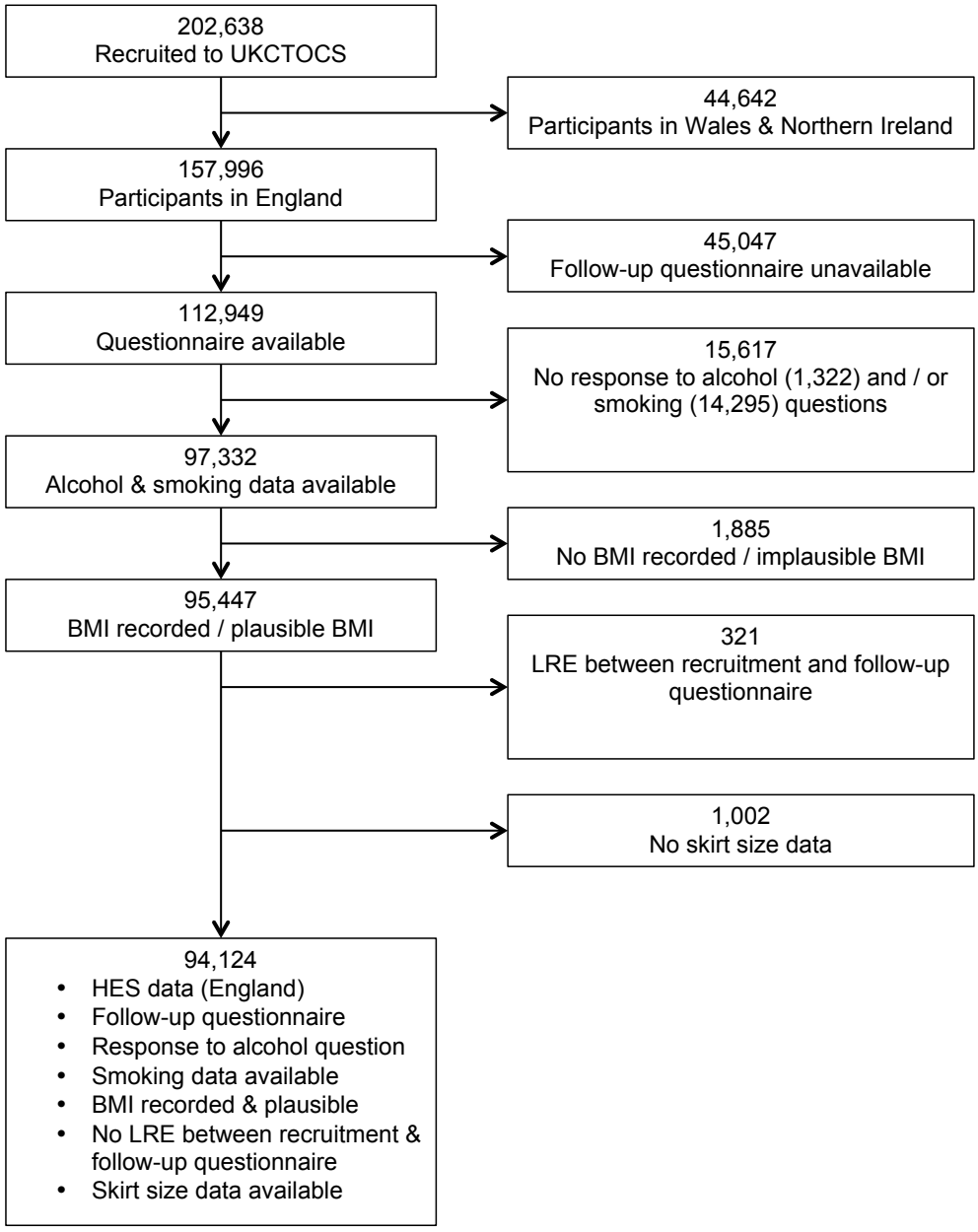


Figure 6.1. Composition of the final study cohort and its derivation from the UKCTOCS cohort

The final cohort comprised 94,124 participants. Participants in Wales and Northern Ireland were excluded due to lack of HES data. Participants who did not return the UKCTOCS questionnaire, did not respond to the alcohol or smoking status questions or in whom there was no skirt size data were removed, as were those who did not record BMI or if the recorded BMI was implausible. Participants who experienced a liver-related event before the start of the study were excluded

BMI, body mass index; HES, Hospital Episode Statistics; LRE, liver-related event; UKCTOCS, United Kingdom Collaborative Trial of Ovarian Cancer Screening

Table 6.1. Baseline characteristics according to BMI category, and for all participants

Data are presented for the entire cohort, comprising 95,124 participants, and categorised according to WHO BMI category. Self-reported comorbidities, smoking and alcohol status and deprivation scores are shown.

Characteristic	BMI category (kg/m^2)			All participants
	<25	25 - <30	≥30	
Total, *n* (% of all participants)	42, 077 (44.7)	34,690 (36.9)	17,260 (18.3)	94,124
Age at questionnaire return, median years (range)	63 (52-80)	64 (53-80)	64 (53-80)	64 (52-80)
IMD, mean (SD)	17.0 (13.1)	18.7 (14.1)	21.3 (15.4)	18.4 (14.0)
Smoker, *n* (%)	14,632 (34.8)	12,511 (36.1)	6,548 (37.7)	33,691 (35.8)
Hypertension, *n* (%)	9,382 (22.3)	11,970 (34.5)	8,307 (47.9)	29,659 (31.5)
Heart disease, *n* (%)	1,698 (4.0)	2,052 (5.9)	1,392 (8.0)	5,142 (5.5)
Hypercholesterolaemia, *n* (%)	7,901 (18.8)	9,044 (26.1)	5,369 (30.9)	22,314 (23.7)
Stroke, *n* (%)	523 (1.2)	552 (1.6)	314 (1.8)	1,389 (1.5)
Type 2 diabetes, *n* (%)	827 (2.0)	1,653 (4.8)	2,221 (12.8)	4,701 (5.0)
Rheumatoid arthritis, *n* (%)	1,592 (3.8)	1,742 (5.0)	1,185 (6.8)	4,519 (4.8)
Osteoarthritis, *n* (%)	5,503 (13.1)	5,822 (16.8)	4,016 (23.1)	15,341 (16.3)
Osteoporosis, *n* (%)	3,808 (9.1)	2,082 (6.0)	770 (4.4)	6,660 (7.1)
Alcohol consumption (units / week), *n* (%)				
None	8,365 (19.9)	8.043 (23.2)	5,432 (31.3)	21,840 (23.2)
<1 – 15	31,567 (75.0)	25,095 (72.3)	11,347 (65.4)	68,009 (72.3)
16 – 20	1,436 (3.4)	1,063 (3.1)	364 (2.1)	2,863 (3.0)
≥21	709 (1.7)	489 (1.4)	214 (1.2)	1,412 (1.5)

BMI, body mass index; IMD, Index of Multiple Deprivation; LRE, liver-related event; WHO, World Health Organization

6.4.2 Frequencies and distributions of BMI and skirt size

Frequencies of skirt sizes ≤16 and ≥18 are shown in table 6.2. 96.6% of participants reported a SS of ≤16 in their 20s, with 3.4% reporting SS ≥18. At the time of questionnaire return, more women reported SS ≥18 (23.2%) with 76.8% reporting SS ≤16. As would be expected, the proportion of participants reporting SS ≥18 at the time of questionnaire return increased with increasing BMI, and this pattern was also seen with the proportion of participants reporting SS ≥18 in their 20s. Overall, 76% of participants reported an increase in SS, 18% reported no change in SS and 7% reported a decrease in SS. This pattern was seen in all BMI groups.

The distributions of BMI, SS when aged in 20s, SS at questionnaire completion and annual change in SS are shown in figure 6.2. Median BMI was 25.57 kg/m^2 (IQR 22.79-28.36), median SS when aged in 20s was 12 (IQR 10-14), median SS at questionnaire completion was 14 (IQR 12-16), and the median change in SS unit per year was 0.0323 (IQR 0.0123-0.0523). This is the equivalent to an increase of one SS unit (e.g. from 12 to 14) every 31 years.

Visual inspection of the histograms (figure 6.2), quantile-quantile plots and box plots for each outcome variable show that each variable was approximately normally distributed, but with right-skewness seen with BMI, SS when aged in 20s and SS at questionnaire completion (BMI – skewness 1.368 (standard error (SE) = 0.008), kurtosis 4.033 (SE = 0.016); SS when aged in 20s – skewness 1.442 (SE=0.008), kurtosis 5.787 (SE = 0.016); SS at questionnaire completion – skewness 0.999 (SE = 0.008), kurtosis 2.415 (SE = 0.016); change in SS per year – skewness 0.470 (SE = 0.008), kurtosis 3.095 (SE = 0.016)).

Table 6.2 Skirt size data for all participants and according to BMI group
Frequencies of participants' self-reported skirt sizes in their 20s, at time of questionnaire return, categorised ≤16 and ≥18, and change in skirt size

Characteristic	BMI category (kg/m^2)			All participants
	<25	25 - <30	≥30	
Skirt size when aged in 20s, *n* (%)				
≤16	41,428 (98.5)	33,835 (97.5)	15,691 (90.4)	90,954 (96.6)
≥18	649 (1.5)	855 (2.5)	1,666 (9.6)	3,170 (3.4)
Skirt size at time of questionnaire completion, *n* (%)				
≤16	40,792 (96.9)	26,982 (77.8)	4,481 (25.8)	72,255 (76.8)
≥18	1,285 (3.1)	7,708 (22.2)	12,876 (74.2)	21,869 (23.2)
Change in skirt size, median (interquartile range)	0.0244 (0.03)	0.0408 (0.04)	0.0667 (0.05)	0.0323 (0.04)
Overall change in skirt size, *n* (%)				
Decrease	4,811 (11.4)	1,153 (3.3)	362 (2.1)	6,326 (6.7)
No change	12,344 (29.3)	3,422 (9.9)	731 (4.2)	16,497 (17.5)
Increase	24,922 (59.2)	30,115 (86.8)	16,264 (93.7)	71,301 (75.8)

BMI, body mass index

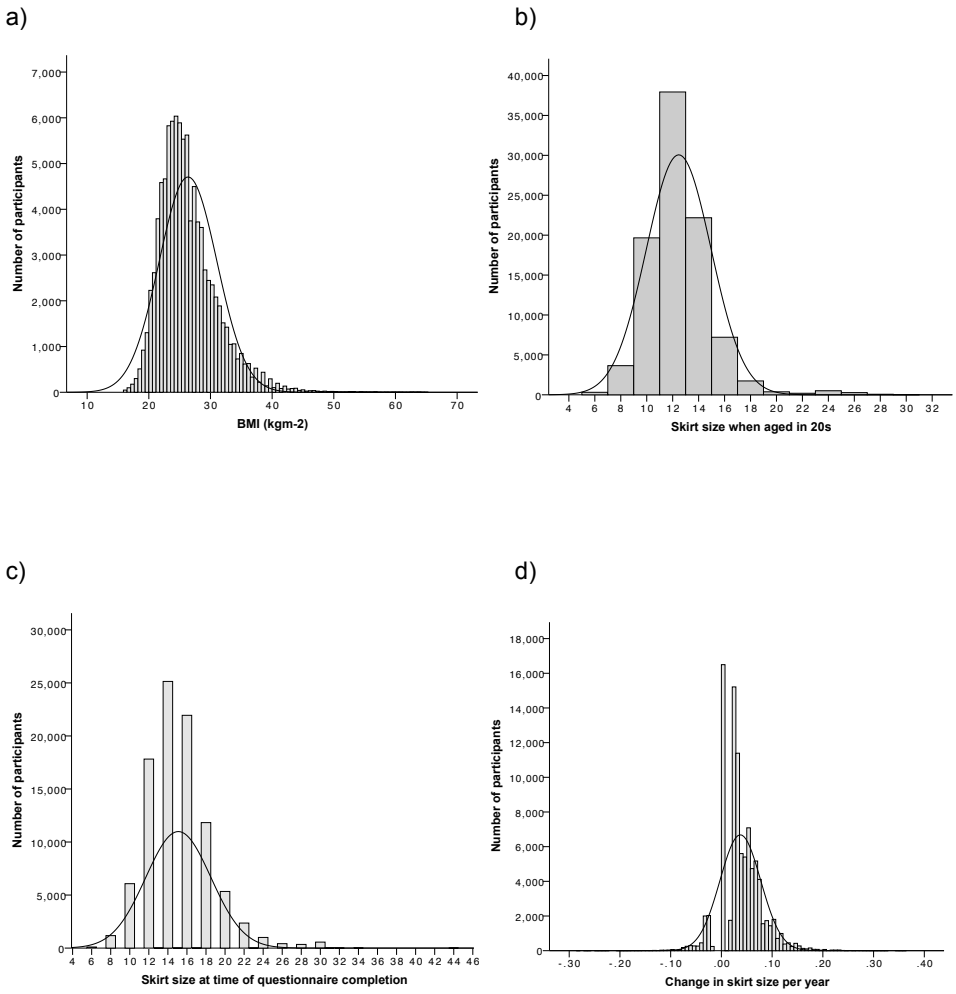


Figure 6.2. Distributions of a) BMI, b) skirt size in 20s, c) skirt size at questionnaire completion, and d) change in skirt size per year
Histograms are shown with frequencies plotted for BMI and skirt size categories for the entire study group, with normal curves overlaid, demonstrating that in all distributions data are approximately normally distributed

6.4.3 Crude event rates

Three hundred and twenty-two (0.34%) women experienced a first LRE over the follow up period. Crude rates of LRE are shown in table 6.3, categorised by BMI, SS when aged in 20s, SS at questionnaire completion and overall change in SS. The most common incident ICD-10 code was K76 (table 6.4).

Table 6.3. Event rate of first LRE for each BMI category, and for skirt size in 20s categories, change in skirt size categories and skirt size at time of questionnaire categories

Data are presented for the entire study cohort, showing event rate estimates, calculated person year follow up and incidence per 1000 person years

Exposure		Number of events Person year follow up Incidence per 1000 person years (95% confidence intervals)
BMI (kg/m^2)	<25	102 225328.2 0.453 (0.369-0.550)
	25-<30	123 186216.9 0.661 (0.549-0.788)
	≥30	97 92919.6 1.044 (0.847-1.273)
Skirt size in 20s	≤16	303 487559.9 0.621 (0.553-0.696)
	≥18	19 16904.8 1.124 (0.677-1.755)
Skirt size at questionnaire completion	≤16	213 387037.4 0.550 (0.479-0.629)
	≥18	109 117427.3 0.928 (0.762-1.120)
Change in skirt size / year	Decrease	13 33616.4 0.387 (0.206-0.661)
	No change	53 88462.8 0.599 (0.449-0.784)
	Increase	256 382385.5 0.669 (0.590-0.757)

BMI, body mass index; LRE, liver-related event

Table 6.4. ICD-10 codes and / or death certificate text of first LREs
The number of codes / death certificate text results is higher than the number of LREs (322) as some participants had more than one code when presenting with first LRE. Numbers of participants with codes of interest are divided by source of the code (hospital admission (HES), outpatient appointment (HES), cancer registration (ONS) and death certification)

Source	Code or text	Number of participants (% of those with LRE)
Hospital admission	K70	15 (4.7)
	K73	9 (2.8)
	K74	45 (14.0)
	K76	180 (56.9)
	C22.0	6 (1.9)
	I85	12 (3.7)
	Z94.4	33 (10.2)
Outpatient appointment	K74	1 (0.3)
	Z94.4	11 (3.4)
Cancer registration	C22.0	12 (3.7)
Death certificate	K70	6 (1.9)
	K74	7 (2.2)
	K76	10 (3.1)
	C22.0	2 (0.6)
	Mention of alcoholic liver disease	8 (2.5)
	Mention of non-alcoholic fatty liver disease	8 (2.5)

HES, Hospital Episode Statistics; ICD-10, International Classification of Diseases, Version 10; LRE, liver-related event; ONS, Office for National Statistics

The rate of LRE increased with increasing BMI. Comparison of rates of LREs in SS categories found a higher incidence in participants with SS ≥18, compared to participants with SS ≤16, both in the SS when aged in 20s group and the SS at questionnaire completion group. In terms of overall change in SS, event rate was lowest in the group where SS decreased. The rate was higher if there was no change, and highest if there was an increase in SS (figure 6.3).

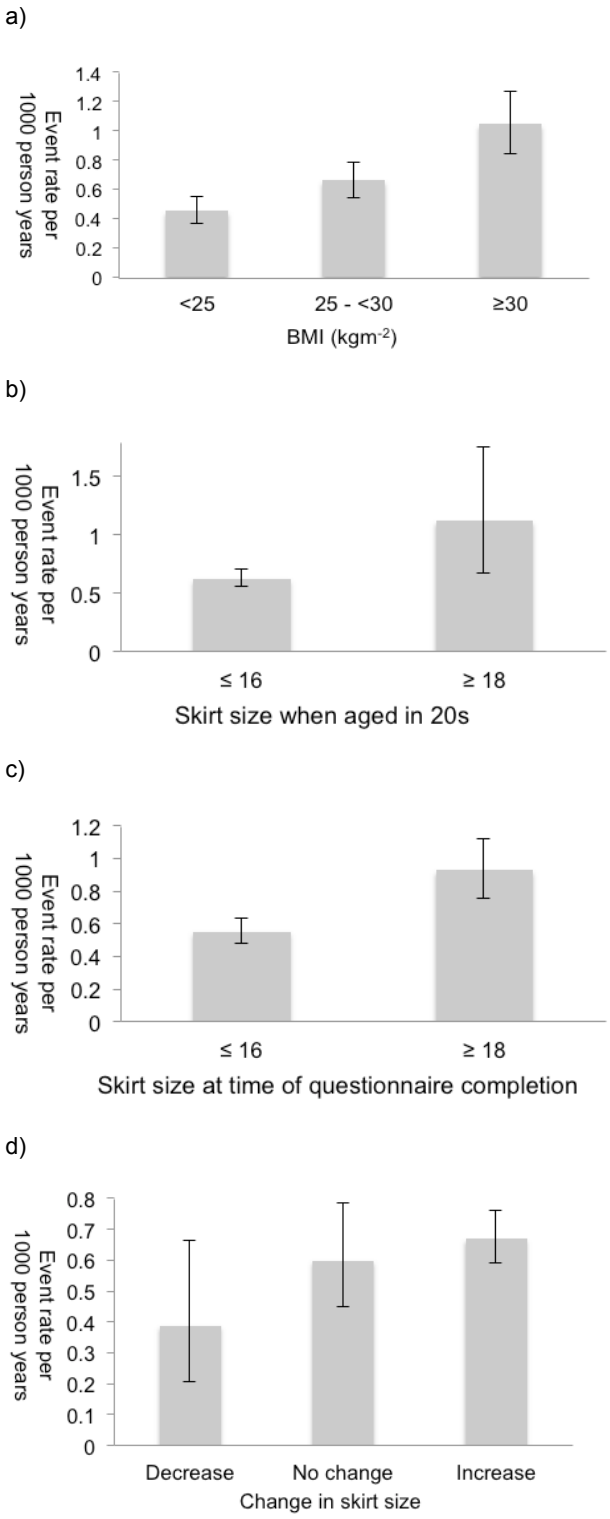


Figure 6.3 Crude rates of LRE per 1000 person years, for a) BMI, b) skirt size in 20s, c) skirt size at questionnaire completion and d) change in skirt size

Event rates are presented for WHO categories of BMI, for skirt size categories ≤16 and ≥18, and three categories of skirts size change (decrease, no change, increase), with 95% confidence intervals

BMI, body mass index; LRE, liver-related event; WHO, World Health Organization

6.4.4 Survival analysis

6.4.4.1 Cox proportional model estimate for each potential confounder

There were significant associations between LRE and smoking, deprivation, BMI, heart disease, hypercholesterolaemia, type 2 diabetes, rheumatoid arthritis, alcohol abstinence and alcohol excess (≥21 units / week) (table 6.5 and figure 6.4). A “J-shaped” relationship between alcohol and risk of CLD is seen, as described in chapter 5.

Table 6.5. Hazard ratio estimates for first liver-related events for covariates

Data are presented for number of events for category of each variable, with associated person year follow up value and incidence per 1000 person years. Associated hazard ratio estimates are shown with 96% confidence intervals

Variable		Number of events Person year follow up Incidence per 1000 person years (95% confidence intervals)	Hazard ratio (95% confidence interval) *p* value
Smoking	No	155 308669.1 0.502 (0.426-0.588)	Reference
	Yes	167 195795.6 0.853 (0.729-0.993)	1.847 (1.484-2.299) *p* <0.0005
IMD	Continuous		1.019 (1.013-1.026) *p* <0.0005
IMD tertile	1	74 168314.9 0.440 (0.345-0.552)	Reference
	2	93 164371.4 0.566 (0.457-0.693)	1.283 (0.945-1.741) *p* = 0.110
	3	150 169135.5 0.887 (0.751-1.041)	2.060 (1.559-2.721) *p* <0.0005
Alcohol	None	100 115320.9 0.867 (0.706-1.055)	Reference
	<1-15 units/week	200 365219.7	0.637 (0.501-0.809)

Variable		Number of events Person year follow up Incidence per 1000 person years (95% confidence intervals)	Hazard ratio (95% confidence interval) *p* value
		0.548 (0.473-0.629)	*p* <0.0005
	16-20 units/week	11 16008.5 0.687 (0.343-1.230)	0.819 (0.440-1.527) *p* = 0.531
	≥21 units/week	11 7915.7 1.390 (0.694-2.486)	1.661 (0.891-3.095) *p* = 0.110
Alcohol ≥21 units/week	No	311 496549.1 0.626 (0.559-0.700)	Reference
	Yes	11 7915.7 1.390 (0.694-2.486)	2.283 (1.251-4.166) *p* = 0.007
Abstinence from alcohol	No	222 389143.8 0.571 (0.498-0.651)	Reference
	Yes	100 115320.9 0.867 (0.706-1.055)	1.505 (1.189-1.906) *p* = 0.001
Hypertension	No	197 347577.0 0.567 (0.490-0.652)	Reference
	Yes	125 156887.7 0.797 (0.663-0.949)	1.391 (1.112-1.740) *p* = 0.004

Variable		Number of events Person year follow up Incidence per 1000 person years (95% confidence intervals)	Hazard ratio (95% confidence interval) *p* value
Heart disease	No	286 477256.3 0.599 (0.532-0.673)	Reference
	Yes	36 27208.4 1.323 (0.927-1.832)	2.201 (1.556-3.112) *p* <0.0005
Hypercholesterolaemia	No	212 386361.2 0.549 (0.477-0.628)	Reference
	Yes	110 118103.5 0.931 (0.766-1.112)	1.679 (1.334-2.114) *p* <0.0005
Stroke	No	314 497206.5 0.632 (0.564-0.705)	Reference
	Yes	8 7258.3 1.102 (0.476-2.172)	1.722 (0.854-3.474) *p* = 0.129
Type 2 diabetes	No	281 479831.4 0.586 (0.519-0.658)	Reference
	Yes	41 24633.3 1.664 (1.194-2.258)	2.810 (2.025-3.899) *p* <0.0005
Rheumatoid arthritis	No	295 480066.7	Reference

Variable		Number of events Person year follow up Incidence per 1000 person years (95% confidence intervals)	Hazard ratio (95% confidence interval) *p* value
		0.615 (0.546-0.689)	
	Yes	27 24398.1 1.107 (0.729-1.610)	1.815 (1.224-2.692) *p* = 0.003
Osteoarthritis	No	256 422497.2 0.606 (0.534-0.685)	Reference
	Yes	66 81967.5 0.805 (0.623-1.024)	1.328 (1.013-1.741) *p* = 0.040
Osteoporosis	No	284 469595.5 0.605 (0.537-0.679)	Reference
	Yes	38 34869.3 1.090 (0.771-1.496)	1.784 (1.272-2.503) *p* = 0.001

IMD, Index of Multiple Deprivation

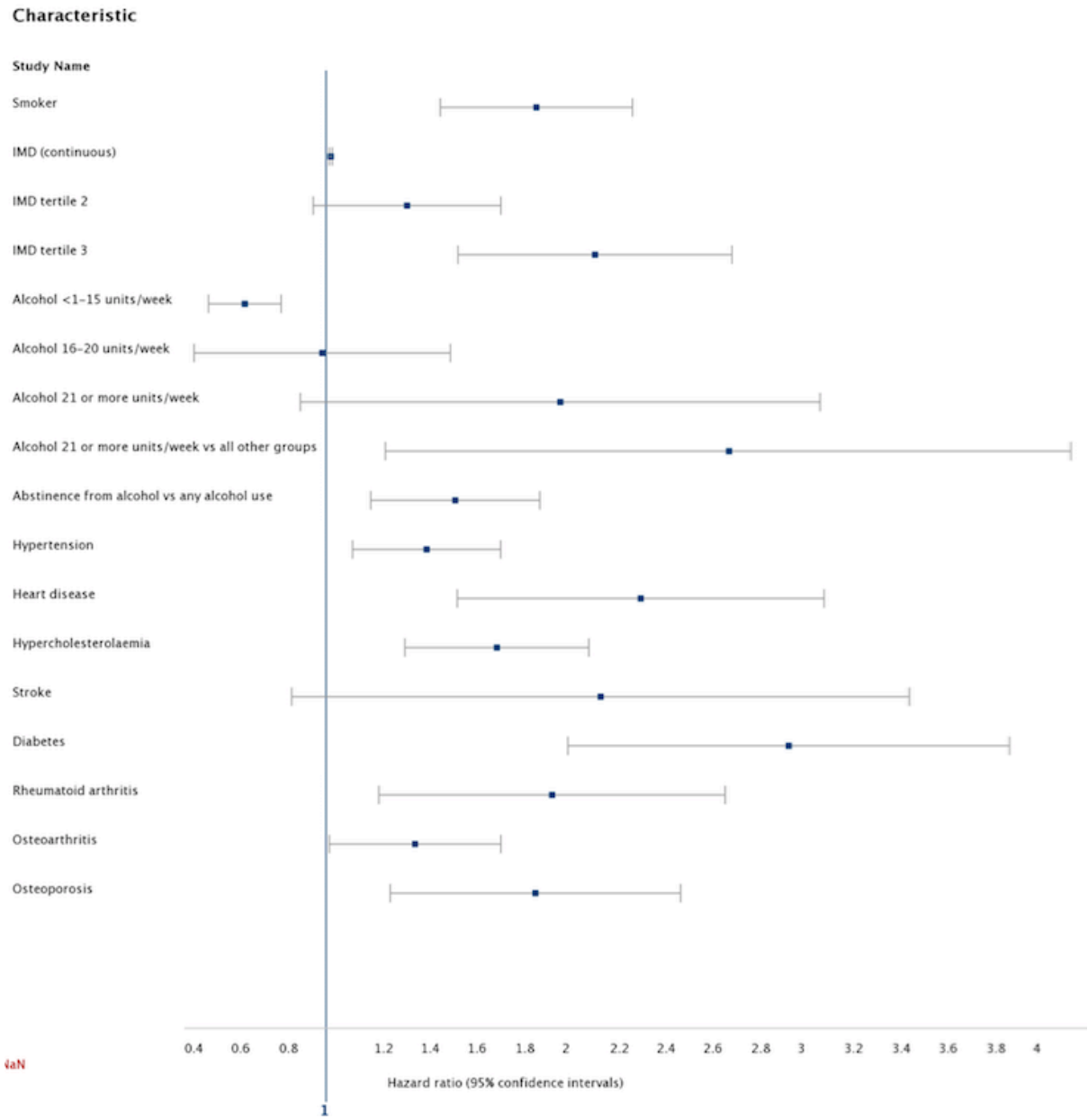


Figure 6.4. Forest plot showing hazard ratio (HR) estimates and 95% confidence intervals for univariate associations between baseline characteristics and liver-related event

Reference HRs for smoking, hypertension, heart disease, hypercholesterolaemia, stroke, type 2 diabetes, rheumatoid arthritis, osteoarthritis and osteoporosis are absence of the covariate. Reference HR for IMD tertiles 2 and 3 is IMD tertile 1. Reference HR for the alcohol groups is abstinence.

IMD, Index of Multiple Deprivation

Forest plot generated using DistillerSR Forest Plot Generator from Evidence Partners (https://www.evidencepartners.com/resources/forest-plot-generator/#forest_plot_5_graph_edit_linebyline)

6.4.4.2 Cox proportional model estimates for each exposure

When SS when aged in 20s ≥18 was compared to ≤16, HR for LRE was increased in the unadjusted (HR = 1.81 (95% CI; 1.14-2.87)), partially adjusted (HR = 1.68 (95% CI; 1.06-2.68)) and fully adjusted (HR = 1.39 (95% CI; 0.87-2.23)) models. The confidence interval for the fully adjusted model crossed unity, suggesting that a component of the risk may be partially attributable to one or more of the metabolic comorbidities (hypertension, hypercholesterolaemia, type 2 diabetes and heart disease) (table 6.6 and figure 6.5). Comparing the two SS groups at questionnaire completion, HRs were again higher in the higher SS group in all models (HR = 1.69 (95% CI; 1.34-2.13) in the unadjusted model, HR = 1.58 (95% CI; 1.25-2.00) in the partially adjusted model, HR = 1.37 (95% CI; 1.07-1.75) in the fully adjusted model).

Compared to women whose SS decreased between their 20s and questionnaire completion, HRs were higher in those whose SS did not change and highest in those whose SS increased (table 6.6 and figure 6.5).

Compared to normal BMI, overweight and obesity were significantly associated with LRE in all models (table 6.6 and figure 6.5).

Table 6.6. Hazard ratios of first liver-related events for skirt size in 20s, skirt size at questionnaire completion, BMI and change in skirt size, for three models

Hazard ratio estimates for each variable are presented for a univariate model, a model adjusted for smoking and deprivation and a model adjusted for age, smoking, deprivation, hypertension, heart disease, hypercholesterolaemia, stroke, type 2 diabetes, rheumatoid arthritis, osteoarthritis, osteoporosis, alcohol abstinence, alcohol ≥21 units/week

Variable				Hazard ratio (95% confidence interval) *p* value
Skirt size when aged in 20s	Univariate	Continuous		1.062 (1.022-1.104) *p* = 0.002
		Categorical	≤16	Reference
			≥18	1.806 (1.136-2.871) *p* = 0.012
	Adjusted for smoking, deprivation		≤16	Reference
			≥18	1.681 (1.057-2.675) *p* = 0.028
	Adjusted for age, smoking, deprivation, hypertension, heart disease, hypercholesterolaemia, stroke, type 2 diabetes, rheumatoid arthritis, osteoarthritis, osteoporosis, alcohol abstinence, alcohol ≥21 units/week		≤16	Reference
			≥18	1.390 (0.868-2.226) *p* = 0.171
Skirt size at time of questionnaire completion	Univariate	Continuous		1.091 (1.062-1.121) *p* <0.0005

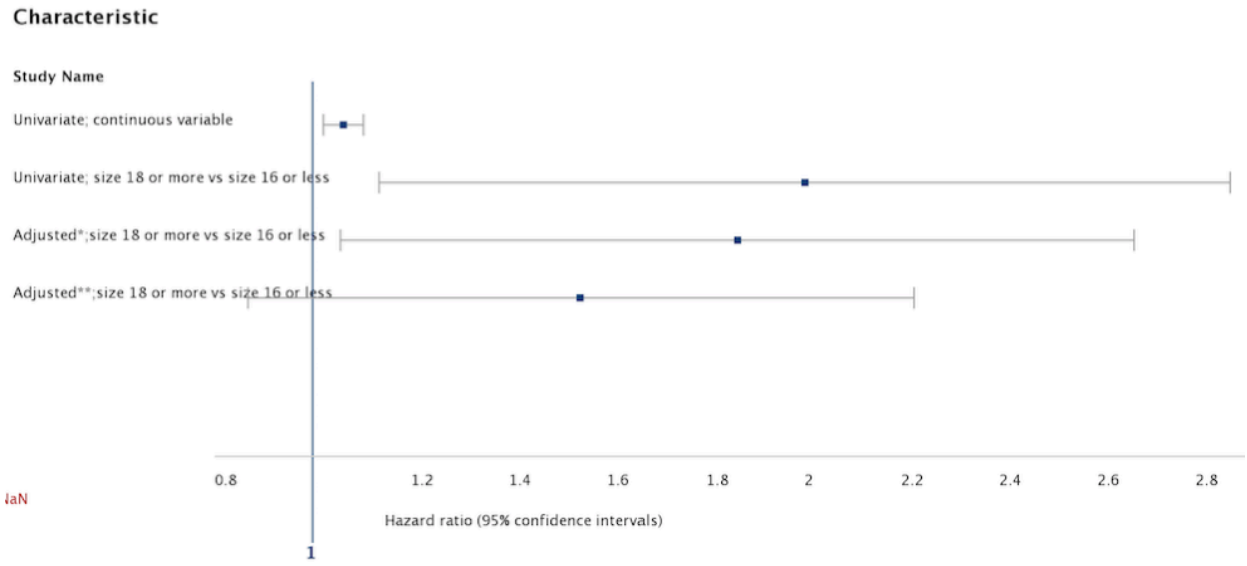
Variable				Hazard ratio (95% confidence interval) *p* value
		Categorical	≤16	Reference
			≥18	1.690 (1.342-2.129) *p* <0.0005
	Adjusted for smoking, deprivation		≤16	Reference
			≥18	1.579 (1.250-1.995) *p* <0.0005
	Adjusted for age, smoking, deprivation, hypertension, heart disease, hypercholesterolaemia, stroke, type 2 diabetes, rheumatoid arthritis, osteoarthritis, osteoporosis, alcohol abstinence, alcohol ≥21 units/week		≤16	Reference
			≥18	1.369 (1.071-1.749) *p* = 0.012
BMI (kg/m^2)	Univariate	Continuous		1.063 (1.044-1.082) *p* <0.0005
		Categorical	<25	Reference
			≥25 - <30	1.461 (1.123-1.899) *p* = 0.005
			≥30	2.308 (1.748-3.047) *p* <0.0005
	Adjusted for smoking, deprivation		<25	Reference

Variable				Hazard ratio (95% confidence interval) *p* value
			≥25 - <30	1.403 (1.076-1.830) *p* = 0.012
			≥30	2.162 (1.631-2.864) *p* <0.0005
	Adjusted for age, smoking, deprivation, hypertension, heart disease, hypercholesterolaemia, stroke, type 2 diabetes, rheumatoid arthritis, osteoarthritis, osteoporosis, alcohol abstinence, alcohol ≥21 units/week		<25	Reference
			≥25 - <30	1.353 (1.034-1.770) *p* = 0.028
			≥30	1.880 (1.395-2.533) *p* <0.0005
Change in skirt size / year	Univariate	Categorical	Decrease	Reference
			No change	1.554 (0.847-2.850) *p* = 0.155
			Increase	1.736 (0.994-3.031) *p* = 0.052
	Adjusted for smoking, deprivation		Decrease	Reference
			No change	1.714 (0.915-3.211)

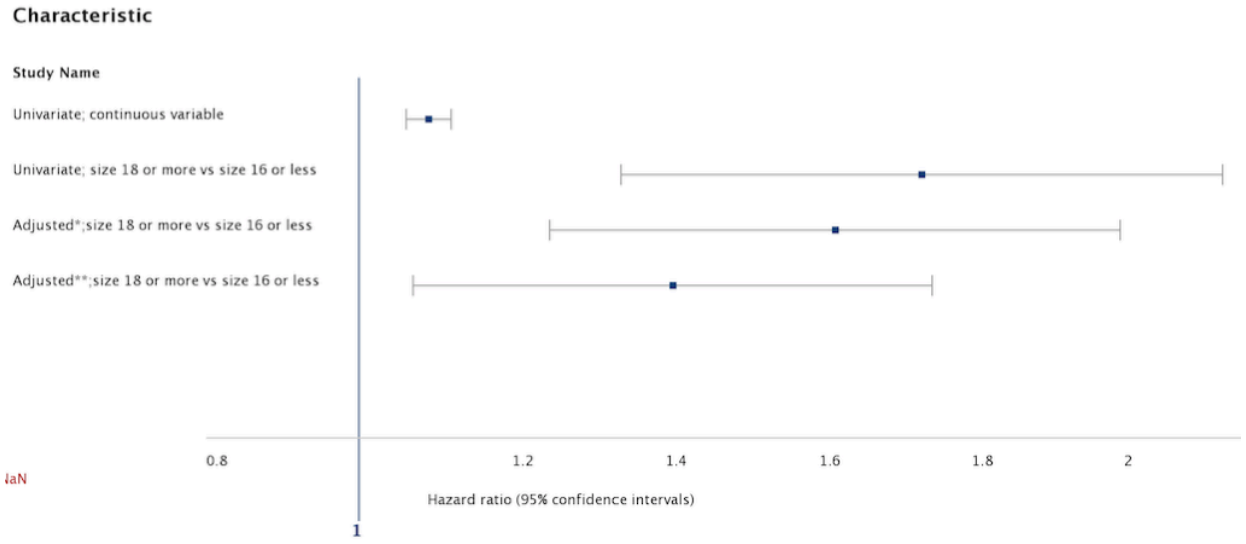
Variable				Hazard ratio (95% confidence interval) *p* value
				p = 0.092
			Increase	1.873 (1.050-3.343) *p* = 0.034
	Adjusted for age, smoking, deprivation, hypertension, heart disease, hypercholesterolaemia, stroke, type 2 diabetes, rheumatoid arthritis, osteoarthritis, osteoporosis, alcohol abstinence, alcohol ≥21 units/week		Decrease	Reference
			No change	1.781 (0.950-3.337) *p* = 0.072
			Increase	1.799 (1.007-3.214) *p* = 0.047

BMI, body mass index

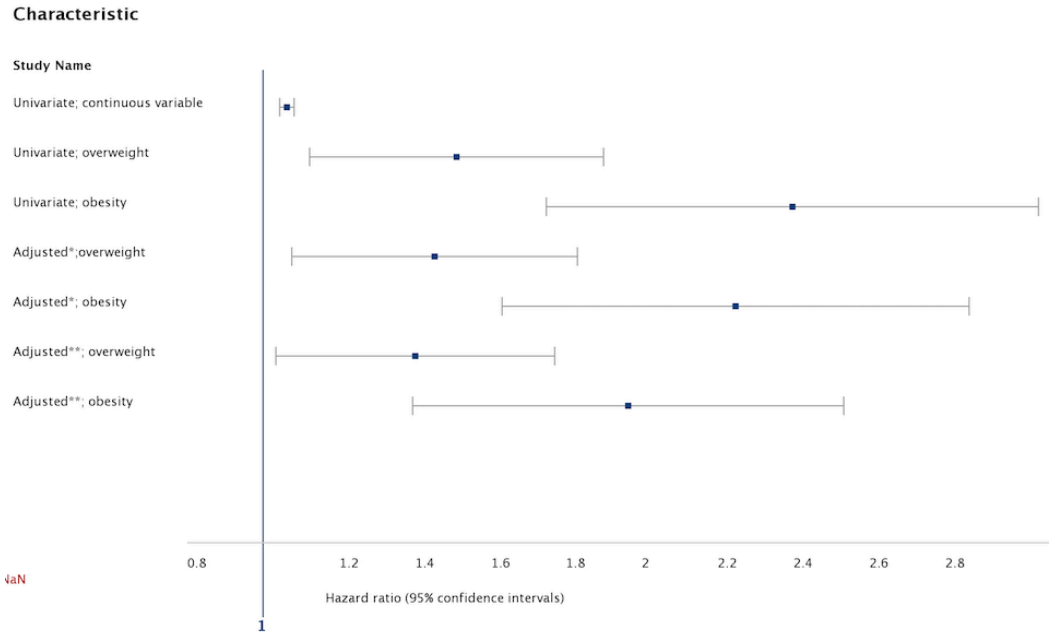
a) skirt size in twenties



b) Skirt size at time of questionnaire completion



c) Body mass index



d) Change in skirt size / year

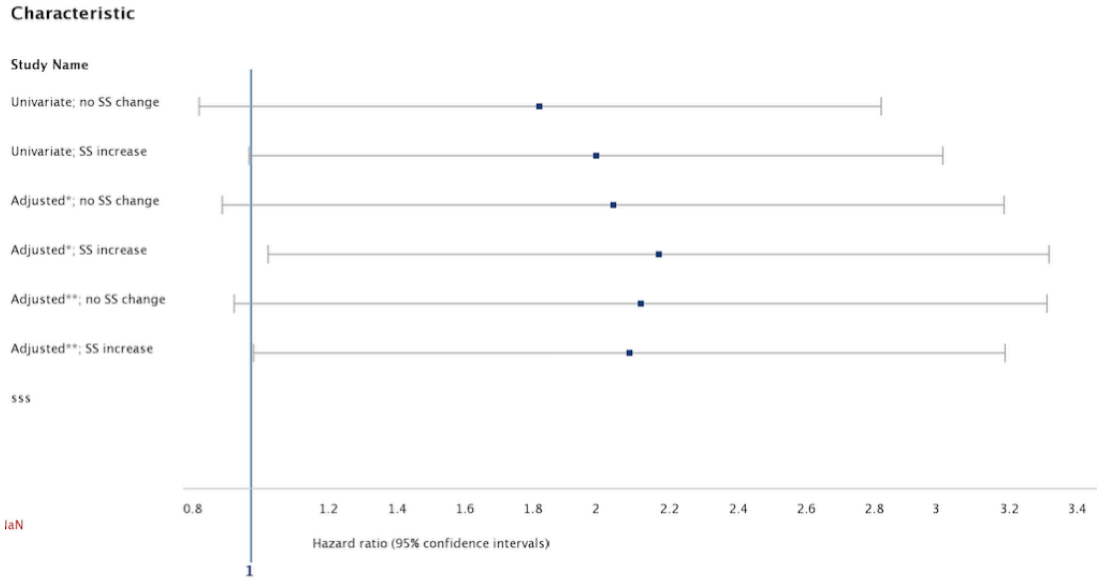


Figure 6.5. Forest plots of hazard ratio (HR) estimates for liver-related events, for univariate models and adjusted models

HR estimates for liver-related events are shown in unadjusted models and in adjusted models. a) skirt size in twenties, b) skirt size at time of questionnaire completion, c) body mass index, d) change in skirt size. * adjusted for smoking and deprivation. ** adjusted for age, smoking, deprivation, hypertension, heart disease, hypercholesterolaemia, stroke, type 2 diabetes, rheumatoid arthritis, osteoarthritis, osteoporosis, alcohol abstinence, alcohol ≥21 units/week.

In plot c, HR estimates for overweight and obesity use normal BMI as the reference. In plot d, HRs for no change in skirt size and increase in skirt size use the decrease in skirt size HR as the reference group (see table 6.6)

Forest plot generated using DistillerSR Forest Plot Generator from Evidence Partners (https://www.evidencepartners.com/resources/forest-plot-generator/#forest_plot_5_graph_edit_linebyline)

6.5 Discussion

6.5.1 Main findings

This study has demonstrated that in the UKCTOCS population larger SS is associated with subsequent risk of LRE, and a SS of ≥18 compared to a SS of ≤16 is associated with a higher HR than that associated with overweight, but less than that associated with obesity when compared to a normal BMI. Although the risks of high SS and high BMI may not be directly comparable, the value of communicating public health messages in terms of SS lies in better understanding amongst the general public compared to communicating the risk of liver disease associated with increased BMI.

76% reported an increase in SS between when aged in 20s and questionnaire completion. This is consistent with previous studies reporting the change in body composition associated with transitioning from pre-menopausal to postmenopausal status, with an increase in central adiposity manifested by increased waist circumference (WC).[221]

When BMI and SS (as continuous variables) were combined, the HR for each was reduced, suggesting that SS (and BMI) is an independent predictor for NAFLD, and that SS may reflect centripetal fat distribution associated with NAFLD better than BMI.

NAFLD is poorly identified in primary care and it is conceivable that a proportion of individuals with LREs that were not associated with an ICD-10 code for fatty liver may have had NAFLD. SS may be a better predictor of NAFLD (obesity) related liver disease than a clinical diagnosis of NAFLD in primary care.

This study aimed to identify an association with SS and CLD in general. The codes or text contributing most commonly to LRE were those representing NAFLD, although those representing alcoholic liver disease contributed to nearly 10% of LREs (table 6.4). Regardless of the aetiology of CLD, the clinicopathological pathway is progressive fibrosis leading to cirrhosis[222] and there may be common pathways through which the liver is damaged, in particular a common pathway for alcohol and BMI.[223] Patterns of alcohol consumption in women are changing; 16% of women in England consume above recommended limits, and this practice is highest in the 55-64 year old group,[224] and the rate of alcohol-related hospital admissions by women increased by over 30% between 2008 and 2015.[225]

6.5.2 Menopause and the role of oestrogens

Menopause is the permanent cessation of menstruation due to loss of ovarian follicular activity. Cardiovascular disease, osteoporotic fractures and Alzheimer's dementia are associated with post-menopause and thought to be related to low oestrogen levels. The corollary of this is that the relatively high oestrogen levels pre-menopause are likely to be protective. For example, rates of cardiovascular disease are lower before menopause, with an exponential rise in incidence post-menopause.[226]

However, these complications are not universal, indicating that they are multifactorial, a combination of environmental and (multiple) genetic factors in common with other complex traits. Although the central factor is oestrogen deficiency, the effects in the individual is likely to be influenced by the interaction of multiple genes. In the context of post-menopause, the 'oestrogen

cassette' has been described, where the concentrations of oestrogens and the different levels of their effects are regulated by a group of genes.[227]

As outlined in chapter two, liver fibrosis has a genetic influence. In females, fibrosis progression is slower than in males until menopause, and this again is thought to be due to the relatively higher levels of oestrogens pre-menopause, and the effects of the 'oestrogen cassette' in regulating the role of oestrogens in extracellular matrix turnover.[5]

In the context of NAFLD, where risk increases after menopause, a reduced risk has been reported in those taking hormone replacement therapy, although data currently must be described as inconclusive. A cross-sectional study investigating the associations of premature menopause and time from menopause with severity of liver fibrosis in postmenopausal women with NAFLD found that, after adjustment for factors including BMI, both premature menopause and time from menopause were significantly associated with more severe fibrosis assessed histologically, adding weight to the importance of risk stratification in this population.[228]

This discussion highlights the need to understand risk of liver disease inherent in the post-menopausal state and the importance of risk reduction where possible in the postmenopausal population. Data presented in this chapter highlight one such risk, that of increasing SS. Although my data do not provide a mechanistic explanation, the increased risk seen with increasing SS is likely to be mediated via the effects of hepatic steatosis, which as described, represents a higher risk for CLD in the post-menopausal state.

6.5.3 Strengths and limitations

6.5.3.1 Strengths

Strengths of this study include the large size of the cohort, the prospective design and the independence of data capture for outcomes. As described in chapter 5, an attempt to maximise the ability to identify liver disease in addition to the standard ICD-10 codes for cirrhosis, codes relating to clinical consequences of advanced cirrhosis, the events defining decompensated liver disease were also used. Evaluation of numerous possible confounders including self-reported known comorbidities and socioeconomic status minimised bias.

6.5.3.2 Limitations

Limitations include the reliance of self-reporting of SS, height and weight and co-morbidities. As previously discussed, there is some evidence supporting the reliability of self-reporting of biometric data including height and weight,[178-182] notably in a longitudinal study of older people.[229]

There was a 30-50 year recall of participants' SS when aged in their 20s, raising the possibility of recall error. Several studies have demonstrated good accuracy in recalled weight, with some data indicating underestimation in those with higher BMI.[230-233] It could be postulated that participants may have a better recollection of their skirt size than their weight or waist size. There was a 25 year age range in participants in my study, and older participants may have had children at a younger age than younger participants, which may have increased their SS.[234]

It is likely that there will be some variability between SS over the period between the two SS estimates. In the UK, there is no requirement for manufacturers to adhere to the standard sizing. In addition, the phenomenon of vanity sizing is recognised, where clothes with the same size label have become larger over recent decades. This has become a common practice of clothing manufacturers, which may potentially impede comparisons of sizes over time.[235] Indeed, the Chief Medical Officer for England has highlighted this 'size inflation' as a risk for society normalising overweight.[236]

As previously discussed reliance on ICD-10 to define events may result in errors due to miscoding, however the use of three independent sources may reduce risk of non-coding.

6.5.4 Other studies

The link between obesity and the risk of NAFLD is strong, with a clear dose-response relationship demonstrated in cross-sectional studies.[237] Data from prospective studies, however, are extremely limited. A significant contribution to the literature was made by a large prospective study which combined data from two resources, the UK Health Improvement Network (THIN) which collects data from general practitioners, and Humedica, a US database with information on over 25 million patients.[238] 1.3 million patients from THIN and 1 million patients from Humedica were followed for first diagnosis of NAFLD or non-alcoholic steatohepatitis using the Read code classification in THIN and the ICD-9 code 571.8 in Humedica, for a median of 5 years (THIN) and 1.5 years (Humedica). The US cohort had a higher average BMI than the UK cohort (28.14 and 26.81 kg/m^2, respectively). Risk of NAFLD/NASH increased with

increasing BMI (figure 6.4). Compared to a BMI of 20-25 kg/m^2, HR for a recorded diagnosis was 8.9 (THIN) and 4.8 (Humedica) for a BMI of 30-32.5 kg/m^2. At a BMI of 37.5-40 kg/m^2, HRs were 14.3 and 9.8, respectively. Risk was higher in males. Presence of type 2 diabetes and BMI 40-46 kg/m^2 resulted in HRs of 24.9 and 21.6, respectively.

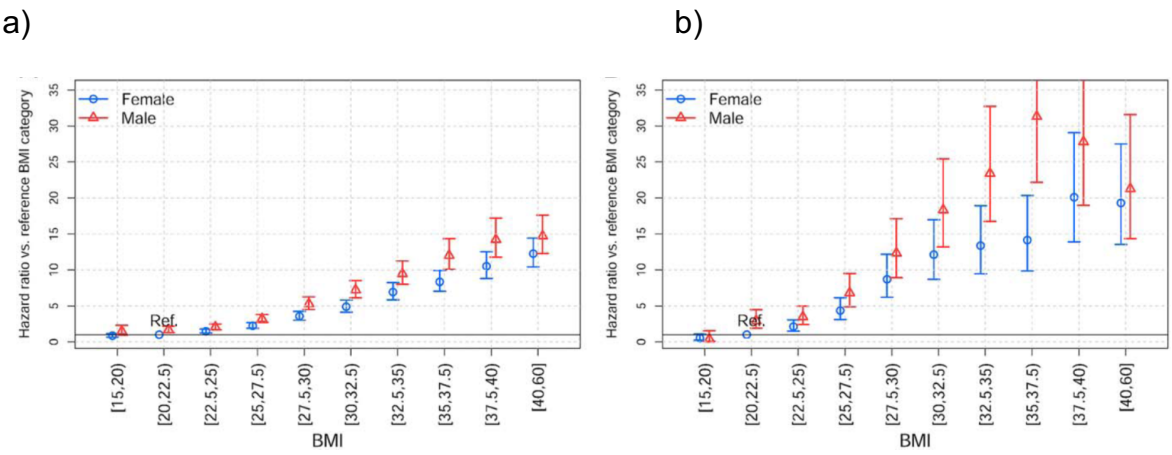


Figure 6.6. BMI and risk of non-alcoholic fatty liver disease
Hazard ratios (95% CI) for BMI categories compared to BMI 20-<22.5 kg/m^2, in a) THIN population and b) Humedica population.
From Loomis, AK *et al*.[238] Reproduced under the terms and conditions of the Creative Commons CC BY license (*https://creativecommons.org/licenses/*)

A prospective study in 6,905 non-obese subjects reported a baseline prevalence of NAFLD of 7%, with a further 9% developing NAFLD over a 5 year follow up period, a reminder that the disease is not entirely confined to the overweight population.[239]

Few studies have investigated the relationship between SS and disease. My study appears to be the only one that has investigated the association between SS and liver disease. The UKCTOCS group demonstrated an increase in risk of breast cancer with increase in SS over time.[218]

A study nested in the Netherlands Cohort Study on Diet and Cancer reported increased risk of endometrial cancer with increasing SS. The correlation between self-reported SS and self-reported WC, self-reported hip circumferences and BMI based on self-reported height and weight in 1334 women, were 0.71, 0.78 and 0.76, respectively.[240]

A study of 293 men and women found that professionally measured WC correlated closely with clothing size in both men and women (r = 0.80 and 0.78, respectively).[241]

Similarly, a study nested in the fourth Glasgow monitoring cardiovascular (MONICA) disease risk factor survey measured height, weight, WC and hip circumference, and obtained dress size in 161 women. Dress size correlated with WC and BMI. Dress size ≥18 was associated with a significantly increased risk of cardiovascular disease.[219]

6.5.5 Body mass index as a 'gold' standard

This study raises the issue of the use of BMI as the 'gold' standard for evaluating overweight and obesity, and highlights the potential use of alternative anthropometric measurements including WC or similar abdominal measurements. As discussed above, SS has not previously been extensively investigated, although WC has. My findings increase the need for a debate and I will discuss these issues here.

BMI was designed to track the weight of populations, but may be less accurate as a marker of individual health. BMI does not take in to account ratios of fat and muscle weight, or body shape, and for example would tend to overestimate risk for people with high muscle mass.

Monitoring of obesity at the population level primarily utilises BMI rather than WC. This may be, historically, due to their similar ability to predict cardiovascular and metabolic disease.[242-244] BMI needs to be calculated, and height must be measured accurately because, as height is squared in the

equation, small errors are exaggerated,[245] and studies such as mine could be susceptible to this.

There is evidence that obesity rates differ between BMI categories and WC categories. Several studies have shown increasing WC where BMI has not changed. For example, a Finnish study found that over a fifteen-year period, there was an insignificant increase in mean BMI from 26.1 to 26.4 kg/m^2 but a significant increase in mean WC by 4.3 cm.[246]

Compared to WC, BMI may be underestimating level of risk in the population. An Australian study, using thresholds for obesity of BMI ≥30 kg/m^2 or WC ≥80 cm found that over 25 years, weight and WC in women increased by 5.4 kg and 10.7 cm, respectively, and 63% of increases in WC were independent of increases in weight. The prevalence of obesity according to BMI and / or WC increased by 25% in women, but the proportion of these cases detected by BMI decreased by 20% and the proportion detected by WC increased by 10%.[247] This suggests that at the population level, BMI may not be the 'gold' standard measurement and may be underestimating levels of population risk.

Thresholds for defining obesity using WC vary by organisation and region. According to the International Diabetes Federation's definition of the metabolic syndrome, abdominal obesity is defined as a WC of ≥80 cm for women.[248] The generally adopted United States threshold for women is ≥88 cm, defined by the Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults, who note that presence of abdominal obesity is more highly correlated with metabolic risk factors than elevated BMI. [249] This is in line with the findings of the Scottish MONICA project that showed that in women a

WC ≥88 cm identified those with BMI ≥30 kg/m^2.[245] The Health Survey for England 2012 adopted the US definition of WC ≥88 cm for women to define obesity, reporting that 10% of women with a normal BMI were obese according to their WC. This suggests that BMI underestimated prevalence of obesity by almost 50% in women.[250] These studies demonstrate the conflicting data in this area but do highlight the need to consider alternative measures of obesity.

6.6 Conclusion

This study has demonstrated that SS in middle age is associated with increased risk of CLD. In postmenopausal women who develop liver disease, there is a significantly higher average SS when aged in their 20s (and in middle age). If these results are confirmed in further population studies, this may provide a simple way for women to stratify their risk of liver disease.

Chapter 7. Performance of the ELF Test in Predicting Clinical Outcomes of CLD in a General Population with Risk Factors

7.1 Introduction

In chapter 5 the association between increasing BMI and risk of CLD in the UKCTOCS population was estimated, where more clinical events attributable to cirrhosis were observed amongst women who were overweight or obese compared to those with a normal BMI. In this chapter, I will focus on those participants in UKCTOCS with risk factors for CLD in the form of elevated BMI and / or hazardous alcohol consumption, and investigate whether a liver fibrosis biomarker can predict liver-related clinical outcomes.

This chapter aims to determine the performance of the ELF test as a prognostic marker predicting liver-related events in individuals identified in the community as having risk factors for CLD. The risk factors considered in this study will be overweight and / or hazardous alcohol consumption.

As discussed in chapter 3, there are limited, but emerging, data describing the performance of non-invasive markers of fibrosis in community settings. Similarly, the use of non-invasive markers to predict long-term outcomes has not been extensively evaluated, and is generally limited to secondary care populations.

NICE guidance for management of NAFLD recommends the use of an ELF test threshold of 10.51 as the determinant for CLD and referral to secondary care.[108] This threshold was also used in a recent study evaluating the ELF test in

individuals with alcohol-related liver disease.[95] However, the manufacturer of the ELF test (Siemens Healthineers) recommends an ELF threshold of 9.8 for the detection of advanced fibrosis and a secondary aim of this study was to explore any consequences of the use of different thresholds in the UKCTOCS cohort.[251]

7.2 Aims and objectives of study

The aim of this study was to evaluate the performance of the ELF test in the UKCTOCS population, and was two-fold; in participants with risk factors for liver disease,

- To demonstrate that, at trial entry, the ELF test could discriminate between those participants who would subsequently experience a liver-related outcome and participants who would not, and;
- To evaluate the performance of the ELF test in predicting liver-related outcomes over time.

The objectives of the study were;

- To measure the ELF test score in UKCTOCS participants with risk factors for CLD, comprising self-reported hazardous alcohol use and / or BMI ≥25 kg/m^2, and;
- Using time-to-event analysis, determine the ability of the ELF test to predict liver-related events at various ELF score thresholds.

7.3 Methods

7.3.1 Ethical approval

Ethical approval for this study was obtained as described in chapter 5, section 5.4.1.

7.3.2 Study design

This study is a nested case-control study using the PRoBE (prospective-specimen-collection, retrospective-blinded-evaluation) design. This design focuses on the prospective collection of samples before ascertainment of the clinical outcomes and the performance of the biomarker assays (i.e. the ELF test) in a blinded fashion on samples from cases and matched controls.[252] The basis of this study design and rationale for use in the context of biomarker evaluation is discussed in chapter two.

7.3.3 Study population and exposures

The study population was a cohort selected and extracted from UKCTOCS, a sub-set of the participants described in chapter five. The cohort studied in this chapter comprised women with risk factors for liver disease; those with a calculated BMI of 25 kg/m^2 or more and / or those with a self-reported high alcohol consumption. The UKCTOCS follow up questionnaire (appendix E) asked participants to estimate weekly alcohol consumption as the number of drinks consumed per week (none, less than 1, 1–3, 4–6, 7–10, 11–15, 16–20 or ≥21 drinks) and as discussed in chapter 5, I made the assumption that one drink equated to one unit of alcohol, although this may under-estimate 'home measures'. The UK Chief Medical Officer's (CMO) guidance is to limit weekly

alcohol intake to no more than 14 units per week (for men and women).[176] This threshold falls within the 11-15 units / week category in the UKCTOCS categories, therefore in this study, this category was included in the definition of 'high alcohol'. Although this may over-estimate 'high alcohol' use it ensures women consuming alcohol over the recommended limit are included.

Participants were included if they had returned the follow-up questionnaire (as this included the alcohol question) and if they resided in England as HES data was only available for participants in England.

This study utilised serum samples previously collected from participants in UKCTOCS as part of the trial protocol. The UKCTOCS biobank is managed by Abcodia, a private company in part funded by UCL Business. I made an application to access samples for ELF testing in this study to the Abcodia / UCL Joint Steering Committee and was granted permission to access samples for ELF testing (appendix H).

The study design is shown in figure 7.1.

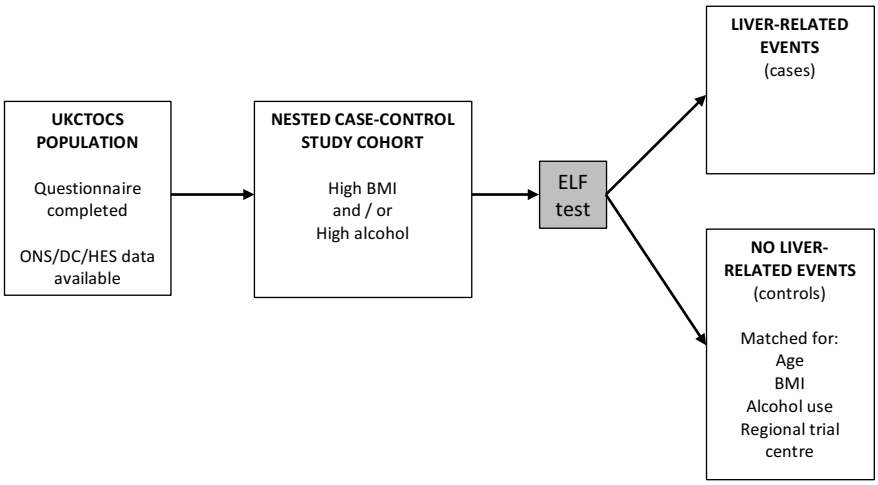


Figure 7.1. Study design for evaluating performance of the ELF test in predicting liver-related events in participants with risk factors
The study is nested within UKCTOCS, comprising participants with BMI ≥25 kg/m^2 and / or self-reported high alcohol use. ELF test is performed and participants who experience LRE during follow up are compared to matched controls who do not experience an LRE
BMI, body mass index; DC, death certification; ELF, enhanced liver fibrosis; HES, Hospital Episode Statistics; LRE, liver-related event; ONS, Office for National Statistics

7.3.4 Covariates

As previously described, the follow-up questionnaire asked participants to report known comorbidities including heart disease, hypercholesterolaemia, hypertension and type 2 diabetes mellitus, and whether they currently smoked (all categorised as yes / no). Socioeconomic status was estimated using the Index of Multiple Deprivation 2007 (IMD) (continuous variable).[177] I extracted these data from the UKCTOCS database for the participants in my study.

7.3.5 Missing data

Missing data was managed as described in chapters 5 and 6. A notable change compared to analyses in chapters 5 and 6 is related to smoking status. As previously discussed, smoking status was ascertained from a 'yes' or 'no' answer in the follow-up questionnaire. In the two previous studies, participants

with no smoking status data (i.e. neither the 'yes' or 'no' box was selected) were excluded from analyses so that the study group had full smoking status data to allow regression analysis. In the present study, there was missing smoking status data for some participants, therefore smoking status was not included in analyses. Participants with no (or implausible) BMI measurements or no alcohol consumption status were excluded from my project when I created my master database as previously described.

As outlined in chapter 5, prevalence of comorbidities was derived from questions where participants only indicated if they had the condition. Therefore, this could possibly underestimate the true prevalence, leading to a conservative estimate.

7.3.6 Follow up

All participants in England and Wales were followed through the 'flagging' study with NHS Digital, as previously described in chapter 5, which provided data on cancer registrations and deaths, with diagnosis / cause of death coded according to the International Classification of Diseases, version 10 (ICD-10); hospital inpatient and outpatient episodes through linkage to the Hospital Episodes Statistics (HES) database; and death certification. The present study was limited to participants in England due to availability of their relevant HES data. Women were included in the study from the time of first blood sample taken.

7.3.7 Selection of cases and controls

Cases comprised of eligible participants with risk factors (BMI ≥25 kg/m^2 and / or self-reported alcohol consumption of ≥11 units per week), with a first presentation of a liver-related event (LRE), defined as first presentation of one or more of the following: a hospital admission, outpatient appointment, cancer registration with, or death from, an ICD-10 code of interest. I included the following codes: K70 (alcoholic liver disease), K73 (chronic hepatitis) and K74 (fibrosis and cirrhosis), consistent with the study in chapter 5 and with other UK studies of cirrhosis.[26,170] Similarly, I also searched for K76.0 (other diseases of the liver, including fat). In addition, codes relating to sequelae of decompensated liver disease were also searched for; I85 (oesophageal varices), Z944 (liver transplant) and C22.0 (hepatocellular carcinoma). In addition to ICD-10 code, death certificates were also searched for any mention in the text of alcoholic liver disease or fatty liver.

Controls were participants with risk factors who did not experience an LRE. I matched controls for age and trial recruitment centre. Each case was matched to two controls.

7.3.8 Sample collection and sample selection

Participants in UKCTOCS gave a blood sample at recruitment.

I selected samples from the central UKCTOCS database. An inherent feature of this study was to measure the ELF score in participants at the time of recruitment to the trial, and to measure the ELF score over time in participants.

In order to collate a database for future studies, I matched controls for all cases (i.e. including those without risk factors), and possible future studies utilising this resource are discussed in chapter 8. In this study, a sub-set of this database was used, described below.

For this study, I identified the sample taken nearest to the time of recruitment (in most cases this was the sample taken at the recruitment visit). Sample 3 was the sample taken nearest to the time of liver-related event. Samples taken up to 6 months prior to the event were not used to reduce risk of the liver event itself influencing the ELF score. I selected a sample taken at approximately the mid-point between sample 1 and sample 3 as sample 2 (figure 7.2).

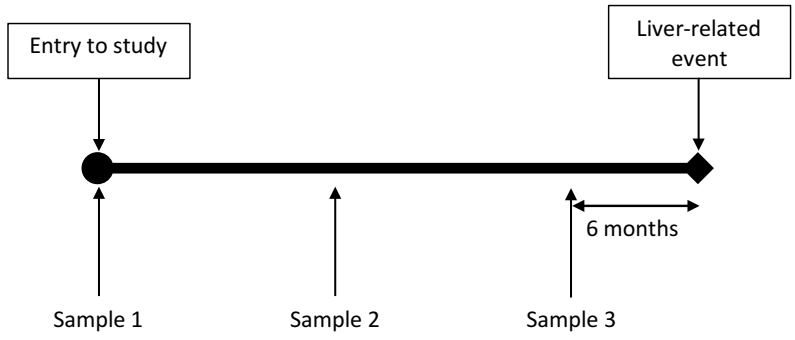


Figure 7.2. Sample time collection points in UKCTOCS ELF test study
The serum sample taken for ELF testing at or near recruitment to UKCTOCS signified the entry point to my study. The third sample was that taken nearest to the time of liver-related event (but not within six months of event). A further sample was selected at the mid-point between the first and third samples

Sample matching

For each case, I selected two controls from the UKCTOCS cohort that had not experienced a liver-related event. The control matching criteria were:

- Age (+/- 5 years) at trial recruitment
- BMI (+/- 2 kg/m^2)
- Alcohol group (i.e. either 11-15, 16–20 or ≥21 units/week)
- Regional trial (recruitment) centre

As part of the UKCTOCS trial protocol, the time from sample collection and transfer from the regional trial centre to the central laboratory where the sample underwent centrifugation was recorded for each sample obtained ('time to spin'). Samples were databased on receipt at the central laboratory allowing the storage time of each sample to be calculated. Cases were not matched for:

- Time to spin
- Storage time of sample

After due consideration by the study team, it was agreed that controls did not need to have a cancer-free history and did not need to be free of other HES codes.

For this study, as serial samples were required, cases were only included if blood samples had been taken at, or near, recruitment and on at least two subsequent occasions. All participants in the UKCTOCS trial provided a blood sample at recruitment, before being randomly assigned to one of three groups; no screening, annual ultrasound scanning, and the multimodal screening arm (annual blood sampling for serum CA125 level with ultrasound scanning as a

second-line test, MMS). Therefore I only included cases from the MMS arm in my study as this was the only arm in which serial blood samples were obtained.

7.3.9 Serum marker testing

7.3.9.1 Sample preparation

The preparation of serum samples used in this study is described in detail in UKCTOCS publications.[172] Samples were collected into Greiner Bio-One gel tubes (Greiner Bio-One Ltd, Stonehouse) and shipped overnight to the central laboratory. The blood was centrifuged at 4000 rpm for 10 minutes and 500 µl aliquots of serum was dispensed into straws (MAPI CryoBioSystem, Cryo Bio System, Paris, France) that were heat-sealed, barcoded, databased and frozen using a two-stage process in which they were placed for 24 hours at -80°C and then transferred to liquid nitrogen (vapour phase at -180°C) for long term storage in alarm monitored freezers. The UKCTOCS serum biobank is stored at Fisher BioServices.

7.3.9.2 ELF testing

Samples were retrieved from the UKCTOCS cryorepository, thawed and immediately aliquoted into 2D barcoded tubes for ELF testing by Fisher BioServices. Serum samples were shipped to the central ELF laboratory (iQur, London), and the assaying supervised by the ELF technician. The iQur laboratory participates in a national quality assurance programme for ELF testing. Serum samples were analysed for levels of the components of the ELF test, HA, TIMP-1 and P3NP using the proprietary assays developed for the ELF test by Siemens Healthineers Inc. These assays are magnetic particle

separation immunoassays, and samples were analysed on an ADVIA Centaur® immunoassay system (Siemens Healthineers Inc., Tarrytown, NY, USA). Results were entered into the manufacturer's published algorithm to derive an ELF score (see chapter two for the algorithm).[77]

7.3.10 Statistical analysis

Independent sample t-tests were performed to compare means between cases and controls. Chi-square tests were performed to compare frequencies between cases and controls; if >20% of the expected counts were <5 in any group, assumptions were considered not to have been violated and Fisher's exact test was used rather than Pearson's Chi-square test.

7.3.10.1 Covariates

As in chapter 5, potential confounding risk factors were included individually in a Cox regression model to estimate their univariate associations with LREs, confirming that deprivation and self-reported hypertension, heart disease, hypercholesterolaemia and type 2 diabetes were all associated with LRE.

7.3.10.2 Analysis of recruitment samples

I used Cox proportional hazards models to evaluate performance of ELF score at recruitment to predict LRE. Univariate models were produced, and then adjusted for the covariates listed above for ELF score thresholds of 9.8 (as recommended by Siemens Healthineers in the ELF test instructions)[77] and 10.51, the threshold recommended in the NICE guidance on management of NAFLD[108] and used in a recent study to stratify patients with alcohol related liver disease.[95]

7.3.10.3 Analysis of serial samples

To evaluate ELF as a time-dependent variable (and to minimise immortal time bias) I performed time-dependent Cox proportional hazards analysis using the same ELF thresholds. The rationale for this statistical approach is outlined in the discussion.

The time during follow up at which ELF reached the threshold was assumed to be the time of the first sample in which ELF was measured at or above that threshold. Both univariate and adjusted models were produced.

7.3.11 Stability of the ELF assay

As outlined above, blood samples obtained from UKCTOCS participants were taken at the trial centres and shipped to the central laboratory for processing, where serum samples were frozen and subsequently thawed for ELF testing. Prior to commencing this study, our group investigated the stability of the ELF assay under a set of storage conditions including medium to long-term storage at −80°C, repeated freeze-thawing and refrigeration 4 °C for four days.[105] The ELF score was stable against these conditions. This study is described in detail in chapter two. These data provide reassurance regarding the utilisation of the UKCTOCS samples in my study.

7.4 Results

Derivation of the study cohort

Following exclusion of samples as outlined above, fifty-eight cases were eligible for inclusion in my study, and therefore 116 controls were selected. Of the 522 samples selected, I excluded one control (3 samples) as per UKCTOCS

protocol due to a diagnosis of ovarian cancer during follow up, and one subsequent sample from a case was not available. The derivation of the study cohort is shown in figure 7.3.

Median interval from recruitment sample to a first presentation of LRE was 3.8 years (IQR 1.5 years). In controls, median follow up with no event was 9.8 years (IQR 2.1 years).

Baseline characteristics

Baseline characteristics of the study cohort are presented in table 7.1. Median recruitment age was 61 years (range 52-74). Median time to LRE or censoring was 8.5 years (range 0.5-11.4).

'High risk' was defined by participants with a high BMI more often than by high alcohol use. This reflects the composition of the UKCTOCS population as a whole, which comprises a higher proportion of participants with high BMI compared to the proportion of participants reporting high alcohol use. In this study group, high alcohol use was reported by 19%, with high BMI in 88%. Figure 7.4 shows the numbers of participants with high BMI or high alcohol or both high BMI and high alcohol.

As per the matching strategy, there was no significant difference in age and there were no significant differences in the proportions of each BMI group and of each alcohol group between the cases and control groups.

Forty four percent of the study cohort were smokers, with no difference between groups. The most prevalent of the comorbidities was hypertension (37%), followed by hypercholesterolaemia (31%), heart disease (12%), type 2 diabetes

(3%) and stroke (0.6%). There were significantly more self-reported diagnoses of type 2 diabetes in the cases compared to controls, but there was no significant difference in prevalence of hypertension, heart disease or stroke between groups. There was a significantly higher mean deprivation score in the cases.

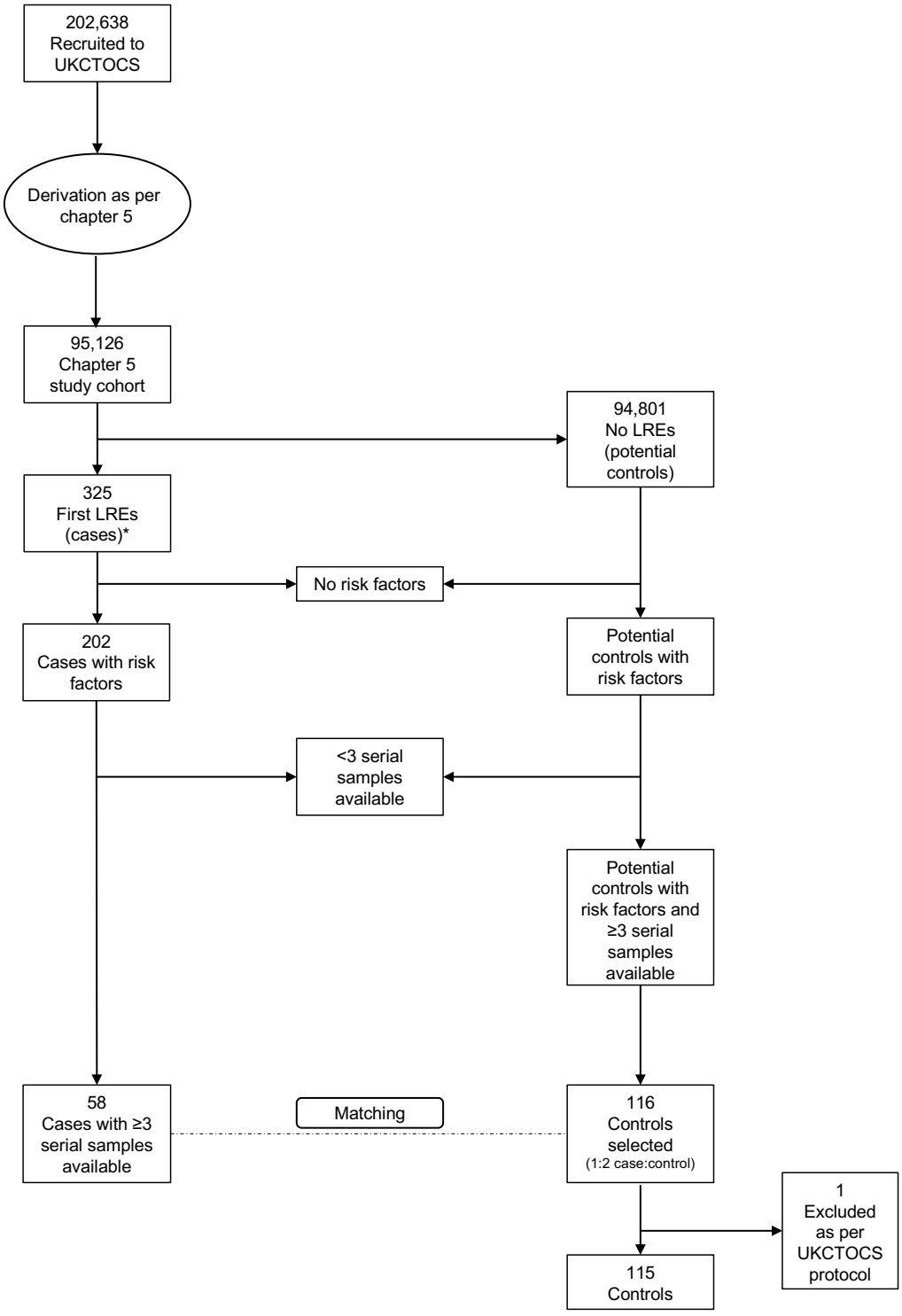


Figure 7.3. Derivation of the study cohort for ELF testing
Derivation of the cases and controls from the study cohort described in chapter five, and the sample selection strategy is shown. Cases were included if a recruitment sample and two subsequent samples were available. For each case, the recruitment sample and two subsequent samples were selected (with the third sample taken at a time point at least six months before the LRE). Samples from each control were selected at the closest equivalent time pints to the respective cases.
LRE, liver-related event. * total of 1,691 samples

Table 7.1. Baseline characteristics of study participants and comparisons between cases and controls

Data are presented for the entire study cohort, categorised in to cases and controls, for mean deprivation score, numbers of self-reported comorbidities, and WHO BMI categories. Tests of statistical differences were applied

Characteristic		Cases	Controls	All participants	χ^2 or F † *p* value *
Participants, *n*		58	115	173	
Age at recruitment, median (range)		60.9 (51.6-74.3)	61.5 (51.8-74.2)	61.0 (51.8-74.3)	0.006[a,d] 0.850
IMD, mean (SD)		25.55 (17.03)	19.86 (15.61)	21.8 (16.3)	0.972[a] 0.031
Hypertension, *n* (%)		26 (44.8)	39 (33.9)	65 (37.6)	1.958[b] 0.162
Heart disease, *n* (%)		8 (13.8)	13 (11.3)	21 (12.1)	0.224[b] 0.636
Hypercholesterolaemia, *n* (%)		22 (37.9)	32 (27.8)	54 (31.2)	1.834[b] 0.176
Type 2 diabetes, *n* (%)		11 (19.0)	5 (4.3)	16 (3.4)	9.815[b] 0.002
Smoker, *n* (%)[e]		33 (57)	38 (33) 13 missing	71 (44.4) 13 missing	5.779[b] 0.016
Stroke, *n* (%)		1 (1.7)	0 (0.0)	1 (0.6)	C 0.335
BMI (kg/m^2) *n*, (%)	<25	7 (12.1)	15 (13.0)	22 (12.7)	0.033[a,d] 0.856
	25 - <30	33 (56.9)	65 (56.5)	98 (56.6)	0.002[a,d] 0.963
	≥30	18 (31.0)	35 (30.4)	53 (30.6)	0.007[a,d] 0.936
Alcohol (units/week) *n,* (%)	None	18 (31.0)	35 (30.4)	53 (30.6)	0.007[a,d] 0.936
	<1-10	29 (50.0)	58 (50.4)	87 (50.3)	0.003[a,d] 0.957
	11-15	7 (12.1)	14 (12.2)	21 (12.1)	<0.0001[a,d] 0.984
	16-20	2 (3.4)	4 (3.5)	6 (3.5)	<0.0001[a,d] 0.992
	≥21	2 (3.4)	4 (3.5)	6 (3.5)	<0.0001[a,d] 0.992

† Chi-squared value or F-statistic for Chi-square test or independent t-test, respectively

* at the 5% level. a, Independent sample t-test; b Pearson's Chi-square test; c, Fisher's exact test; d, Matched variable; e, variable excluded form regression analyses. BMI, body mass index; IMD, Index of Multiple Deprivation

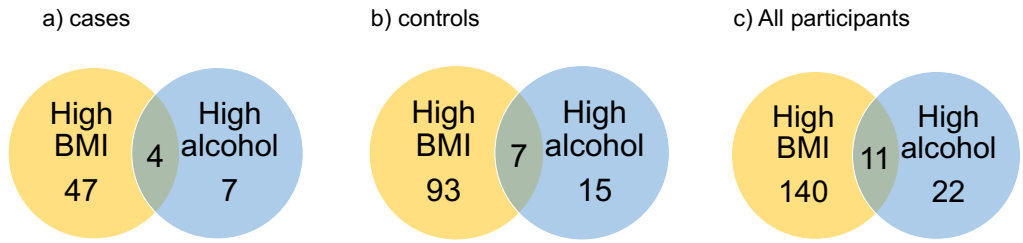


Figure 7.4. Venn diagrams showing contributions from BMI and alcohol to the study group
The diagrams show the number of participants in the study with high BMI, high alcohol or both risk factors, divided in to a) cases, b) controls and c) in the whole study group. As high BMI was more prevalent than high alcohol use in the UKCTOCS population, high BMI is the bigger contributor to the high-risk population in this study

ELF scores

The mean concentrations for the three components of the ELF assay for cases and controls are shown in table 7.2. Table 7.3 shows the mean ELF scores for the cases and controls for each sample type.

The mean ELF score in the recruitment samples was higher in the cases compared to the controls (9.36 and 8.96, respectively (p = 0.007)). In both the cases and controls, mean ELF score was higher in the combined subsequent samples compared to the recruitment samples (10.02 and 9.63, respectively (p <0.001)). The change in mean ELF score between recruitment and second sample was 0.57 in the cases and 0.63 in the controls, and between recruitment and third sample was 0.75 in the cases and 0.71 in the controls (figure 7.5).

Table 7.2 Assay results for individual components of the ELF test and calculated ELF test score for cases and controls

Mean concentrations for each assay component with corresponding standard deviation, median and interquartile range values, and the mean calculated ELF score are presented for recruitment samples and combined subsequent samples

Participant type		Assay			
		HA (ng/ml)	P3NP (ng/ml)	TIMP-1 (ng/ml)	ELF score
Recruitment samples					
Cases	Mean	93.4	9.7	244.8	9.36
	SD	197.2	4.4	80.9	1.14
	Median	38.8	8.3	227.0	9.10
	IQR	54.4	3.9	64.1	1.53
Controls	Mean	45.8	8.0	226.1	8.96
	SD	37.4	2.6	46.0	0.75
	Median	34.7	7.6	224.1	9.05
	IQR	32.8	2.5	52.7	0.81
Subsequent samples					
Cases	Mean	147.6	11.7	277.9	10.02
	SD	258.6	8.1	108.7	1.14
	Median	88.2	9.3	253.5	9.89
	IQR	94.8	6.1	77.9	1.30
Controls	Mean	88.6	8.9	248.7	9.63
	SD	68.9	3.6	54.6	0.80
	Median	66.6	8.3	242.1	9.58
	IQR	77.9	3.2	58.0	1.06

ELF, enhanced liver fibrosis; HA, hyaluronic acid; IQR, interquartile range; P3NP, aminoterminal propeptide of procollagen type III; DS, standard deviation; TIMP-1, tissue inhibitor of matrix metalloproteinase-1

Table 7.3. Mean ELF scores for cases and controls in recruitment samples, subsequent samples and in the combined subsequent samples

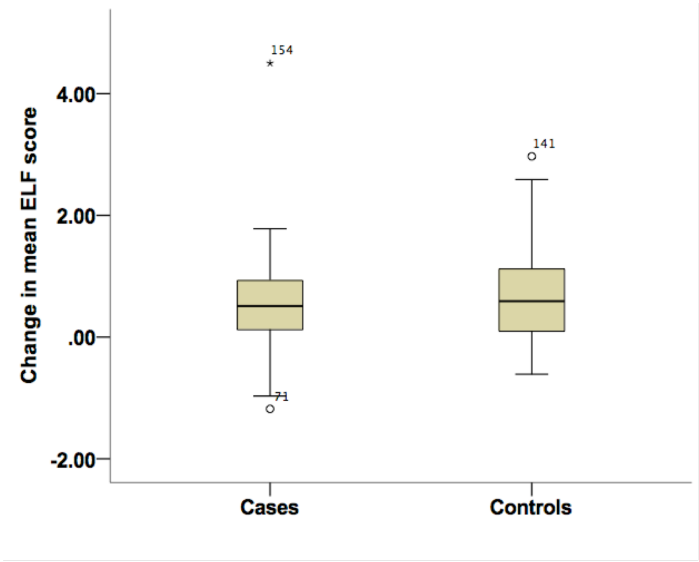
Numbers of participants in each group are shown with corresponding mean ELF test score, for first sample (recruitment samples), subsequent sample 1 (second samples) and subsequent sample 2 (third samples). Differences between cases and controls for each sample type were calculated using the Chi squared test and associated *p* values are presented

Case / control	Sample type							
	Recruitment sample		Subsequent sample 1		Subsequent sample 2		Combined subsequent samples	
	N	Mean ELF score (SD)	N	Mean ELF score (SD)	N	Mean ELF score (SD)	N	Mean ELF score (SD)
Cases	58	9.355 (1.136)	58	9.901 (1.198)	57	10.143 (1.017)	115	10.022 (1.138)
Controls	115	8.959 (0.743)	115	9.588 (0.798)	115	9.669 (0.807)	230	9.628 (0.802)
χ^2 value		10.996		4.997		4.675		9.702
***p* value** *		0.007		0.030		0.002		<0.001

* at the 5% level

ELF, enhanced liver fibrosis; SD, standard deviation; χ^2, Chi squared

a)



b)

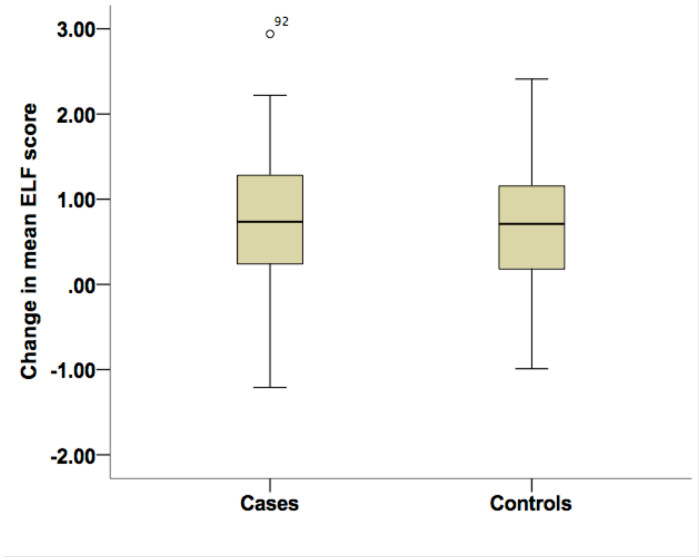


Figure 7.5. Box plots showing changes and distributions of change in mean ELF scores from recruitment to a) second sample and b) third sample, in cases and controls

Fifteen (25.9%) cases had a recruitment ELF score of ≥9.8 compared to 14 (12.2%) controls. Nine (15.5%) cases had a recruitment ELF score of ≥10.51 compared to 2 (1.7%) controls.

HRs for recruitment ELF are shown in Table 7.4. With an ELF threshold of 9.8, HR for LRE was 2.21 in the unadjusted model and 2.18 in the adjusted model. At the threshold of 10.51, HR in the unadjusted model was 4.88 and in the adjusted model HR was 4.62. Cumulative hazards for both models are shown in figures 7.6 and 7.7.

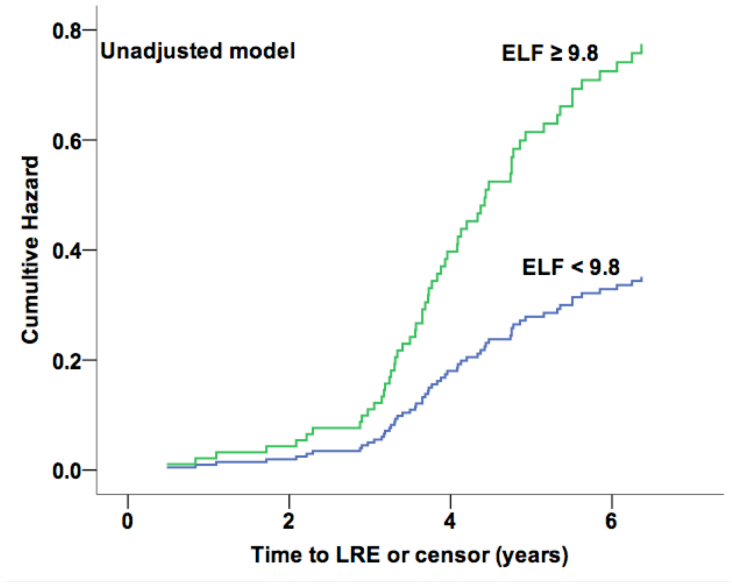
In the time-dependent Cox models, HRs at an ELF threshold of 9.8 are 1.85 and 1.80 in the unadjusted and adjusted models, respectively, and at a threshold of 10.51, HRs are 1.94 and 2.05 in the unadjusted and adjusted models, respectively.

Table 7.4. Hazard ratio estimates for liver-related event at ELF thresholds of 9.8 and 10.51
Hazard ratio estimates are presented using standard Cox proportional hazards and using time-dependent Cox analysis for liver-related event, at two ELF thresholds. Hazard ratio estimates are shown in unadjusted models and in models adjusted for deprivation, hypertension, heart disease, hypercholesterolaemia and type 2 diabetes

ELF threshold	Unadjusted / adjusted*	Cox		Time-dependent Cox	
		HR (95% CI)	*p* value†	HR (95% CI)	*p* value†
9.8	Unadjusted	2.205 (1.224-3.971)	0.008	1.854 (1.092-3.148)	0.022
	Adjusted	2.184 (1.189-4.013)	0.012	1.804 (1.041-3.126)	0.035
10.51	Unadjusted	4.880 (2.374-10.029)	<0.0001	1.935 (1.104-3.391)	0.021
	Adjusted	4.617 (2.115-10.081)	<0.0001	2.053 (1.157-3.644)	0.014

† at the 5% level
ELF, enhanced liver fibrosis; HR, hazard ratio

a)



b)

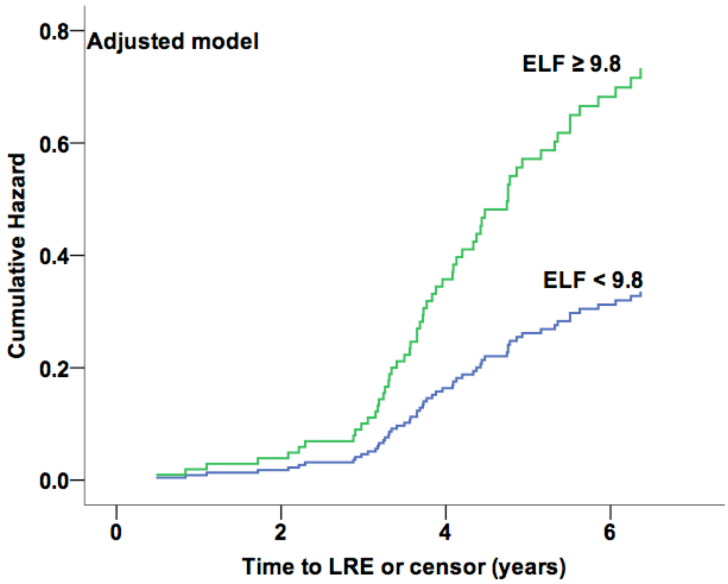
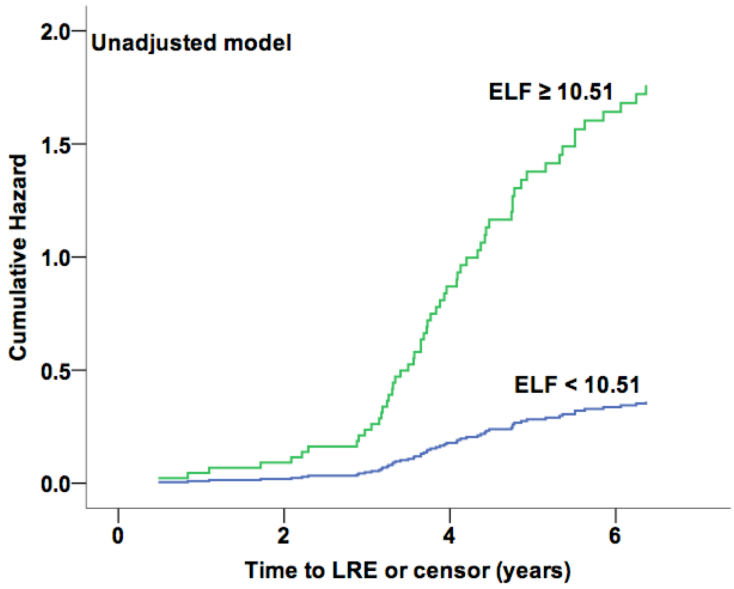


Figure 7.6. Cumulative hazards for LRE using ELF threshold of 9.8
Cumulative hazards plots for liver-related event according to ELF test threshold of 9.8 are shown, for a) an unadjusted model and b) a model adjusted for deprivation, hypertension, heart disease, hypercholesterolaemia and type 2 diabetes
ELF, enhanced liver fibrosis; LRE, liver-related event

a)



b)

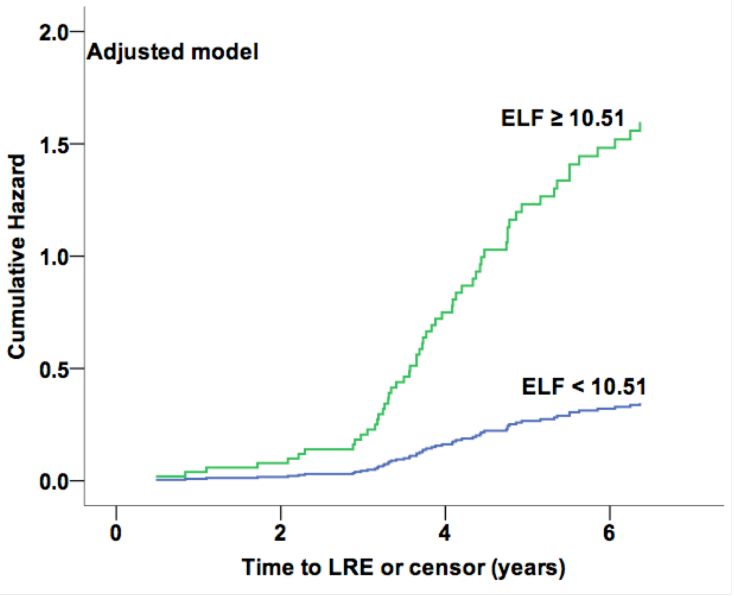


Figure 7.7. Cumulative hazards for LRE using ELF threshold of 10.51
Cumulative hazards plots for liver-related event according to ELF test threshold of 10.51 are shown, for a) an unadjusted model and b) a model adjusted for deprivation, hypertension, heart disease, hypercholesterolaemia and type 2 diabetes
ELF, enhanced liver fibrosis; LRE, liver-related event

Outcomes

The LREs for the cases are shown in table 7.5. The most common ICD-10 code for the study definition of LRE was K76, 'other diseases of the liver'. As with the studies in chapters 5 and 6, the ICD-10 codes were selected to include codes for liver disease and codes for complications of liver disease (I85, Z944 and / or C22.0) which may signify more advanced liver disease in these participants.

When cases with LREs coding for complications of liver disease (I85, Z944 and / or C22.0) were compared to cases with any other LRE code, mean recruitment ELF score, first subsequent ELF score and second subsequent ELF score were not significantly different (9.409 v 9.350, $p = 0.890$; 9.540 v 9.966, $p = 0.341$; 9.646 v 10.161, $p = 0.204$, respectively).

Table 7.5. ICD-10 codes and / or death certificate text of first LREs for the cases

The number of codes / death certificate text results is higher than the number of LREs (58) as some participants had more than one code when presenting with first LRE. Numbers of participants with codes of interest are divided by source of the code (hospital admission (HES), outpatient appointment (HES), cancer registration (ONS) and death certification)

Source	Code or text	Number of participants (% of those with LRE)
Hospital admission	K70	6 (10.3)
	K73	2 (3.4)
	K74	4 (6.9)
	K76	28 (48.3)
	I859	1 (1.7)
	Z94.4	4 (6.9)
Outpatient appointment	Z94.4	3 (5.2)
Cancer registration	C22.0	1 (1.7)
Death certificate	K70	1 (1.7)
	K74	3 (5.2)
	K76	5 (8.6)
	Mention of alcoholic liver disease	2 (3.4)
	Mention of non-alcoholic fatty liver disease	4 (6.9)

HES, Hospital Episode Statistics; ICD-10, International Classification of Diseases, Version 10; LRE, liver-related event; ONS, Office for National Statistics

7.5 Discussion

This prospective case-control study demonstrates that the ELF test can identify individuals at two-fold greater risk of liver related outcomes compared to other people with the same risk factors for CLD. The ELF test predicts liver-related outcomes in a general population of middle-aged women with risk factors for liver disease in the form of high BMI, high alcohol consumption or both, with a hazard ratio for event of 2 compared to women who do not develop CLD. This

is the first study to evaluate the predictive performance of the ELF test in a general population.

As discussed in chapter two, the asymptomatic nature of progressive liver fibrosis often results in individuals presenting with advanced cirrhosis and life-threatening complications. There is a need to;

- Appreciate the silent nature of liver disease
- Understand that standard liver chemistry blood tests do not identify individuals with cirrhosis
- Identify individuals with risk factors (and that these are to be found in community settings)
- Discriminate between individuals with risk factors who would and would not subsequently develop CLD

The key to addressing the rising mortality rates of liver disease is early diagnosis. This will permit strategies for risk modification and reversal of liver disease before irreversible and life-threatening complications develop. Early identification and management of liver disease would not only prevent morbidity and reduced quality of life in the individual, it would result in health economic benefits where resources could be channelled elsewhere, for example health promotion, and a healthier population.

This study shows that a simple blood test of liver fibrosis may discriminate between those with common risk factors for liver disease who could develop CLD from those who are less likely to, offering the possibility to direct health promotion strategies to those in most need. This will be discussed in more detail in the next chapter.

In this population, there was no significant difference in ELF score in those participants who would present with an LRE defined by a code for a complication of cirrhosis compared to those who would present with a code for liver disease. The conclusion from these observations is that the ELF test predicts development of liver disease but does not discriminate by mode of presentation. The clinical utility of the ELF test in predicting CLD early is the clear message from this study.

There was a small increase in mean ELF score over time in both cases and controls. Whilst this could suggest that ELF score increases with time, it could also represent progressive liver fibrosis in both groups or an effect of sample storage. This observation will require further study.

7.5.1 Strengths and limitations

7.5.1.1 Strengths

Strengths of this study include the prospective design and the independence of data capture for outcomes. As discussed in chapters 5 and 6, ICD-10 codes for cirrhosis that have been used in other studies were used, but in an attempt to maximise the ability to identify liver disease codes relating to the clinical consequences of advanced cirrhosis, the events defining decompensated liver disease, were also included. Evaluation of numerous possible confounders including self-reported known comorbidities and socioeconomic status minimised bias.

ELF tests were performed in one central laboratory, ensuring quality control and consistency, using the proprietary ELF assays.

Time-dependent Cox regression analysis

Time-independent Cox regression analysis compares survival distributions between those with a high ELF score to those with a low ELF score, based on a single ELF score (which could be at any time in the follow up period). It assumes that those with a high ELF score had a high ELF score from time zero, which is not necessarily the case. Time-independent Cox analysis using, for example recruitment ELF score assumes that participants with a low recruitment ELF score will never develop a high ELF score. However, the covariate status (ELF score) may change with time. Time-dependent Cox analysis compares risk of LRE between high and low ELF scores at each event time, but re-evaluates which risk group each participant belongs to, based on their most recent ELF score. It reduces the immortal time bias, i.e. the effect of considering the period of follow up during which a high ELF score had not yet occurred.

Strictly speaking, time-dependent Cox analysis is not a proportional hazards model as the HR changes with time. However, risk is reported in the same way as with time-independent Cox regression.[253]

I used time-dependent Cox analysis to miminise time-dependent bias. Immortal time bias refers to the period of follow up during which the study outcome could not have occurred. It occurs with the passing of time before a participant is subject to the exposure or defined level of the covariate (e.g. ELF ≥9.8). The period is considered immortal because participants necessarily had to remain event-free until the time of 'exposure' (in this case a high ELF score) to be classified as 'exposed'. An incorrect consideration of this 'unexposed' time

period will lead to immortal time bias.[254] It has been reported that that time-dependent bias is common in survival analyses published in leading clinical journals.[255]

7.5.1.2 Limitations

Limitations, as described in previous chapters, include the reliance on self-reporting of height and weight and co-morbidities in the UKCTOCS questionnaire, and reliance on ICD-10 to define events, that may result in errors due to miscoding. Three independent sources were used in an attempt to reduce risk of non-coding. Further, HES data may not capture some areas of healthcare, including the private sector. Finally, although attempts were made to ensure UKCTOCS was representative of the general population, there was a 'healthy volunteer effect' on overall and cause-specific mortality, which may affect the generalisability the findings.[177] As previously discussed, the UKCTOCS alcohol categories do not align with the CMO's threshold for hazardous drinking of 14 units / week and therefore including the UKCTOCS threshold of 11-15 units / week may have over-estimated 'high alcohol use'; however excluding this category in this study would have risked excluding women with hazardous alcohol consumption, and the tendency to under-report alcohol use was given due consideration.

7.5.2 Other studies investigating the performance of biomarkers to predict clinical outcomes

The ELF test has been shown to predict clinical outcomes in several secondary care or disease-specific contexts. Participants in the original ELF study were followed up for liver-related outcomes (ascites, hepatic encephalopathy,

oesophageal variceal haemorrhage, liver transplant or hepatocellular carcinoma, or liver-related death).[83] When the ELF score was divided in to tertiles, compared to the low ELF score (4.14 - 8.33), adjusted HRs for the subsequent three tertiles were: 4.9 in the lower half of the middle tertile, 19.8 in the upper half of the middle tertile and 75.7 in the highest tertile. ELF was superior to liver biopsy in predicting liver-related outcomes, and predicted outcomes at 6 years with AUROC of 0.88. Subsequently, in a cohort of patients with primary biliary cholangitis event-free survival (development of varices, variceal haemorrhage, ascites, hepatic encephalopathy, liver transplant, liver-related death) was significantly lower in those with a high baseline ELF score.[79]

A large retrospective study based in Tayside followed up 95,977 patients with no known liver disease in whom LFTs had been measured (outlined in chapter two).[41] Over a median of 3.7 years, 1.14% developed liver disease. Elevated transaminases were strongly associated with development of liver disease, for example HR for mildly elevated ALT was 4.27 and for severe ALT 12.67, compared to a normal ALT. Interestingly, GGT which is not universally part of the standard “LFT” panel, HRs were 2.54 and 13.44, respectively. These individual markers had low sensitivity for predicting events, with high specificity (i.e. effective at ruling in liver disease, but less effective in ruling out disease). These data were used to inform the development and validation of an algorithm to predict all-cause mortality in patients with no apparent liver disease (over one year) for use in primary care to aid decision making (the Algorithm for Liver Function Investigations, ALFI).[256] The investigators found that inclusion of LFTs in the model improved the sensitivity of the algorithm, but did not improve the (low) PPV of the (high) NPV. The authors point out that the model is most

effective at lower thresholds of predicted probabilities of mortality, providing good sensitivity and NPV, but the low PPV of the algorithm may result in (subsequent) over-investigation of patients.

The prognostic value of a number of biomarkers was reported in a meta-analysis which found good performance in disease-specific cohorts.[257] Of note, the performances of estimating 5-year survival without liver-related death in disease-specific cohorts for those markers where there was more than one validation study, in terms of AUROC, values were 0.88 for FibroTest, 0.73 for Fib-4 and 0.66 for APRI. A subsequent study, again based in secondary care, evaluated the performance of the simple serum panels NAFLD fibrosis score, AST:ALT, Fib-4 and BARD to predict liver-related complications and death or liver transplantation in a cohort of 320 patients with biopsy-proven NAFLD.[258] There was an even distribution of fibrosis stages, and over a median follow up of 105 months, there were events in 14% of patients (the most common events being development of oesophageal varices, ascites or hepatic encephalopathy), death or liver transplantation in 13%. This was a retrospective study using case notes (as opposed to HES codes) to identify events. Using Cox regression, adjusted HRs for liver-related event for indeterminate risk and high-risk scores, using low risk score as the reference were 7.7 and 34.2 for NAFLD fibrosis score, 8.8 and 20.9 for APRI, 0.92 and 14.6 for Fib-4, and 6.2 and 6.6 for BARD. The longitudinal arm of a study described in chapter 2 evaluated the ability of a number of non-invasive markers to predict outcome in patients with NAFLD.[64] Over follow up of 5-9 years, of APRI, Fib-4, Hepascore, FibroMeter and TE, FibroMeter and TE performed well in predicting both all-cause mortality and liver-related deaths.

The studies described above have investigated the prognostic performance of non-invasive tests in disease-specific groups, rather than in screening studies. A Canadian study evaluated the performance of TE in predicting LRE in a cohort of patients with a range of liver disease aetiologies.[259] Following TE, patients were followed up for a median period of 5.6 months, for complications of liver disease (ascites, spontaneous bacterial peritonitis, jaundice, hepatic encephalopathy, variceal haemorrhage, hepatorenal syndrome, hepatocellular carcinoma, liver transplantation). HR estimates for LRE, compared to TE <10 kPa, were 3.2, 7.0 and 12.5 for TE 10-19.9 kPa, 20-39.9 kPa and ≥40 kPa, respectively. As a comparison to my results, the authors considered TE ≥12.5 kPa as the threshold for cirrhosis, at this threshold the HR estimate for LRE was 7.6.

An Australian study of 300 subjects with CLD who underwent liver biopsy and ELF testing, previously described in chapter 2, were followed up for a median of 6 years for LRE.[87] 19% of subjects with an ELF score ≥9.8 experienced an LRE compared to <1% in those with an ELF score <9.8. A unit increase in ELF score was associated with a 2.5-fold increase of LRE, adjusted for age and stage of fibrosis. A limitation of this study was that clinical outcome data was taken by review of hospital medical records rather than from linked databases, therefore there may have been an underestimation of events.

Long term community studies using biomarkers to stratify patients and to determine the value of stratification to predict liver events are under way.[260] As these data become available, they will need to be considered in the context of the practical aspects of each modality (table 7.6).

Table 7.6. Benefits and limitations of non-invasive markers of liver fibrosis to be considered during evaluation as a prognostic marker

Non-invasive test	Benefits	Limitations
ELF test	Low failure rate CE marked & central laboratory processing Can be added to 'routine' blood tests No 'start-up' costs	Not yet universally commissioned
TE	Instant result Truly 'non-invasive' Good inter- and intra-observer agreement	Failure rate 2-9%[261] Unreliable in 16%[168] (Minimal) training required to operate machine Capital and maintenance or rental costs Affected by ascites, oedema, eating
Simple markers	Inexpensive No 'start-up' costs	Non-liver specific – affected by comorbidities

ELF, enhanced liver fibrosis; TE, transient elastography

7.6 Conclusion

This study builds on the growing body of evidence supporting the clinical utility of the ELF test in community settings for diagnosis of liver disease. In this chapter, it has been demonstrated that the ELF test has the ability to predict clinically significant liver-related events in middle-aged women with risk factors. Serum transaminases are a widely used but inaccurate measure of chronic liver disease and the absence of symptoms or signs of early liver disease combined with increasing liver-related deaths highlight the need for accurate, reproducible tests of liver disease, in particular clinical endpoints. Stratification in those with risk factors is a potentially valuable strategy. My study population are of particular interest. Patterns of alcohol consumption in women are changing,

with 16% of women in England consuming above recommended limits, and this practice is highest in the 55-64-year-old group.[224] Further work is required to demonstrate the generalisability of these findings to other community-based populations.

Chapter 8. DISCUSSION

8.1 Statement of main findings

8.1.1 Validation of the ELF test in chronic hepatitis B and comparison with transient elastography

In this study, which was the first to validate the ELF test in an external cohort of subjects with chronic hepatitis B, my work has shown that the ELF test is able to discriminate between fibrosis stages, and accurately quantify fibrosis at all stages, with AUROC values above 0.8. To compare results of this study with other similar studies of serum-based markers of fibrosis, the DANA method was applied to calculate adjusted uniform AUROC, assuming equal prevalence of all fibrosis stages and revealed enhanced performance.

The performance of the ELF test was compared to that of TE in the same subjects. Performance of the two methods was similar, with TE performing marginally better than the ELF test in the diagnosis of severe fibrosis.

A model comprising both modalities in a serial algorithm in which the first non-invasive marker classifies subjects in to ‘disease’, ‘no disease’ or ‘indeterminate’ with the indeterminate cases proceeding to the second marker with a single threshold (‘disease’ or ‘no disease’) resulted in increased overall sensitivity and specificity without substantial loss of diagnostic accuracy.

This study not only validated the ELF test in the context of chronic hepatitis B, it demonstrated that its performance is similar to TE. TE is being evaluated in community settings as a screening tool, and therefore ELF may too have a role.

8.1.2 Defining the incidence of chronic liver disease in a general population

In the UKCTOCS cohort, comprising a group of women aged 50-74 recruited from the general population, the incidence of chronic liver disease, defined by first presentation with cirrhosis or a clinical consequence of advanced liver disease, was 0.64 events per 1000 person years. Several other studies have reported incidence rates of liver disease. The population sharing the most similarity to the UKCTOCS population is that of the Million Women Study, where incidence of CLD was found to be 0.24 events per 1000 person years. However, the data definition used in the Million Women Study covered a narrower group of ICD-10 codes than used for the UKCTOCS study, possibly leading to an underestimate of advanced liver disease. This highlights the caution needed in comparing studies. This is explored in more detail later.

8.1.3 Evaluating the risks of alcohol and above normal body mass index on the risk of chronic liver disease

Alcohol and overweight and obesity are known to be significant risk factors for development of CLD, but their precise influence is not conclusive. This study found that in postmenopausal women, the risk of CLD increased with increasing BMI. Crude rate of first CLD event in the normal BMI group was 0.45 per 1000 person years, 0.65 per 1000 person years in the overweight group, and 1.06 per 1000 person years in the obese group. Using survival analysis, the hazard ratio estimates, when adjusted for smoking, alcohol use, deprivation, and the metabolic risk factors of heart disease, hypercholesterolaemia and type 2

diabetes, were found to be, compared to normal BMI, 1.31 in the overweight group and 1.85 in the obese group.

The crude event rate was lowest in the women consuming between <1 unit / week and 15 units of alcohol weekly. There was an increase in events with increasing reported alcohol use. Compared to this group, there was an increased number of events in abstainers. The adjusted hazard ratio estimates followed the same pattern. One explanation for the J-shaped relationship could be that the abstainer group contains previous heavy drinkers who have subsequently developed liver disease, and this theory is supported by a number of abstainers presenting with alcohol-related ICD-10 codes.

8.1.4 Exploring the interaction between alcohol and body mass index and the risk of chronic liver disease

The interaction between alcohol and BMI on the risk of CLD was explored by dividing the group in to twelve, each group's exposure being defined as a combination of alcohol use and BMI. The lowest event rate was in the normal BMI group consuming <1-15 units of alcohol weekly, with the highest rate in the overweight group consuming over 21 units of alcohol weekly. In all alcohol groups, the general trend was increasing event rate with increasing BMI. In each BMI group, there was a J-shaped relationship with increasing alcohol use, with the lowest rate in the <1-15 units / week group in all the BMI groups. Survival analysis showed that the hazard ratio estimates followed a similar pattern. Compared to the normal BMI group consuming no alcohol, the hazard ratio estimates in the overweight and obese groups consuming over 21 units per week were over 3. Although a statistical synergistic effect could not be

demonstrated, these data clearly show a substantial risk to health in those who are obese and consume excess alcohol.

8.1.5 Evaluating the association between skirt size and chronic liver disease

In this study, UKCTOCS participants' skirt sizes in their early 20s, their postmenopausal skirt sizes, and the change in skirt size were analysed to determine any association with development of chronic liver disease. The rate of liver-related events was higher in those women with a skirt size in their 20s of 18 or more compared to those with a skirt size of 16 or less. This pattern was again seen with participants' postmenopausal skirt size. The change in skirt size was also associated with liver-related outcome; the rate of event was highest in group where skirt size increased with time, and lowest in the group where skirt size decreased.

Compared to a postmenopausal skirt size of less than 16, survival analysis demonstrated a significant association between skirt size of more than 18 and a liver-related event. The hazard ratio estimate was between the hazard ratio for overweight and that for obesity. The same pattern was seen with skirt size in participants' early 20s, however the 5% significance was lost in the adjusted model, raising the possibility that a component of the risk may be attributable to metabolic comorbidities.

These data raise the possibility of using skirt size as a way for women to stratify their risk of liver disease.

8.1.6 Determining the performance of the ELF test to predict liver-related clinical outcomes in postmenopausal women with risk factors

In this study it was found that, in women with risk factors for chronic liver disease in the form of overweight or obesity and / or hazardous alcohol use, the ELF score at recruitment to the UKCTOCS trial was higher in those who subsequently experienced events related to cirrhosis. Time-dependent survival analysis demonstrated that a high ELF score predicted liver-related outcomes over the follow-up period, with a hazard ratio of 2. These data suggest a role for the ELF test in stratifying patients at risk of liver disease in primary care.

8.1.7 Overall statement of findings

This work has validated the performance of the ELF test in chronic hepatitis B, demonstrating good performance, and has shown that the ELF test has comparable performance to transient elastography in this context. The incidence of liver disease in postmenopausal women, an under-evaluated demographic, has been elucidated. In addition, interesting and novel findings related to the influence of two major causes of liver disease, alcohol and overweight and obesity have been presented. For the first time, the association between skirt size and liver disease has been described, raising the possibility for its role in stratifying risk in a straightforward way. Finally, a well-established marker of liver fibrosis has been shown, for the first time, to predict development of chronic liver disease in a community setting.

8.2 Strengths and limitations

The strengths and limitations relating to each study presented in this thesis have been described in detail in the relevant chapters. However, there are a number of general strengths and limitations that should be highlighted and discussed.

8.2.1 Strengths

A major strength of this thesis is the utilisation of two large cohorts. The first, a cohort of subjects with chronic hepatitis B, which was carefully and methodically selected in accordance with the study design's aims and objectives. This allowed for high quality data to be available. At recruitment, subjects underwent liver biopsy, transient elastography and serum sampling on the same day. The ELF testing was performed at the central ELF laboratory under the direction of the personnel involved in the data analysis.

Collaboration with the UKCTOCS team allowed me to gain an understanding of the way that longitudinal cohorts are managed and to tap in to the knowledge and skills of the custodians of this programme. I learnt a great deal about the management of large databases and developed new skills in searching routine data sources.

The main strength of the UKCTOCS studies is the large number of trial participants involved. The UKCTOCS population represent an important but under-investigated group in terms of liver disease. Rates of hazardous alcohol use are rising in this demographic, and on the background of the increasing burden of NAFLD, an understanding of the epidemiology of liver disease is

crucial. The ability to statistically control for a large number of potential confounders was a particular strength.

The work has added to the knowledge gained from other studies of middle-aged women, and provided new insights in to the relationship between two important risks of liver disease, alcohol and overweight / obesity. Planning of community health services and public health strategies depends on knowledge of levels of disease in the target populations. My work, by using a variety of data sources in a large representative population through a prospective study has added to this knowledge.

The finding that skirt size strongly predicts future liver disease represents a novel opportunity to convey a simple public health message that is backed by robust data.

My work has shown that the ELF test has excellent performance in predicting liver-related clinical outcomes in middle-aged women, and is the first study to demonstrate the ability of the ELF test to predict outcome in a community setting. The success of this study was in no small part due to being a member of the ELF research group, which under the direction of Professor Rosenberg, the ‘founder’ of the ELF test, has more cumulative knowledge on the ELF test than can be found anywhere else.

8.2.2 Limitations

There are a number of general limitations to this thesis that should be rehearsed in these concluding remarks.

8.2.2.1 Healthy volunteer effect

Although UKCTOCS was a large trial, the participants comprised women who chose to take part. The UKCTOCS investigators selected eligible women to take part by requesting electronic files containing details of 2000 – 10000 women on a three-monthly basis from each participating primary care trust, and then sent women personal invitations. Of the 1,243,282 invited, 205,090 were recruited.[175] It may be that individuals who choose to take part in trials are more health conscious. The 'healthy volunteer' effect has been described in the UKCTOCS cohort[177] and in other clinical trials, as discussed in chapter 5.

8.2.2.2 Self-reported alcohol questionnaire

The self-reported alcohol consumption question, as part of the UKCTOCS follow-up questionnaire (appendix E) was sent before this thesis was conceived and therefore not specifically designed for my study. The lack of a standard approach to use of alcohol consumption questionnaires in research (and the difficulty in validating questionnaires[262]) limits the ability to compare studies. My hope is that the findings from my work will inform public health strategies, however the thresholds for alcohol use in my work does not directly correlate with those used by the Department of Health.[176] An additional complication is the controversy around how many units there are in an alcoholic drink and the variability in the size of measures and these issues are discussed in chapter 5.

8.2.3 Missing data

As expected in a large trial involving return of questionnaires and self-reporting of data, there is a degree of missing data in the UKCTOCS dataset. The missing data relevant to my dataset related to self-reporting of alcohol consumption,

smoking status and skirt size. These all required an answer to be documented on the questionnaire.

Of the participants who returned the follow-up questionnaire (a requirement to be included in my dataset), data were missing for alcohol and / or smoking in 14% (see figures 5.1 and 6.1 which show the derivations of the study groups). Given that the questions relate to potentially hazardous lifestyle activities, there is a possibility that the data were 'missing not at random' (MNAR) and therefore raises the possibility of bias by excluding these participants. However, there is no routinely available software to deal with MNAR, so I considered the data to be 'missing at random' (MAR). I repeated the analyses after using the 'multiple imputation' function of SPSS. Results were similar to those obtained from the dataset with the 14% of participants with missing values, therefore I chose to use this smaller dataset. Justification for adopting 'complete case analysis' is outlined in chapter 5.

The other potential source of missing data was related to reporting comorbidities. Although there was a box to check if participants had none of the listed medical conditions, it is possible that conditions were under-reported. The potential significance of this discussed in chapter 5.

There were also problems related to self-reported height and weight, requiring a pragmatic approach to deciding what were considered acceptable values. This is discussed in the method and discussion sections of chapter 5.

8.2.3.1 Coding

The limitations related to coding have been discussed in chapters 5-7, but should be mentioned here as a general limitation, given that coding was a major feature of the thesis. Although a strength of the study was the use of three independent data sources, all three relied on ICD-10 coding. There are some areas of healthcare where coding may be less accessible, for example in the private sector. However, it is unlikely that this would substantially impact on my results. It has been reported that some diagnoses rates change as new ICD versions are introduced. For example, an increase in recorded liver deaths after 1979 when ICD-9 was introduced, followed by a reduction when ICD-10 was introduced.[24]

A study in hospital inpatients identified a number of sources of error in the patient pathway in terms of data gathering, including quality of information available at the time of hospital admission, communication between doctor and patient and clinician's knowledge of the presenting illness, in addition to experience of the coder.[263]

Miscoding is a potential source of error in all prospective cohort studies that rely on routine data sources, and the use of three sources in my study, in addition to use of a group of codes that included not only cirrhosis but complications of cirrhosis may have helped to reduce error.

8.3 Implications of research findings for clinical practice

This thesis has focused on postmenopausal women and highlights the importance of recognising this cohort as an at-risk group. Due to the rising

prevalence of CLD, in particular driven by NAFLD and obesity, there is much potential in targeted case-finding strategies in the community. General practitioners should be aware of this group but need to be given tools to identify liver fibrosis and diagnostic tests for early identification of liver disease.

Although hazardous alcohol use is declining in the younger age groups, this is not the case in the over 50s.[264] Particularly worrying is the increasing proportion of women drinking later in life, driven by life events including bereavement, changes in personal circumstances and retirement.[265] Therefore not only is it important for screening and treatment of alcohol misuse in the community, it is crucial to focus on the older population.

The clinical concept of "both alcoholic and non-alcoholic fatty liver disease" (BAFLD) is not yet commonly used, but is an attractive concept initially proposed by the public health group in Southampton (Parkes *et al*, personal communication) to provide a diagnosis in patients found to have hepatic steatosis on ultrasound imaging and where the clinical history comprises elements of both hazardous alcohol use and risks for NAFLD. As I have demonstrated, the risk to liver health is high when both risk factors are present (i.e. alcohol excess and high BMI) in postmenopausal women, therefore this diagnostic description is particularly relevant to this group and if used more widely in community practice may improve early diagnosis.

My work offers the possibility to intervene in the development of liver disease at several points. Awareness of the risk of large skirt size at a young (and older age) provides a way to 'self-stratify' risk and modify behaviour and lifestyle to prevent liver disease. A second point of intervention, with the use of the ELF

test, which predicts development of liver disease in those who have accumulated the risk factors, offers the potential to develop targeted public health strategies, again with the hope of preventing progression to established liver disease and complications of cirrhosis.

8.3.1 Hepatology services for NAFLD and alcohol and pathways of care

8.3.1.1 NAFLD

Despite the increasing prevalence of NAFLD, there are little data related to identification, diagnosis and referral of patients with NAFLD. The disease is poorly recognised in primary care. In a US study, 251 patients with elevated ALT (in the absence of positive hepatitis B or C markers, or history of alcohol excess) were randomly selected, and primary care records were used to identify NAFLD.[266] 39% had documentation of abnormal ALT, 22% had NAFLD as a possible diagnosis, 15% had been given lifestyle advice, and 10% had been referred to a specialist. Only 3% of those at high risk of fibrosis, based on a high NAFLD fibrosis score calculated by the investigators had been referred to a specialist. In an Australian survey of 108 primary care clinicians, 51% considered prevalence of NAFLD in the general population to be ≤10% (compared to a population prevalence of 30%[125]).[267] Although most agreed that liver enzymes, platelet count, albumin, prothrombin time, NAFLD fibrosis score, liver ultrasound and TE could help diagnose NAFLD, 63% and 64% were unsure whether Fib-4 or the ELF test, respectively, could aid in identifying advanced fibrosis or cirrhosis. The large majority (94%) stated they would provide information on optimising diet and exercise, but 71% indicated that they would use abnormal liver chemistry as a deciding factor in referring to

secondary care. As discussed in previous chapters, many cases of liver disease are diagnosed when an individual presents to hospital with a life-threatening complication due to decompensated cirrhosis. These patients can expect to be subsequently followed up by a liver specialist. However, due to the asymptomatic nature of progressive fibrosis and cirrhosis, many individuals with liver disease are unidentified and therefore are not followed for the complications of CLD, in particular variceal and HCC screening.

The lack of secondary care hepatology expertise outside of major tertiary liver units has been highlighted, but the need to improve liver care and screening in primary care has also been recognised.[22] In terms of NAFLD, there is a need to stratify patients thought to have NAFLD so that those with low risk for advanced fibrosis remain in primary care to receive lifestyle advice or referral to a metabolic clinic, and those at high risk of advanced fibrosis are referred to a hepatologist for assessment. As described in chapter 2, the role of the ELF test has been embedded in NICE guidelines. However, the NICE guidelines are often not used. I have been involved, with colleagues, in developing a pathway for general practitioners to stratify patients with a diagnosis of NAFLD (see chapters 2 and 3, and list of publications).[97,115] Rather than use the ELF test for all patients, a two-stage process is adopted; the use of a simple panel first (Fib-4 score) stratifies patients in to three groups – low risk of advanced fibrosis, high risk of advanced fibrosis and an indeterminate group. The ELF test is performed in those in the indeterminate group and those with a high ELF score are referred to secondary care along with those with a high NAFLD fibrosis score. Implementation of this pathway has improved stratification of patients, reducing overall referrals but increasing referrals of individuals with cirrhosis.

As discussed in chapter 3, work from Nottingham has demonstrated the value of a community-based targeted approach to increasing the rate of detection of participants with CLD. A two-stage approach, similar to that described above was offered to individuals at risk (alcohol misuse, type 2 diabetes or abnormal liver enzymes) identified using general practice electronic records. Initial screening was with a simple marker panel, followed by TE for those with high values. Overall prevalence of cirrhosis in participants was 2%.[117] In an extension to the study, patients with high alcohol use and / or type 2 diabetes were invited to undergo TE directly, and 3% were found to have cirrhosis.[121]

As discussed in chapter two, lifestyle change, including dietary modification and physical activity are the first-line interventions in management of NAFLD. Weight reduction of 10% or more can result in resolution of non-alcoholic steatohepatitis and an improvement in fibrosis by at least one stage. There is also strong evidence of the risk of overweight and obesity in development of many cancers, representing another important public health message related to BMI (discussed below). In addition to obesity, there must also be a focus on control of other modifiable metabolic risk factors including type 2 diabetes, hyperlipidaemia and hypertension. Differences in study design and outcome measures makes comparison of studies difficult, and therefore recommendations are inconclusive.[268] The key to success is likely in large part to be in changing individuals' behaviour and motivation. The 5As model (ask, advise, assess, assist, arrange) is a behavioural counselling intervention tool that has been applied to other areas of healthcare including smoking, and could be used in primary care when managing patients with NAFLD. It includes

support on how to advise patients to modify behaviour, assess their interest in doing so, assist their effort to change and arrange appropriate follow up.[269]

NICE guidelines on obesity include recommendations for general practitioners on referral of overweight and obese patients for lifestyle and weight loss programmes. A key element is the measurement of BMI and delivering weight loss advice in a sensitive manner, stressing the benefits of weight loss in conjunction with changes to dietary habits and physical activity levels.[270] However, there is evidence that general practitioners are reluctant to advise on weight loss, diet and exercise and consequently appropriate patients are often not referred.[271] Barriers to delivering advice include lack of time, fear of causing offence and a belief that interventions are ineffective.[272] A study of 366 patients surveyed in a general practice waiting room found that 49% of obese patients, 24% of overweight patients and 12% of non-overweight or obese patients had discussed weight with their doctor, and most patients, particularly those who were overweight or obese, wanted more help with weight management than they were receiving.[273] An alternative strategy for intervention is screening and brief intervention, in a similar structure to alcohol (and smoking cessation) advice. I found only one trial in this area. In this UK study, patients attending their general practitioners were screened for obesity using BMI. 1882 patients were enrolled, and randomised to either an advice and weight loss programme group or an advice only (control) group.[274] The control intervention comprised a 30 second advice session on the health benefits of weight loss. The intervention comprised the offer of referral to a 12-week weight management programme, each session lasting one hour. 77% of the intervention group agreed to attend the weight management programme, and 40% of these

attended. At one year, mean weight loss was 1.04 kg in the advice group and 2.43 kg in the intervention group. 81% felt that the advice and / or intervention was helpful and appropriate, which was similar between the groups.

Association between cancer and BMI

Although not directly related to my work, the association between cancer and BMI should be acknowledged and represents an area that requires increased public awareness.

The positive association between BMI and cancer is long established. A number of studies have provided more clarity by estimating the magnitude of effect. The association between BMI and incidence of cancer was examined in the Million Women Study.[275] BMI was calculated using self-reported height and weight at recruitment to the study. 1,222,630 women, with a mean age of 55.9 years, were followed up for a mean of 5.4 years for incidence and 7.0 years for cancer mortality, using National Health Service cancer registries. 17 types of cancer were searched, using ICD-10 codes. Increasing BMI was associated with increased risk of 10 of the 17 cancers studied. The relation between BMI and incidence was similar to that between BMI and cancer-related mortality. The overall relative risk for developing cancer (adjusted for age, socioeconomic status, alcohol use and physical activity) was 1.12. The highest risk associated with obesity was seen in incidence of endometrial cancer and adenocarcinoma of the oesophagus, with relative risks of 2.73 and 2.54, respectively. Menopausal status was associated with risk related to BMI, compared to premenopausal status, of both cancers known to be hormone-related and some other cancers. For example, risk of breast cancer associated with increasing

BMI was reduced in premenopausal women and increased in postmenopausal women. Risk of endometrial cancer was increased in both pre- and postmenopausal women, but the risk was higher in postmenopausal women. Risk of colorectal cancer increased with increasing BMI in premenopausal women but decreased in postmenopausal women.

More recently, the association between BMI and the twenty-two most common site-specific cancers has been investigated in a general population. Using the earliest BMI recorded (usually measured in general practice surgeries), 5.24 million individuals were followed up for a mean of 7.5 years by interrogating the Clinical Practice Research Datalink for first presentation of any of the cancers of interest, and hazard ratio estimates calculated using Cox proportional hazards analysis.[276] Overweight or obesity, compared to normal BMI, was associated with higher risk of development of 13 of the 22 cancers. The highest risk was for uterine cancer (hazard ratio of 1.62). Of particular note, hazard ratio for liver cancer was 1.19.

These studies suggest that BMI is associated with increased risk of a large number of cancers, and therefore may represent a significant modifiable risk factor for developing cancer. The Million Women Study suggests a possible interaction between BMI and menopausal status that differs between types of cancer.

8.3.1.2 Alcohol

In terms of community management of alcohol misuse, the focus must be on screening and intervention. The NICE quality standard (Alcohol-use disorders: diagnosis and management) defines screening in this context as;[277]

> "identifying people who are not seeking treatment for alcohol problems but who may have an alcohol-use disorder"

The aim of the quality standard was to increase the identification of alcohol misuse (harmful drinking and alcohol dependence) and encompassed statements relating to training of health and social care staff, screening and brief intervention and appropriate referral of individuals to specialist alcohol services.

Screening, as a form of case-finding, can be universally applied or targeted to at-risk populations. Established alcohol screening tools include CAGE, AUDIT and MAST. The alcohol use disorders identification test (AUDIT) has been validated for use as an outcome measure in primary care. A score of ≥8 out of 40 indicates hazardous or harmful drinking or the likelihood of dependent drinking, with a sensitivity of 92% and a specificity of 94%.[278,279] The older CAGE questionnaire is briefer and was designed to identify only alcohol use disorders. Although AUDIT identifies individuals with risky drinking (who may benefit from intervention), i.e. may be too long for use in busy primary care settings. The AUDIT – Consumption (AUDIT-C) tool comprises the first three questions of the AUDIT tool and pertains to alcohol consumption and can therefore be used as a screening test for alcohol use disorders or risky drinking. It has been validated in primary care.[280]

Kaner *et al* describe brief intervention;

> "brief intervention is grounded in social-cognitive theory and … incorporates … the following elements:

- Feedback on the person's alcohol use and any alcohol-related harm clarification as to what constitutes low risk alcohol consumption;
- Information on the harms associated with risky alcohol use;
- Benefits of reducing intake;
- Advice on how to reduce intake;
- Motivational enhancement; analysis of high risk situations for drinking and coping strategies; and
- The development of a personal plan to reduce consumption."

There is evidence for the efficacy and cost-effectiveness of brief interventions. A Cochrane systematic review and meta-analysis in 2007 identified twenty-one randomised controlled trials and found that participants receiving brief intervention reduced their alcohol consumption compared to the control group, with a mean difference of -41 grammes / week. Interestingly sub-group analysis confirmed the benefit of brief intervention in males at one year of follow up (-57 grammes / week) but not in females (-10 grammes / week).[281] This Cochrane review has recently been updated, comprising sixty-nine studies. Most studies were in primary care (55%) or in emergency departments (39%).[282] The meta-analysis showed that those who received brief intervention consumed less alcohol at one year than those receiving minimal or no intervention, with a mean difference of -20 grammes / week. In this, larger, study it was demonstrated that both males and females reduced alcohol consumption after brief intervention.

The way that brief intervention is delivered may be important. A recent randomised controlled trial has evaluated the effectiveness of three different interventions; patient leaflet, five-minute brief advice and 20 minutes of lifestyle counselling. The outcome was hazardous or harmful dinking assessed using the AUDIT test at six months.[279] The proportions of AUDIT-negative subjects at six months was 36%, 29% and 29%, respectively, representing no difference between the groups.

Opportunities for brief intervention exist outside of primary care and emergency settings. The acute medical admissions unit represents one such area. A large study based in a hospital acute medical unit, between July 2011 and March 2014 screened nearly 50,000 admissions.[283] The screening tool was a modified version of the Paddington Alcohol Test and recorded type of alcohol consumed, frequency and maximum daily amount. Those at increasing risk of alcohol harm were given a brief intervention and those at high risk were assessed by an alcohol specialist nurse and subject to the AUDIT questionnaire. 2.3% were classified as 'increasing' risk and 4% as high risk of alcohol harm. Of the high-risk group completing the AUDIT test 81% had a score over 20 including 38% with the maximum score. These data suggest a role for identifying individuals at risk of alcohol-related harm who may not use primary care services.

There is some evidence that brief intervention is enhanced when coupled with staging of liver disease. A study based in primary care identified hazardous and harmful drinkers using the AUDIT tool (score $\geq$8), finding 24% of those completing the tool in this category.[284] These individuals were alerted to their general practitioner for referral to alcohol services and offered an assessment of liver fibrosis using the 'Southampton Traffic Light' (STL) test, a serum panel

comprising HA and P3NP combined with platelet count. Thirty-eight percent of the hazardous or harmful drinkers attended for liver fibrosis assessment using the STL for this test and their level of liver disease was fed back to them. Follow up AUDIT data showed that 50% with liver damage reduced alcohol use by at least one category compared to 35% in the group with no liver damage. Although this study may suggest that individuals' knowledge of their liver damage may increase the motivation to reduce alcohol use, there are a number of limitations to the methodology. The study does not report the numbers of participants who were referred and attended alcohol services and therefore the contribution of this intervention cannot be evaluated. Unlike other non-invasive markers of liver fibrosis, the STL test has not been extensively validated and was used in this study merely to categorically assign participants to have, or not have, liver disease.

The data presented in this section support the use of screening and intervention for risk factors for liver disease, but there are some barriers to delivery in clinical practice. Later in the chapter I will suggest a targeted integrated approach that could be developed to screen for risk factors for liver disease.

8.4 Application of this work to the general population – Health Survey for England

Comparisons of my study population to populations used in other studies have been outlined in discussions of these studies in previous chapters. However, the Health Survey for England (HSE)[224] is an observational study that warrants particular comparison with my findings.

The HSE is an annual survey that looks at changes in the health and lifestyles of the population of England. Data are gathered from questionnaires and interviews with nurses where anthropometric data is also gathered from physical examination. In addition to core questions and measurements including height, weight and blood pressure, additional topics are covered that vary year to year. Data from the 2015 HSE are available for analysis, the sample comprised 8,034 adults of which 5,378 had a nurse visit. In this section I will compare and contrast elements of baseline data in my study population with women of an equivalent age in the HSE.

8.4.1 Alcohol consumption

Figure 8.1 shows self-reported alcohol consumption in females in HSE and shows that the median number of alcohol units reportedly consumed per week is highest in the 55-65-year-old group, highlighting the importance of focusing on this demographic. In the 55-64 and 65-74-year-old groups, 17% and 21% reported consuming no alcohol in the past twelve months, similar to the prevalence of participants reporting no alcohol intake of 23% in my study group. However, the prevalence of women consuming over 21 units of alcohol per week was higher in the HSE compared to women in my study group who reported consuming 21 units or more per week (10% and 5% in HSE compared to 1.5% in my study group). The categories, however, were not directly comparable as in HSE those reporting consuming 21 units per week were in the 14-21 units/week category.

Estimated weekly alcohol	Age group								Total
	16-24	25-34	35-44	45-54	55-64	65-74	75-84	85+	
	%	%	%	%	%	%	%	%	%
Non drinker/did not drink in last 12 months[1]	25	18	18	17	17	21	29	40	21
1 unit or less (lower risk)	11	16	14	14	17	19	23	25	16
More than 1, up to 7 units (lower risk)	37	40	40	34	28	32	30	25	35
More than 7, up to 14 units (lower risk)	13	14	13	15	14	12	8	6	13
More than 14, up to 21 units (increasing risk)	8	6	7	8	9	5	6	4	7
More than 21, up to 28 units (increasing risk)	3	3	3	4	6	4	2	1	4
More than 28, up to 35 units (increasing risk)	2	1	2	4	4	2	1	-	2
More than 35, up to 50 units (higher risk)	0	1	2	2	3	2	1	-	2
More than 50 units (higher risk)	1	0	2	2	3	1	-	-	1
Mean number of units[3]	10.5	6.6	8.9	10.2	11.7	8.3	5.8	3.6	8.9
Standard error of mean	2.89	0.47	0.59	0.62	0.92	0.65	0.50	0.56	0.40
Median number of units	3.9	3.0	4.2	5.3	6.0	3.2	2.3	1.2	3.8

Figure 8.1. Self-reported alcohol consumption in women in the Health Survey for England 2015

The Health Survey for England 2015 covered areas related to alcohol including frequency of drinking in the last 12 months, frequency of drinking different types of drink, number of drinking days per week and maximum amount drunk on any day in the last week. Data presented here is estimated weekly alcohol consumption, by age group

Taken from Health Survey for England 2015.[224] Reproduced under the terms of the Open Government Licence v3.0

8.4.2 Body mass index and waist circumference

The mean BMI in women in the HSE was 29.1 kg/m^2 in the 55-64-year-old group and 28.7 kg/m^2 in the 65-74-year-old group. The mean BMI in my study group was similar at 26.4 kg/m^2. The HSE prevalence of (measured) normal BMI, overweight and obesity was 18%, 44% and 37%, respectively in the 55-65-year-old group, and 22%, 44% and 33%, respectively in the 65-74-year-old group. These estimates contrast with the findings in my study population (45%, 37% and 19%, respectively). Although the larger prevalence of normal BMI may

be related to case ascertainment in that UKCTOCS participants have been shown to be generally healthier, the HSE is also a voluntary study. The lower rate of obesity in my study population compared to HSE is difficult to explain.

The mean WC in females in the HSE was 91.7cm in the 55-64-year-old group and 93cm in the 6-74-year-old group. These means approximate to UK sizes 16 and 18, respectively. The median SS in my study group was 14.

8.4.3 Smoking

HSE reported a prevalence of current smokers of 14% in the 55-64-year-old group and 11% in the 65-74-year-old group. This is much lower than the 36% reported in my group. This to some extent may relate to excluding those with no response to the smoking question in my analysis (which was 9% of the participants in England in UKCTOCS).

8.4.4 Hypertension and diabetes

The prevalence of measured hypertension in HSE (defined as systolic blood pressure ≥140 mmHg or diastolic blood pressure of ≥90 mmHg, or taking medication for hypertension) was 23%, compared to a higher prevalence of self-reported hypertension in my study cohort of 32%.

Prevalence of hypertension increased with increasing BMI in my study group (22%, 35% and 40% in normal BMI, overweight and obesity, respectively), compared to 18%, 24% and 37%, respectively in all females in HSE. Prevalence of hypertension in the HSE increased with increasing WC. WC was categorised as <80cm, 80-88cm and >88 cm (the cut off of 80cm was selected by HSE due to the observation of increased metabolic complications associated

with obesity in a Dutch population study).[285] Prevalence of hypertension was 16%, 20% and 31%, respectively.

Prevalence of diabetes (defined as HbA1c ≥48 mM or known diabetes) in females in HSE was 7%, similar to the 5% prevalence seen in my study group. Prevalence of diabetes increased with increasing BMI in the HSE (3%, 5% and 14% with normal BMI, overweight and obesity, respectively), similar to the pattern seen in my study group (2%, 3% and 13%, respectively).

8.4.5 Summary

Overall, the baseline characteristics of my study group are broadly similar to the women of similar age in the HSE. This suggests that the findings in my sample could be applied to women of similar age in the general population.

8.5 Future research

This work has demonstrated the value and rewards of collaboration. UKCTOCS continues to collect data and extension of the follow up of the UKCTOCS cohort may provide further granularity in this dataset, and ultimately more knowledge translatable to clinical practice. The existing data output could be further enriched by examining outcomes using different data definitions. I have already retrieved outcome data for a number of codes not used in this thesis, and as commented by Ratib *et al*,[24] incidence of liver disease varies substantially depending on data definitions. Collaboration with investigators of different datasets could allow direct comparison and collation of data to increase understanding of liver disease epidemiology in a larger, more representative sample. Measurement of other biochemical parameters in samples from

UKCTOCS, e. g. ALT, AST and platelet count, would allow calculation of simple non-invasive markers of liver fibrosis to explore their role in this population.

The discussion above outlines the potential benefits to patients and society of early intervention and early identification in liver disease. I have demonstrated the clinical utility of a simple measure (skirt size) and an easy to use blood test (ELF test) which could be incorporated in to existing strategies for stratification for NAFLD and brief intervention for alcohol. However, with BAFLD being an important diagnosis in this demographic, strategies to incorporate both risks should be considered. My work has shown that the ELF test accurately predicts outcomes in these individuals, making it well-placed to play a part. In addition, when considering simple strategies to take place in community settings, compared to FibroScan which requires training and expensive equipment and is associated with high failure rate, the ELF test is an attractive option.

8.5.1 Screening for liver disease

Screening can be defined as;

> "the application of a test to detect a potential disease in an individual who has no signs of symptoms of the disease, in order to detect the disease earlier".

Screening for a disease is appropriate if;

- The prevalence of the disease is high among the population targeted for screening
- The screening occurs at a point where intervention can be effective.

In the case of liver disease, the pre-clinical phase is long and therefore does offer a period of time to intervene.

A screening test should be;

- Simple
- Inexpensive
- Valid and reproducible[286]

My work has shown the potential use of the ELF test in those with defined risk factors. This leads to the question of the role of the ELF test as a screening tool in the general population and could again be investigated within the UKCTOCS cohort. This proposed study would evaluate the role of ELF in identifying those at risk of liver disease. The low risk cohort would be defined as participants self-reporting low alcohol use and normal BMI. The ELF test would be measured in those who subsequently experienced a liver-related event, and in matched controls (no risk factors and no liver-related event) (figure 8.2). Using a study design similar to the study in chapter 7, this would be a case-control study defined by the PRoBE criteria,[252] and statistical analysis would again comprise time-dependent Cox models to measure the performance of ELF to predict events with time.

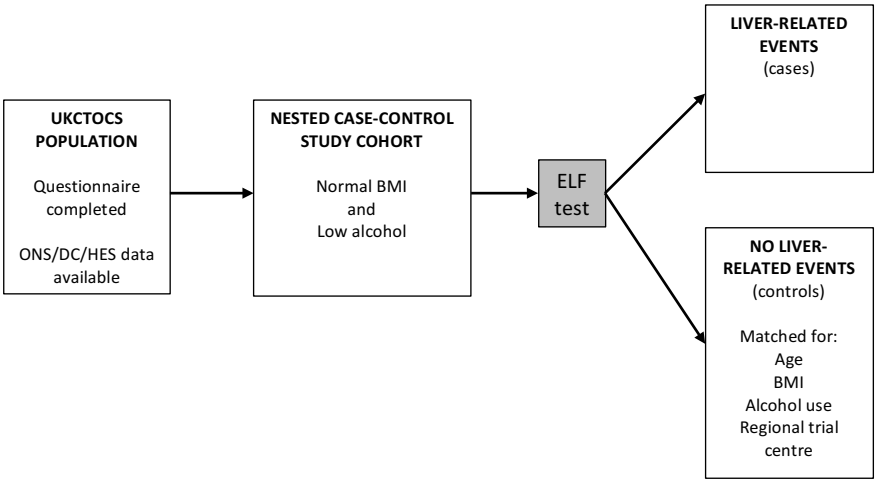


Figure 8.2. Study design to evaluate performance of the ELF test in low risk participants, to evaluate its role as a screening tool
The proposed study is nested within UKCTOCS, comprising participants with BMI <25 kg/m^2 and self-reported alcohol consumption <16 units / week. ELF test would be performed and participants who experience LRE during follow up compared to matched controls who do not experience an LRE
BMI, body mass index; DC, death certification; ELF, enhanced liver fibrosis; HES, Hospital Episode Statistics; LRE, liver-related event; ONS, Office for National Statistics

Exploration of the ELF test in a low risk population would provide information on the population screening value of the ELF test. However, a two-stage stratification pathway may represent a more cost-effective strategy to target this demographic (figure 8.3). This proposal is based on an effective public health message regarding the risks of liver disease in middle-aged women, highlighting the main risks, alcohol and obesity, and incorporates the concept of combining tests described in chapter four, in particular the strategy of employing, in a serial fashion, several tests with an initial test which is inexpensive. As previously discussed, BMI is not easy for individuals to measure (and may not be as predictive as a measure of waist or hip circumference). A simple self-stratification guide could include combining skirt

size and the AUDIT tool to inform individuals about a need to consider lifestyle changes. A subsequent targeted stratification could comprise assessment of alcohol use and BMI to assess risk and for those at high risk, an ELF test would stratify those at risk of future liver disease who may benefit from enhanced lifestyle intervention, including a multidisciplinary approach to managing metabolic risk factors. This pathway could be cost effective, with a low cost self-stratification, an intermediate cost targeted stratification with the ELF test, with a high cost intervention in those at highest risk. The algorithm could be based within an at-risk population as a targeted screening strategy, as the initial test, skirt size, as a screening tool is simple and inexpensive, although would need evaluating for validity and reproducibility. Qualitative work to inform future service development would be required in the first instance, in conjunction with initiatives to increase the awareness of liver disease in the community, in particular NAFLD, and the availability of the diagnostic tools.

Any screening strategy would need evaluation, by measuring outcomes, for example cause-specific (e.g. liver-related) outcomes. A case-control study could be employed, where individuals with and without the disease are compared with respect to whether they underwent screening or not. However, a randomised controlled trial would be the gold standard design. This would reduce volunteer bias as after participants agree to enter the study, those that are screened are selected at random.

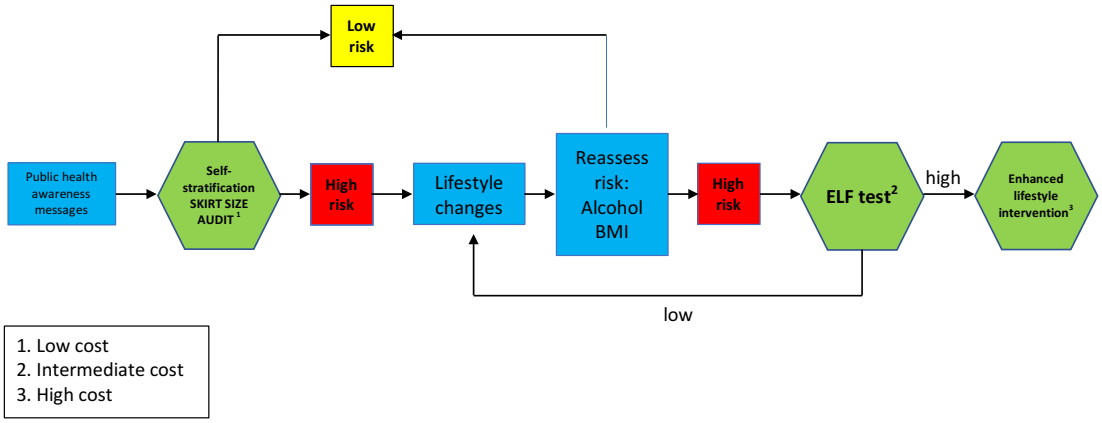


Figure 8.3. Proposed two-stage liver disease risk stratification pathway for middle-aged women in the community
The first stage of the algorithm comprises risk stratification using skirt size and the AUDIT questionnaire. Those screened as high risk will be given lifestyle advice and reassessed using BMI and an alcohol questionnaire. Those at high risk would be subjected to the ELF test to identify those with or without liver disease.
AUDIT, Alcohol Use Disorders Identification Test; BMI, body mass index; ELF, enhanced liver fibrosis

8.6 Overall summary

Non-invasive markers of liver fibrosis have become established in clinical practice and have the potential to revolutionise the assessment of liver disease, particularly in the field of screening and in other settings where it has been difficult to perform liver fibrosis assessment, for example in primary care. Although there are a number of deficiencies in the information provided by non-invasive markers compared to liver biopsy, for example in the assessment of other features of liver disease including inflammation, serum markers are safe, reliable tests that may be the key to screening for liver disease in the general population. The ability to repeat testing with non-invasive markers will allow clinicians to monitor those patients with or at risk of liver disease.

I have shown that the ELF test has the potential to fill this unmet need. There is a pressing need to provide health care professionals with an easy to use and

easily interpretable test that supports decision making and, for example, can provide the necessary stratification tool for general practitioners to use when deciding whether to refer patients to secondary care. In a climate where appropriate use of resources is paramount, my work shows the benefit of serum markers in clinical practice and their potential to improve individual patient care whilst having a positive health economic impact.

Alcohol and fat are leading drivers of chronic liver disease and require particular attention when addressing the burden of liver disease. The risk to health from obesity and heavy alcohol use, particularly in combination, in postmenopausal women has been highlighted in this thesis and I have shown how the ELF test could be used to identify those who are more likely to develop liver disease and its complications amongst those with risk factors. General practitioners are able to identify individuals with risk factors, for example those attending primary care diabetes services, but require a tool to identify individuals from these higher risk groups who are at the highest risk of developing liver disease. My work provides a platform from which work in this area can be developed. Although more work is needed to correlate non-invasive markers of liver fibrosis with clinical endpoints, as we aim to move fibrosis assessment beyond specialist environments, they offer a clinical tool that is easily used, interpretable and understood in a wider clinical setting.

8.7 Overall conclusions

In this thesis, I present novel data on the performance of the Enhanced Liver Fibrosis test, providing data to show that it accurately detects liver fibrosis in

patients with chronic hepatitis B, and also that it has a similar performance to an alternative modality, transient elastography.

I have added clarity to the epidemiology of liver disease in postmenopausal women, defining the incidence of liver disease and elucidating the contribution of two important risk factors for liver disease, alcohol and overweight and obesity. For the first time, the association between skirt size and liver disease has been described, offering the opportunity to use this simple tool in public health messages about liver disease.

I have demonstrated the performance of the ELF test to predict clinically significant liver-related outcomes in those women with risk factors for liver disease.

This work has outlined the building blocks that could be used to create a pathway for stratifying individuals at risk of liver disease in the community. Both early detection of liver disease and appropriate risk stratification will be the keys to managing the growing burden of liver disease to ensure safe clinical care of those with or at risk of liver disease and effective and efficient utilisation of resources. I hope that the data presented in this thesis may contribute to initiatives that support general practitioners to incorporate stratification of liver disease risk in to prediction algorithms of patients with metabolic and other risk factors.

REFERENCES

1. Tortora GJ, Grabowski SR. Principles of Anatomy and Physiology (8th ed.): John Wiley & Sons; 1996.

2. Schuppan D, Afdhal NH. Liver cirrhosis. *Lancet* 2008; **371**(9615): 838-51.

3. Perz JF, Armstrong GL, Farrington LA, Hutin YJ, Bell BP. The contributions of hepatitis B virus and hepatitis C virus infections to cirrhosis and primary liver cancer worldwide. *J Hepatol* 2006; **45**(4): 529-38.

4. Rosenberg WM. Rating fibrosis progression in chronic liver diseases. *J Hepatol* 2003; **38**(3): 357-60.

5. Pinzani, M. Genetic predisposition to hepatic fibrosis. In: Genetic in the diagnosis and treatment of patients with chronic liver injury (pre-meeting course); Italian Association for the Study of the Liver: 2009. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.434.2884&rep=rep1&type=pdf - page=77 Accessed 27 February 2019.

6. Piano S, Tonon M, Angeli P. Management of ascites and hepatorenal syndrome. *Hepatol Int* 2018; **12**(Suppl 1): 122-34.

7. Planas R, Montoliu S, Balleste B, et al. Natural history of patients hospitalized for management of cirrhotic ascites. *Clin Gastroenterol Hepatol* 2006; **4**(11): 1385-94.

8. Moore KP, Aithal GP. Guidelines on the management of ascites in cirrhosis. *Gut* 2006; **55 Suppl 6**: vi1-12.

9. Tripathi D, Stanley AJ, Hayes PC, et al. U.K. guidelines on the management of variceal haemorrhage in cirrhotic patients. *Gut* 2015; **64**(11): 1680-704.

10. Bajaj JS. Review article: the modern management of hepatic encephalopathy. *Aliment Pharmacol Ther* 2010; **31**(5): 537-47.

11. Ferenci P, Lockwood A, Mullen K, Tarter R, Weissenborn K, Blei AT. Hepatic encephalopathy--definition, nomenclature, diagnosis, and quantification: final report of the working party at the 11th World Congresses of Gastroenterology, Vienna, 1998. *Hepatology* 2002; **35**(3): 716-21.

12. White DL, Kanwal F, El-Serag HB. Association between nonalcoholic fatty liver disease and risk for hepatocellular cancer, based on systematic review. *Clin Gastroenterol Hepatol* 2012; **10**(12): 1342-59 e2.

13. Fattovich G, Stroffolini T, Zagni I, Donato F. Hepatocellular carcinoma in cirrhosis: incidence and risk factors. *Gastroenterology* 2004; **127**(5 Suppl 1): S35-50.

14. Wong CR, Nguyen MH, Lim JK. Hepatocellular carcinoma in patients with non-alcoholic fatty liver disease. *World J Gastroenterol* 2016; **22**(37): 8294-303.

15. Llovet JM, Bru C, Bruix J. Prognosis of hepatocellular carcinoma: the BCLC staging classification. *Semin Liver Dis* 1999; **19**(3): 329-38.

16. Schmucker DL, Sanchez H. Liver regeneration and aging: a current perspective. *Curr Gerontol Geriatr Res* 2011; **2011**: 526379.

17. Kim IH, Kisseleva T, Brenner DA. Aging and liver disease. *Curr Opin Gastroenterol* 2015; **31**(3): 184-91.

18. Park SH, Jeon WK, Kim SH, et al. Prevalence and risk factors of non-alcoholic fatty liver disease among Korean adults. *J Gastroenterol Hepatol* 2006; **21**(1 Pt 1): 138-43.

19. Amarapurkar D, Kamani P, Patel N, et al. Prevalence of non-alcoholic fatty liver disease: population based study. *Ann Hepatol* 2007; **6**(3): 161-3.

20. Noureddin M, Yates KP, Vaughn IA, et al. Clinical and histological determinants of nonalcoholic steatohepatitis and advanced fibrosis in elderly patients. *Hepatology* 2013; **58**(5): 1644-54.

21. Frith J, Day CP, Henderson E, Burt AD, Newton JL. Non-alcoholic fatty liver disease in older people. *Gerontology* 2009; **55**(6): 607-13.

22. Williams R, Aspinall R, Bellis M, et al. Addressing liver disease in the UK: a blueprint for attaining excellence in health care and reducing premature mortality from lifestyle issues of excess consumption of alcohol, obesity, and viral hepatitis. *Lancet* 2014; **384**(9958): 1953-97.

23. International Classification of Diseases (ICD). World Health Organization. Available at: http://www.who.int/classifications/icd/en/.

24. Ratib S, West J, Fleming KM. Liver cirrhosis in England-an observational study: are we measuring its burden occurrence correctly? *BMJ Open* 2017; **7**(7): e013752.

25. Ratib S, West J, Crooks CJ, Fleming KM. Diagnosis of liver cirrhosis in England, a cohort study, 1998-2009: a comparison with cancer. *Am J Gastroenterol* 2014; **109**(2): 190-8.

26. Leon DA, McCambridge J. Liver cirrhosis mortality rates in Britain from 1950 to 2002: an analysis of routine data. *Lancet* 2006; **367**(9504): 52-6.

27. Jepsen P, Vilstrup H, Sorensen HT. Alcoholic cirrhosis in Denmark - population-based incidence, prevalence, and hospitalization rates between 1988 and 2005: a descriptive cohort study. *BMC Gastroenterol* 2008; **8**: 3.

28. Thampanitchawong P, Piratvisuth T. Liver biopsy: complications and risk factors. *World J Gastroenterol* 1999; **5**(4): 301-4.

29. Gilmore IT, Burroughs A, Murray-Lyon IM, Williams R, Jenkins D, Hopkins A. Indications, methods, and outcomes of percutaneous liver biopsy in England and Wales: an audit by the British Society of Gastroenterology and the Royal College of Physicians of London. *Gut* 1995; **36**(3): 437-41.

30. Myers RP, Fong A, Shaheen AA. Utilization rates, complications and costs of percutaneous liver biopsy: a population-based study including 4275 biopsies. *Liver Int* 2008; **28**(5): 705-12.

31. Regev A, Berho M, Jeffers LJ, et al. Sampling error and intraobserver variation in liver biopsy in patients with chronic HCV infection. *Am J Gastroenterol* 2002; **97**(10): 2614-8.

32. Rousselet MC, Michalak S, Dupre F, et al. Sources of variability in histological scoring of chronic viral hepatitis. *Hepatology* 2005; **41**(2): 257-64.

33. Ratziu V, Charlotte F, Heurtier A, et al. Sampling variability of liver biopsy in nonalcoholic fatty liver disease. *Gastroenterology* 2005; **128**(7): 1898-906.

34. Goodman ZD. Grading and staging systems for inflammation and fibrosis in chronic liver diseases. *J Hepatol* 2007; **47**(4): 598-607.

35. Ishak K, Baptista A, Bianchi L, et al. Histological grading and staging of chronic hepatitis. *J Hepatol* 1995; **22**(6): 696-9.

36. Scheuer PJ. Classification of chronic viral hepatitis: a need for reassessment. *J Hepatol* 1991; **13**(3): 372-4.

37. Bedossa P, Poynard T. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. *Hepatology* 1996; **24**(2): 289-93.

38. Brunt EM, Janney CG, Di Bisceglie AM, Neuschwander-Tetri BA, Bacon BR. Nonalcoholic steatohepatitis: a proposal for grading and staging the histological lesions. *Am J Gastroenterol* 1999; **94**(9): 2467-74.

39. 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): 1313-21.

40. Asselah T, Bieche I, Sabbagh A, et al. Gene expression and hepatitis C virus infection. *Gut* 2009; **58**(6): 846-58.

41. McLernon DJ, Donnan PT, Ryder S, et al. Health outcomes following liver function testing in primary care: a retrospective cohort study. *Fam Pract* 2009; **26**(4): 251-9.

42. Skelly MM, James PD, Ryder SD. Findings on liver biopsy to investigate abnormal liver function tests in the absence of diagnostic serology. *J Hepatol* 2001; **35**(2): 195-9.

43. Mofrad P, Contos MJ, Haque M, et al. Clinical and histologic spectrum of nonalcoholic fatty liver disease associated with normal ALT values. *Hepatology* 2003; **37**(6): 1286-92.

44. Johnston DE. Special considerations in interpreting liver function tests. *Am Fam Physician* 1999; **59**(8): 2223-30.

45. Ratib S, Fleming KM, Crooks CJ, Aithal GP, West J. 1 and 5 year survival estimates for people with cirrhosis of the liver in England, 1998-2009: a large population study. *J Hepatol* 2014; **60**(2): 282-9.

46. Scaglione S, Kliethermes S, Cao G, et al. The Epidemiology of Cirrhosis in the United States: A Population-based Study. *J Clin Gastroenterol* 2015; **49**(8): 690-6.

47. Chin JL, Pavlides M, Moolla A, Ryan JD. Non-invasive Markers of Liver Fibrosis: Adjuncts or Alternatives to Liver Biopsy? *Front Pharmacol* 2016; **7**: 159.

48. Adams LA, Angulo P, Lindor KD. Nonalcoholic fatty liver disease. *CMAJ* 2005; **172**(7): 899-905.

49. Ekstedt M, Franzen LE, Mathiesen UL, et al. Long-term follow-up of patients with NAFLD and elevated liver enzymes. *Hepatology* 2006; **44**(4): 865-73.

50. Day CP, James OF. Steatohepatitis: a tale of two "hits"? *Gastroenterology* 1998; **114**(4): 842-5.

51. Buzzetti E, Pinzani M, Tsochatzis EA. The multiple-hit pathogenesis of non-alcoholic fatty liver disease (NAFLD). *Metabolism* 2016; **65**(8): 1038-48.

52. Vilar-Gomez E, Martinez-Perez Y, Calzadilla-Bertot L, et al. Weight Loss Through Lifestyle Modification Significantly Reduces Features of Nonalcoholic Steatohepatitis. *Gastroenterology* 2015; **149**(2): 367-78 e5; quiz e14-5.

53. Lassailly G, Caiazzo R, Buob D, et al. Bariatric Surgery Reduces Features of Nonalcoholic Steatohepatitis in Morbidly Obese Patients. *Gastroenterology* 2015; **149**(2): 379-88; quiz e15-6.

54. Sanyal AJ, Chalasani N, Kowdley KV, et al. Pioglitazone, vitamin E, or placebo for nonalcoholic steatohepatitis. *N Engl J Med* 2010; **362**(18): 1675-85.

55. Miller ER, 3rd, Pastor-Barriuso R, Dalal D, Riemersma RA, Appel LJ, Guallar E. Meta-analysis: high-dosage vitamin E supplementation may increase all-cause mortality. *Ann Intern Med* 2005; **142**(1): 37-46.

56. Ratziu V, Harrison SA, Francque S, et al. Elafibranor, an Agonist of the Peroxisome Proliferator-Activated Receptor-alpha and -delta, Induces Resolution of Nonalcoholic Steatohepatitis Without Fibrosis Worsening. *Gastroenterology* 2016; **150**(5): 1147-59 e5.

57. Neuschwander-Tetri BA, Loomba R, Sanyal AJ, et al. Farnesoid X nuclear receptor ligand obeticholic acid for non-cirrhotic, non-alcoholic

steatohepatitis (FLINT): a multicentre, randomised, placebo-controlled trial. *Lancet* 2015; **385**(9972): 956-65.

58. Friedman SL, Ratziu V, Harrison SA, et al. A randomized, placebo-controlled trial of cenicriviroc for treatment of nonalcoholic steatohepatitis with fibrosis. *Hepatology* 2018; **67**(5): 1754-67.

59. Loomba R, Lawitz E, Mantry PS, et al. The ASK1 inhibitor selonsertib in patients with nonalcoholic steatohepatitis: A randomized, phase 2 trial. *Hepatology* 2017.

60. Sebastiani G, Gkouvatsos K, Plebani M. Non-invasive assessment of liver fibrosis: it is time for laboratory medicine. *Clin Chem Lab Med* 2011; **49**(1): 13-32.

61. Ratziu V, Massard J, Charlotte F, et al. Diagnostic value of biochemical markers (FibroTest-FibroSURE) for the prediction of liver fibrosis in patients with non-alcoholic fatty liver disease. *BMC Gastroenterology* 2006; **6**: 6.

62. Adams LA, Bulsara M, Rossi E, et al. Hepascore: an accurate validated predictor of liver fibrosis in chronic hepatitis C infection. *Clin Chem* 2005; **51**(10): 1867-73.

63. Angulo P, Hui JM, Marchesini G, et al. The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD. *Hepatology* 2007; **45**(4): 846-54.

64. Boursier J, Vergniol J, Guillet A, et al. Diagnostic accuracy and prognostic significance of blood fibrosis tests and liver stiffness measurement by FibroScan in non-alcoholic fatty liver disease. *J Hepatol* 2016; **65**(3): 570-8.

65. Shah AG, Lydecker A, Murray K, Tetri BN, Contos MJ, Sanyal AJ. Comparison of noninvasive markers of fibrosis in patients with nonalcoholic fatty liver disease. *Clin Gastroenterol Hepatol* 2009; **7**(10): 1104-12.

66. Adams LA, George J, Bugianesi E, et al. Complex non-invasive fibrosis models are more accurate than simple models in non-alcoholic fatty liver disease. *J Gastroenterol Hepatol* 2011; **26**(10): 1536-43.

67. Petta S, Wong VW, Camma C, et al. Serial combination of non-invasive tools improves the diagnostic accuracy of severe liver fibrosis in patients with NAFLD. *Aliment Pharmacol Ther* 2017; **46**(6): 617-27.

68. Vernon G, Baranova A, Younossi ZM. Systematic review: the epidemiology and natural history of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis in adults. *Aliment Pharmacol Ther* 2011; **34**(3): 274-85.

69. Tanwar S, Trembling PM, Guha IN, et al. Validation of terminal peptide of procollagen III for the detection and assessment of nonalcoholic steatohepatitis in patients with nonalcoholic fatty liver disease. *Hepatology* 2013; **57**(1): 103-11.

70. Sørensen TI, Orholm M, Bentsen KD, Høybye G, Eghøje K, Christoffersen P. Prospective evaluation of alcohol abuse and alcoholic liver injury in men as predictors of development of cirrhosis. *Lancet* 1984; **324**: 241-4.

71. Reid MC, Fiellin DA, O'Connor PG. Hazardous and harmful alcohol consumption in primary care. *Arch Intern Med* 1999; **159**(15): 1681-9.

72. Parkes J, Guha IN, Harris S, Rosenberg WM, Roderick PJ. Systematic review of the diagnostic performance of serum markers of liver fibrosis in alcoholic liver disease. *Comp Hepatol* 2012; **11**(1): 5.

73. Rosenberg WMC, Voelker M, Thiel R, et al. Serum markers detect the presence of liver fibrosis: A cohort study. *Gastroenterology* 2004; **127**(6): 1704-13.

74. Halfon P, Bourliere M, Penaranda G, et al. Accuracy of hyaluronic acid level for predicting liver fibrosis stages in patients with hepatitis C virus. *Comp Hepatol* 2005; **4**: 6.

75. Bhasin S, He EJ, Kawakubo M, et al. N-terminal propeptide of type III procollagen as a biomarker of anabolic response to recombinant human GH and testosterone. *J Clin Endocrinol Metab* 2009; **94**(11): 4224-33.

76. Arthur MJ, Iredale JP, Mann DA. Tissue inhibitors of metalloproteinases: role in liver fibrosis and alcoholic liver disease. *Alcohol Clin Exp Res* 1999; **23**(5): 940-3.

77. ADVIA Centaur Enhanced Liver Fibrosis (ELF) Test Specification Sheet. Siemens Healthineers. Available at: https://www.healthcare.siemens.co.uk/clinical-specialities/liver-disease/elf-test-now-avail.

78. Guha IN, Parkes J, Roderick P, et al. Noninvasive markers of fibrosis in nonalcoholic fatty liver disease: Validating the European Liver Fibrosis Panel and exploring simple markers. *Hepatology* 2007; **47**(2): 455-60.

79. Mayo MJ, Parkes J, Adams-Huet B, et al. Prediction of clinical outcomes in primary biliary cirrhosis by serum enhanced liver fibrosis assay. *Hepatology* 2008; **48**(5): 1549-57.

80. Nobili V, Parkes J, Bottazzo G, et al. Performance of ELF Serum Markers in Predicting Fibrosis Stage in Pediatric Non-Alcoholic Fatty Liver Disease. *Gastroenterology* 2009; **136**(1): 160-7.

81. Parkes J, Guha IN, Roderick P, et al. Enhanced Liver Fibrosis (ELF) test accurately identifies liver fibrosis in patients with chronic hepatitis C. *J Viral Hepat* 2011; **18**(1): 23-31.

82. Vesterhus M, Hov JR, Holm A, et al. Enhanced liver fibrosis score predicts transplant-free survival in primary sclerosing cholangitis. *Hepatology* 2015; **62**(1): 188-97.

83. Parkes J, Roderick P, Harris S, et al. Enhanced liver fibrosis test can predict clinical outcomes in patients with chronic liver disease. *Gut* 2010; **59**(9): 1245-51.

84. Puigvehi M, Hernandez J, Broquetas T, et al. Diagnostic Accuracy of the Enhanced Liver Fibrosis (ELF(R)) Score Using HCV-Infected Serum Samples Cryopreserved for up to 25 Years. *PLoS One* 2016; **11**(12): e0164883.

85. Peters MG, Bacchetti P, Boylan R, et al. Enhanced liver fibrosis marker as a noninvasive predictor of mortality in HIV/hepatitis C virus-coinfected women from a multicenter study of women with or at risk for HIV. *AIDS* 2016; **30**(5): 723-9.

86. Fagan KJ, Pretorius CJ, Horsfall LU, et al. ELF score >/=9.8 indicates advanced hepatic fibrosis and is influenced by age, steatosis and histological activity. *Liver Int* 2015; **35**(6): 1673-81.

87. Irvine KM, Wockner LF, Shanker M, et al. The Enhanced liver fibrosis score is associated with clinical outcomes and disease progression in patients with chronic liver disease. *Liver Int* 2016; **36**(3): 370-7.

88. Loo WM, Goh GB, Wang Y, et al. Enhanced Liver Fibrosis Score as a Predictor of Hepatocellular Carcinoma. *Clin Chem* 2018; **64**(9): 1404-5.

89. Hametner S, Ferlitsch A, Ferlitsch M, et al. The VITRO Score (Von Willebrand Factor Antigen/Thrombocyte Ratio) as a New Marker for Clinically Significant Portal Hypertension in Comparison to Other Non-Invasive Parameters of Fibrosis Including ELF Test. *PLoS One* 2016; **11**(2): e0149230.

90. Sandahl TD, McGrail R, Moller HJ, et al. The macrophage activation marker sCD163 combined with markers of the Enhanced Liver Fibrosis (ELF) score predicts clinically significant portal hypertension in patients with cirrhosis. *Aliment Pharmacol Ther* 2016; **43**(11): 1222-31.

91. Mauro E, Crespo G, Montironi C, et al. Portal pressure and liver stiffness measurements in the prediction of fibrosis regression after sustained virological response in recurrent hepatitis C. *Hepatology* 2018; **67**(5): 1683-94.

92. Lopez IC, Aroca FG, Bernal MDF, et al. Utility of the ELF Test for Detecting Steatohepatitis in Morbid Obese Patients with Suspicion of Nonalcoholic Fatty Liver Disease. *Obes Surg* 2017; **27**(9): 2347-53.

93. Crespo G, Fernandez-Varo G, Marino Z, et al. ARFI, FibroScan, ELF, and their combinations in the assessment of liver fibrosis: a prospective study. *J Hepatol* 2012; **57**(2): 281-7.

94. Degos F, Perez P, Roche B, et al. Diagnostic accuracy of FibroScan and comparison to liver fibrosis biomarkers in chronic viral hepatitis: A multicenter prospective study (the FIBROSTIC study). *J Hepatol* 2010; **53**(6): 1013-21.

95. Thiele M, Madsen BS, Hansen JF, Detlefsen S, Antonsen S, Krag A. Accuracy of the Enhanced Liver Fibrosis Test vs Fibrotest, Elastography and Indirect Markers in Detection of Advanced Fibrosis in Patients with Alcoholic Liver Disease. *Gastroenterology* 2018.

96. Tanwar S, Trembling PM, Hogan BJ, et al. Noninvasive markers of liver fibrosis: on-treatment changes of serum markers predict the outcome of antifibrotic therapy. *Eur J Gastroenterol Hepatol* 2017; **29**(3): 289-96.

97. Srivastava A, Gailer R, Demma S, et al. Primary care sequential use of FIB-4 and the enhanced liver fibrosis test to stratify patients with nonalcoholic fatty liver acide doubles cirrhosis detection and reduces referrals of patients with mild disease [abstract]. *J Hepatol* 2016; **64 (Suppl 1)**: S474-S5.

98. Patel P, Hossain F, Horsfall LU, et al. A Pragmatic Approach Identifies a High Rate of Nonalcoholic Fatty Liver Disease With Advanced Fibrosis in Diabetes Clinics and At-Risk Populations in Primary Care. *Hepatol Commun* 2018; **2**(8): 893-905.

99. Ragazzo TG, Paranagua-Vezozzo D, Lima FR, et al. Accuracy of transient elastography-FibroScan(R), acoustic radiation force impulse (ARFI)

imaging, the enhanced liver fibrosis (ELF) test, APRI, and the FIB-4 index compared with liver biopsy in patients with chronic hepatitis C. *Clinics (Sao Paulo)* 2017; **72**(9): 516-25.

100. Harkisoen S, Boland GJ, van den Hoek JA, van Erpecum KJ, Hoepelman AI, Arends JE. ELF-test less accurately identifies liver cirrhosis diagnosed by liver stiffness measurement in non-Asian women with chronic hepatitis B. *J Clin Virol* 2014; **61**(4): 503-8.

101. Yoo EJ, Kim BK, Kim SU, et al. Normal enhanced liver fibrosis (ELF) values in apparently healthy subjects undergoing a health check-up and in living liver donors in South Korea. *Liver Int* 2013; **33**(5): 706-13.

102. Lichtinghagen R, Pietsch D, Bantel H, Manns MP, Brand K, Bahr MJ. The Enhanced Liver Fibrosis (ELF) score: normal values, influence factors and proposed cut-off values. *J Hepatol* 2013; **59**(2): 236-42.

103. Duffy MJ, Sturgeon CM, Soletormos G, et al. Validation of new cancer biomarkers: a position statement from the European group on tumor markers. *Clin Chem* 2015; **61**(6): 809-20.

104. Pepe MS, Etzioni R, Feng Z, et al. Phases of biomarker development for early detection of cancer. *J Natl Cancer Inst* 2001; **93**(14): 1054-61.

105. Kennedy OJ, Parkes J, Tanwar S, Trembling PM, Rosenberg WM. The Enhanced Liver Fibrosis (ELF) panel: analyte stability under common sample storage conditions used in clinical practice. *J Appl Lab Med* 2017; **1(6)**: 720-8.

106. Dellavance A, Fernandes F, Shimabokuro N, et al. Enhanced liver fibrosis (ELF) score: Analytical performance and distribution range in a large cohort of blood donors. *Clin Chim Acta* 2016; **461**: 151-5.

107. Jabor A, Kubicek Z, Frankova S, Senkerikova R, Franekova J. Enhanced liver fibrosis (ELF) score: Reference ranges, biological variation in healthy subjects, and analytical considerations. *Clin Chim Acta* 2018; **483**: 291-5.

108. National Institute for Health and Care Excellence (2016) Non-alcoholic fatty liver disease (NAFLD): assessment and management (NICE Guideline 49). Available at: https://www.nice.org.uk/guidance/ng49/resources.

109. Enhanced Liver fibrosis (ELF) test to Uncover Cirrhosis as an Indication for Diagnosis and Action for Treatable Events. ISRCTN74815110. Available at: http://www.isrctn.com/ISRCTN74815110.

110. National Institute for Health and Care Excellence (2016) Cirrhosis in over 16s: assessment and management (NICE Guideline 50). Available at: https://www.nice.org.uk/guidance/ng50.

111. Harris R, Harman DJ, Card TR, Aithal GP, Guha IN. Prevalence of clinically significant liver disease within the general population, as defined by non-invasive markers of liver fibrosis: a systematic review. *Lancet Gastroenterol Hepatol* 2017; **2**(4): 288-97.

112. Critical Appraisal Skills Programme (2017). CASP Systematic Review Checklist. Available at: http://www.casp-uk.net/checklists Accessed: 3 October 2017.

113. Crombie, IK. The Pocket Guide to Critical Appraisal. London: BMJ Publishing Group; 1996.

114. Center for Evidence Based Management (July, 2014), Critical Appraisal Checklist for a Meta-Analysis or Systematic Review. Available at: https://www.cebma.org/ Accessed 20 March 2017.

115. Srivastava A, Gailer R, Tanwar S, et al. A one year retrospective review of new patient attendances at a tertiary hepatology centre highlighting the increasing challenge of NAFLD and the need to develop clinical pathways [abstract]. *J Hepatol* 2015; **62(Suppl 2)**: S740.

116. Roulot D, Roudot-Thoraval F, G NK, et al. Concomitant screening for liver fibrosis and steatosis in French type 2 diabetic patients using Fibroscan. *Liver Int* 2017; **37**(12): 1897-906.

117. Harman DJ, Ryder SD, James MW, et al. Direct targeting of risk factors significantly increases the detection of liver cirrhosis in primary care: a cross-sectional diagnostic study utilising transient elastography. *BMJ Open* 2015; **5**(4): e007516.

118. Caballeria L, Pera G, Arteaga I, et al. Early detection of chronic liver disease with fibrosis among presumed healthy adults using transient elastography (TE). A population-based study [abstract]. *Hepatology* 2016; **64 (Supplement 1)**.

119. Patel PJ, Stuart KA, Horsfall LU, et al. Enhanced Liver Fibrosis test and liver stiffness measurements identify high prevalence of advanced fibrosis in

people with type 2 diabetes [abstract]. *J Gastroenterol Hepatol* 2017; **32 (Supplement 2)**: 103.

120. Doycheva I, Cui J, Nguyen P, et al. Non-invasive screening of diabetics in primary care for NAFLD and advanced fibrosis by MRI and MRE. *Aliment Pharmacol Ther* 2016; **43**(1): 83-95.

121. Harman DJ, Ryder SD, James MW, et al. Obesity and type 2 diabetes are important risk factors underlying previously undiagnosed cirrhosis in general practice: a cross-sectional study using transient elastography. *Aliment Pharmacol Ther* 2018; **47**(4): 504-15.

122. Cheng PN, Chiu YC, Chiu HC, Chien SC. The Application of Liver Stiffness Measurement in Residents Without Overt Liver Diseases Through a Community-Based Screening Program. *Medicine (Baltimore)* 2016; **95**(12): e3193.

123. Morling JR, Fallowfield JA, Guha IN, et al. Clinically significant chronic liver disease in people with Type 2 diabetes: the Edinburgh Type 2 Diabetes Study. *QJM* 2016; **109**(4): 249-56.

124. Chen Y, Xu M, Wang T, et al. Advanced fibrosis associates with atherosclerosis in subjects with nonalcoholic fatty liver disease. *Atherosclerosis* 2015; **241**(1): 145-50.

125. Koehler EM, Plompen EP, Schouten JN, et al. Presence of diabetes mellitus and steatosis is associated with liver stiffness in a general population: The Rotterdam study. *Hepatology* 2016; **63**(1): 138-47.

126. Conti F, Vukotic R, Foschi FG, et al. Transient elastography in healthy subjects and factors influencing liver stiffness in non-alcoholic fatty liver disease: An Italian community-based population study. *Dig Liver Dis* 2016; **48**(11): 1357-63.

127. Donato F, Gelatti U, Tagger A, et al. Intrahepatic cholangiocarcinoma and hepatitis C and B virus infection, alcohol intake, and hepatolithiasis: a case-control study in Italy. *Cancer Causes Control* 2001; **12**(10): 959-64.

128. Hsing AW, Zhang M, Rashid A, et al. Hepatitis B and C virus infection and the risk of biliary tract cancer: a population-based study in China. *Int J Cancer* 2008; **122**(8): 1849-53.

129. EASL Clinical Practice Guidelines: management of chronic hepatitis B. *J Hepatol* 2009; **50**(2): 227-42.

130. Lok ASF, McMahon BJ. Chronic hepatitis B: Update 2009. *Hepatology* 2009; **50**(3): 661-2.

131. Poynard T, Ngo Y, Munteanu M, Thabut D, Ratziu V. Noninvasive Markers of Hepatic Fibrosis in Chronic Hepatitis B. *Curr Hepat Rep* 2011; **10**(2): 87-97.

132. Viganò M, Paggi S, Lampertico P, et al. Dual cut-off transient elastography to assess liver fibrosis in chronic hepatitis B: a cohort study with internal validation. *Aliment Pharmacol Ther* 2011; **34**(3): 353-62.

133. Raftopoulos SC, George J, Bourliere M, et al. Comparison of noninvasive models of fibrosis in chronic hepatitis B. *Hepatol Int* 2011; **6**(2): 457-67.

134. Bonnard P, Sombie R, Lescure FX, et al. Comparison of Elastography, Serum Marker Scores, and Histology for the Assessment of Liver Fibrosis in Hepatitis B Virus (HBV)-Infected Patients in Burkina Faso. *Am J Trop Med Hyg* 2010; **82**(3): 454-8.

135. Marcellin P, Ziol M, Bedossa P, et al. Non-invasive assessment of liver fibrosis by stiffness measurement in patients with chronic hepatitis B. *Liver Int* 2009; **29**(2): 242-7.

136. Sporea I, Sirli R, Deleanu A, et al. Liver stiffness measurements in patients with HBV vs HCV chronic hepatitis: a comparative study. *World J Gastroenterol* 2010; **16**(38): 4832-7.

137. Lesmana CRA, Salim S, Hasan I, et al. Diagnostic accuracy of transient elastography (FibroScan) versus the aspartate transaminase to platelet ratio index in assessing liver fibrosis in chronic hepatitis B: The role in primary care setting. *J Clin Pathol* 2011; **64**(10): 916-20.

138. Ogawa E, Furusyo N, Murata M, et al. Longitudinal assessment of liver stiffness by transient elastography for chronic hepatitis B patients treated with nucleoside analog. *Hepatol Res* 2011; **41**(12): 1178-88.

139. Mbaye PS, Sarr A, Sire J-M, et al. Liver Stiffness Measurement and Biochemical Markers in Senegalese Chronic Hepatitis B Patients with Normal ALT and High Viral Load. *PLoS ONE* 2011; **6**(7): e22291.

140. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988; **44**(3): 837-45.

141. Poynard T, Bedossa P, Opolon P. Natural history of liver fibrosis progression in patients with chronic hepatitis C. The OBSVIRC, METAVIR, CLINIVIR, and DOSVIRC groups. *Lancet* 1997; **349**(9055): 825-32.

142. Fattovich G, Bortolotti F, Donato F. Natural history of chronic hepatitis B: special emphasis on disease progression and prognostic factors. *J Hepatol* 2008; **48**(2): 335-52.

143. Rigamonti C, Donato MF, Fraquelli M, et al. Transient elastography predicts fibrosis progression in patients with recurrent hepatitis C after liver transplantation. *Gut* 2008; **57**(6): 821-7.

144. Poynard T, Halfon P, Castera L, et al. Standardization of ROC curve areas for diagnostic evaluation of liver fibrosis markers based on prevalences of fibrosis stages. *Clin Chem* 2007; **53**(9): 1615-22.

145. Poynard T, Ngo Y, Marcellin P, Hadziyannis S, Ratziu V, Benhamou Y. Impact of adefovir dipivoxil on liver fibrosis and activity assessed with biochemical markers (FibroTest-ActiTest) in patients infected by hepatitis B virus. *J Viral Hepat* 2009; **16**(3): 203-13.

146. Obuchowski NA. Estimating and Comparing Diagnostic Tests' Accuracy When the Gold Standard Is Not Binary. *Acad Radiol* 2005; **12**(9): 1198-204.

147. Lambert J, Halfon P, Penaranda G, Bedossa P, Cacoub P, Carrat F. How to Measure the Diagnostic Accuracy of Noninvasive Liver Fibrosis Indices: The Area Under the ROC Curve Revisited. *Clin Chem* 2008; **54**(8): 1372-8.

148. Carrión JA, Fernández–Varo G, Bruguera M, et al. Serum Fibrosis Markers Identify Patients With Mild and Progressive Hepatitis C Recurrence After Liver Transplantation. *Gastroenterology* 2010; **138**(1): 147-58.e1.

149. Martinez SM, Fernández-Varo G, González P, et al. Assessment of liver fibrosis before and after antiviral therapy by different serum marker panels in patients with chronic hepatitis C. *Aliment Pharmacol Ther* 2011; **33**(1): 138-48.

150. Gumusay O, Ozenirler S, Atak A, et al. Diagnostic potential of serum direct markers and non-invasive fibrosis models in patients with chronic hepatitis B. *Hepatol Res* 2013; **43**(3): 228-37.

151. Cardoso AC, Carvalho-Filho RJ, Stern C, et al. Direct comparison of diagnostic performance of transient elastography in patients with chronic hepatitis B and chronic hepatitis C. *Liver Int* 2012; **32**(4): 612-21.

152. Musso G, Gambino R, Cassader M, Pagano G. Meta-analysis: Natural history of non-alcoholic fatty liver disease (NAFLD) and diagnostic accuracy of non-invasive tests for liver disease severity. *Ann Med* 2011; **43**(8): 617-49.

153. Castera L, Foucher J, Bernard PH, et al. Pitfalls of liver stiffness measurement: a 5-year prospective study of 13,369 examinations. *Hepatology* 2010; **51**(3): 828-35.

154. Oliveri F, Coco B, Ciccorossi P, et al. Liver stiffness in the hepatitis B virus carrier: a non-invasive marker of liver disease influenced by the pattern of transaminases. *World J Gastroenterol* 2008; **14**(40): 6154-62.

155. Chan HL, Wong GL, Choi PC, et al. Alanine aminotransferase-based algorithms of liver stiffness measurement by transient elastography (Fibroscan) for liver fibrosis in chronic hepatitis B. *J Viral Hepat* 2009; **16**(1): 36-44.

156. Verveer C, Zondervan PE, ten Kate FJ, Hansen BE, Janssen HL, de Knegt RJ. Evaluation of transient elastography for fibrosis assessment compared with large biopsies in chronic hepatitis B and C. *Liver Int* 2012; **32**(4): 622-8.

157. Fraquelli M, Rigamonti C, Casazza G, et al. Etiology-related determinants of liver stiffness values in chronic viral hepatitis B or C. *J Hepatol* 2011; **54**(4): 621-8.

158. Fraquelli M, F. B. The role of transient elastography in patients with hepatitis B viral disease. *Dig Liver Dis* 2011; **43**(Suppl 1): S25-31.

159. Fraquelli M, Rigamonti C, Casazza G, et al. Reproducibility of transient elastography in the evaluation of liver fibrosis in patients with chronic liver disease. *Gut* 2007; **56**(7): 968-73.

160. Kim BK, Kim HS, Park JY, et al. Prospective Validation of ELF Test in Comparison with Fibroscan and FibroTest to Predict Liver Fibrosis in Asian Subjects with Chronic Hepatitis B. *PLoS ONE* 2012; **7**(7): e41964.

161. Wahl K, Rosenberg W, Vaske B, et al. Biopsy-controlled liver fibrosis staging using the enhanced liver fibrosis (ELF) score compared to transient elastography. *PLoS One* 2012; **7**(12): e51906.

162. Wong GL, Chan HL, Choi PC, et al. Non-invasive algorithm of enhanced liver fibrosis and liver stiffness measurement with transient elastography for

advanced liver fibrosis in chronic hepatitis B. *Aliment Pharmacol Ther* 2014; **39**(2): 197-208.

163. Heo JY, Kim BK, Park JY, et al. Combination of Transient Elastography and an Enhanced Liver Fibrosis Test to Assess the Degree of Liver Fibrosis in Patients with Chronic Hepatitis B. *Gut Liver* 2018; **12**(2): 190-200.

164. Kim BK, Kim HS, Yoo EJ, et al. Risk assessment of clinical outcomes in Asian patients with chronic hepatitis B using enhanced liver fibrosis test. *Hepatology* 2014; **60**(6): 1911-9.

165. Palaniyappan N, Cox E, Bradley C, et al. Non-invasive assessment of portal hypertension using quantitative magnetic resonance imaging. *J Hepatol* 2016; **65**(6): 1131-9.

166. Crespo G, Gambato M, Millan O, et al. Early non-invasive selection of patients at high risk of severe hepatitis C recurrence after liver transplantation. *Transpl Infect Dis* 2016; **18**(3): 471-9.

167. de Vries EMG, Farkkila M, Milkiewicz P, et al. Enhanced liver fibrosis test predicts transplant-free survival in primary sclerosing cholangitis, a multi-centre study. *Liver Int* 2017; **37**(10): 1554-61.

168. Lejealle C, Castera L. Non-invasive fibrosis testing in patients with chronic hepatitis B. *Curr Hepatol Rep* 2018; **17**(4): 492-501.

169. Trembling PM, Lampertico P, Parkes J, et al. Performance of Enhanced Liver Fibrosis test and comparison with transient elastography in the identification of liver fibrosis in patients with chronic hepatitis B infection. *J Viral Hepat* 2014; **21**(6): 430-8.

170. Liu B, Balkwill A, Reeves G, Beral V. Body mass index and risk of liver cirrhosis in middle aged UK women: prospective study. *BMJ* 2010; **340**: c912.

171. Hart CL, Batty GD, Morrison DS, Mitchell RJ, Smith GD. Obesity, overweight and liver disease in the Midspan prospective cohort studies. *Int J Obes* 2010; **34**(6): 1051-9.

172. Menon U, Gentry-Maharaj A, Hallett R, et al. Sensitivity and specificity of multimodal and ultrasound screening for ovarian cancer, and stage distribution of detected cancers: results of the prevalence screen of the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS). *Lancet Oncol* 2009; **10**(4): 327-40.

173. Menon U, Gentry-Maharaj A, Ryan A, et al. Recruitment to multicentre trials--lessons from UKCTOCS: descriptive study. *BMJ* 2008; **337**: a2079.

174. Menon U, Ryan A, Kalsi J, et al. Risk Algorithm Using Serial Biomarker Measurements Doubles the Number of Screen-Detected Cancers Compared With a Single-Threshold Rule in the United Kingdom Collaborative Trial of Ovarian Cancer Screening. *J Clin Oncol* 2015; **33**(18): 2062-71.

175. Jacobs IJ, Menon U, Ryan A, et al. Ovarian cancer screening and mortality in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial. *Lancet* 2016; **387**(10022): 945-56.

176. Drinkaware. Alcohol unit guidelines. Drinkaware, London. Available at: http://www.drinkaware.co.uk/.

177. Burnell M, Gentry-Maharaj A, Ryan A, et al. Impact on mortality and cancer incidence rates of using random invitation from population registers for recruitment to trials. *Trials* 2011; **12**: 61.

178. Bolton-Smith C, Woodward M, Tunstall-Pedoe H, Morrison C. Accuracy of the estimated prevalence of obesity from self reported height and weight in an adult Scottish population. *J Epidemiol Community Health* 2000; **54**(2): 143-8.

179. Hill A, Roberts J. Body mass index: a comparison between self-reported and measured height and weight. *J Public Health Med* 1998; **20**(2): 206-10.

180. Nieto-Garcia FJ, Bush TL, Keyl PM. Body mass definitions of obesity: sensitivity and specificity using self-reported weight and height. *Epidemiology* 1990; **1**(2): 146-52.

181. Roberts RJ. Can self-reported data accurately describe the prevalence of overweight? *Public Health* 1995; **109**(4): 275-84.

182. Kuskowska-Wolk A, Karlsson P, Stolt M, Rossner S. The predictive validity of body mass index based on self-reported weight and height. *Int J Obes* 1989; **13**(4): 441-53.

183. Midanik L. The validity of self-reported alcohol consumption and alcohol problems: a literature review. *Br J Addict* 1982; **77**(4): 357-82.

184. Romelsjo A, Leifman H, Nystrom S. A comparative study of two methods for the measurement of alcohol consumption in the general population. *Int J Epidemiol* 1995; **24**(5): 929-36.

185. Parker DR, Derby CA, Usner DW, Gonzalez S, Lapane KL, Carleton RA. Self-reported alcohol intake using two different question formats in southeastern New England. *Int J Epidemiol* 1996; **25**(4): 770-4.

186. Goddard E. Estimating alcohol consumption from survey data: updated method of converting volumes to units. National Statistics Methodological Series No. 37. Office of National Statistics. London, HMSO. 2007.

187. Pollock DA, Boyle CA, DeStefano F, Moyer LA, Kirk ML. Underreporting of alcohol-related mortality on death certificates of young US Army veterans. *JAMA* 1987; **258**(3): 345-8.

188. Pinsky PF, Miller A, Kramer BS, et al. Evidence of a healthy volunteer effect in the prostate, lung, colorectal, and ovarian cancer screening trial. *Am J Epidemiol* 2007; **165**(8): 874-81.

189. Public Health England. Hepatitis C in the UK 2015 report. London: Public Health England, 2015. July 2015.

190. Ajmera VH, Terrault NA, Harrison SA. Is moderate alcohol use in nonalcoholic fatty liver disease good or bad? A critical review. *Hepatology* 2017; **65**(6): 2090-9.

191. Hagstrom H, Nasr P, Ekstedt M, et al. Low to moderate lifetime alcohol consumption is associated with less advanced stages of fibrosis in non-alcoholic fatty liver disease. *Scand J Gastroenterol* 2017; **52**(2): 159-65.

192. Sogabe M, Okahisa T, Taniguchi T, et al. Light alcohol consumption plays a protective role against non-alcoholic fatty liver disease in Japanese men with metabolic syndrome. *Liver Int* 2015; **35**(6): 1707-14.

193. Davies MJ, Baer DJ, Judd JT, Brown ED, Campbell WS, Taylor PR. Effects of moderate alcohol intake on fasting insulin and glucose concentrations and insulin sensitivity in postmenopausal women: a randomized controlled trial. *JAMA* 2002; **287**(19): 2559-62.

194. Askgaard G, Gronbaek M, Kjaer MS, Tjonneland A, Tolstrup JS. Alcohol drinking pattern and risk of alcoholic liver cirrhosis: a prospective cohort study. *J Hepatol* 2015; **62**(5): 1061-7.

195. Ioannou GN, Weiss NS, Kowdley KV, Dominitz JA. Is obesity a risk factor for cirrhosis-related death or hospitalization? A population-based cohort study. *Gastroenterology* 2003; **125**(4): 1053-9.

196. Hart CL, Morrison DS, Batty GD, Mitchell RJ, Davey Smith G. Effect of body mass index and alcohol consumption on liver disease: analysis of data from two prospective cohort studies. *BMJ* 2010; **340**: c1240.

197. Naveau S, Giraud V, Borotto E, Aubert A, Capron F, Chaput JC. Excess weight risk factor for alcoholic liver disease. *Hepatology* 1997; **25**(1): 108-11.

198. Yi SW, Hong JS, Yi JJ, Ohrr H. Impact of alcohol consumption and body mass index on mortality from nonneoplastic liver diseases, upper aerodigestive tract cancers, and alcohol use disorders in Korean older middle-aged men: Prospective cohort study. *Medicine (Baltimore)* 2016; **95**(39): e4876.

199. Gronbaek M, Deis A, Sorensen TI, Becker U, Schnohr P, Jensen G. Mortality associated with moderate intakes of wine, beer, or spirits. *BMJ* 1995; **310**(6988): 1165-9.

200. Klatsky AL, Armstrong MA, Friedman GD. Alcohol and mortality. *Ann Intern Med* 1992; **117**(8): 646-54.

201. Shaper AG, Wannamethee G, Walker M. Alcohol and mortality in British men: explaining the U-shaped curve. *Lancet* 1988; **2**(8623): 1267-73.

202. Gronbaek M, Deis A, Sorensen TI, et al. Influence of sex, age, body mass index, and smoking on alcohol intake and mortality. *BMJ* 1994; **308**(6924): 302-6.

203. Becker U, Gronbaek M, Johansen D, Sorensen TI. Lower risk for alcohol-induced cirrhosis in wine drinkers. *Hepatology* 2002; **35**(4): 868-75.

204. Becker U, Deis A, Sorensen TI, et al. Prediction of risk of liver disease by alcohol intake, sex, and age: a prospective population study. *Hepatology* 1996; **23**(5): 1025-9.

205. Dunn W, Xu R, Schwimmer JB. Modest wine drinking and decreased prevalence of suspected nonalcoholic fatty liver disease. *Hepatology* 2008; **47**(6): 1947-54.

206. Dunn W, Sanyal AJ, Brunt EM, et al. Modest alcohol consumption is associated with decreased prevalence of steatohepatitis in patients with non-alcoholic fatty liver disease (NAFLD). *J Hepatol* 2012; **57**(2): 384-91.

207. Hayashi PH, Harrison SA, Torgerson S, Perez TA, Nochajski T, Russell M. Cognitive lifetime drinking history in nonalcoholic fatty liver disease: some cases may be alcohol related. *Am J Gastroenterol* 2004; **99**(1): 76-81.

208. Ruhl CE, Everhart JE. Joint effects of body weight and alcohol on elevated serum alanine aminotransferase in the United States population. *Clin Gastroenterol Hepatol* 2005; **3**(12): 1260-8.

209. Loomba R, Bettencourt R, Barrett-Connor E. Synergistic association between alcohol intake and body mass index with serum alanine and aspartate aminotransferase levels in older adults: the Rancho Bernardo Study. *Aliment Pharmacol Ther* 2009; **30**(11-12): 1137-49.

210. Loomba R, Yang HI, Su J, Brenner D, Iloeje U, Chen CJ. Obesity and alcohol synergize to increase the risk of incident hepatocellular carcinoma in men. *Clin Gastroenterol Hepatol* 2010; **8**(10): 891-8, 8 e1-2.

211. Loomba R, Yang HI, Su J, et al. Synergism between obesity and alcohol in increasing the risk of hepatocellular carcinoma: a prospective cohort study. *Am J Epidemiol* 2013; **177**(4): 333-42.

212. Saunders JB, Walters JR, Davies AP, Paton A. A 20-year prospective study of cirrhosis. *Br Med J (Clin Res Ed)* 1981; **282**(6260): 263-6.

213. Ludviksdottir D, Skulason H, Jakobsson F, et al. Epidemiology of liver cirrhosis morbidity and mortality in Iceland. *Eur J Gastroenterol Hepatol* 1997; **9**(1): 61-6.

214. Gunnarsdottir SA, Olsson R, Olafsson S, et al. Liver cirrhosis in Iceland and Sweden: incidence, aetiology and outcomes. *Scand J Gastroenterol* 2009; **44**(8): 984-93.

215. Fleming K, Aithal G, Solaymanidodaran M, Card T, West J. Incidence and prevalence of cirrhosis in the United Kingdom, 1992–2001: A general population-based study. *J Hepatol* 2008; **49**(5): 732-8.

216. Fialla AD, Schaffalitzky de Muckadell OB, Lassen AT. Incidence, etiology and mortality of cirrhosis: a population-based cohort study. *Scand J Gastroenterol* 2012; **47**(6): 702-9.

217. James PT, Leach R, Kalamara E, Shayeghi M. The worldwide obesity epidemic. *Obes Res* 2001; **9 Suppl 4**: 228S-33S.

218. Fourkala EO, Burnell M, Cox C, et al. Association of skirt size and postmenopausal breast cancer risk in older women: a cohort study within the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS). *BMJ Open* 2014; **4**(9): e005400.

219. Han TS, Gates E, Truscott E, Lean ME. Clothing size as an indicator of adiposity, ischaemic heart disease and cardiovascular risks. *J Hum Nutr Diet* 2005; **18**(6): 423-30.

220. British Standards Institution. BS 3666:1982 Specification for size designation of women's wear. London.

221. Donato GB, Fuchs SC, Oppermann K, Bastos C, Spritzer PM. Association between menopause status and central adiposity measured at different cutoffs of waist circumference and waist-to-hip ratio. *Menopause* 2006; **13**(2): 280-5.

222. Garcia-Tsao G, Friedman S, Iredale J, Pinzani M. Now there are many (stages) where before there was one: In search of a pathophysiological classification of cirrhosis. *Hepatology* 2010; **51**(4): 1445-9.

223. Diehl AM. Obesity and alcoholic liver disease. *Alcohol* 2004; **34**(1): 81-7.

224. Health and Social Care Information Centre. Health Survey for England 2015: Adult alcohol consumption. HSIC, 2016.

225. Department of Health. Adult alcohol statistics from the National Drug Treatment Monitoring System (NDTMS) 1 April 2013 to 31 March 2014. London: Department of Health; 2014.

226. Shah T, Palaskas N, Ahmed A. An update on gender disparities in coronary heart disease care. *Curr Atheroscler Rep* 2016; **18**(5): 28.

227. Massart F, Reginster JY, Brandi ML. Genetics of menopause-associated diseases. *Maturitas* 2001; **40**(2): 103-16.

228. Klair JS, Yang JD, Abdelmalek MF, et al. A longer duration of estrogen deficiency increases fibrosis risk among postmenopausal women with nonalcoholic fatty liver disease. *Hepatology* 2016; **64**(1): 85-91.

229. Dahl AK, Hassing LB, Fransson EI, Pedersen NL. Agreement between self-reported and measured height, weight and body mass index in old age--a longitudinal study with 20 years of follow-up. *Age Ageing* 2010; **39**(4): 445-51.

230. Stevens J, Keil JE, Waid LR, Gazes PC. Accuracy of current, 4-year, and 28-year self-reported body weight in an elderly population. *Am J Epidemiol* 1990; **132**(6): 1156-63.

231. Dahl AK, Reynolds CA. Accuracy of recalled body weight--a study with 20-years of follow-up. *Obesity (Silver Spring)* 2013; **21**(6): 1293-8.

232. Perry GS, Byers TE, Mokdad AH, Serdula MK, Williamson DF. The validity of self-reports of past body weights by U.S. adults. *Epidemiology* 1995; **6**(1): 61-6.

233. Must A, Willett WC, Dietz WH. Remote recall of childhood height, weight, and body build by elderly subjects. *Am J Epidemiol* 1993; **138**(1): 56-64.

234. Office for National Statistics. Live births in England and Wales by characteristics of mother 1, 2010. ONS, London. Available at: http://www.ons.gov.uk/ons/rel/vsob1/characteristics-of-Mother-1--england-and-wales/2010/index.html.

235. Aydinoğlu NZ, Krishna A. Imagining thin: Why vanity sizing works. *J Consumer Psychol* 2012; **22**: 575-2.

236. Howard SJ, Davies SC. Chief medical officer urges action to tackle overweight and obesity. *BMJ* 2014; **348**: g2375.

237. Li L, Liu DW, Yan HY, Wang ZY, Zhao SH, Wang B. Obesity is an independent risk factor for non-alcoholic fatty liver disease: evidence from a meta-analysis of 21 cohort studies. *Obes Rev* 2016; **17**(6): 510-9.

238. Loomis AK, Kabadi S, Preiss D, et al. Body mass index and risk of nonalcoholic fatty liver disease: two electronic health record prospective studies. *J Clin Endocrinol Metab* 2016; **101**(3): 945-52.

239. Xu C, Yu C, Ma H, Xu L, Miao M, Li Y. Prevalence and risk factors for the development of nonalcoholic fatty liver disease in a nonobese Chinese population: the Zhejiang Zhenhai Study. *Am J Gastroenterol* 2013; **108**(8): 1299-304.

240. Hughes LA, Schouten LJ, Goldbohm RA, van den Brandt PA, Weijenberg MP. Self-reported clothing size as a proxy measure for body size. *Epidemiology* 2009; **20**(5): 673-6.

241. Battram DS, Beynon C, He M. The reliability and validity of using clothing size as a proxy for waist circumference measurement in adults. *Appl Physiol Nutr Metab* 2011; **36**(2): 183-90.

242. Huxley R, Mendis S, Zheleznyakov E, Reddy S, Chan J. Body mass index, waist circumference and waist:hip ratio as predictors of cardiovascular risk--a review of the literature. *Eur J Clin Nutr* 2010; **64**(1): 16-22.

243. Cheong KC, Ghazali SM, Hock LK, et al. The discriminative ability of waist circumference, body mass index and waist-to-hip ratio in identifying metabolic syndrome: Variations by age, sex and race. *Diabetes Metab Syndr* 2015; **9**(2): 74-8.

244. Seo DC, Choe S, Torabi MR. Is waist circumference >/=102/88cm better than body mass index >/=30 to predict hypertension and diabetes development

regardless of gender, age group, and race/ethnicity? Meta-analysis. *Prev Med* 2017; **97**: 100-8.

245. Lean ME, Han TS, Morrison CE. Waist circumference as a measure for indicating need for weight management. *BMJ* 1995; **311**(6998): 158-61.

246. Lahti-Koski M, Harald K, Mannisto S, Laatikainen T, Jousilahti P. Fifteen-year changes in body mass index and waist circumference in Finnish adults. *Eur J Cardiovasc Prev Rehabil* 2007; **14**(3): 398-404.

247. Gearon E, Tanamas SK, Stevenson C, Loh VHY, Peeters A. Changes in waist circumference independent of weight: Implications for population level monitoring of obesity. *Prev Med* 2018; **111**: 378-83.

248. International Diabetes Federation (2006). The IDF consensus worldwide definition of the metabolic syndrome. Available at: https://www.idf.org/e-library/consensus-statements/60-idfconsensus-worldwide-definitionof-the-metabolic-syndrome.html.

249. Expert Panel on Detection EToHBCiA. Executive summary of The Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). *JAMA* 2001; **285**(19): 2486-97.

250. Health Survey for England 2011. The Information Centre for Health and Social Care. 2012.

251. Day J, Patel P, Parkes J, Rosenberg W. Derivation and performance of standardized Enhanced Liver Fibrosis (ELF) test thresholds for the detection

and prognosis of liver fibrosis. *J Appl Lab Med* 2018. DOI: 10.1373/jalm.2018.027359 (Epub ahead of print).

252. Pepe MS, Feng Z, Janes H, Bossuyt PM, Potter JD. Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. *J Natl Cancer Inst* 2008; **100**(20): 1432-8.

253. Fisher LD, Lin DY. Time-dependent covariates in the Cox proportional-hazards regression model. *Annu Rev Public Health* 1999; **20**: 145-57.

254. Suissa S. Immortal time bias in pharmaco-epidemiology. *Am J Epidemiol* 2008; **167**(4): 492-9.

255. van Walraven C, Davis D, Forster AJ, Wells GA. Time-dependent bias was common in survival analyses published in leading clinical journals. *J Clin Epidemiol* 2004; **57**(7): 672-82.

256. McLernon DJ, Dillon JF, Sullivan FM, et al. The utility of liver function tests for mortality prediction within one year in primary care using the algorithm for liver function investigations (ALFI). *PLoS ONE* 2012; **7**(12): e50965.

257. Poynard T, Ngo Y, Perazzo H, et al. Prognostic value of liver fibrosis biomarkers: a meta-analysis. *Gastroenterol Hepatol* 2011; **7**(7): 445-54.

258. Angulo P, Bugianesi E, Bjornsson ES, et al. Simple noninvasive systems predict long-term outcomes of patients with nonalcoholic fatty liver disease. *Gastroenterology* 2013; **145**(4): 782-9 e4.

259. Pang JX, Zimmer S, Niu S, et al. Liver stiffness by transient elastography predicts liver-related complications and mortality in patients with chronic liver disease. *PLoS One* 2014; **9**(4): e95776.

260. Nottingham Community Liver Biomarkers Cohort. Available at: https://clinicaltrials.gov/ct2/show/NCT02037867.

261. Castera L, Forns X, Alberti A. Non-invasive evaluation of liver fibrosis using transient elastography. *J Hepatol* 2008; **48**(5): 835-47.

262. Midanik LT. Validity of self-reported alcohol use: a literature review and assessment. *Br J Addict* 1988; **83**(9): 1019-30.

263. O'Malley KJ, Cook KF, Price MD, Wildes KR, Hurdle JF, Ashton CM. Measuring diagnoses: ICD code accuracy. *Health Serv Res* 2005; **40**(5 Pt 2): 1620-39.

264. Adult drinking habits in Great Britain: 2005 to 2016. Office for National Statistics. 2017. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/drugusealcoholandsmoking/bulletins/opinionsandlifestylesurveyadultdrinkinghabitsingreatbritain/2005to2016.

265. Rao R, Roche A. Substance misuse in older people. *BMJ* 2017; **358**: j3885.

266. Blais P, Husain N, Kramer JR, Kowalkowski M, El-Serag H, Kanwal F. Nonalcoholic fatty liver disease is underrecognized in the primary care setting. *Am J Gastroenterol* 2015; **110**(1): 10-4.

267. Patel PJ, Banh X, Horsfall LU, et al. Underappreciation of non-alcoholic fatty liver disease by primary care clinicians: limited awareness of surrogate markers of fibrosis. *Intern Med J* 2018; **48**(2): 144-51.

268. Romero-Gomez M, Zelber-Sagi S, Trenell M. Treatment of NAFLD with diet, physical activity and exercise. *J Hepatol* 2017; **67**(4): 829-46.

269. Whitlock EP, Orleans CT, Pender N, Allan J. Evaluating primary care behavioral counseling interventions: an evidence-based approach. *Am J Prev Med* 2002; **22**(4): 267-84.

270. National Institute for Health and Care Excellence (2014) Weight management: lifestyle services for overweight or obese adults (Public health guideline PH53). Available at: https://www.nice.org.uk/guidance/PH53/chapter/recommendations - recommendation-4-raise-awareness-of-lifestyle-weight-management-services-among-health-and-social-care-professionals.

271. Shiffman S, Sweeney CT, Pillitteri JL, Sembower MA, Harkins AM, Wadden TA. Weight management advice: what do doctors recommend to their patients? *Prev Med* 2009; **49**(6): 482-6.

272. Foster GD, Wadden TA, Makris AP, et al. Primary care physicians' attitudes about obesity and its treatment. *Obes Res* 2003; **11**(10): 1168-77.

273. Potter MB, Vu JD, Croughan-Minihane M. Weight management: what patients want from their primary care physicians. *J Fam Pract* 2001; **50**(6): 513-8.

274. Aveyard P, Lewis A, Tearne S, et al. Screening and brief intervention for obesity in primary care: a parallel, two-arm, randomised trial. *Lancet* 2016; **388**(10059): 2492-500.

275. Reeves GK, Pirie K, Beral V, et al. Cancer incidence and mortality in relation to body mass index in the Million Women Study: cohort study. *BMJ* 2007; **335**(7630): 1134.

276. Bhaskaran K, Douglas I, Forbes H, dos-Santos-Silva I, Leon DA, Smeeth L. Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5.24 million UK adults. *Lancet* 2014; **384**(9945): 755-65.

277. National Institute for Health and Care Excellence (2011) Alcohol-use disorders: diagnosis, assessment and management of harmful drinking and alcohol dependence (NICE Guideline CG115). Available at: https://www.nice.org.uk/guidance/cg115.

278. Saunders JB, Aasland OG, Babor TF, Fuente JRDL, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption - II. *Addiction* 1993; **88**: 791-804.

279. Kaner E, Bland M, Cassidy P, et al. Effectiveness of screening and brief alcohol intervention in primary care (SIPS trial): pragmatic cluster randomised controlled trial. *BMJ* 2013; **346**: e8501.

280. Frank D, DeBenedetti AF, Volk RJ, Williams EC, Kivlahan DR, Bradley KA. Effectiveness of the AUDIT-C as a screening test for alcohol misuse in three race/ethnic groups. *J Gen Intern Med* 2008; **23**(6): 781-7.

281. Kaner EF, Beyer F, Dickinson HO, et al. Effectiveness of brief alcohol interventions in primary care populations. *Cochrane Database Syst Rev* 2007; (2): CD004148.

282. Kaner EF, Beyer FR, Muirhead C, et al. Effectiveness of brief alcohol interventions in primary care populations. *Cochrane Database Syst Rev* 2018; **2**: CD004148.

283. Westwood G, Meredith P, Atkins S, Greengross P, Schmidt PE, Aspinall RJ. Universal screening for alcohol misuse in acute medical admissions is feasible and identifies patients at high risk of liver disease. *J Hepatol* 2017; **67**(3): 559-67.

284. Sheron N, Moore M, O'Brien W, Harris S, Roderick P. Feasibility of detection and intervention for alcohol-related liver disease in the community: the Alcohol and Liver Disease Detection study (ALDDeS). *Br J Gen Pract* 2013; **63**(615): e698-705.

285. World Health Organization, 2000. Obesity: preventing and managing the global epidemic (No. 894). World Health Organization.

286. Herman CR, Gill HK, Eng J, Fajardo LL. Screening for preclinical disease: test and disease characteristics. *AJR Am J Roentgenol* 2002; **179**(4): 825-31.

APPENDICES

Appendix A
Publications

Appendix B
Search strategies for structured literature review

Appendix C
P values for pairwise comparisons among group means for variables in chapter four

Appendix D
UKCTOCS baseline questionnaire

Appendix E
UKCTOCS follow up questionnaire

Appendix F
ICD-10 codes for studies cited in this thesis. Those marked * were the codes used in my studies

Appendix G
Comparisons between study cohort and cohort comprising remainder of participants in England who returned follow-up questionnaire

Appendix H
Permission to access UKCTOCS samples

APPENDIX A
Publications

Trembling PM, et al. J Viral Hepat. 2014;21:430-8

Reproduced with permission

Trembling PM, et al. BMC Public Health. 2017;17:603

Trembling PM, et al. BMC Public Health. 2018;18:409

JOURNAL OF VIRAL HEPATITIS
JVH

Journal of Viral Hepatitis, 2014, **21**, 430–438 doi:10.1111/jvh.12161

Performance of Enhanced Liver Fibrosis test and comparison with transient elastography in the identification of liver fibrosis in patients with chronic hepatitis B infection

P. M. Trembling,[1] P. Lampertico,[2] J. Parkes,[3] S. Tanwar,[1] M. Viganò,[2,4] F. Facchetti,[2] M. Colombo[2] and W. M. Rosenberg[1] [1]*Institute for Liver and Digestive Health, Division of Medicine, University College London, London, UK;* [2]*AM and A Migliavacca Center for Liver Disease, 1st Division of Gastroenterology, Department of Medicine, Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico, Università degli Studi di Milano, Milan, Italy;* [3]*Public Health Sciences and Medical Statistics, Faculty of Medicine, University of Southampton, Southampton, UK; and* [4]*UO Epatologia, Ospedale San Giuseppe, Università degli Studi di Milano, Milan, Italy*

Received January 2013; accepted for publication July 2013

SUMMARY. Assessment of liver fibrosis is important in determining prognosis, disease progression and need for treatment in patients with chronic hepatitis B (CHB). Limitations to the use of liver biopsy in assessing fibrosis are well recognized, and noninvasive tests are being increasingly evaluated including transient elastography (TE) and serum markers such as the Enhanced Liver Fibrosis (ELF) test. We assessed performance of ELF and TE in detecting liver fibrosis with reference to liver histology in a cohort of patients with CHB ($n = 182$), and compared the performance of these modalities. Median age was 46 and mean AST 70 IU/L. Cirrhosis was reported in 20% of liver biopsies. Both modalities performed well in assessing fibrosis at all stages. Area under receiver operator characteristic (AUROC) curves for detecting METAVIR fibrosis stages $F \geq 1$, $F \geq 2$, $F \geq 3$ and F4 were 0.77, 0.82, 0.80 and 0.83 for ELF and 0.86, 0.86, 0.90 and 0.95 for TE. TE performed significantly better in the assessment of severe fibrosis (AUROC 0.80 for ELF and 0.90 for TE, $P < 0.01$) and cirrhosis (0.83 for ELF and 0.95 for TE, $P < 0.01$). This study demonstrates that ELF has good performance in detection of liver fibrosis in patients with CHB, and when compared, TE performs better in detection of severe fibrosis/cirrhosis.

Keywords: chronic hepatitis B, Enhanced Liver Fibrosis test, liver fibrosis, noninvasive markers, transient elastography.

INTRODUCTION

Chronic hepatitis B (CHB) caused by infection with the hepatitis B virus (HBV) is characterized by periods of continuous or fluctuating inflammation of the liver, leading to fibrosis, which may remain occult, with no clinical signs or symptoms at the time of diagnosis of CHB. Morbidity and mortality in patients with CHB are related to persistence of viral replication and the development of liver fibrosis that may progress to cirrhosis and its complications, particularly portal hypertension and liver cancers including hepatocellular cancer, and an increased risk of intra- and extrahepatic biliary cancer [1,2]. The assessment of liver fibrosis is therefore an essential component in the initial evaluation of patients with CHB and informs the decision to commence antiviral therapy. Liver fibrosis assessment using invasive or noninvasive tests is a key feature of international guidelines [3,4]. Continued monitoring of fibrosis is critical to determine changes in fibrosis over time and to assess the efficacy of therapy and the necessity for interventions to manage portal hypertension and screen for liver cancer and progression to cirrhosis.

The traditional method for assessing liver fibrosis has been needle biopsy of the liver, however this is expensive, frequently painful and potentially hazardous for the patient, and subject to sampling error and variation in interpretation [5,6]. While many patients with CHB can be persuaded to undergo a first biopsy, most will be reluctant to accept subsequent follow-up biopsies to evaluate disease progression or response to treatment. Noninvasive methods of assessing

Abbreviations: ALT, alanine aminotransferase; AST, aspartate transaminase; AUROC, area under receiver operator characteristic curves; CHB, chronic hepatitis B; DANA, difference between advanced and nonadvanced fibrosis stages; ELF, Enhanced Liver Fibrosis test; HA, hyaluronic acid; HBV, hepatitis B virus; NPV, negative predictive value; PPV, positive predictive value; SR, success rate; TE, transient elastography; ULN, upper limit of normal.

Correspondence: Dr. Paul Trembling, Institute for Liver and Digestive Health, Division of Medicine, University College London, Royal Free Hospital, Rowland Hill Street, London, NW3 2PF, UK. E-mail: p.trembling@ucl.ac.uk

The copyright line for this article was changed on 23 July 2014 after original online publication.

liver fibrosis in a range of chronic liver diseases are being explored. Principal among these are transient elastography (TE) and serum markers, and these are now being evaluated in patients with CHB [7–9]. The Enhanced Liver Fibrosis (ELF) test (Siemens Healthcare Diagnostics Inc., Tarrytown, New York, USA) is a panel of biomarkers comprising hyaluronic acid (HA), tissue inhibitor of matrix metalloproteinase-1 (TIMP-1) and aminoterminal propeptide of procollagen type III (PIIINP), derived from studies in patients with a range of chronic liver diseases including CHB [10].

Previous studies comparing the performance of noninvasive markers of liver fibrosis in CHB have reported contradictory results. Performance defined by the area under the receiver operator curve (AUROC) of TE to identify $F \geq 2$ has been reported in several studies to range from 0.61 to 0.87 [11–16].

The aim of this primary study was to evaluate and validate the performance of ELF in a cohort of patients with CHB and to compare ELF to a different noninvasive modality, TE, in the assessment of liver fibrosis defined by histological staging of liver biopsies.

MATERIALS AND METHODS

Patients

Subjects were recruited at a single Italian centre. Among 224 treatment-naïve patients with CHB who were consecutively referred for a liver biopsy and TE evaluation to the Liver Center, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan [8], those with a stored serum sample available for ELF testing were included. Patients with hepatitis C virus, hepatitis delta virus and human immunodeficiency virus coinfections, other concomitant liver diseases, current or previous hepatic decompensation, current or previous antiviral treatment and/or an absolute contraindication to liver biopsy (platelet count $<60 \times 10^9$/L, INR > 1.35) were excluded. In all patients, serum sampling, liver biopsy and TE were performed on the same day. All patients gave their written consent to the study, which was approved by the local ethics committee.

Blood markers

Serum samples were analysed for levels of HA, TIMP-1 and PIIINP using the proprietary assays developed for the ELF test by Siemens Healthcare Diagnostics Inc. These assays are magnetic particle separation immunoassays, and samples were analysed on an ADVIA Centaur® immunoassay system (Siemens Medical Solutions Diagnostics Inc., Tarrytown, NY, USA). Results were entered into the manufacturer's published algorithm to derive an ELF score.

Quantitative polymerase chain reaction amplification for HBV DNA was performed using Amplicor HBV Monitor® (Roche Diagnostics, Branchburg, NJ, USA), and serology for HBeAg status was assessed with standard assays, and serum alanine aminotransferase (ALT) and aspartate transaminase (AST) were measured using standard enzymatic immunoassays.

Liver biopsy

All patients underwent an ultrasound-guided liver biopsy with a semiautomatic modified Menghini system (16G, Bio-Mol, Hospital Service, Pomezia, Italy, Philips iU22, Bothell, WA, USA) to stage severity of hepatitis. All the procedures were carried out by two highly experienced hepatologists. Liver specimens were considered of adequate size if longer than 2 cm. Patients with a smaller specimen underwent a repeat procedure during the same session. Five-micron thick sections of formalin-fixed, paraffin-embedded liver tissue were stained with haematoxylin–eosin and Masson trichrome, and read by a single liver pathologist blind to TE and clinical data. Grading and staging were evaluated according to METAVIR (staging F0 = fibrosis absent, F1 = portal fibrosis without septa, F2 = portal fibrosis with few septa, F3 = severe fibrosis, F4 = cirrhosis) [17].

Transient elastography

After an overnight fast, patients underwent a FibroScan® (Echosens, Paris, France) utilizing a 5-MHz ultrasound transducer probe mounted on the axis of a vibrator that was operated by three experienced hepatologists who were blind to clinical, biochemical and histological data [18,19]. Briefly, mild amplitude and low-frequency vibrations (50 Hz) are transmitted to the liver, thus inducing an elastic shear wave propagating through the underlying liver tissue. Velocity of the wave is directly related to tissue stiffness. The tip of the transducer was covered with a drop of gel and placed perpendicularly in the intercostal space with the patient lying in dorsal decubitus with the right arm in maximal abduction. Under control time motion and A-mode, the operator chose a liver portion within the right liver lobe at least 6 cm thick, free of large vascular structures and gallbladder. Ten successful acquisitions were performed on each patient. The success rate (SR) was calculated as the ratio of the number of successful acquisitions over the total number of acquisitions. The median value, expressed in kPa, was kept as representative of the liver stiffness. The manufacturer recommends that liver stiffness measurements are considered reliable using the following criteria: (i) number of valid acquisitions at least 10, (ii) SR at least 60% and an interquartile range of the median of 30% or less.

Statistical analysis

Statistical analyses were performed using SPSS for Windows (version 19, SPSS Inc, Chicago, IL, USA), Stata Statistical Software (StataCorp 2007. Release 10. College

Station, TX, USA: StataCorp LP) and R (version 2.11.1, R Foundation for Statistical Computing, Vienna, Austria). Median values and interquartile ranges for each diagnostic test were determined for each fibrosis stage. The diagnostic performances of ELF and TE were assessed by deriving the area under receiver operator characteristic (AUROC) curves. AUROC and 95% confidence intervals of AUROC were calculated. Comparisons of AUROC values for ELF and TE were determined for each stage of fibrosis using the DeLong method to calculate the chi-squared value for the comparison and expressed as the significance of difference (*P* value) [20].

Optimal cut-off values for discriminating positive and negative cases at each fibrosis stage for ELF and TE were determined by identifying the point of maximum sensitivity and specificity on the ROC curve, and sensitivity, specificity, positive and negative predictive values (PPV and NPV), and positive and negative likelihood ratios calculated. The clinical utility of each test was evaluated by analysing performance by selecting an upper threshold with high specificity, therefore high PPV to 'rule in' fibrosis and a low threshold with high sensitivity and therefore high NPV to 'rule out' fibrosis.

Logistic regression analysis was conducted to further investigate the relationship both between individual modalities and fibrosis, and within a model combining both ELF and TE.

Recently, several methodological issues have been raised in relation to the application of ROC curve analysis to compare noninvasive tests with liver biopsy. The spectrum effect (the differences in the distributions of fibrosis stages in the sample and reference populations) may result in the performance of a noninvasive test varying between the populations giving rise to apparent differences in performance of tests between different sample populations. In addition, ROC analysis assumes the reference standard to be binary, whereas the METAVIR scoring system employs a five-stage ordinal scale. To overcome these potential flaws, the difference between advanced and nonadvanced (DANA) fibrosis stages [21] and Obuchowski [22] methods of correcting for spectrum effect were applied. The results are presented of applying the Obuchowski measure using previously described penalty functions [23] to correct for the degree of difference between the histological stages ascribed by pathological staging and conversion of ELF test scores.

RESULTS

Of the 224 subjects consecutively recruited, 188 had a stored serum sample. TE acquisition was unsuccessful in six of these subjects (3%); therefore, paired ELF and TE data were available for 182 subjects. Replacing values for missing TE results by both imputation of simple mean and expectation maximization methods did not change the significance of difference between ELF and TE in ROC analysis, therefore, only subjects with paired results were used in the analysis.

Baseline characteristics are shown in Table 1. All patients had a diagnosis of CHB and were treatment-naïve. Median age was 46 years, 71% were male, and 71% were HBeAg negative. 79 (43%) were overweight (BMI $\geq$ 25 kg/m^2). Biopsies reported any fibrosis (METAVIR F $\geq$ 1) in 90.1%, moderate fibrosis (METAVIR F2) in 25.8% and severe fibrosis/cirrhosis (METAVIR F $\geq$ 3, equivalent to Ishak stage 4–6) in 36.8%.

Both ELF and TE discriminated different fibrosis stages well with linear progression (Fig. 1), and both modalities performed well in predicting fibrosis stage. The AUROC for the diagnosis of each stage of fibrosis for ELF and TE is shown in Table 2. The AUROC for the diagnosis of any fibrosis for ELF and TE was 0.77 and 0.86, respectively (P = 0.09). The AUROC for the diagnosis of severe fibrosis/

Table 1 Baseline subject characteristics

		By METAVIR stage				
Characteristic	All subjects	F0	F1	F2	F3	F4
Number of subjects	182	18 (9.9)	50 (27.5)	47 (25.8)	31 (17.0)	36 (19.8)
Age, median (range)	46 (18–67)	32.5 (21–54)	44.0 (18–65)	46 (20–67)	55 (27–65)	50 (29–65)
AST (IU/L), mean (SD)	69.7 (64.1)	47.3 (31.9)	49.2 (31.6)	66.4 (38.5)	86.6 (71.1)	97.1 (105.5)
ALT (IU/L), mean (SD)	110.3 (103.4)	86.4 (78.1)	86.7 (72.0)	110.2 (68.0)	148.1 (167.0)	122.6 (112.4)
HBeAg + (n)	53	7	12	10	12	12
− (n)	129	11	38	37	19	24
HBV DNA, $\log_{10}$ mean	7.96	7.97	7.82	8.07	7.93	7.98

AST, aspartate transaminase; ALT, alanine aminotransferase; SD, standard deviation; HBV, hepatitis B virus.

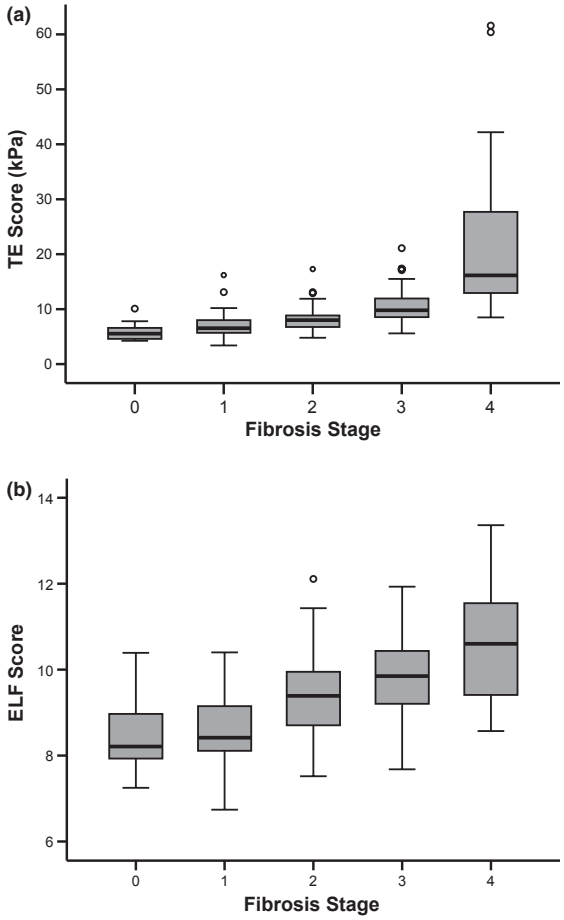


Fig. 1 Box plots showing median and quartiles for (a) TE and (b) ELF scores for diagnosing METAVIR fibrosis stages.

cirrhosis for ELF and TE was 0.80 and 0.90, respectively ($P < 0.01$). The AUROC for the diagnosis of cirrhosis (METAVIR F4) was 0.83 and 0.95, respectively ($P < 0.01$).

Table 3 shows the sensitivities, specificities, predictive values and diagnostic odds ratios of ELF and TE predicting severe fibrosis/cirrhosis and cirrhosis. If two thresholds with high sensitivity and specificity are used to 'rule in' fibrosis (upper threshold with high specificity, therefore high positive predictive value) or 'rule out' fibrosis (lower threshold with high sensitivity, therefore, high negative predictive value), the clinical utility of each modality can be evaluated. For example, using ELF to identify severe fibrosis at data-derived thresholds of 9.08 and 9.94 (sensitivity and specificity of 85%, respectively), 60% of patients would have correctly avoided liver biopsy and 16% would have incorrectly avoided biopsy. 24% would have had an indeterminate result – a value between the thresholds. Using TE to identify severe fibrosis with thresholds of 8.75 and 8.95 (sensitivity and specificity of 85%, respectively)

Table 2 Median scores and diagnostic performance of ELF and TE according to METAVIR fibrosis stage

	ELF score (n = 182)			TE (kPa) (n = 182)			
Fibrosis stage	Median (IQR)	AUROC (95% CI)	Adjusted AUROC	Median (IQR)	AUROC (95% CI)	Adjusted AUROC	P value*
0 *vs* 1–4	8.21 (1.08) *vs* 9.39 (1.81)	0.77 (0.67–0.87)	0.81	5.55 (2.08) *vs* 8.50 (5.93)	0.86 (0.78–0.94)	0.90	0.09
0,1 *vs* 2–4	8.35 (1.13) *vs* 9.82 (1.53)	0.82 (0.76–0.88)	0.86	6.30 (2.47) *vs* 9.80 (6.43)	0.86 (0.80–0.91)	0.89	0.34
0–2 *vs* 3,4	8.75 (1.35) *vs* 10.06 (1.83)	0.80 (0.73–0.87)	0.83	6.90 (2.60) *vs* 13.00 (11.10)	0.90 (0.85–0.95)	0.94	<0.01
0–3 *vs* 4	9.01 (1.61) *vs* 10.60 (2.16)	0.83 (0.76–0.90)	0.86	7.60 (2.93) *vs* 16.15 (14.77)	0.95 (0.91–0.98)	0.96	<0.01

IQR, interquartile range; AUROC, area under receiver operator characteristic curve; CI, confidence interval.
*Significance of comparison of observed ELF and TE AUROC values.

Table 3 Diagnostic performance indices for ELF and TE in the identification of severe fibrosis (F3,4) and cirrhosis (F4) at a range of thresholds

Modality	Threshold	Sensitivity%	Specificity%	PPV%	NPV%	LR +	LR −	DOR
Severe fibrosis (prevalence = 37%)								
ELF	8.02	96	17	40	86	1.10	0.24	4.58
	8.45	93	41	48	90	1.58	0.17	9.29
	8.96	85	56	53	86	1.93	0.27	7.15
	9.39	73	70	58	82	2.43	0.39	6.23
	9.88	60	83	67	78	3.53	0.48	7.35
	10.41	45	95	83	75	9.00	0.58	15.52
TE	6.85	96	50	52	95	1.92	0.08	24.00
	7.70	91	60	57	92	2.28	0.15	15.20
	8.45	88	77	69	92	3.83	0.16	23.94
	9.35	79	87	78	88	6.08	0.24	25.33
	10.15	64	90	80	81	6.40	0.40	16.00
	11.95	57	96	88	79	14.25	0.45	31.67
Cirrhosis (prevalence = 20%)								
ELF	8.61	94	39	28	97	1.54	0.15	10.27
	9.43	72	64	34	90	2.00	0.44	4.55
	9.66	69	72	38	90	2.46	0.43	5.72
	9.99	67	81	47	91	3.53	0.41	8.61
	10.34	61	87	54	90	4.69	0.45	10.42
	10.68	44	95	70	87	8.80	0.59	14.92
TE	9.70	94	80	54	98	4.70	0.08	58.75
	10.30	89	86	62	97	6.36	0.13	48.92
	11.85	83	90	67	96	8.30	0.19	43.68
	12.95	75	92	71	94	9.38	0.27	34.74
	14.15	61	95	74	91	12.20	0.41	29.76
	15.45	50	95	72	88	10.00	0.53	18.87

PPV, positive predictive value; NPV, negative predictive value; LR +, positive likelihood ratio; LR −, negative likelihood ratio; DOR, diagnostic odds ratio.

would have resulted in biopsy correctly being avoided in 82% and incorrectly avoided in 15%, with an indeterminate result in 3%, shown in Fig. 2 and in Table S1. A model for predicting any fibrosis is also shown. At higher sensitivity and specificity, the proportion avoiding biopsy decreases. For example, if sensitivity and specificity thresholds are increased to 90%, the proportion of incorrectly classified cases (i.e. the false positive and false negative rates) substantially decreases to around 10% for both modalities for diagnosis of both severe and any fibrosis. However, this is at the cost of increased proportions of indeterminate cases.

Logistic regression analysis found that in a model combining both modalities, in the prediction of METAVIR $F \geq 1$ and F4, ELF was a nonsignificant predictor. In the prediction of $F \geq 2$ and $F \geq 3$, ELF significantly improved the prediction of fibrosis when combined with TE. Respective ELF and TE odds ratios in the combined models were as follows: 1.45 (95% CI 0.75–2.83) and 1.99 (1.31–3.02), 2.47 (1.55–3.94) and 1.54 (1.25–1.90), 1.61 (1.03–2.51) and 1.55 (1.31–1.83), 1.32 (0.75–2.32) and 1.44 (1.23–1.68) for $F \geq 1$, $F \geq 2$, $F \geq 3$ and F4, respectively (Table S2). Combining the two tests results in AUROC values of 0.87, 0.88, 0.90 and 0.95 for diagnosis of $F \geq 1$, $F \geq 2$, $F \geq 3$ and F4 stages, respectively.

A subanalysis of the performance in HBeAg-negative patients showed similar performance of ELF and TE to that for the whole cohort. AUROC values for ELF and TE for $F \geq 1$, $F \geq 2$, $F \geq 3$ and F4 stages were 0.71, 0.80, 0.79, 0.81 and 0.81, 0.83, 0.90, 0.95, respectively, with a significant difference in performance at $F \geq 3$ and F4.

The effect of ALT on test performance was assessed. Diagnostic accuracy appears to be maintained with both modalities when ALT is 3 or 5 times above the upper limit of normal (ULN). In the diagnosis of severe fibrosis, both modalities maintained their performance in all categories of ALT. The AUROC values indicate that in the diagnosis of any fibrosis, ELF is less accurate when ALT is below the ULN compared with when ALT is above the ULN, and accuracy of TE improves when ALT is below the ULN. The 95% confidence interval for ELF in diagnosing any fibrosis is very large in this small cohort. When ALT is above 3 or

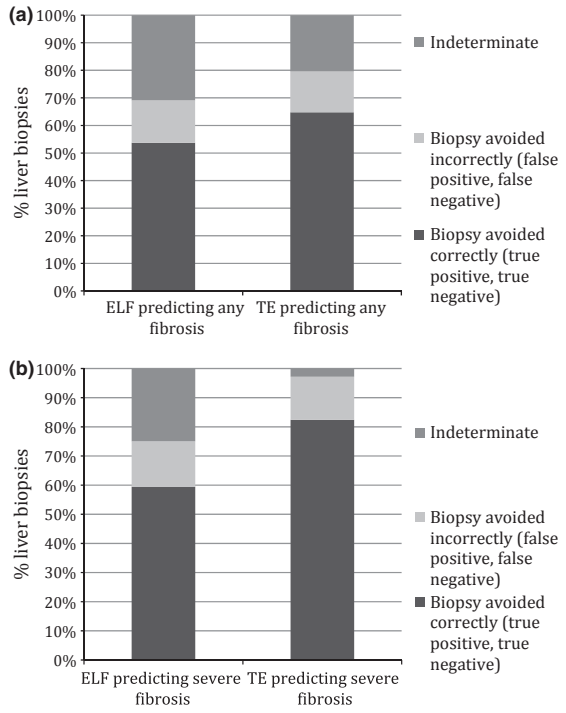


Fig. 2 Clinical utility model for ELF and TE predicting (a) any fibrosis and (b) severe fibrosis with sensitivity and specificity of 85%.

5 times the ULN, diagnostic accuracy appears to be maintained with both modalities (Table S3).

DISCUSSION

This study has demonstrated that the ELF test accurately assesses liver fibrosis severity in patients with CHB. Comparison of TE and ELF demonstrated good performance of both modalities, with TE performing significantly better in the identification of severe fibrosis/cirrhosis.

The ELF test has been validated in external disease-specific cohorts of patients with nonalcoholic fatty liver disease, primary biliary cirrhosis and chronic hepatitis C [24–29]. It predicts liver-related outcomes at 7 years at least as well as biopsy, with a unit change in ELF associated with a doubling of risk [30]. Of the 25 patients with CHB followed up for over 7 years in that study, 2 died of a liver-related cause and one experienced a nonfatal liver-related outcome by 7 years (median for the whole cohort) after biopsy and ELF test. In all 3 cases, the incident ELF score exceeded 7.8. The median ELF score was 8.63 for the whole cohort of CHB patients that were followed up.

This study reports the external validation of the ELF test in subjects with CHB. Performance in patients with CHB in the original cohort ($n = 44$) was good at all fibrosis stages and maintained in this validation cohort. Logistic regression, which included age and simple biochemical parameters (AST, ALT), did not improve performance. These data suggest a role for ELF in the assessment of patients with CHB and in informing the decision-making process when antiviral therapy is being considered.

A recent study [31] reporting the performance of ELF in 58 patients with CHB used the published algorithm [24] but not the immune assays that have been specifically developed for the ELF test. AUROC values for predicting Ishak fibrosis stages 1–6 and 2–6 (equivalent to METAVIR $F \geq 1$) were 0.66 and 0.59, respectively, lower than the values we report. AUROC for predicting Ishak stages 3–6 was 0.83, similar to our findings. The inferior performance of the test in this cohort is likely to be attributable to the use of assays that were not specifically developed for the ELF test and failure to use the appropriate autoanalyser.

Recently, the performance of ELF and TE has been studied in a cohort of Asian subjects with CHB [32]. AUROC values for predicting $F \geq 2$, $F \geq 3$ and F4 were 0.90, 0.86 and 0.86 for ELF and 0.94, 0.96 and 0.96 for TE, respectively. TE was significantly better than ELF for predicting $F \geq 3$ and F4.

In the present study, TE performed as well or better than in other studies in patients with CHB. For example, in the detection of F4 fibrosis, AUROC values in other studies range from 0.88 [11] to 0.94 [33]. A meta-analysis of noninvasive tests for liver disease severity in nonalcoholic fatty liver disease [34] found that the collective performance of TE in detecting $F \geq 2$ and $F \geq 3$ fibrosis was 0.84 (95% CI 0.79–0.90) and 0.94 (95% CI 0.86–0.99), respectively. Regression analysis found that success was unaffected by the severity of inflammation or steatosis, but obesity was an independent predictor of failure of TE.

The rate of TE failure (3%) was very low in this study; a major review of clinical performance found a failure rate of 18.9% [35]. Studies investigating TE in patients with CHB report success rates for acquiring valid TE results ranging between 79% and 99.6% [11,16,35–39]. TE reproducibility has been shown to be excellent for both inter-and intra-observer agreement, but this is reduced at lesser stages of fibrosis and in patients with steatosis, high body mass index and in particular waist circumference [40,41]. All 6 patients in our study excluded due to TE failure were overweight ($n = 2$) or obese ($n = 4$).

Both ELF and TE represent alternative and potentially complimentary approaches to assessing liver fibrosis and are associated with minimal discomfort and hazard to the patient when compared with biopsy. Logistic regression analysis suggests that the performance of ELF is improved with the addition of TE, although TE does not improve with the addition of ELF.

Both modalities track fibrosis stage linearly, with TE having superior discrimination and closer correlation with histological staging, particularly at higher fibrosis stages.

The performance of TE predicting $F \geq 2$ fibrosis in this study was superior to most of the previous studies assessing TE in CHB. The diagnostic performance of each modality was evaluated at various sensitivities and specificities; the median diagnostic odds ratio for ELF for detecting severe fibrosis between sensitivity and specificity of 95% was 7.3 and for TE 24.0. Clinical utility modelling supports a role for these modalities in the assessment of patients and in treatment decisions.

Applying previously published thresholds to our data allows for some generalizability of the model. Recent studies investigating ELF and TE both in a heterogeneous population [42] and in CHB [32] did not report dual thresholds, making comparison difficult. However, using thresholds reported in separate studies allows some comparisons to be drawn. A study of TE in CHB [8] reported that cut-off values of 9.4 and 6.2 which had sensitivity and specificity of >90% ruled in and ruled out $F \geq 2$ in 56% of cases, with 90% accuracy. Applying these thresholds to our data, 57% of patients would have $F \geq 2$ ruled in or ruled out, with 91% accuracy. Data from patients with chronic hepatitis C [27] found that using ELF cut-off values of 9.59 and 10.22, with sensitivity and specificity of 85%, 81% of patients could avoid biopsy by having severe fibrosis ($F \geq 3$) ruled in or ruled out, with 81% accuracy. Applying these thresholds to our data, 77% of patients would avoid biopsy, with 86% accuracy.

Using the DANA method to calculate the adjusted uniform AUROC, diagnostic performance increased at all fibrosis stages with both modalities. This method assumes equal prevalence in all fibrosis stages, which may not be reflective of true prevalence and may overestimate prevalence at the extremes of fibrosis stage. Further, the coefficient in the equation was developed using a population of patients with chronic hepatitis C, and with a different noninvasive test, although it has been employed subsequently in a cohort of CHB patients [43]. Further validation of this method is required. Adjustment using the Obuchowski method showed that the overall mean accuracy (unweighted Obuchowski measure) was 0.91 for ELF and 0.95 for TE. For diagnosis between F3 and F4, performance was 0.59 for ELF and 0.73 for TE (Table S4).

Strengths of this study include the method of data collection. Liver biopsy, TE and serum sampling were all performed on the same day. ELF tests were performed in one central laboratory, ensuring quality control and consistency. It is important to note that the present study used the proprietary ELF assays in accordance with the manufacturer's instructions rather than a 'homebrew' combination of substitute assays performed on other platforms as reported in other studies [31]. There are several potential limitations to this study. The low failure rate of TE in this study was at odds with much larger reports of clinical practice. The relatively high prevalence of fibrosis in this cohort means that the findings may not be reliably applied to lower prevalence populations such as the primary care setting, where the positive predictive value of the test will be lower.

This study has demonstrated that the performance of ELF in detection of liver fibrosis in subjects with CHB is good and is reproducible. Both ELF and TE perform well in the prediction of fibrosis at all stages, with TE superior at detecting severe fibrosis and cirrhosis in this cohort that contained a high prevalence of severe fibrosis. Further analyses in cohorts of subjects with CHB are required.

CONFLICTS OF INTEREST

Pietro Lampertico has speaker bureau roles for Roche, Gilead, Bristol-Myers Squibb and GlaxoSmithKline. Massimo Colombo has served as an advisory committee member for Merck, Roche, Novartis, Bayer, Bristol-Myers Squibb, Gilead Sciences, Tibotec, Vertex, Janssen Cilag and Achillion. He has served as a speaker and teacher for Tibotec, Roche, Novartis, Bayer, Bristol-Myers Squibb, Gilead Sciences and Vertex and has received grant and research support from Merck, Roche, Bristol-Myers Squibb and Gilead Sciences. William Rosenberg is CEO of iQur Ltd and receives grant funding form Siemens Healthcare Diagnostics Inc. He has speaker bureau roles for Roche and Gilead and is an advisory committee member for Roche, Gilead Sciences, MSD and GlaxoSmithKline. Paul Trembling and Sudeep Tanwar have received educational grant support from Janssen, MSD, Gilead Sciences, Novartis and Roche.

REFERENCES

1 Donato F, Gelatti U, Tagger A *et al.* Intrahepatic cholangiocarcinoma and hepatitis C and B virus infection, alcohol intake, and hepatolithiasis: a case-control study in Italy. *Cancer Causes Control* 2001; 12: 959–964.

2 Hsing AW, Zhang M, Rashid A *et al.* Hepatitis B and C virus infection and the risk of biliary tract cancer: a population-based study in China. *Int J Cancer* 2008; 122: 1849–1853.

3 European Association for the Study of the Liver. EASL clinical practice guidelines management of chronic hepatitis B. *J Hepatol* 2009; 50: 227–242.

4 Lok ASF, McMahon BJ. Chronic hepatitis B: update 2009. *Hepatology* 2009; 50: 661–662.

5 Thampanitchawong P, Piratvisuth T. Liver biopsy: complications and risk factors. *World J Gastroenterol* 1999; 5: 301–304.

6 Regev A, Berho M, Jeffers LJ *et al.* Sampling error and intraobserver variation in liver biopsy in patients with chronic HCV infection. *Am J Gastroenterol* 2002; 97: 2614–2618.

7 Poynard T, Ngo Y, Munteanu M, Thabut D, Ratziu V. Noninvasive markers of hepatic fibrosis in chronic hepatitis B. *Curr Hepat Rep* 2011; 10: 87–97.

8 Viganò M, Paggi S, Lampertico P. *et al.* Dual cut-off transient elastography to assess liver fibrosis in chronic hepatitis B: a cohort study with internal validation. *Aliment Pharmacol Ther* 2011; 34: 353–362.

9 Raftopoulos SC, George J, Bourliere M *et al.* Comparison of noninvasive models of fibrosis in chronic hepatitis B. *Hepatol Int* 2011; 6: 457–467.

10 Rosenberg WMC, Voelker M, Thiel R *et al.* Serum markers detect the presence of liver fibrosis: a cohort study. *Gastroenterology* 2004; 127: 1704–1713.

11 Bonnard P, Sombie R, Lescure FX *et al.* Comparison of elastography, serum marker scores, and histology for the assessment of liver fibrosis in hepatitis B Virus (HBV)-infected patients in Burkina Faso. *Am J Trop Med Hyg* 2010; 82: 454–458.

12 Marcellin P, Ziol M, Bedossa P *et al.* Non-invasive assessment of liver fibrosis by stiffness measurement in patients with chronic hepatitis B. *Liver Int* 2009; 29: 242–247.

13 Sporea I, Sirli R, Deleanu A *et al.* Liver stiffness measurements in patients with HBV vs HCV chronic hepatitis: a comparative study. *World J Gastroenterol* 2010; 16: 4832–4837.

14 Lesmana CR, Salim S, Hasan I *et al.* Diagnostic accuracy of transient elastography (FibroScan) versus the aspartate transaminase to platelet ratio index in assessing liver fibrosis in chronic hepatitis B: the role in primary care setting. *J Clin Pathol* 2011; 64: 916–920.

15 Ogawa E, Furusyo N, Murata M *et al.* Longitudinal assessment of liver stiffness by transient elastography for chronic hepatitis B patients treated with nucleoside analog. *Hepatol Res* 2011; 41: 1178–1188.

16 Mbaye PS, Sarr A, Sire J-M *et al.* Liver stiffness measurement and biochemical markers in Senegalese chronic hepatitis B patients with normal ALT and high viral load. *PLoS ONE* 2011; 6: e22291.

17 Poynard T, Bedossa P, Opolon P. Natural history of liver fibrosis progression in patients with chronic hepatitis C. The OBSVIRC, METAVIR, CLINIVIR, and DOSVIRC groups. *Lancet* 1997; 349: 825–832.

18 Fattovich G, Bortolotti F, Donato F. Natural history of chronic hepatitis B: special emphasis on disease progression and prognostic factors. *J Hepatol* 2008; 48: 335–352.

19 Rigamonti C, Donato MF, Fraquelli M *et al.* Transient elastography predicts fibrosis progression in patients with recurrent hepatitis C after liver transplantation. *Gut* 2008; 57: 821–827.

20 DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988; 44: 837–845.

21 Poynard T, Halfon P, Castera L *et al.* Standardization of ROC curve areas for diagnostic evaluation of liver fibrosis markers based on prevalences of fibrosis stages. *Clin Chem* 2007; 53: 1615–1622.

22 Obuchowski NA. Estimating and comparing diagnostic tests' accuracy when the gold standard is not binary. *Acad Radiol* 2005; 12: 1198–1204.

23 Lambert J, Halfon P, Penaranda G, Bedossa P, Cacoub P, Carrat F. How to measure the diagnostic accuracy of noninvasive liver fibrosis indices: the area under the ROC curve revisited. *Clin Chem* 2008; 54: 1372–1378.

24 Guha IN, Parkes J, Roderick P *et al.* Noninvasive markers of fibrosis in nonalcoholic fatty liver disease: validating the European liver fibrosis panel and exploring simple markers. *Hepatology* 2007; 47: 455–460.

25 Nobili V, Parkes J, Bottazzo G *et al.* Performance of ELF serum markers in predicting fibrosis stage in pediatric non-alcoholic fatty liver disease. *Gastroenterology* 2009; 136: 160–167.

26 Mayo MJ, Parkes J, Adams-Huet B *et al.* Prediction of clinical outcomes in primary biliary cirrhosis by serum enhanced liver fibrosis assay. *Hepatology* 2008; 48: 1549–1557.

27 Parkes J, Guha IN, Roderick P *et al.* Enhanced liver fibrosis (ELF) test accurately identifies liver fibrosis in patients with chronic hepatitis C. *J Viral Hepat* 2011; 18: 23–31.

28 Carrión JA, Fernández–Varo G, Bruguera M *et al.* Serum fibrosis markers identify patients with mild and progressive hepatitis C recurrence after liver transplantation. *Gastroenterology* 2010; 138(147–158): e1.

29 Martinez SM, Fernández-Varo G, González P *et al.* Assessment of liver fibrosis before and after antiviral therapy by different serum marker panels in patients with chronic hepatitis C. *Aliment Pharmacol Ther* 2011; 33: 138–148.

30 Parkes J, Roderick P, Harris S *et al.* Enhanced liver fibrosis test can predict clinical outcomes in patients with chronic liver disease. *Gut* 2010; 59: 1245–1251.

31 Gümüsay O, Ozenirler S, Atak A *et al.* Diagnostic potential of serum direct markers and non-invasive fibrosis models in patients with chronic hepatitis B. *Hepatol Res* 2013; 43: 228–237.

32 Kim BK, Kim HS, Park JY *et al.* Prospective validation of ELF test in comparison with Fibroscan and Fibro test to predict liver fibrosis in Asian subjects with chronic hepatitis B. *PLoS ONE* 2012; 7: e41964.

33 Cardoso AC, Carvalho-Filho RJ, Stern C *et al.* Direct comparison of diagnostic performance of transient elastography in patients with chronic hepatitis B and chronic hepatitis C. *Liver Int* 2012; 32: 612–621.

34 Musso G, Gambino R, Cassader M, Pagano G. Meta-analysis: natural history of non-alcoholic fatty liver disease (NAFLD) and diagnostic accuracy of non-invasive tests for liver disease severity. *Ann Med* 2011; 43: 617–649.

35 Castera L, Foucher J, Bernard PH *et al.* Pitfalls of liver stiffness measurement: a 5-year prospective study of 13,369 examinations. *Hepatology* 2010; 51: 828–835.

36 Fraquelli M, Rigamonti C, Casazza G *et al.* Etiology-related determinants of liver stiffness values in chronic viral hepatitis B or C. *J Hepatol* 2011; 54: 621–628.

37 Oliveri F, Coco B, Ciccorossi P *et al.* Liver stiffness in the hepatitis B virus carrier: a non-invasive marker of liver disease influenced by the pattern of transaminases. *World J Gastroenterol* 2008; 14: 6154–6162.
38 Chan HL, Wong GL, Choi PC *et al.* Alanine aminotransferase-based algorithms of liver stiffness measurement by transient elastography (Fibroscan) for liver fibrosis in chronic hepatitis B. *J Viral Hepat* 2009; 16: 36–44.
39 Verveer C, Zondervan PE, ten Kate FJ, Hansen BE, Janssen HL, de Knegt RJ. Evaluation of transient elastography for fibrosis assessment compared with large biopsies in chronic hepatitis B and C. *Liver Int* 2012; 32: 622–628.
40 Fraquelli M, Branchi F. The role of transient elastography in patients with hepatitis B viral disease. *Dig Liver Dis* 2011; 43: S25–S31.
41 Fraquelli M, Rigamonti C, Casazza G *et al.* Reproducibility of transient elastography in the evaluation of liver fibrosis in patients with chronic liver disease. *Gut* 2007; 56: 968–973.
42 Crespo G, Fernandez-Varo G, Marino Z *et al.* ARFI, FibroScan, ELF, and their combinations in the assessment of liver fibrosis: a prospective study. *J Hepatol* 2012; 57: 281–287.
43 Poynard T, Ngo Y, Marcellin P, Hadziyannis S, Ratziu V, Benhamou Y. Impact of adefovir dipivoxil on liver fibrosis and activity assessed with biochemical markers (FibroTest-ActiTest) in patients infected by hepatitis B virus. *J Viral Hepat* 2009; 16: 203–213.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Table S1. Sensitivities and specificities of ELF and TE identifying severe fibrosis (METAVIR $\geq$ F3) and any fibrosis (F $\geq$1) using thresholds with sensitivity and specificity of 85%.

Table S2. Odds ratios calculated by logistic regression for prediction of fibrosis using models comprising ELF, TE and ELF and TE.

Table S3. Diagnostic performance of ELF and TE according to categories of ALT.

Table S4. Obuchowski measures for ELF and TE for each fibrosis stage pair.

Trembling et al. BMC Public Health (2017) 17:603
DOI 10.1186/s12889-017-4518-y

BMC Public Health

RESEARCH ARTICLE Open Access

Risk of chronic liver disease in post-menopausal women due to body mass index, alcohol and their interaction: a prospective nested cohort study within the United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS)

Paul M Trembling[1*], Sophia Apostolidou[2], Aleksandra Gentry-Maharaj[2], Julie Parkes[3], Andy Ryan[2], Sudeep Tanwar[1], Matthew Burnell[2], Ian Jacobs[2,4], Usha Menon[2†] and William M. Rosenberg[1†]

Abstract

Background: We investigated the risk of chronic liver disease (CLD) due to alcohol consumption and body mass index (BMI) and the effects of their interaction in a prospective cohort study of women recruited to the UKCTOCS trial.

Methods: 95,126 post-menopausal women without documented CLD were stratified into 12 groups defined by combinations of BMI (normal, overweight, obese) and alcohol consumption (none, <1–15, 16–20 and ≥21 units/week), and followed for an average of 5.1 years. Hazard ratios (HR) were calculated for incident liver-related events (LRE).

Results: First LREs were reported in 325 (0.34%) participants. Compared to women with normal BMI, HR = 1.44 (95% CI; 1.10–1.87) in the overweight group and HR = 2.25 (95% CI; 1.70–2.97) in the obese group, adjusted for alcohol and potential confounders. Compared to those abstinent from alcohol, HR = 0.70 (95% CI; 0.55–0.88) for <1–15 units/week, 0.93 (95% CI; 0.50–1.73) for 16–20 units/week and 1.82 (95% CI; 0.97–3.39) for ≥21 units/week adjusted for BMI and potential confounders. Compared to women with normal BMI drinking no alcohol, HR for LRE in obese women consuming ≥21 units/week was 2.86 (95% CI; 0.67–12.42), 1.58 (95% CI; 0.96–2.61) for obese women drinking <1–15 units/week and 1.93 (95% CI; 0.66–5.62) in those with normal BMI consuming ≥21 units/week after adjustment for potential confounders. We found no significant interaction between BMI and alcohol.

Conclusion: High BMI and alcohol consumption and abstinence are risk factors for CLD in post-menopausal women. However, BMI and alcohol do not demonstrate significant interaction in this group.

Trial registration: UKCTOCS is registered as an International Standard Randomised Controlled Trial, number ISRCTN22488978. Registered 06/04/2000.

Keywords: Chronic liver disease, Cirrhosis, Alcohol, Body mass index, Obesity

* Correspondence: p.trembling@ucl.ac.uk
†Equal contributors
[1]Institute for Liver and Digestive Health, Division of Medicine, University College London, Royal Free Hospital, Rowland Hill Street, London NW3 2PF, UK
Full list of author information is available at the end of the article

Background

Chronic liver disease (CLD) is the 5th commonest cause of death in the UK, and the only rising major cause of mortality and morbidity. 60,000 people in England and Wales have cirrhosis [1–3]. Recent data estimates that over 600,000 adults in the USA have CLD, with over half of affected individuals unaware of the diagnosis [4]. Overweight and alcohol consumption are major causes of CLD [5–7]. Non-alcoholic fatty liver disease (NAFLD) can be considered the pathological manifestation in the liver of the metabolic syndrome, of which high BMI is a key feature [8]. NAFLD comprises a spectrum of disease, from steatosis, through inflammation (steatoheaptitis) to fibrosis and cirrhosis. The precise influence of body mass index (BMI) on the risk of liver disease in women, however, is not conclusive and previous studies using smaller subsets of ICD-10 codes to identify liver-related morbidity and mortality may have underestimated the impact of BMI and alcohol [9, 10]. Further, interaction between alcohol and BMI and risk of liver disease is not well understood. Regardless of the etiology of liver disease, the clinicopathological outcome in those who develop CLD is cirrhosis [11] and the there may be common pathways in which alcohol and high BMI damage the liver [12]. A synergistic interaction between steatosis and alcohol consumption in the progression of fibrosis in patients with chronic hepatitis C has been demonstrated in histological studies [13].

Both alcohol consumption and NAFLD are common. Moderate alcohol consumption is associated with decreased mortality, largely due to reduced cardiovascular-related disease, but there are no guidelines related to alcohol use in NAFLD and these factors, in addition to rising levels of liver disease and the high prevalence of excess alcohol consumption, coupled with the worldwide increase in obesity demonstrate the need to further understand the roles of alcohol and BMI and their interaction in CLD.

In a large cohort of women we investigated incidence of CLD and its relationship to alcohol and BMI, and examined the interaction between these two risk factors.

Methods

Study population

This prospective cohort study was nested in the United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS) [14]. UKCTOCS is a multi-center UK-based randomised controlled trial designed to define the effect of ovarian cancer screening on mortality. Between April 2001 and October 2005, 202,638 post-menopausal women aged 50–74 were recruited in England, Wales and Northern Ireland. Participants were invited at random from 27 local authority registers. Exclusion criteria included bilateral oophorectomy, increased risk of familial ovarian cancer, previous ovarian cancer and active non-ovarian cancer. The trial design and detailed eligibility criteria have been described elsewhere [14–16]. This study is nested within UKCTOCS, comprising of participants in England.

UKCTOCS was approved by the UK North West Multicentre Research Ethics Committee (North West MREC 00/8/34), with site-specific approval from the local regional ethics committees and the Caldicott guardians (data controllers) of the participating primary care trusts. Written informed consent was obtained from all volunteers.

Exposures

The exposures of interest were BMI and current weekly alcohol consumption. Participants completed a questionnaire at recruitment, which included self-reported height and weight. BMI was calculated (BMI (kg/m^2) = weight (kg)/(height (m))^2) and categorised according to the World Health Organisation's definitions; normal (<25 kg/m^2), overweight (25- < 30 kg/m^2) or obese (≥30 kg/m^2). As there are no existing population estimates for the range of BMI a pragmatic approach was adopted to selecting patients with plausible BMI values. Participants who recorded a height outside the range 140-210 cm, or a weight outside the range 25-200 kg, or where the BMI was outside the range 16–65 kg/m^2 were excluded.

Via a follow-up questionnaire 3–5 years after randomisation, participants estimated their current alcohol consumption as the number of drinks consumed per week (none, less than 1, 1–3, 4–6, 7–10, 11–15, 16–20 or ≥21 drinks), assuming one drink is a glass of wine, half a pint of beer or cider, or a measure of spirits. Alcohol units were calculated using the convention that one drink is the equivalent of 1 UK unit (10 ml or 8 g of pure alcohol) [17]. Participants were categorised in the following groups; none, <1–15, 16–20 and ≥21 units/week, and those with no alcohol response were excluded.

Covariates

The follow-up questionnaire asked participants to report known comorbidities including heart disease, hypercholesterolaemia, hypertension and diabetes mellitus, and whether they currently smoked (all categorised as yes/no). Socioeconomic status was estimated using the Index of Multiple Deprivation 2007 (IMD) (continuous variable) [18]. This ascribes a deprivation score to participants based on their postcode, with a higher score indicating higher deprivation.

Follow up

All participants are followed through a 'flagging' study with the NHS Information Centre for Health and Social

Care in England and Wales which provided data on cancer registrations and deaths, with diagnosis/cause of death coded according to the International Classification of Diseases, version 10 (ICD-10). 99.98% of UKCTOCS participants were successfully flagged. In addition, hospital inpatient and outpatient episode data for 2001–10 were available through linkage to the Hospital Episodes Statistics (HES) database. Each HES record reports a main diagnosis and up to 19 (inpatient admissions) and 11 (outpatient appointments) further diagnoses and each death record reports the primary death code and additional diagnoses recorded on the death certificate. As HES data were only available for participants in England, only participants in England were included in this study. Women were included in the study from the point of return of questionnaire. Women with known pre-existing liver disease were not included, by excluding those where a code of interest had been registered between recruitment to UKCTOCS and return of questionnaire.

Outcome

The main outcome measure was first liver-related event (LRE), defined as first presentation of either a hospital admission, outpatient appointment, cancer registration with, or death from, an ICD-10 code of interest. The following codes for liver disease were searched for: K70 (alcoholic liver disease), K73 (chronic hepatitis) and K74 (fibrosis and cirrhosis). These codes are consistent with other UK studies of cirrhosis [1, 9]. We also included K76 (other diseases of liver, including fat) in order to widen the search for liver disease beyond cirrhosis to include fatty liver disease. In addition, codes relating to sequelae of decompensated liver disease were also searched for; I85 (oesophageal varices), Z94.4 (liver transplant) and C22.0 (hepatocellular carcinoma). In addition to ICD-10 codes, death certificates were also searched for any mention of alcoholic liver disease or non-alcoholic fatty liver disease.

Statistical analysis

Crude incidence rates of first LRE were calculated using person-years of follow-up as the denominators, for each BMI group, each alcohol group and each BMI/alcohol combination. For each participant, person-years of follow-up were accrued from date that the follow-up questionnaire was returned (as this was the date that current alcohol use was ascertained), to the censorship date (February 1, 2013), date of first presentation with LRE, or death from any other cause. Participants who experienced a LRE at any time from randomisation to return of questionnaire were excluded.

Separate influences of BMI and of alcohol on incident liver disease

Cox proportional hazards models were used to calculate hazard ratios (HRs) of first LRE in three categories of BMI using normal BMI as the reference. Similar analysis was performed for alcohol with no alcohol consumption as the reference. The proportional hazards models were adjusted for BMI, or alcohol respectively.

All potential confounding risk factors (smoking, IMD, hypertension, heart disease, hypercholesterolaemia, diabetes) were included individually in a Cox model to calculate univariate HRs for LREs, to guide their utility in the models evaluating risk due to BMI and alcohol.

Influences of combinations of BMI and alcohol

HRs were calculated for twelve BMI and alcohol combinations using the normal BMI/no alcohol consumption category as the reference, adjusted for potential confounders with significant HRs for LRE, and then adjusted only for factors associated with the metabolic syndrome (hypertension, hypercholesterolemia, heart disease and diabetes). The proportional hazards assumption was checked by examining the log minus log plot.

Interaction between alcohol and BMI

Interaction between alcohol using several thresholds and BMI (as a continuous variable) was analysed by calculating the interaction term from the Cox regression models.

All analyses were performed using SPSS (version 19, SPSS Inc., Chicago, IL, USA) and STATA statistical software (StataCorp 2007. Release 10. College Station, TX, USA: StataCorp LP).

Results

Of the 157,996 UKCTOCS participants resident in England, 62,870 were excluded including 321 women who experienced an LRE or died between recruitment and return of questionnaire and 14,295 (9%) with no data on smoking. The final cohort comprised 95,126 participants (Fig. 1).

Baseline characteristics are shown in Table 1. 97.1% of the participants were white. 36% were smokers. 55% were either overweight (37%) or obese (19%). 23.4% reported drinking no alcohol and 1.5% reported drinking more than 21 units/week. Increasing BMI correlated with increased reporting of hypertension, heart disease, hypercholesterolaemia and diabetes.

Three hundred twenty five (0.34%) women experienced a first LRE over a total of 509,561 person-years of follow-up (mean 5.1 years), equivalent to 0.64 first events per 1000 person-years (3.3 per 1000 women over 5 years). The most common ICD-10 code signaling a first presentation of LRE was K76 (Additional file 1: Table S1). Only 763 (0.8%) of participants were

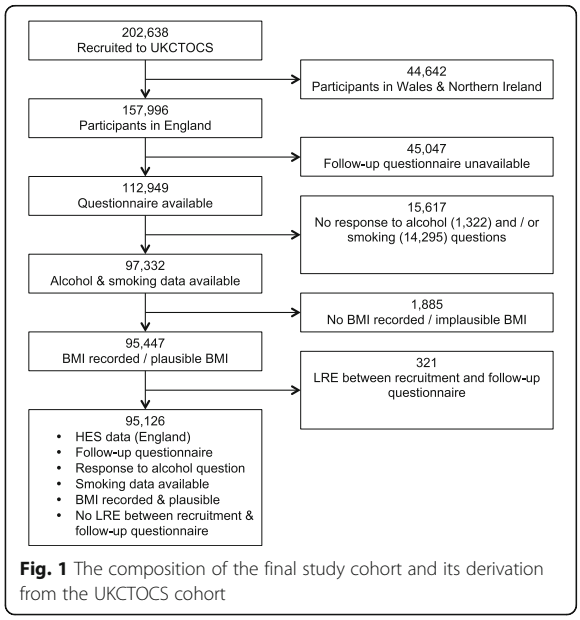


Fig. 1 The composition of the final study cohort and its derivation from the UKCTOCS cohort

underweight (BMI <18·5 kg/m^2) and in this group there were only 4 LREs, therefore this group was combined with the normal BMI group. 1237 (7% of the obese group) women could be classified as morbidly obese (BMI ≥ 40 kgm^{-2}) and in this group, the event rate was highest (1.98 events per 1000 person years (95% CI; 1.05–3.38)). There were 2713 (2.9%) deaths from any cause.

Risk of liver-related events due to potential confounders

Other covariates also demonstrated independent association with liver-related events (Table 1). Significant HRs were seen with smoking, hypertension, heart disease, hypercholesterolaemia, diabetes and IMD.

BMI and risk of liver-related events

Crude rates of LREs increased with rising BMI. HRs for LRE were significantly higher in both overweight (1.44, 95% CI; 1.10–1.87) and obese categories (2.25, 95% CI; 1.70–2.97) compared to the normal BMI group. A fully adjusted model is presented incorporating adjustment for confounders with significant HRs (Table 2).

Alcohol consumption and risk of liver-related events

The rate of LRE was lowest in the group drinking <1–15 units weekly and increased with abstinence and increasing alcohol use. This tendency towards a "J-shaped" relationship between LRE and alcohol consumption was preserved after adjustment for BMI, with lowest HRs in the <1–15 units/week group, although the there was no statistically significant difference between the HRs for this group and the reference group. A fully adjusted model is shown, adjusted for variables with significant HRs for LRE (Table 2).

In the group reporting no alcohol consumption the proportion of LREs that were alcohol-related was 3.96% compared to 11.16% in those drinking any alcohol.

Risk of liver-related events in participants grouped in to combinations of BMI and alcohol use

Participants were grouped according to combinations of BMI and alcohol consumption. Table 3 shows the rates of LRE in each group. The fully adjusted Cox model shows that the lowest risk is in those with normal BMI consuming <1–15 units/week. Within the normal BMI group, abstinence or drinking >16 units/week increases the risk of LRE, although there are wide confidence intervals.

Among overweight and obese women, the nadirs of risk were in the <1–15 units/week groups, and as in the normal BMI group, the risk was highest in the highest alcohol group (HR 3.32, 95% CI; 1.25–8.81; and HR 2.86, 95% CI; 0.67–12.21 respectively).

To estimate the effect of cardiovascular disease and diabetes on the morbidity associated with fatty liver disease, HRs were adjusted for confounding factors associated with the metabolic syndrome. When elements of the metabolic syndrome were controlled for, risk of LRE attributable to heavier drinking increased. This suggests that the risk of liver disease attributable to BMI in patients with, or at risk of, metabolic syndrome is not entirely accounted for by hypertension, heart disease, hypercholesterolemia or diabetes, but may be partly attributable to steatosis itself.

When separated by BMI group, the trend to a "J-shaped" relationship of risk of LRE remains in all BMI groups, with risk highest in the abstainers and heavier drinkers, compared to those in the <1–15 units/week alcohol groups.

Interaction between alcohol and BMI

Interaction terms were calculated for BMI (continuous) and alcohol, using thresholds for high alcohol of ≥16 units/week and ≥21 units/week. There was no significant interaction between BMI and high alcohol use. Similarly, no interaction was seen with BMI and any alcohol use.

Discussion

Main findings

The most striking finding of this study is the risk of liver disease associated with overweight/obesity in post-menopausal women. While the association between alcohol consumption and CLD is well established, there is still much to characterise in the natural history of non-alcoholic fatty liver disease (NAFLD) [3]. Furthermore the

Table 1 Baseline characteristics, number of first events according to BMI category and in all participants, and hazard ratios for LRE for potential confounders (continuous[a] and categorical[b] variables)

Characteristic	BMI category (kg/m^2)			All participants	Hazard ratio (95% confidence intervals)
	<25	25 - < 30	≥30		
Total, *n* (%)	42,452 (44.6)	35,073 (36.9)	17,601 (18.5)	95,126	
Recruitment age, median (years)	60.0 (50–74)	61.0 (50–74)	60.0 (50–74)	60.0 (50–74)	1.01[a] (0.99–1.02)
Smoker, *n* (%)	14,740 (34.7)	12,616 (36.0)	6621 (37.6)	33,977 (35.7)	1.89[b] (1.52–2.35)
Hypertension, *n* (%)	9477 (22.3)	12,116 (34.5)	8440 (48.0)	30,033 (31.6)	1.38[b] (1.11–1.73)
Heart disease, *n* (%)	1721 (4.1)	2086 (5.9)	1416 (8.0)	5223 (5.5)	2.17[b] (1.53–3.06)
Hypercholesterolemia, *n* (%)	8001 (18.8)	9148 (26.1)	5440 (30.9)	22,589 (23.7)	1.68[b] (1.33–2.11)
Diabetes, *n* (%)	836 (2.0)	1689 (2.6)	2263 (12.9)	4788 (5.0)	2.76[b] (1.99–3.83)
IMD, mean	17.0	18.7	21.3	18.5	1.09[a] (1.01–1.03)
Alcohol consumption (units/week)					
None	8479 (20.0)	8189 (23.3)	5547 (31.5)	22,215 (23.4)	1[b] (reference)
< 1–15	31,811 (74.9)	25,324 (72.2)	11,473 (65.2)	68,608 (72.1)	0.64[b] (0.51–0.82)
16–20	1448 (3.4)	1067 (3.0)	366 (2.1)	2881 (3.0)	0.82[b] (0.44–1.53)
≥ 21	714 (1.7)	493 (1.4)	215 (1.2)	1422 (1.5)	1.66[b] (0.89–3.09)
Alcohol consumption (units/week)	Number of first LREs				
None	23	36	42	101	
< 1–15	71	77	55	202	
16–20	17	10	3	11	
≥ 21	4	5	2	11	
Total	102	123	100	325	

BMI body mass index, *IMD* Index of Multiple Deprivation, *LRE* liver-related event

study supports the adverse impact of heavy drinking compounding the effects of overweight and obesity. Strategies for preventing and detecting liver disease should be developed accommodating these findings.

This study suggests that in women aged 50–74, those consuming <1–15 units/week are at lowest risk of liver disease. Those drinking 16–20 units/week are only marginally more at risk. The UK Institute of Alcohol Studies defines hazardous drinking as more than 14 units/week and harmful drinking as >35 units/week which would be consistent with the observations in our study population.

Those that are overweight or obese have an increased risk of liver disease. Women of normal BMI who drink <1–15 units/week are at lowest risk, compared to those who drink more or who abstain. It is possible, however, that some abstainers had previously been heavy drinkers. This is supported by our data

Table 2 Event rates and adjusted hazard ratios of first liver-related events, according to BMI category and according to alcohol category

BMI and alcohol categories	First event rate per 1000 person years (95% confidence intervals)	Hazard ratio (95% confidence intervals)[a]	Hazard ratio (95% confidence intervals)[b]
BMI category (kg/m^2)			
< 25	0.45 (0.4–0.5)	1 (reference)	1 (reference)
25 - <30	0.65 (0.5–0.8)	1.44 (1.10–1.87)	1.31 (1.01–1.72)
≥ 30	1.06 (0.9–1.3)	2.25 (1.70–2.97)	1.85 (1.38–2.48)
Alcohol category (units/week)			
None	0.86 (0.7–1.0)	1 (reference)	1 (reference)
< 1–15	0.55 (0.5–0.6)	0.70 (0.55–0.88)	0.78 (0.61–1.00)
16–20	0.68 (0.3–1.2)	0.93 (0.50–1.73)	0.97 (0.52–1.82)
≥ 21	1.37 (0.7–2.5)	1.82 (0.97–3.39)	1.83 (0.97–3.44)

[a]Adjusted for BMI (continuous variable) or alcohol category as appropriate [b]Adjusted for BMI (continuous variable) or alcohol category as appropriate and smoking, hypertension, heart disease, hypercholesterolaemia, diabetes and IMD

Table 3 Event rates and hazard ratios of first liver-related event according to various BMI and alcohol combinations

BMI category (kg/m^2)	Alcohol category (units/week)			
	None	<1–15	16–20	≥21
	First event rate per 1000 person years (95% confidence intervals)			
<25	0.52 (0.3–0.8)	0.42 (0.3–0.5)	0.50 (0.1–1.3)	1.00 (0.3–2.6)
25 - <30	0.83 (0.6–1.2)	0.57 (0.4–0.7)	0.84 (0.3–2.0)	1.81 (0.6–4.2)
≥30	1.43 (1.0–1.9)	0.88 (0.7–1.1)	0.98 (0.1–3.5)	1.68 (0.2–6.1)
	Hazard ratio (95% confidence intervals) adjusted for smoking, hypertension, heart disease, hypercholesterolaemia, diabetes and IMD			
<25	1 (reference)	0.91 (0.56–1.47)	1.03 (0.35–2.99)	1.93 (0.66–5.62)
25 - <30	1.46 (0.85–2.50)	1.34 (0.71–1.83)	1.61 (0.61–4.26)	3.32 (1.25–8.81)
≥30	2.28 (1.35–3.86)	1.58 (0.96–2.61)	1.67 (0.39–7.15)	2.86 (0.67–12.21)
	Hazard ratio (95% confidence intervals) adjusted for hypertension, heart disease, hypercholesterolemia and diabetes			
<25	1 (reference)	0.85 (0.53–1.37)	1.07 (0.37–3.09)	2.13 (0.74–6.17)
25 - <30	1.51 (0.89–2.55)	1.11 (0.69–1.76)	1.74 (0.66–4.57)	3.69 (1.40–9.72)
≥30	2.35 (1.40–3.95)	1.59 (0.97–2.60)	1.89 (0.44–8.01)	3.16 (0.74–13.41)

showing that 4% of LREs in the abstainers were alcohol related.

When combinations of risk are considered, compared to a baseline of normal BMI and abstinence, higher BMI (≥30 kg/m^2) confers a greater risk than higher alcohol consumption (≥21 units/week). The highest risk is in those who are overweight or obese and drink the most alcohol.

After adjustment for confounding due to metabolic risk factors, HRs in the two highest alcohol categories increased in all BMI groups, suggesting that these factors may contribute to the risk of CLD. It is biologically plausible that diabetes, hypercholesterolaemia and hypertension may contribute to liver disease over and above that caused by fatty liver disease and alcoholic liver toxicity. The corollary is that obesity can cause liver morbidity and mortality in the absence of the metabolic syndrome, providing evidence that case ascertainment cannot be restricted to overweight or obese patients with features of the metabolic syndrome and challenging the "two hit" and "three hit" hypotheses [19].

Strengths and limitations

Strengths of this study include the size and duration of follow-up, the prospective design and the independence of the data capture for outcomes. This study was also able to adjust for confounding factors, which has not been possible in other cross-sectional studies. In an effort to capture all morbidity and mortality attributable to liver disease, rather than just cirrhosis, we selected ICD-10 codes that encompass a clinically relevant group of diseases including codes for CLD and those relating to the consequences of decompensated liver disease. This was designed to maximise the ability to detect liver disease.

Limitations include reliance on self-reporting of alcohol consumption, co-morbidities, height and weight, which may be a factor in the wide confidence intervals seen for all HR estimates. However, good reliability of self-reporting height and weight [20–24], and alcohol [25–27], has been demonstrated in other studies.

Height and weight were reported at recruitment, and alcohol consumption reported later, on the follow-up questionnaire. Participants were asked to report current alcohol use, rather than lifetime patterns. Changes in drinking patterns would not have been identified, and this method of data collection may fail to identify episodic ("binge") drinkers. We used the convention that one drink is equivalent to 1 unit of alcohol. However assumptions about alcohol content are difficult to make as measures of volume are likely to vary depending on where the alcohol is consumed, and the alcohol content of drinks continues to change. There is evidence that the number of units in alcoholic drinks in the UK have been undercounted [28], however we have used the standard 1 drink = 1 unit as this remains a widely used convention, particularly in public health promotion.

Reliance on ICD-10 to define events may result in errors due to mis-coding. We used additional codes to those used to define cirrhosis in order to maximize the capture of cases, but these may also be subject to mis-coding. We attempted to reduce the risk of non-coding of events by using 3 independent sources, and in the case of death certification also used hand searching of key words in the text of death certificates. Also, the HES database may not capture some areas of healthcare, for example the private sector. The number of LREs that included ICD-10 Z94.4 is surprising (Additional file 1: Table S1). This may be

because participants with liver transplants are engaged in hospital care and are easily identified and coded.

Only post-menopausal women aged 50–74 were included with 97% being white. The loss to follow up rate in UKCTOCS was very small (0.02%). The acceptance rate was 23%. However, despite attempts to ensure that UKCTOCS was representative of the general population [15] there was a 'healthy volunteer effect' [29] on both overall and cause-specific mortality, which may have an effect on the generalisability of findings [18]. Although the health section of the follow-up questionnaire did not specifically ask about liver disease, we excluded those who had a code of interest recorded between recruitment to UKCTOCS and the start of this study. However, exclusion of all participants with known CLD could not be guaranteed.

It is unlikely that viral hepatitis made a significant contribution to LRE based on low prevalence in the demographic of women in this study [30]. During the follow-up period in our study, only 21 (0.02%) of participants had a code for viral hepatitis recorded.

Other studies

A number of studies have demonstrated a reduced risk of liver disease in patients with NAFLD who consume low or moderate amounts of alcohol [31–33], and it has been suggested that these levels of alcohol use may be associated with beneficial effects of insulin sensitivity in post-menopausal women [34]. However, at higher extremes of BMI and alcohol use, data is not conclusive. Previous studies have attributed a lower incidence of CLD to BMI and alcohol, and as expected a lower incidence of CLD when only alcoholic cirrhosis is examined [35]. However these have relied on cirrhosis codes alone, ignoring complications characterising decompensated cirrhosis that are indicative of CLD and clearly associated with BMI and alcohol included in the present study. This study is in broad agreement with some other studies including the National Health and Nutrition Examination Survey (NHANES) [6] which found increasing risk with both increasing BMI and alcohol, but no excess risk in overweight or obese drinkers or in abstainers. A Scottish prospective study reported increasing risk with increasing BMI in men, but not in women [10]. A sub-analysis of men found the lowest risk of CLD in abstainers with normal BMI with a supra-additive interaction between BMI and alcohol [36]. The UK-based Million Women Study [9] used a limited range of ICD-10 codes to identify cirrhosis and reported a rate of hospital admission or death from liver disease less than half that found in our study. However, as in our study, highest risk was in overweight or obese women consuming the most alcohol. In a study of patients with a history of alcohol excess who were admitted to hospital with an alcohol-related problem, risk of cirrhosis was twice as high among the overweight group as those with normal BMI [37]. A recent prospective study of 107,735 middle-aged males used self-reported BMI and alcohol use to assess liver-related mortality ascertained form record linkage, using ICD-10 codes K70-K76, demonstrating a U-shaped relationship between alcohol and mortality and BMI and mortality. Although there was evidence of synergy between low BMI and high alcohol, as in our study there was no evidence of interaction between high BMI and high alcohol use [38].

Our finding of increased risk in abstainers has precedent but is controversial. Previous studies have demonstrated the "J-shaped" relationship between alcohol and risk of mortality [39–42] or CLD [43, 44]. Some prospective studies have found that men but not women abstainers were at increased risk [9, 44], in contrast to the present study that provides a more comprehensive insight into the effects of weight and alcohol. Using raised aminotransferase levels to diagnose suspected NAFLD in men and women in NHANES the highest risk was seen in non-drinkers compared to modest drinkers [45], and in biopsy-proven NAFLD, moderate drinkers had lower risk of steatohepatitis compared to non-drinkers [46]. A prospective Danish study investigating risk of alcohol-related cirrhosis in over 30,000 participants found a dose-dependent increase in risk of cirrhosis with increasing alcohol in women, rather than a "J-shaped" relation which they observed in males [43].

We have confirmed this relationship with risk of CLD in our cohort, and also have demonstrated that the trend towards a "J-shape" relationship remains, irrespective of BMI group.

The increased risk of alcoholic cirrhosis in abstainers compared to light drinkers may be due, in part, to this group containing previous drinkers who raise the overall risk in the abstainer group, rather than due to a true protective effect of alcohol in the light drinkers. One prospective study [35] demonstrated the loss of the "J-shaped" curve when lifetime abstainers were separated from current abstainers. In a small study of patients with biopsy-proven NAFLD, a comprehensive alcohol history was obtained and found to be higher than the original estimate at diagnosis in some patients, suggesting that some of these patients may have had alcohol-related liver disease rather than NAFLD [47].

We found alcohol-related LREs in abstainers (although at less than half the rate seen in drinkers) which, although may partly be a function of miscoding, provides further evidence that this group comprises some ex-drinkers.

Interaction between higher levels of alcohol consumption and NAFLD may result in greater risk of liver disease. A study measuring aminotransferase activity found

that increased BMI potentiates the harmful effect of alcohol on the liver [48]. Increased aminotransferase levels were associated with higher alcohol consumption and BMI. In those with normal BMI there was no association between alcohol and raised aminotransferase levels, but in the overweight and obese groups, alcohol increased risk of elevated aminotransferases. A study of an older population also found risk of elevated aminotransferases with increased BMI and increased alcohol consumption (with lowest risk in abstainers), and a large synergistic effect in the obese group consuming more than three drinks/day [49]. This group also examined the risk of hepatocellular carcinoma in people with chronic hepatitis B, finding synergism between obesity and alcohol [50, 51].

Implications

Our results suggest a substantial influence of both elevated BMI and alcohol on risk of CLD. Although no significant interaction between BMI and alcohol was seen and this lack of synergy is reassuring, the compelling risk in the overweight and obese groups adds to the evidence that rising BMI and increasing alcohol use are risk factors for liver disease among women.

By considering the clinical consequences of liver disease beyond the diagnosis of cirrhosis we revealed a greater burden of disease than previously recognised. Currently much CLD goes undiagnosed until complications of cirrhosis result in serious morbidity and mortality. Earlier identification of those at risk could avert illness and reduce costs by targeted interventions. While the risks associated with heavy alcohol consumption are frequently publicised these data emphasise the importance of disseminating awareness of the risks of liver disease associated with BMI, particularly in light of the growing prevalence of overweight and obesity throughout the world [52]. Public health policy and health education and awareness campaigns should take these facts into account.

Conclusion

This study of post-menopausal women suggests that elevated BMI and high alcohol intake are independent risk factors for liver disease. Strategies for detecting liver disease and public health strategy should recognise the importance of BMI as well as alcohol when confronting the growing burden of liver disease.

Additional file

Additional file 1: Figure S1. Crude rates of first liver-related events (per 1000 person years) according to a) BMI category and b) alcohol consumption category over mean follow-up of 5.1 years. **Table S1.** ICD-10 codes and death certificate text of first LREs. (DOCX 120 kb)

Abbreviations

BMI: Body mass index; CLD: Chronic liver disease; HES: Hospital episode statistics; HR: Hazard ratio; ICD-10: International classification of disease 10th revision; IMD: Index of multiple deprivation; LRE: Liver-related event; NAFLD: Non-alcoholic fatty liver disease; NHANES: National health and nutrition examination survey; UKCTOCS: United Kingdom Collaborative Trial of Ovarian Cancer Screening

Acknowledgements

We are grateful to the women who participated in the trial and to the entire medical, nursing, and administrative staff who contributed to the implementation of UKCTOCS.

Funding

No external funding.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Authors' contributions

All authors were involved in study design. PT, SA, AGM, UM and WR were involved in data collection. PT, UM and WR drafted the manuscript and figs. PT undertook the literature search. PT, JP, ST, UM, WR and MB performed the statistical analysis. All authors critically revised the manuscript and approved the final version.

Ethics approval and consent to participate

UKCTOCS was approved by the UK North West Multicentre Research Ethics Committee (North West MREC 00/8/34), with site-specific approval from the local regional ethics committees and the Caldicott guardians (data controllers) of the participating primary care trusts. Written informed consent was obtained from all volunteers.

Consent for publication

Not applicable.

Competing interests

UM and IJ have a financial interest through Abcodia Ltd. in the third party exploitation of the trial biobank. During part of the UKCTOCS trial IJ had a consultancy arrangement with Becton Dickinson in the field of tumour markers. None of the other authors declared any conflicts of interest.

Publisher's Note

Author details

[1]Institute for Liver and Digestive Health, Division of Medicine, University College London, Royal Free Hospital, Rowland Hill Street, London NW3 2PF, UK. [2]Gynaecological Cancer Research Centre, University College London Elizabeth Garrett Anderson Institute for Women's Health, University College London, London, UK. [3]Public Health Sciences and Medical Statistics, Faculty of Medicine, University of Southampton, Southampton, UK. [4]Office of the President and Vice-Chancellor, The University of New South Wales, UNSW Sydney, Sydney, Australia.

Received: 19 October 2016 Accepted: 20 June 2017
Published online: 28 June 2017

References

1. Leon DA, McCambridge J. Liver cirrhosis mortality rates in Britain from 1950 to 2002: an analysis of routine data. Lancet. 2006;367:52–6.
2. Department of Health. Safe, sensible, social – consultation on further action. London: Department of Health; 2008.
3. Williams R, Aspinall R, Bellis M, Camps-Walsh G, Cramp M, Dhawan A, et al. Addressing liver disease in the UK: a blueprint for attaining excellence in health care and reducing premature mortality from lifestyle issues of excess consumption of alcohol, obesity, and viral hepatitis. Lancet. 2014;384:1953–97.

4. Scaglione S, Kliethermes S, Cao G, Shoham D, Durazo R, Luke A, et al. The epidemiology of cirrhosis in the United States: a population-based study. J Clin Gastroenterol. 2015;49:690–6.
5. Whitlock G, Lewington S, Sherliker P, Clarke R, Emberson J, Halsey J, et al. Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies. Lancet. 2009;373:1083–96.
6. Ioannou GN, Weiss NS, Kowdley KV, Dominitz JA. Is obesity a risk factor for cirrhosis-related death or hospitalization? A population-based cohort study. Gastroenterology. 2003;125:1053–9.
7. Batty GD, Shipley MJ, Kivimaki M, Barzi F, Smith GD, Mitchell R, et al. Obesity and overweight in relation to liver disease mortality in men: 38 year follow-up of the original Whitehall study. Int J Obes. 2008;32:1741–4.
8. Marchesini G, Brizi M, Bianchi G, Tomassetti S, Bugianesi E, Lenzi M, et al. Nonalcoholic fatty liver disease: a feature of the metabolic syndrome. Diabetes. 2001;50:1844–50.
9. Liu B, Balkwill A, Reeves G, Beral V. Body mass index and risk of liver cirrhosis in middle aged UK women: prospective study. BMJ. 2010;340:c912.
10. Hart CL, Batty GD, Morrison DS, Mitchell RJ, Smith GD. Obesity, overweight and liver disease in the Midspan prospective cohort studies. Int J Obes. 2010;34:1051–9.
11. Garcia-Tsao G, Friedman S, Iredale J, Pinzani M. Now there are many (stages) where before there was one: in search of a pathophysiological classification of cirrhosis. Hepatology. 2010;51:1445–9.
12. Diehl AM. Obesity and alcoholic liver disease. Alcohol. 2004;34:81–7.
13. Serfaty L, Poujol-Robert A, Carbonell N, Chazouillères O, Poupon RE, Poupon R. Am J Gastroenterol. 2002;97:1807–12.
14. Menon U, Gentry-Maharaj A, Hallett R, Ryan A, Burnell M, Sharma A, et al. Sensitivity and specificity of multimodal and ultrasound screening for ovarian cancer, and stage distribution of detected cancers: results of the prevalence screen of the UK collaborative trial of ovarian cancer screening (UKCTOCS). Lancet Oncol. 2009;10:327–40.
15. Menon U, Gentry-Maharaj A, Ryan A, Sharma A, Burnell M, Hallett R, et al. Recruitment to multicentre trials–lessons from UKCTOCS: descriptive study. BMJ. 2008;337:a2079.
16. Menon U, Ryan A, Kalsi J, Gentry-Maharaj A, Dawnay A, Habib M, et al. Risk algorithm using serial biomarker measurements doubles the number of screen-detected cancers compared with a single-threshold rule in the United Kingdom collaborative trial of ovarian cancer screening. J Clin Oncol. 2015;33:2062–71.
17. Drinkaware. Alcohol unit guidelines. Drinkaware, London. www.drinkaware.co.uk. Accessed 21 February 2017.
18. Burnell M, Gentry-Maharaj A, Ryan A, Apostolidou S, Habib M, Kalsi J, et al. Impact on mortality and cancer incidence rates of using random invitation from population registers for recruitment to trials. Trials. 2011;12:61.
19. Day CP, James OFW. Steatohepatitis: a tale of two "hits"? Gastroenterology. 1998;114:842–5.
20. Bolton-Smith C, Woodward M, Tunstall-Pedoe H, Morrison C. Accuracy of the estimated prevalence of obesity from self reported height and weight in an adult Scottish population. J Epidemiol Community Health. 2000;54:143–8.
21. Hill A, Roberts J. Body mass index: a comparison between self-reported and measured height and weight. J Public Health Med. 1998;20:206–10.
22. Nieto-Garcia FJ, Bush TL, Keyl PM. Body mass definitions of obesity: sensitivity and specificity using self-reported weight and height. Epidemiology. 1990;1:146–52.
23. Roberts RJ. Can self-reported data accurately describe the prevalence of overweight? Public Health. 1995;109:275–84.
24. Kuskowska-Wolk A, Karlsson P, Stolt M, Rössner S. The predictive validity of body mass index based on self-reported weight and height. Int J Obes. 1989;13:441–53.
25. Midanik L. The validity of self-reported alcohol consumption and alcohol problems: a literature review. Br J Addict. 1982;77:357–82.
26. Romelsjö A, Leifman H, Nyström S. A comparative study of two methods for the measurement of alcohol consumption in the general population. Int J Epidemiol. 1995;24:929–36.
27. Parker DR, Derby CA, Usner DW, Gonzalez S, Lapane KL, Carleton RA. Self-reported alcohol intake using two different question formats in southeastern New England. Int J Epidemiol. 1996;25:770–4.
28. Goddard E. Estimating alcohol consumption from survey data: updated method of converting volumes to units. National Statistics methodological series no. 37. Office of National Statistics. London: HMSO; 2007.
29. Pinsky PF, Miller A, Kramer BS, Church T, Reding D, Prorok P, et al. Evidence of a healthy volunteer effect in the prostate, lung, colorectal, and ovarian cancer screening trial. Am J Epidemiol. 2007;165:874–81.
30. Hepatitis C in the UK 2015 report. London: Public Health England; 2015.
31. Ajmera VH, Terrault NA, Harrison SA. Is moderate alcohol use in non-alcoholic fatty liver disease good or bad? A critical review. Hepatology. 2017; doi:10.1002/hep.29055. [Epub ahead of print]
32. Hagström H, Nasr P, Ekstedt M, Kechagias S, Önnerhag K, Nilsson E, et al. Low to moderate lifetime alcohol consumption is associated with less advanced stages of fibrosis in non-alcoholic fatty liver disease. Scand J Gastroenterol. 2017;52:159–65.
33. Sogabe M, Okahisa T, Taniguchi T, Tomonari T, Tanaka T, Tanaka H, et al. Light alcohol consumption plays a protective role against non-alcoholic fatty liver disease in Japanese men with metabolic syndrome. Liver Int. 2015;35:1707–14.
34. Davies MJ, Baer DJ, Judd JT, Brown ED, Campbell WS, Taylor PR. Effects of moderate alcohol intake on fasting insulin and glucose concentrations and insulin sensitivity in postmenopausal women: a randomized controlled trial. JAMA. 2002;287:2559–62.
35. Askgaard G, Grønbæk M, Kjær MS, Tjønneland A, Tolstrup JS. Alcohol drinking pattern and risk of alcoholic liver cirrhosis: a prospective cohort study. J Hepatol. 2015;62:1061–7.
36. Hart CL, Morrison DS, Batty GD, Mitchell RJ, Davey SG. Effect of body mass index and alcohol consumption on liver disease: analysis of data from two prospective cohort studies. BMJ. 2010;340:c1240.
37. Naveau S, Giraud V, Borotto E, Aubert A, Capron F, Chaput JC. Excess weight risk factor for alcoholic liver disease. Hepatology. 1997;25:108–11.
38. Yi S-W, Hong J-S, Yi J-J, Ohrr H. Impact of alcohol consumption and body mass index on mortality from nonneoplastic liver diseases, upper aerodigestive tract cancers, and alcohol use disorders in Korean older middle-aged men. Medicine (Baltimore). 2016;95:e4876.
39. Grønbæk M, Deis A, Sørensen TI, Becker U, Schnohr P, Jensen G. Mortality associated with moderate intakes of wine, beer, or spirits. BMJ. 1995;310:1165–9.
40. Klatsky AL, Armstrong MA, Friedman GD. Alcohol and mortality. Ann Intern Med. 1992;117:646–54.
41. Shaper AG, Wannamethee G, Walker M. Alcohol and mortality in British men: explaining the U-shaped curve. Lancet. 1988;2:1267–73.
42. Grønbæk M, Deis A, Sørensen TI, Becker U, Borch-Johnsen K, Müller C, et al. Influence of sex, age, body mass index, and smoking on alcohol intake and mortality. BMJ. 1994;308:302–6.
43. Becker U, Grønbæk M, Johansen D, Sørensen TI. Lower risk for alcohol-induced cirrhosis in wine drinkers. Hepatology. 2002;35:868–75.
44. Becker U, Deis A, Sørensen TI, Grønbæk M, Borch-Johnsen K, Müller CF, et al. Prediction of risk of liver disease by alcohol intake, sex, and age: a prospective population study. Hepatology. 1996;23:1025–9.
45. Dunn W, Xu R, Schwimmer JB. Modest wine drinking and decreased prevalence of suspected nonalcoholic fatty liver disease. Hepatology. 2008;47:1947–54.
46. Dunn W, Sanyal AJ, Brunt EM, Unalp-Arida A, Donohue M, McCullough AJ, et al. Modest alcohol consumption is associated with decreased prevalence of steatohepatitis in patients with non-alcoholic fatty liver disease (NAFLD). J Hepatol. 2012;57:384–91.
47. Hayashi PH, Harrison SA, Torgerson S, Perez TA, Nochajski T, Russell M. Cognitive lifetime drinking history in nonalcoholic fatty liver disease: some cases may be alcohol related. Am J Gastroenterol. 2004;99:76–81.
48. Ruhl CE, Everhart JE. Joint effects of body weight and alcohol on elevated serum alanine aminotransferase in the United States population. Clin Gastroenterol Hepatol. 2005;3:1260–8.
49. Loomba R, Bettencourt R, Barrett-Connor E. Synergistic association between alcohol intake and body mass index with serum alanine and aspartate aminotransferase levels in older adults: the rancho Bernardo study. Aliment Pharmacol Ther. 2009;30:1137–49.
50. Loomba R, Yang HI, Su J, Brenner D, Iloeje U, Chen CJ. Obesity and alcohol synergize to increase the risk of incident hepatocellular carcinoma in men. Clin Gastroenterol Hepatol. 2010;8:891–8.
51. Loomba R, Yang HI, Su J, Brenner D, Barrett-Connor E, Iloeje U, et al. Synergism between obesity and alcohol in increasing the risk of hepatocellular carcinoma: a prospective cohort study. Am J Epidemiol. 2013;177:333–42.
52. James PT, Leach R, Kalamara E, Shayeghi M. The worldwide obesity epidemic. Obes Res. 2001;9(Suppl 4):228S–33S.

Trembling et al. BMC Public Health (2018) 18:409
https://doi.org/10.1186/s12889-018-5308-x

BMC Public Health

RESEARCH ARTICLE Open Access

Association between skirt size and chronic liver disease in post-menopausal women: a prospective cohort study within the United Kingdom Trial of Ovarian Cancer Screening (UKCTOCS)

P. M. Trembling[1*], S. Apostolidou[2], A. Gentry-Maharaj[2], J. Parkes[3], A. Ryan[2], S. Tanwar[1], M. Burnell[2], U. Menon[2†] and W. M. Rosenberg[1†]

Abstract

Background: We investigated the association between self-reported skirt size (SS) and change in SS, and incidence of chronic liver disease (CLD) in a prospective cohort study of women recruited to the UKCTOCS trial.

Methods: Women recruited to UKCTOCS in England without documented CLD self-reported their current UK SS during trial participation and were asked to recall their SS when aged in 20s (via completion of a questionnaire 3–5 years after recruitment). Participants were followed up via electronic health record linkage and hazard ratios (HR) calculated for incident liver-related events (LRE).

Results: Three hundred twenty-two (0.3%) of 94,124 women experienced a first LRE. Compared to SS ≤ 16, rates of LRE were higher in the SS ≥ 18 groups (both when aged in 20s and at questionnaire completion). Event rates were higher if there was no change in SS or an increase in SS, compared to a decrease in SS.
In the models adjusted for potential confounders, HRs for LRE were higher in the groups of women reporting SS ≥ 18 both when aged in 20s (HR = 1.39 (95% CI; 0.87–2.23)) and at questionnaire completion (HR = 1.37 (95% CI; 1.07–1.75)). Compared to a decrease in SS, HRs were higher in the no change (HR = 1.78 (95% CI; 0.95–3.34)) and increase (HR = 1.80 (95% CI; 1.01–3.21)) groups.

Conclusion: CLD is associated with high SS and an increase in SS over time. These data suggest SS can be used in simple public health messages about communicating the risk of liver disease.

Trial Registration: UKCTOCS is registered as an International Standard Randomised Controlled Trial, number ISRCTN22488978. Registered 06/04/2000.

Keywords: Chronic liver disease, Cirrhosis, Skirt size, Body mass index, Obesity, UKCTOCS

* Correspondence: p.trembling@ucl.ac.uk
†Equal contributors
[1]Institute for Liver and Digestive Health, Division of Medicine, University College London, Royal Free Hospital, Rowland Hill Street, NW3 2PF, London, UK
Full list of author information is available at the end of the article

Background

Chronic liver disease (CLD) is a leading cause of death in the UK. It is estimated that 60,000 people in the UK have cirrhosis [1, 2] but over half of those affected are unaware of the diagnosis [3]. The main causes of CLD are alcohol, non-alcoholic fatty liver disease (NAFLD) and viral hepatitis.

NAFLD describes the process of hepatic accumulation of fat, ranging from benign steatosis, via liver inflammation (steatohepatitis) to progressive liver fibrosis and eventually cirrhosis, and can be considered the pathological manifestation in the liver of the metabolic syndrome. In addition to type 2 diabetes, hypertension and hypercholesterolaemia, high body mass index (BMI) is a significant driver for NAFLD, and is associated with increased risk of heart disease and stroke [4]. Although BMI is commonly used as a measure of body fat, it has been demonstrated that waist and hip measurements may be stronger predictors of body fat than BMI [5, 6].

We have previously demonstrated the association between increasing BMI and risk of CLD. In a large cohort of post-menopausal women we observed more clinical events attributable to cirrhosis amongst women who were overweight or obese compared to those with a normal BMI. Although there was no evidence of significant interaction between alcohol and BMI, the highest risk of liver disease was seen in women who were overweight or obese and consumed the most alcohol [7].

Skirt size (SS) could be an easily understood surrogate for BMI to communicate public health messages about the risks of obesity. Increase in self-reported SS in participants in the United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS) has been shown to be associated with increased breast cancer risk. A unit increase in UK SS (e.g. 12 to 14) every 10 years between 25 and postmenopausal age is associated with postmenopausal breast cancer risk of 33%. Validation of these results could provide women with a simple and easy to understand message, using SS [8]. We now explore the association between SS and change in SS and the incidence of CLD.

Methods

Study population

This prospective cohort study was nested in UKCTOCS, a UK-based randomised controlled trial investigating the effect of ovarian cancer screening on mortality. The trial design is described elsewhere [7–11]; briefly, between April 2001 and October 2005, post-menopausal women aged 50–74 in England, Wales and Northern Ireland were invited at random and 202,638 participants recruited to the trial. Participants were randomly allocated to one of three arms (no screening, annual serum CA125 measurement and then transvaginal ultrasound as a second line test, or ultrasound only). Recent data from the trial have demonstrated the predictive value of changes in CA125 levels to predict ovarian cancer [12], and reduced mortality in the multimodal arm [13].

UKCTOCS was approved by the UK North West Multicentre Research Ethics Committee (North West MREC 00/8/34). All women provided written consent. The current study was approved by the National Research Ethics Service (NRES) Committee London - Bentham (Ref: 05/Q0505/57) on 10th August 2011.

Exposures

The exposures of interest in this study were BMI and SS of participants. At the time of recruitment, participants completed a questionnaire, which included self-reported height and weight. BMI was calculated (BMI (kg/m^2) = weight (kg)/height $(m)^2$) and categorised according to the World Health Organization's definitions; normal (< 25 kg/m^2), overweight (25- < 30 kg/m^2) or obese (≥30 kg/m^2). There were some extreme values in self-reported data and as there are no existing population estimates for the range of BMI we applied a pragmatic approach in order to include participants with biologically plausible BMI values. Therefore, participants were excluded if their reported height lay outside the range 140–210 cm, or their weight lay outside the range 25–200 kg, or where the calculated BMI was outside the range 16–65 kg/m^2.

Participants were asked to complete a follow-up questionnaire 3–5 years post randomisation, and were asked to estimate their UK SS when they were in their early twenties and to report their current SS. Using the two SS responses overall change in SS and change in SS per year were calculated. In the UK SS range comprises of even numbers, for example in increase in SS from 12 to 14 is an increase in one UK SS.

Categorisation of exposure variables

BMI was categorised according to World Health Organization classification as normal, overweight, or obese. SS was categorised using UK dress sizes as ≤16 and ≥ 18; the latter cut-off selected because of its association with an increased risk of cardiovascular morbidity [14]. The British Standards Institution defines UK size 16 as 100–104 cm, and size 18 as 105–109 cm, measured at the hips [15]. Change in SS was categorised as decrease, no change or increase in SS between when participants were in their early 20s and at their current age.

Covariates

Participants reported, via the follow-up questionnaire, known comorbidities, comprising hypertension, heart disease, hypercholesterolaemia, stroke, diabetes, rheumatoid arthritis, osteoarthritis, osteoporosis ("do you have/are you being treated for any of the following conditions?"), and whether they currently smoked, all categorised as yes/no.

Participants were asked "approximately how much alcohol on average do you drink each week, assuming one drink = a glass of wine, half a pint of lager or cider, a measure of spirits?" This was then categorised as none, < 1–15 units/week, 16–20 units/week and ≥ 21 units/week, assuming one drink is equivalent to one UK unit (10 ml or 8 g of pure alcohol) [16].

Participants were assigned a deprivation score using the Index of Multiple Deprivation 2007 (IMD) (continuous variable) [17], with a higher score indicating higher deprivation.

Follow up

Participants in this study were followed through a 'flagging' study with NHS Digital which provided data on cancer registrations and deaths, with diagnosis and/or cause of death coded according to the International Classification of Diseases, version 10 (ICD-10). Hospital inpatient and outpatient data for 2001–10 were also available through linkage to the Hospital Episodes Statistics (HES) database. Each inpatient HES episode record reports a main diagnosis and up to 19 additional diagnoses. Outpatient records report a main diagnosis and up to 11 further diagnoses. Death records report the primary death code and additional diagnoses documented on the death certificate, comprising both ICD-10 codes and free text. Only participants in England were included in this study, due to availability of their relevant HES data. Participants entered the study at the point of return of the follow-up questionnaire, as this was the date that current comorbidities and SS data were ascertained. Women with pre-existing liver disease were excluded if a code of interest had been registered between recruitment to UKCTOCS and return of follow-up questionnaire.

Outcome

First liver-related event (LRE) was deemed the main outcome measure. LRE was defined as a participant's first presentation of a hospital admission, outpatient appointment or cancer registration with, or death from, a relevant ICD-10 code. These codes were K70 (alcoholic liver disease), K73 (chronic hepatitis) and K74 (fibrosis and cirrhosis) and are consistent with codes employed in other UK studies of cirrhosis [1, 18]. In addition K76 (other diseases of the liver, including fat) and codes related to decompensated liver disease (I85 (oesophageal varices), Z94.4 (liver transplant) and C22.0 (hepatocellular carcinoma)) were included. Death certificates were interrogated for ICD-10 codes of interest and free text relating to alcoholic liver disease or fatty liver.

Statistical analysis

For the incidence analyses person-years of follow-up was used as the denominator. Participants contributed person-years until the date of censoring (February 1, 2013), date of first presentation with an LRE or death from any other cause.

Crude incidence was calculated for each BMI group, each SS when aged in 20s group, each SS at questionnaire completion group, and change in SS group.

Survival analysis

Potential confounding risk factors including self-reported comorbidities were analysed in univariate Cox proportional hazards models to determine their individual risks in liver disease.

Cox proportional hazards models were used to calculate hazard ratios (HRs) of first LRE, with 95% confidence intervals (CI). For each exposure described above, BMI, SS when aged in 20s and SS at questionnaire completion were analysed as continuous coavariates, and then BMI, SS when aged in 20s, SS at questionnaire completion and overall change in SS as categorical covariates. For each outcome, univariate models were produced. Smoking and deprivation were then added (partially adjusted), and then all covariates listed above were added, with abstinence and alcohol consumption ≥21 units/week as individual indicator variables (fully adjusted).

All analyses were performed using SPSS (version 22, SPSS Inc., Chicago, IL, USA).

Results

Sample characteristics

Of the 157,996 UKCTOCS participants resident in England, 62,870 were excluded including 321 women who experienced an LRE or died between recruitment and return of questionnaire and 14,295 (9%) with no data on smoking. There was some missing SS data, and the resulting effective sample size for this study was 94,124 (Fig. 1).

Overall, 97% of the participants were Caucasian, 36% were smokers, 55% were overweight (37%) or obese (18%). Median age at completion of the questionnaire was 64 years. Baseline characteristics are shown in Table 1.

Distributions of BMI and skirt size

The distributions of BMI, SS when aged in 20s, SS at questionnaire completion and annual change in SS are shown in Fig. 2. Median BMI was 25.57 kg/m^2 (IQR 22.79–28.36), median SS when aged in 20s was 12 (IQR 10–14), median SS at questionnaire completion was 14 (IQR 12–16), and the median change in SS unit per year was 0.0323 (IQR 0.0123–0.0523). This is the equivalent to an increase of one SS unit (e.g. from 12 to 14) every 31 years.

Visual inspection of the histograms (Fig. 2), quantile-quantile plots and box plots for each outcome variable showed that each variable was approximately normally

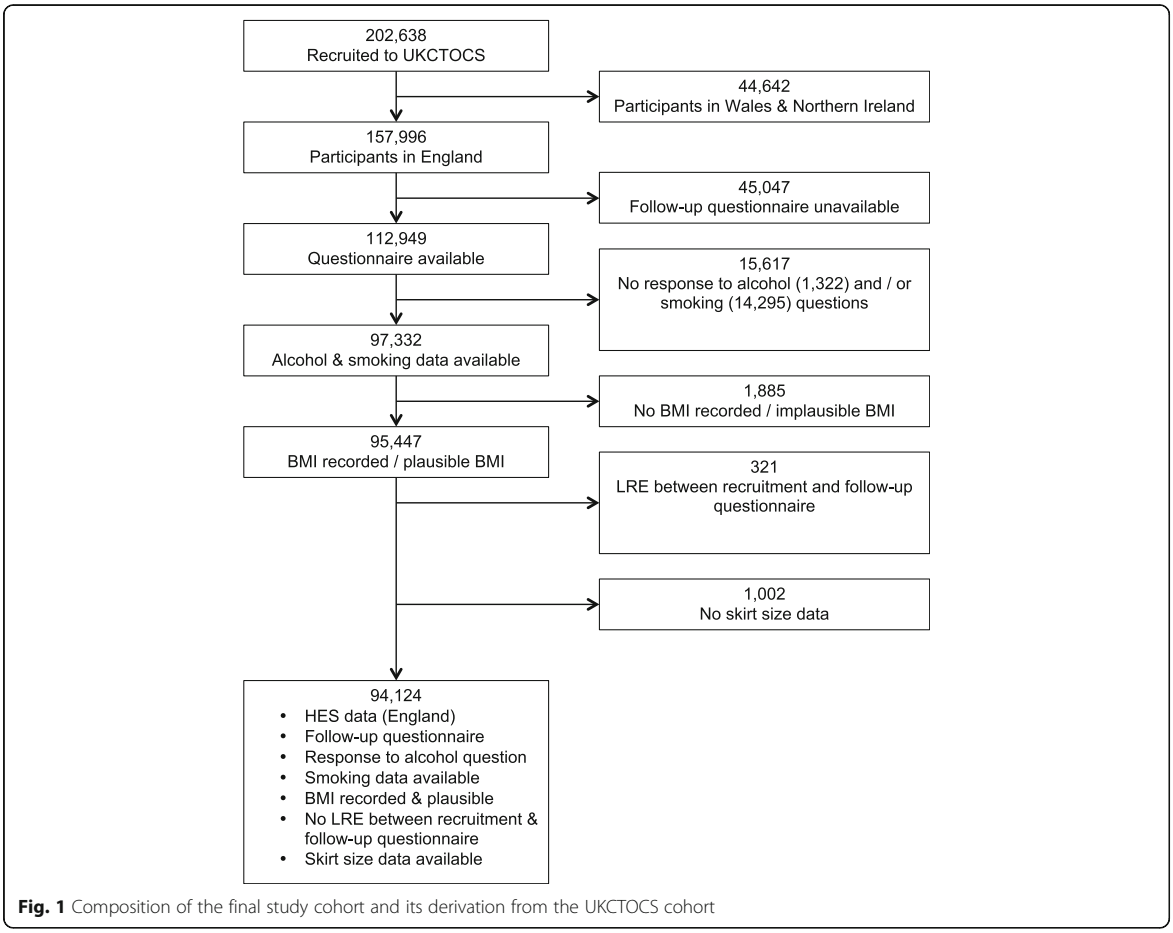


Fig. 1 Composition of the final study cohort and its derivation from the UKCTOCS cohort

distributed, but with right-skewness seen with BMI, SS when aged in 20s and SS at questionnaire completion (BMI – skewness 1.368 (SE = 0.008), kurtosis 4.033 (SE = 0.016); SS when aged in 20s – skewness 1.442 (SE = 0.008), kurtosis 5.787 (SE = 0.016); SS at questionnaire completion – skewness 0.999 (SE = 0.008), kurtosis 2.415 (SE = 0.016); change in SS per year – skewness 0.470 (SE = 0.008), kurtosis 3.095 (SE = 0.016)).

Crude event rates

Three hundred and twenty two (0.34%) women experienced a first LRE over the follow up period. Crude rates of LRE are shown in Table 2, categorised by BMI, SS when aged in 20s, SS at questionnaire completion and overall change in SS. The most common incident ICD-10 code was K76 (Additional file 1: Table S1).

The rate of LRE increased with increasing BMI. Comparison of rates of LREs in SS categories found a higher incidence in participants with SS ≥ 18, compared to participants with SS ≤ 16, both in the SS when aged in 20s group and the SS at questionnaire completion group. In terms of overall change in SS, event rate was lowest in the group where SS decreased. The rate was higher if there was no change, and highest if there was an increase in SS (Fig. 3).

Survival analysis

Cox proportional model estimates for each potential confounder

There were significant associations between LRE and smoking, deprivation, BMI, heart disease, hypercholesterolaemia, diabetes, rheumatoid arthritis, alcohol abstinence and alcohol excess (≥21 units/week) (Additional file 1: Table S2). A "J-shaped" relationship between alcohol and risk of CLD is seen, and we have previously explored this finding in the UKCTOCS population [7].

Cox proportional model estimates for each exposure

When SS when aged in 20s ≥18 was compared to ≤16, HR for LRE was increased in the unadjusted (HR = 1.81

Table 1 Baseline characteristics and number of first events according to BMI category, and for all participants

Characteristic	BMI category (kg/m^2)			All participants
	< 25	25 - < 30	≥18	
Total, *n* (% of all participants)	42, 077 (44.7)	34,690 (36.9)	17,260 (18.3)	94,124
LRE, *n* (% of all participants)	102 (31.7)	123 (38.2)	97 (30.1)	322
Age at questionnaire return, median years (range)	63 (52–80)	64 (53–80)	64 (53–80)	64 (52–80)
IMD, mean (SD)	17.0 (13.1)	18.7 (14.1)	21.3 (15.4)	18.4 (14.0)
Smoker, *n* (%)	14,632 (34.8)	12,511 (36.1)	6548 (37.7)	33,691 (35.8)
Hypertension, *n* (%)	9382 (22.3)	11,970 (34.5)	8307 (47.9)	29,659 (31.5)
Heart disease, *n* (%)	1698 (4.0)	2052 (5.9)	1392 (8.0)	5142 (5.5)
Hypercholesterolaemia, *n* (%)	7901 (18.8)	9044 (26.1)	5369 (30.9)	22,314 (23.7)
Stroke, *n* (%)	523 (1.2)	552 (1.6)	314 (1.8)	1389 (1.5)
Diabetes, *n* (%)	827 (2.0)	1653 (4.8)	2221 (12.8)	4701 (5.0)
Rheumatoid arthritis, *n* (%)	1592 (3.8)	1742 (5.0)	1185 (6.8)	4519 (4.8)
Osteoarthritis, *n* (%)	5503 (13.1)	5822 (16.8)	4016 (23.1)	15,341 (16.3)
Osteoporosis, *n* (%)	3808 (9.1)	2082 (6.0)	770 (4.4)	6660 (7.1)
Alcohol consumption (units/week), *n* (%)				
None	8365 (19.9)	8.043 (23.2)	5432 (31.3)	21,840 (23.2)
< 1–15	31,567 (75.0)	25,095 (72.3)	11,347 (65.4)	68,009 (72.3)
16–20	1436 (3.4)	1063 (3.1)	364 (2.1)	2863 (3.0)
≥ 21	709 (1.7)	489 (1.4)	214 (1.2)	1412 (1.5)
Skirt size when aged in 20s, *n* (%)				
≤ 16	41,428 (98.5)	33,835 (97.5)	15,691 (90.4)	90,954 (96.6)
≥ 18	649 (1.5)	855 (2.5)	1666 (9.6)	3170 (3.4)
Skirt size at time of questionnaire completion, *n* (%)				
≤ 16	40,792 (96.9)	26,982 (77.8)	4481 (25.8)	72,255 (76.8)
≥ 18	1285 (3.1)	7708 (22.2)	12,876 (74.2)	21,869 (23.2)
Change in skirt size, median (interquartile range)	0.0244 (0.03)	0.0408 (0.04)	0.0667 (0.05)	0.0323 (0.04)
Overall change in skirt size, *n* (%)				
Decrease	4811 (11.4)	1153 (3.3)	362 (2.1)	6326 (6.7)
No change	12,344 (29.3)	3422 (9.9)	731 (4.2)	16,497 (17.5)
Increase	24,922 (59.2)	30,115 (86.8)	16,264 (93.7)	71,301 (75.8)

(95% CI; 1.14–2.87)), partially adjusted (HR = 1.68 (95% CI; 1.06–2.68)) and fully adjusted (HR = 1.39 (95% CI; 0.87–2.23)) models. The confidence interval for the fully adjusted model crossed unity, suggesting that a component of the risk may be partially attributable to one or more of the metabolic comorbidities (hypertension, hypercholesterolaemia, diabetes and heart disease) (Table 3). Comparing the two SS groups at questionnaire completion, HRs were again higher in the higher SS group in all models (HR = 1.69 (95% CI; 1.34–2.13) in the unadjusted model, HR = 1.58 (95% CI; 1.25–2.00) in the partially adjusted model, HR = 1.37 (95% CI; 1.07–1.75) in the fully adjusted model).

Compared to women whose SS decreased between their 20s and questionnaire completion, HRs were higher in those whose SS did not change and highest in those whose SS increased (Table 3).

Compared to normal BMI, overweight and obesity were significantly associated with LRE in all models (Table 3).

Discussion

Main findings

We have demonstrated in a cohort of post-menopausal women that a larger SS is associated with subsequent risk of LRE, and a SS of ≥18 compared to a SS of ≤16 is associated with a higher HR than that associated with overweight, but less than that associated with obesity when compared to a normal BMI. Although the risks of high SS and high BMI may not be directly comparable, the value of communicating public health messages in terms of SS lies in better understanding amongst the general public compared to communicating the risk of liver disease associated with increased BMI.

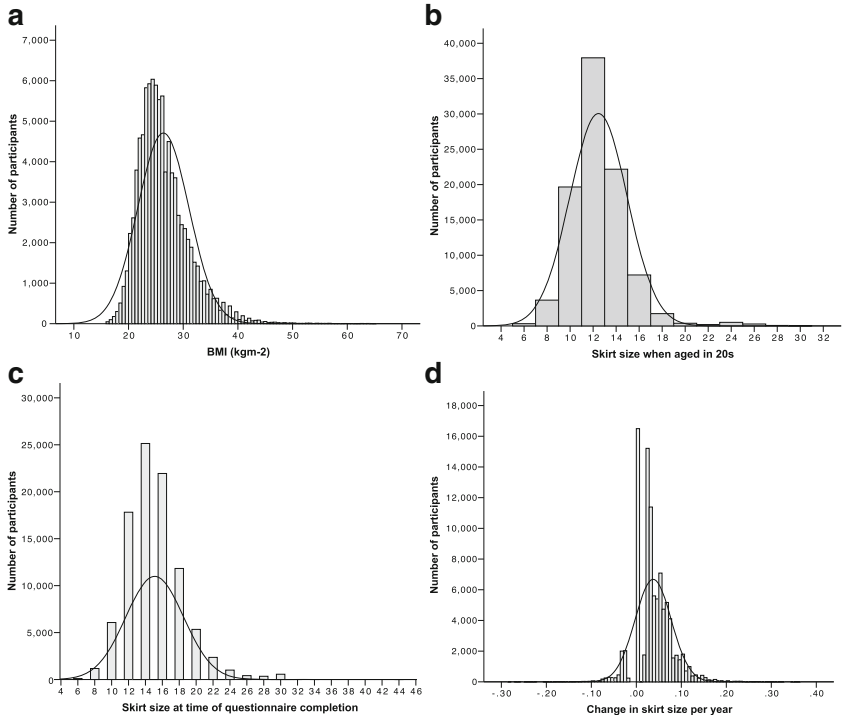


Fig. 2 Distributions of **a**) BMI, **b**) skirt size in 20s, **c**) skirt size at questionnaire completion, and **d**) change in skirt size per year

In our cohort, 76% reported an increase in SS between when aged in 20s and questionnaire completion. This is consistent with previous studies reporting the change in body composition associated with transitioning from pre-menopausal to post-menopausal status, with an increase in central adiposity manifested by increased waist circumference (WC) [19].

When BMI and SS (as continuous variables) were combined, the HR for each was reduced, suggesting that SS (and BMI) is an independent predictor for NAFLD, and that SS may reflect centripetal fat distribution associated with NAFLD better than BMI.

Table 2 Crude rates of liver-related event. Events per 1000 participant years (95% confidence intervals)

Exposure		Event rate
BMI (kg/m^2)	< 25	0.453 (0.369–0.550)
	25- < 30	0.661 (0.549–0.788)
	≥30	1.044 (0.847–1.273)
Skirt size in 20s	≤ 16	0.621 (0.553–0.696)
	≥ 18	1.124 (0.677–1.755)
Skirt size at questionnaire completion	≤ 16	0.550 (0.479–0.629)
	≥ 18	0.928 (0.762–1.120)
Change in skirt size/year	Decrease	0.3867 (0.206–0.661)
	No change	0.599 (0.449–0.784)
	Increase	0.669 (0.590–0.757)

NAFLD is poorly identified in primary care and it is conceivable that a proportion of individuals with LREs that were not associated with an ICD-10 code for fatty liver may have had NAFLD. SS may be a better predictor of NAFLD (obesity) related liver disease than a clinical diagnosis of NAFLD in primary care.

Although the codes or text contributing most commonly to LRE were those representing NAFLD, those representing alcoholic liver disease contributed to nearly 10% of LREs (Additional file 1: Table S1). Regardless of the aetiology of CLD, the clinicopathological pathway is progressive fibrosis leading to cirrhosis [20] and there may be common pathways in which alcohol and BMI damage the liver [21]. Patterns of alcohol consumption in women are changing; 16% of women in England consume above recommended limits, and this practice is highest in the 55–64 year old group [22], and the rate of alcohol-related hospital admissions by women increased by over 30% between 2008 and 2015 [23].

Strengths and limitations

Strengths of this study include the large size of the cohort, the prospective design and the independence of data capture for outcomes. We used ICD-10 codes for cirrhosis

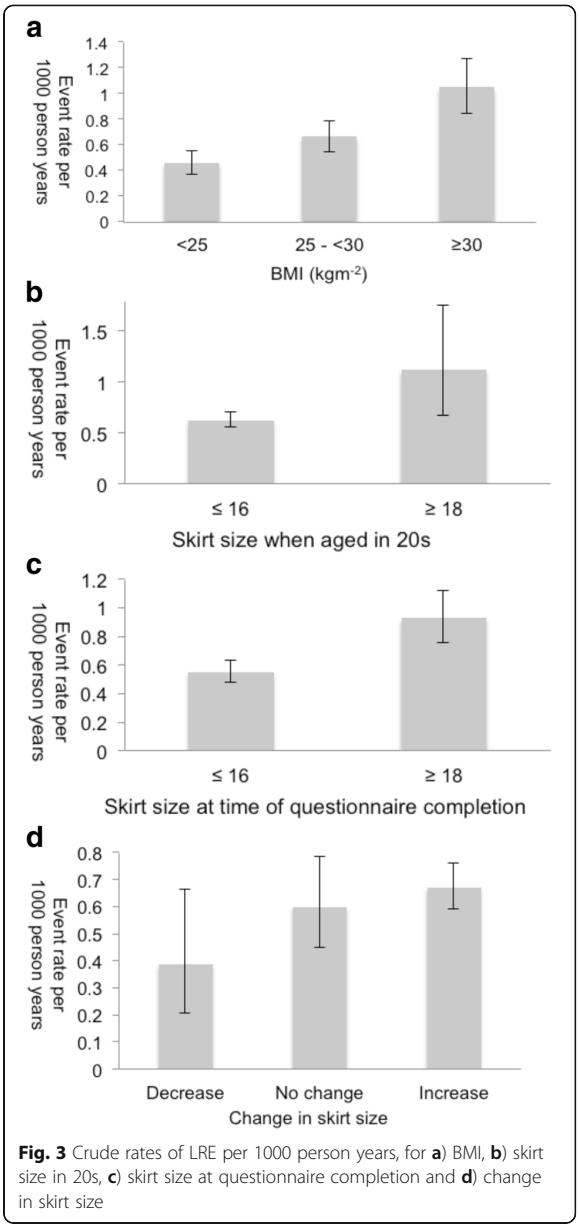


Fig. 3 Crude rates of LRE per 1000 person years, for **a**) BMI, **b**) skirt size in 20s, **c**) skirt size at questionnaire completion and **d**) change in skirt size

that have been used in other studies, but in an attempt to maximise the ability to identify liver disease we also included codes relating to clinical consequences of advanced cirrhosis, the events defining decompensated liver disease. Evaluation of numerous possible confounders including self reported known comorbidities and socioeconomic status minimised bias.

Limitations include the reliance of self-reporting of SS, height and weight and co-morbidities. There is some evidence supporting the reliability of self-reporting of biometric data including height and weight [24–28], notably in a longitudinal study of older people [29]. There was a 30–50 year recall of participants' SS when aged in their 20s, raising the possibility of recall error. Several studies have demonstrated good accuracy in recalled weight, with some data indicating underestimation in those with higher BMI [30–33]. We postulate that participants may have a better recollection of their skirt size than their weight or waist size. There was a 25 year age range in participants, and older participants may have had children at a younger age than younger participants, which may have increased their SS [34].

It is likely that there will be some variability between SS over the period between the two SS estimates. In the UK there is no requirement for manufacturers to adhere to the standard sizing. In addition the phenomenon of vanity sizing is recognised, where clothes with the same size label have become larger over recent decades. This has become a common practice of clothing manufacturers, which may potentially impede comparisons of sizes over time [35]. Indeed, the Chief Medical Officer for England has highlighted this 'size inflation' as a risk for society normalising overweight [36].

Reliance on ICD-10 to define events may result in errors due to mis-coding. We used three independent sources in an attempt to reduce risk of non-coding. Further, HES data may not capture some areas of healthcare, including the private sector. Finally, although attempts were made to ensure UKCTOCS was representative of the general population, there was a 'healthy volunteer effect' on overall and cause-specific mortality, which may affect the generalisability of our findings [17].

Other studies

The link between obesity and the risk of NAFLD is strong, with a clear dose-response relationship demonstrated in cross-sectional studies [37], although data from prospective studies are limited [38, 39].

However, few studies have investigated the relationship between SS and disease. Ours is the only study we are aware of that has investigated the association between SS and liver disease. The UKCTOCS group demonstrated an increase in risk of breast cancer with increase in SS over time [12].

A study nested in the Netherlands Cohort Study on Diet and Cancer reported increased risk of endometrial cancer with increasing SS. The correlation between self-reported SS and self-reported WC, self-reported hip circumferences and BMI based on self-reported height and weight in 1334 women, were 0.71, 0.78 and 0.76 respectively [40].

A study of 293 men and women found that professionally measured WC correlated closely with clothing size in both men and women ($r = 0.80$ and 0.78, respectively) [41].

Table 3 Hazard ratios of first events for skirt size in 20s, skirt size at questionnaire completion, BMI and change in skirt size (95% confidence intervals and *p* values)

Variable				Hazard ratio (95% CI)
Skirt size when aged in 20s	Univariate	Continuous		1.062 (1.022–1.104) $p = 0.002$
		Categorical	≤ 16	Reference
			≥ 18	1.806 (1.136–2.871) $p = 0.012$
	Adjusted for smoking, deprivation		≤ 16	Reference
			≥ 18	1.681 (1.057–2.675) $p = 0.028$
	Adjusted for age, smoking, deprivation, hypertension, heart disease, hypercholesterolaemia, stroke, diabetes, rheumatoid arthritis, osteoarthritis, osteoporosis, alcohol abstinence, alcohol ≥21 units/week		≤ 16	Reference
			≥ 18	1.390 (0.868–2.226) $p = 0.171$
Skirt size at time of questionnaire completion	Univariate	Continuous		1.091 (1.062–1.121) $p < 0.0005$
		Categorical	≤ 16	Reference
			≥ 18	1.690 (1.342–2.129) $p < 0.0005$
	Adjusted for smoking, deprivation		≤ 16	Reference
			≥ 18	1.579 (1.250–1.995) $p < 0.0005$
	Adjusted for age, smoking, deprivation, hypertension, heart disease, hypercholesterolaemia, stroke, diabetes, rheumatoid arthritis, osteoarthritis, osteoporosis, alcohol abstinence, alcohol ≥21 units/week		≤ 16	Reference
			≥ 18	1.369 (1.071–1.749) $p = 0.012$
BMI (kg/m^2)	Univariate	Continuous		1.063 (1.044–1.082) $p < 0.0005$
		Categorical	< 25	Reference
			≥ 25 - < 30	1.461 (1.123–1.899) $p = 0.005$
			≥ 30	2.308 (1.748–3.047) $p < 0.0005$
	Adjusted for smoking, deprivation		< 25	Reference
			≥ 25 - < 30	1.403 (1.076–1.830) $p = 0.012$
			≥ 30	2.162 (1.631–2.864) $p < 0.0005$
	Adjusted for age, smoking, deprivation, hypertension, heart disease, hypercholesterolaemia, stroke, diabetes, rheumatoid arthritis, osteoarthritis, osteoporosis, alcohol abstinence, alcohol ≥21 units/week		< 25	Reference
			≥ 25 - < 30	1.353 (1.034–1.770) $p = 0.028$
			≥ 30	1.880 (1.395–2.533) $p < 0.0005$
Change in skirt size/year	Univariate	Categorical	Decrease	Reference
			No change	1.554 (0.847–2.850) $p = 0.155$
			Increase	1.736 (0.994–3.031) $p = 0.052$
	Adjusted for smoking, deprivation		Decrease	Reference
			No change	1.714 (0.915–3.211) $p = 0.092$

Table 3 Hazard ratios of first events for skirt size in 20s, skirt size at questionnaire completion, BMI and change in skirt size (95% confidence intervals and *p* values) *(Continued)*

Variable			Hazard ratio (95% CI)
		Increase	1.873 (1.050–3.343) $p = 0.034$
	Adjusted for age, smoking, deprivation, hypertension, heart disease, hypercholesterolaemia, stroke, diabetes, rheumatoid arthritis, osteoarthritis, osteoporosis, alcohol abstinence, alcohol ≥21 units/week	Decrease	Reference
		No change	1.781 (0.950–3.337) $p = 0.072$
		Increase	1.799 (1.007–3.214) $p = 0.047$

Similarly, a study nested in the fourth Glasgow monitoring cardiovascular (MONICA) disease risk factor survey measured height, weight, WC and hip circumference, and obtained SS in 161 women. Dress size correlated with WC and BMI. Dress size ≥18 was associated with a significantly increased risk of cardiovascular disease [14].

Conclusion

We have demonstrated that SS in middle age is associated with increased risk of CLD. In post-menopausal women who develop liver disease, there is a significantly higher average SS when aged in their 20s (and in middle age). If these results are confirmed in further population studies, this may provide a simple way for women to stratify their risk of liver disease.

Additional file

Additional file 1: Table S1. ICD-10 codes and death certificate text of first LREs. Summary of the ICD-10 code(s) representing first presentation of liver-related event. **Table S2.** Hazard ratios for liver-related events for potential confounders (95% confidence intervals and *p* values). Univariate hazard ratios for liver-related events for smoking, deprivation, alcohol categories, alcohol ≥21 units/week, abstinence from alcohol, BMI, hypertension, heart disease, hypercholesterolaemia, stroke, diabetes, rheumatoid arthritis, osteoarthritis, osteoporosis. (DOCX 95 kb)

Abbreviations

BMI: Body mass index; CLD: Chronic liver disease; HES: Hospital episode statistics; HR: Hazard ratio; ICD-10: International classification of disease 10th revision; IMD: Index of multiple deprivation; IQR: Interquartile range; LRE: Liver-related event; NAFLD: Non-alcoholic fatty liver disease; NHS: National Health Service; SE: Standard error; SS: Skirt size; UKCTOCS: United Kingdom Collaborative Trial of Ovarian Cancer Screening; WC: Waist circumference

Acknowledgements

The authors are particularly grateful to the women throughout the UK who are participating in the trial and to the entire medical, nursing and administrative staff who worked on UKCTOCS.

Funding

UKCTOCS was core funded by the Medical Research Council, Cancer Research UK and National Institute of Health Research (NIHR), with additional support from the Special Trustees of Bart's and the London, the Special Trustees of University College London Hospitals (UCLH) and the Eve Appeal. Researchers at University College London are supported by the NIHR University College London Hospitals (UCLH) Biomedical Research Centre. WMR is a NIHR Senior Investigator. The funding source or the sponsor had no role in data collection, data analysis, data interpretation or writing of the report. The researchers are independent from the funders. No external funding for this nested study.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Authors' contributions

All authors were involved in study design. PMT, SA, AGM, UM and WMR were involved in data collection. PMT, UM and WMR drafted the manuscripts and figs. PMT undertook the literature search. PMT, JP, ST, UM, WMR and MB performed the statistical analysis. All authors critically revised the manuscript and approved the final version.

Ethics approval and consent to participate

UKCTOCS was approved by the UK North West Multicentre Research Ethics Committee (North West MREC 00/8/34). All women provided written consent. The current study was approved by the National Research Ethics Service (NRES) Committee London - Bentham (Ref: 05/Q0505/57) on 10th August 2011.

Consent for publication

Not applicable.

Competing interests

UM has a financial interest through Abcodia Ltd in the third party exploitation of the UKCTOCS biobank. None of the other authors declared any conflicts of interest.

Publisher's Note

Author details

[1]Institute for Liver and Digestive Health, Division of Medicine, University College London, Royal Free Hospital, Rowland Hill Street, NW3 2PF, London, UK. [2]Gynaecological Cancer Research Centre, Department of Women's Cancer, Institute for Women's Health, University College London, London, UK. [3]Public Health Sciences and Medical Statistics, Faculty of Medicine, University of Southampton, Southampton, UK.

Received: 25 September 2017 Accepted: 13 March 2018
Published online: 27 March 2018

References

1. Leon DA, McCambridge J. Liver cirrhosis mortality rates in Britain from 1950 to 2002: an analysis of routine data. Lancet. 2006;367:52–6.
2. Williams R, Aspinall R, Bellis M, Camps-Walsh G, Cramp M, Dhawan A, et al. Addressing liver disease in the UK: a blueprint for attaining excellence in health care and reducing premature mortality from lifestyle issues of excess consumption of alcohol, obesity, and viral hepatitis. Lancet. 2014;384:1953–97.
3. Scaglione S, Kliethermes S, Cao G, Shoham D, Durazo R, Luke A, et al. The epidemiology of cirrhosis in the United States: a population-based study. J Clin Gastroenterol. 2015;49:690–6.

4. Marchesini G, Brizi M, Bianchi G, Tomassetti S, Bugianesi E, Lenzi M, et al. Nonalcoholic fatty liver disease: a feature of the metabolic syndrome. Diabetes. 2001;50:1844–50.
5. Chang H, Simonsick EM, Ferrucci L, Cooper JA. Validation study of the body adiposity index as a predictor of percent body fat in older individuals: findings from the BLSA. J Gerontol A Biol Sci Med Sci. 2014;69:1069–75.
6. Lean ME, Han TS, Morrison CE. Waist circumference as a measure for indicating need for weight management. BMJ. 1995;311(6998):158–61.
7. Trembling PM, Apostolidou S, Gentry-Maharaj A, Parkes J, Ryan A, Tanwar S, et al. Risk of chronic liver disease in post-menopausal women due to body mass index, alcohol and their interaction: a prospective nested cohort study within the United Kingdom collaborative trial of ovarian Cancer screening (UKCTOCS). BMC Public Health. 2017;17(1):603.
8. Fourkala EO, Burnell M, Cox C, Ryan A, Salter LC, Gentry-Maharaj A, et al. Association of skirt size and postmenopausal breast cancer risk in older women: a cohort study within the UK collaborative trial of ovarian Cancer screening (UKCTOCS). BMJ Open. 2014;24(9):e005400.
9. Menon U, Gentry-Maharaj A, Hallett R, Ryan A, Burnell M, Sharma A, et al. Sensitivity and specificity of multimodal and ultrasound screening for ovarian cancer, and stage distribution of detected cancers: results of the prevalence screen of the UK collaborative trial of ovarian Cancer screening (UKCTOCS). Lancet Oncol. 2009;10:327–40.
10. Menon U, Gentry-Maharaj A, Ryan A, Sharma A, Burnell M, Hallett R, et al. Recruitment to multicentre trials–lessons from UKCTOCS: descriptive study. BMJ. 2008;337:a2079.
11. Gentry-Maharaj A, Glazer C, Burnell M, Ryan A, Berry H, Kalsi J, et al. Changing trends in reproductive/lifestyle factors in UK women: descriptive study within the UK collaborative trial of ovarian Cancer screening (UKCTOCS). BMJ Open. 2017;7:e011822.
12. Menon U, Ryan A, Kalsi J, Gentry-Maharaj A, Dawnay A, Habib M, et al. Risk algorithm using serial biomarker measurements doubles the number of screen-detected cancers compared with a single-threshold rule in the United Kingdom collaborative trial of ovarian Cancer screening. J Clin Oncol. 2015;33:2062–71.
13. Jacobs IJ, Menon U, Ryan A, Gentry-Maharaj A, Burnell M, Kalsi JK. Ovarian cancer screening and mortality in the UK collaborative trial of ovarian Cancer screening (UKCTOCS): a randomised controlled trial. Lancet. 2016; 387(10022):945–56.
14. Han TS, Gates E, Truscott E, Lean ME. Clothing size as an indicator of adiposity, ischaemic heart disease and cardiovascular risks. J Hum Nutr Diet. 2005;18:423–30.
15. British Standards Institution. BS 3666:1982 Specification for size designation of women's wear. London.
16. Drinkaware. Alcohol unit guidelines. Drinkaware, London. www.drinkaware.co.uk Accessed 21 Feb 2017.
17. Burnell M, Gentry-Maharaj A, Ryan A, Apostolidou S, Habib M, Kalsi J, et al. Impact on mortality and cancer incidence rates of using random invitation from population registers for recruitment to trials. Trials. 2011;12:61.
18. Liu B, Balkwill A, Reeves G, Beral V. Body mass index and risk of liver cirrhosis in middle aged UK women: prospective study. BMJ. 2010;340:c912.
19. Donato GB, Fuchs SC, Oppermann K, Bastos C, Spritzer PM. Association between menopause status and central adiposity measured at different cutoffs of waist circumference and waist-to-hip ratio. Menopause. 2006;13:280–5.
20. Garcia-Tsao G, Friedman S, Iredale J, Pinzani M. Now there are many (stages) where before there was one: in search of a pathophysiological classification of cirrhosis. Hepatology. 2010;51:1445–9.
21. Diehl AM. Obesity and alcoholic liver disease. Alcohol. 2004;34:81–7.
22. Health and Social Care Information Centre. Health Survey for England 2015: Adult alcohol consumption. HSIC, 2016.
23. Department of Health. Adult alcohol statistics from the National Drug Treatment Monitoring System (NDTMS) 1 April 2013 to 31 march 2014. London: Department of Health; 2014.
24. Bolton-Smith C, Woodward M, Tunstall-Pedoe H, Morrison C. Accuracy of the estimated prevalence of obesity from self reported height and weight in an adult Scottish population. J Epidemiol Community Health. 2000;54:143–8.
25. Hill A, Roberts J. Body mass index: a comparison between self-reported and measured height and weight. J Public Health Med. 1998;20(2):206–10.
26. Nieto-Garcia FJ, Bush TL, Keyl PM. Body mass definitions of obesity: sensitivity and specificity using self-reported weight and height. Epidemiology. 1990;1:146–52.
27. Roberts RJ. Can self-reported data accurately describe the prevalence of overweight? Public Health. 1995;109(4):275–84.
28. Kuskowska-Wolk A, Karlsson P, Stolt M, Rössner S. The predictive validity of body mass index based on self-reported weight and height. Int J Obes. 1989;13:441–53.
29. Dahl AK, Hassing LB, Fransson EI, Pedersen NL. Agreement between self-reported and measured height, weight and body mass index in old age–a longitudinal study with 20 years of follow-up. Age Ageing. 2010;39:445–51.
30. Stevens J, Keil JE, Waid LR, Gazes PC. Accuracy of current, 4-year, and 28-year self-reported body weight in an elderly population. Am J Epidemiol. 1990;132:1156–63.
31. Dahl AK, Reynolds CA. Accuracy of recalled body weight–a study with 20-years of follow-up. Obesity (Silver Spring). 2013;21:1293–8.
32. Perry GS, Byers TE, Mokdad AH, Serdula MK, Williamson DF. The validity of self-reports of past body weights by U.S. adults. Epidemiology. 1995;6:61–6.
33. Must A, Willett WC, Dietz WH. Remote recall of childhood height, weight, and body build by elderly subjects. Am J Epidemiol. 1993;138:56–64.
34. Office for National Statistics. Live births in England and Wales by characteristics of mother 1, 2010. ONS, London. www.ons.gov.uk/ons/rel/vsob1/characteristics-of-Mother-1%2D-england-and-wales/2010/index.html. Accessed 31 Jan 2018.
35. Aydinoğlu NZ, Krishna A. Imagining thin: why vanity sizing works. J Consum Psychol. 2012;22:565–72.
36. Howard SJ, Davies SC. Chief medical officer urges action to tackle overweight and obesity. BMJ. 2014;348:g2375.
37. Li L, Liu DW, Yan HY, Wang ZY, Zhao SH, Wang B. Obesity is an independent risk factor for non-alcoholic fatty liver disease: evidence from a meta-analysis of 21 cohort studies. Obes Rev. 2016;17:510–9.
38. Xu C, Yu C, Ma H, Xu L, Miao M, Li Y. Prevalence and risk factors for the development of nonalcoholic fatty liver disease in a nonobese Chinese population: the Zhejiang Zhenhai study. Am J Gastroenterol. 2013;108:1299–304.
39. Loomis AK, Kabadi S, Preiss D, Hyde C, Bonato V, St Louis M, et al. Body mass index and risk of nonalcoholic fatty liver disease: two electronic health record prospective studies. J Clin Endocrinol Metab. 2016;101:945–52.
40. Hughes LA, Schouten LJ, Goldbohm RA, van den Brandt PA, Weijenberg MP. Self-reported clothing size as a proxy measure for body size. Epidemiology. 2009;20:673–6.
41. Battram DS, Beynon C, He M. The reliability and validity of using clothing size as a proxy for waist circumference measurement in adults. Appl Physiol Nutr Metab. 2011;36:183–90.

APPENDIX B
Search strategies for structured literature review

Taken and adapted from Harris et al.[111]

Database: Ovid MEDLINE(R) from 30 January 2015 to 17 March 2018

Search Strategy:

1 exp Liver Cirrhosis/di [Diagnosis] (1200)

2 exp Fatty Liver/di [Diagnosis] (680)

3 exp Liver Diseases, Alcoholic/di [Diagnosis] (143)

4 (hepatic fibrosis or Chronic liver disease* or advanced fibrosis or non alcoholic fatty liver disease* or NAFLD or NAFL or alcoholic liver disease* or ALD or liver fibrosis* or hepatic cirrhos* or liver cirrhos* or fatty liver disease* or fatty liver or advanced fibrosis).mp. (16068)

5 exp Biological Markers/ (101489)

6 exp Elasticity Imaging Techniques/ (2176)

7 exp Diagnostic Imaging/ (237199)

8 (non invasive biomarker* or non invasive biological marker* or non invasive marker* or fibroscan or liver stiffness or transient elastography or ultrasound abdomen or ARFI or liver function test* or LFT* or fibrotest* or fib4 or Lok or FORNS or APRI or ELF or NFS or BAAT or BARD or noninvasive biomarker* or noninvasive biological marker* or noninvasive marker* or elastogram* or sonoelastograph* or imaging tissue elastic or elasticity imaging technique*).mp. (6204)

9 exp Family Practice/ or exp General Practice/ (4212)

10 exp Primary Health Care/ (20738)

11 exp Community Health Services/ (23021)

12 (gp or general practice* or family practice* or primary care or communit* or outreach).mp. (90232)

13 1 or 2 or 3(1856)

14 5 or 6 or 7 or 8 (331128)

15 4 and 14 (4578)

16 13 and 14 (1034)

17 9 or 10 or 11 or 12 (115149)

18 15 or 16 (1856)

19 17 and 18 (62)

Database: Embase from 30 January 2015 to 17 March 2018

Search Strategy:

1 exp Liver Cirrhosis/di [Diagnosis] (1153)

2 exp Fatty Liver/di [Diagnosis] (1305)

3 exp Liver Diseases, Alcoholic/di [Diagnosis] (195)

4 (hepatic fibrosis or Chronic liver disease* or advanced fibrosis or non alcoholic fatty liver disease* or NAFLD or NAFL or alcoholic liver disease* or ALD or liver fibrosis* or hepatic cirrhos* or liver cirrhos* or fatty liver disease* or fatty liver or advanced fibrosis).mp. (54576)

5 exp Biological Markers/ (76748)

6 exp Elasticity Imaging Techniques/ (7206)

7 exp Diagnostic Imaging/ (29177)

8 (non invasive biomarker* or non invasive biological marker* or non invasive marker* or fibroscan or liver stiffness or transient elastography or ultrasound abdomen or ARFI or liver function test* or LFT* or fibrotest* or fib4 or LOk or FORNS or APRI or ELF or NFS or BAAT or BARD or noninvasive biomarker* or noninvasive biological marker* or noninvasive marker* or elastogram* or sonoelastograph* or imaging tissue elastic or elasticity imaging technique*).mp. (17542)

9 exp Family Practice/ or exp General Practice/ (6959)

10 exp Primary Health Care/ (27610)

11 exp Community Health Services/ (12546)

12 (gp or general practice* or family practice* or primary care or communit* or outreach).mp. (179369)

13 exp chronic liver disease/di [Diagnosis] (150)

14 exp early diagnosis/ (16932)

15 exp liver fibrosis/ (11017)

16 exp diagnosis/ (1062220)

17 exp non invasive measurement/ (2489)

18 chronic liver disease/ (4117)

19 exp liver cirrhosis/ (27302)

20 exp fatty liver/ (18377)

21 exp nonalcoholic fatty liver/ or exp alcohol liver disease/ (15596)

22 4 or 15 or 18 or 19 or 20 or 21 (56427)

23 5 or 6 or 7 or 8 or 14 or 17 (141889)

APPENDIX C

P values for pairwise comparisons among group means for variables in chapter four

One-way analysis of variance (ANOVA) was performed to evaluate the null hypothesis that there is no difference between liver fibrosis groups of a number of baseline characteristics. ANOVA was significant for age and AST but not ALT or HBV DNA level.

Post hoc comparisons to evaluate pairwise differences among group means were performed using the Tukey test. The pairwise differences are shown in the SPSS output below, and significant differences are highlighted with *.

Dependent Variable: age

Tukey HSD

(I) stagmet	(J) stagmet	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
0	1	-9.55778*	2.83918	.008	-17.3829	-1.7327
	2	-11.86288*	2.86307	.001	-19.7538	-3.9720
	3	-18.06810*	3.06084	.000	-26.5041	-9.6321
	4	-15.00000*	2.98174	.000	-23.2180	-6.7820
1	0	9.55778*	2.83918	.008	1.7327	17.3829
	2	-2.30511	2.09851	.807	-8.0888	3.4786
	3	-8.51032*	2.36122	.004	-15.0181	-2.0026
	4	-5.44222	2.25773	.117	-11.6648	.7803
2	0	11.86288*	2.86307	.001	3.9720	19.7538
	1	2.30511	2.09851	.807	-3.4786	8.0888
	3	-6.20522	2.38989	.075	-12.7920	.3816
	4	-3.13712	2.28770	.647	-9.4422	3.1680
3	0	18.06810*	3.06084	.000	9.6321	26.5041
	1	8.51032*	2.36122	.004	2.0026	15.0181
	2	6.20522	2.38989	.075	-.3816	12.7920
	4	3.06810	2.53084	.744	-3.9072	10.0434
4	0	15.00000*	2.98174	.000	6.7820	23.2180
	1	5.44222	2.25773	.117	-.7803	11.6648
	2	3.13712	2.28770	.647	-3.1680	9.4422
	3	-3.06810	2.53084	.744	-10.0434	3.9072

***. The mean difference is significant at the 0.05 level.**

Dependent Variable: AST

Tukey HSD

(I) METAVIR	(J) METAVIR	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
0	1	-1.92222	16.98154	1.000	-48.7251	44.8807
	2	-19.10520	17.12440	.798	-66.3018	28.0914
	3	-41.39964	18.30733	.163	-91.8565	9.0572
	4	-49.86111*	17.83419	.045	-99.0140	-.7083
1	0	1.92222	16.98154	1.000	-44.8807	48.7251
	2	-17.18298	12.55151	.648	-51.7762	17.4103
	3	-39.47742*	14.12279	.045	-78.4013	-.5536
	4	-47.93889*	13.50383	.004	-85.1568	-10.7210
2	0	19.10520	17.12440	.798	-28.0914	66.3018
	1	17.18298	12.55151	.648	-17.4103	51.7762
	3	-22.29444	14.29425	.525	-61.6909	17.1020
	4	-30.75591	13.68305	.167	-68.4678	6.9560
3	0	41.39964	18.30733	.163	-9.0572	91.8565
	1	39.47742*	14.12279	.045	.5536	78.4013
	2	22.29444	14.29425	.525	-17.1020	61.6909
	4	-8.46147	15.13733	.981	-50.1815	33.2586
4	0	49.86111*	17.83419	.045	.7083	99.0140
	1	47.93889*	13.50383	.004	10.7210	85.1568
	2	30.75591	13.68305	.167	-6.9560	68.4678
	3	8.46147	15.13733	.981	-33.2586	50.1815

*. The mean difference is significant at the 0.05 level.

Dependent Variable: ALT

Tukey HSD

(I) METAVIR	(J) METAVIR	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
0	1	-.23556	28.09825	1.000	-77.6772	77.2061
	2	-23.76832	28.33463	.918	-101.8615	54.3248
	3	-61.68459	30.29194	.253	-145.1723	21.8031
	4	-36.16667	29.50907	.736	-117.4967	45.1633
1	0	.23556	28.09825	1.000	-77.2061	77.6772
	2	-23.53277	20.76816	.789	-80.7719	33.7064
	3	-61.44903	23.36806	.069	-125.8538	2.9557
	4	-35.93111	22.34390	.494	-97.5132	25.6510
2	0	23.76832	28.33463	.918	-54.3248	101.8615
	1	23.53277	20.76816	.789	-33.7064	80.7719
	3	-37.91627	23.65177	.497	-103.1030	27.2704
	4	-12.39835	22.64044	.982	-74.7977	50.0010
3	0	61.68459	30.29194	.253	-21.8031	145.1723
	1	61.44903	23.36806	.069	-2.9557	125.8538
	2	37.91627	23.65177	.497	-27.2704	103.1030
	4	25.51792	25.04675	.846	-43.5135	94.5493
4	0	36.16667	29.50907	.736	-45.1633	117.4967
	1	35.93111	22.34390	.494	-25.6510	97.5132
	2	12.39835	22.64044	.982	-50.0010	74.7977
	3	-25.51792	25.04675	.846	-94.5493	43.5135

Dependent Variable: HBV_DNA

Tukey HSD

(I) METAVIR	(J) METAVIR	Mean Difference (I-J)	Std. Error	Sig.	95% ... Lower Bound
0	1	27299812.8	67205559.1	.994	-1.5793E+8
	2	-25277426	67770937.6	.996	-2.1206E+8
	3	8440474.90	72452426.9	1.000	-1.9125E+8
	4	-2309734.3	70579954.5	1.000	-1.9684E+8
1	0	-27299813	67205559.1	.994	-2.1253E+8
	2	-52577238	49673405.2	.827	-1.8948E+8
	3	-18859338	55891872.2	.997	-1.7290E+8
	4	-29609547	53442279.0	.981	-1.7690E+8
2	0	25277425.6	67770937.6	.996	-1.6151E+8
	1	52577238.4	49673405.2	.827	-84327747
	3	33717900.5	56570435.6	.976	-1.2220E+8
	4	22967691.3	54151546.5	.993	-1.2628E+8
3	0	-8440474.9	72452426.9	1.000	-2.0813E+8
	1	18859337.9	55891872.2	.997	-1.3518E+8
	2	-33717900	56570435.6	.976	-1.8963E+8
	4	-10750209	59906962.7	1.000	-1.7586E+8
4	0	2309734.25	70579954.5	1.000	-1.9222E+8
	1	29609547.0	53442279.0	.981	-1.1768E+8
	2	-22967691	54151546.5	.993	-1.7221E+8
	3	10750209.1	59906962.7	1.000	-1.5436E+8

APPENDIX D
UKCTOCS baseline questionnaire

United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS)

Please complete this form in BLACK INK and in BLOCK CAPITALS. It will help us check that you are eligible for the study as well as collect some data regarding your risk of developing ovarian cancer. If the personal or GP details are incorrect please insert the correct details into the "Amended details" box.

Your Personal Details

Amended details

Your GP Details

Dr

NHS No. D.o.B. AP ID

Your Home Telephone No.

Your Work Telephone No.

ELIGIBILITY DETAILS (use black ink and BLOCK CAPITALS or place a cross "X" in the appropriate boxes)

1. When was your last period? **(dd/mm/yyyy)** / /

2. Are you currently on Hormone Replacement Therapy (HRT)? Yes No

   If Yes then when did you start taking HRT? **(dd/mm/yyyy)** / /

3. Have you had both your ovaries removed? Yes No

4. Have you ever had cancer diagnosed (except skin cancer)? Yes No Go to Q5

   If yes, what cancer was it?

   Ovary Breast Bowel Lung Other

   When was it diagnosed? **(dd/mm/yyyy)** / /

5. Have you had any treatment for any cancer (including surgery, chemotherapy, radiotherapy) in the last 12 months(not including tamoxifen)? Yes No

4999524552

6. If any of the following relatives have had **OVARIAN CANCER** please write the number of affected relatives in the approprate box. Please enter 0 for no affected relatives.
(e.g. 0 Mother, 2 Sister, 0 Daughter).

☐ Mother ☐ Daughter ☐ Sister ☐ Aunt ☐ GrandMother ☐ GrandDaughter

7. If any of the following relatives have had **BREAST CANCER** please write the number of affected relatives in the approprate box. Please enter 0 for no affected relatives.
(e.g. 0 Mother, 2 Sister, 1 Daughter).

☐ Mother ☐ Daughter ☐ Sister ☐ Aunt ☐ GrandMother ☐ GrandDaughter

8. Are you currently taking part in any other ovarian cancer screening trial? Yes ☐ No ☐

If yes what is your study reference number? ☐☐☐☐☐☐☐☐☐☐

ADDITIONAL INFORMATION

ID ☐☐☐☐☐☐

9. Your height (cm) ☐☐☐ Your weight (kg) ☐☐☐

Or (in) ☐☐☐ Or (lb) ☐☐☐

10. Country of birth (please place an "X" as appropriate)

England ☐ Northern Ireland ☐ Scotland ☐ Irish Republic ☐ Wales ☐ Elsewhere ☐

11. Ethnic group, please place an "X" in the appropriate box. (If you are descended from more than one ethnic or racial group, please sellect the group you consider you belong to or choose "Any other ethnic origin")

☐ White ☐ Indian ☐ Pakistani ☐ Chinese ☐ Bangladeshi

☐ Black-African ☐ Black-Caribbean ☐ Black-other ☐ Any other ethnic origin

12. At what age did you first have your period? ☐☐

13. How many pregnancies have you had which ended before they reached 6 months (including miscarriages, ectopic pregnancies)? ☐☐

14. How many pregnancies have you had which lasted beyond 6 months (including all deliveries - both term and preterm)? ☐☐

15. Have you ever taken the oral contraceptive pill? Yes ☐ No ☐

If yes, how many years in total did you take the pill? Years ☐☐

16. Have you ever had a hysterectomy (removal of the womb)? Yes ☐ No ☐

17. Have you had a sterilisation operation (To block your tubes)? Yes ☐ No ☐

18. Have you ever had any treatment for infertility? Yes ☐ No ☐

9236524551

APPENDIX E
UKCTOCS follow up questionnaire

0175097515 1 of 4

Volunteer Ref:

U K C T O C S

Private and Confidential

United Kingdom Collaborative Trial of Ovarian Cancer Screening

For the information that you supply to be useful to our ovarian cancer screening study, it is important that you complete all of the questions in "bold"

Please use a BLACK pen and place a cross INSIDE the box ☒

Please return both sheets using the Free Post envelope supplied

Follow up questionnaire

We would be very grateful if you could answer the following questions. If you are not sure about exact details/dates an approximate answer is better than none. If there are any relevant details you wish to include, please use an additional sheet. Please use a BLACK biro or ink pen.

General questions about you

What qualification(s) do you have from school, college or the equivalent?

(please place a cross inside the most appropriate box(es))

☐ "O" level or equivalent
☐ Nursing or teaching
☐ "A" level or equivalent
☐ College/ university degree (or equivalent)
☐ Clerical or commercial qualification (e.g. secretarial, hairdressing etc)
☐ None of these

Approximately how much alcohol on average do you drink each week? (One drink = a glass of wine, half a pint of lager or cider, a measure of spirits). Average number of drinks of alcohol each week :

☐ None ☐ Less than 1 ☐ 1-3 ☐ 4-6 ☐ 7-10 ☐ 11-15 ☐ 16-20 ☐ 21+

Have you ever been a smoker? ☐ Yes ☐ No

If you answered yes to the above please answer the following questions:

How many years in total have you smoked for? ☐☐

During those years how many cigarettes on average did you smoke per day? ☐☐

What was your skirt size when you were in your early twenties?

☐ 6 ☐ 8 ☐ 10 ☐ 12 ☐ 14 ☐ 16 ☐ 18 ☐ 20 ☐ 22 ☐ 24 ☐ 26 ☐ 28 ☐ 30

What is your skirt size now?

☐ 6 ☐ 8 ☐ 10 ☐ 12 ☐ 14 ☐ 16 ☐ 18 ☐ 20 ☐ 22 ☐ 24 ☐ 26 ☐ 28 ☐ 30

Are you currently taking HRT? ☐ No ☐ Yes

Have you used any of the following to relieve menopausal symptoms? ☐ Yes ☐ No

☐ Herbal remedies e.g. Black cohosh
☐ Homeopathic remedies
☐ Phytoestrogens or soy products
☐ Aromatherapy, reflexology or acupuncture
☐ Vitamins e.g. Menopace, vitamin E
☐ Life style changes e.g. relaxation, exercise
☐ Other medical treatments e.g. Venlafaxine, Megace

We are interested to know more about how women deal with the menopause. If you are 50-60 years old would you be willing to complete a survey? ☐ Yes ☐ No

Questions about your outlook on life

Using the scale below, please indicate the extent to which you agree with each item:

1. The future seems to me to be hopeful, and I believe that things are changing for the better.

☐ Absolutely agree ☐ Somewhat agree ☐ Cannot say ☐ Somewhat disagree ☐ Absolutely disagree

2. I feel that it is possible to reach the goals I would like to strive for.

☐ Absolutely agree ☐ Somewhat agree ☐ Cannot say ☐ Somewhat disagree ☐ Absolutely disagree

Questions about your Health

Do you have/are you being treated for any of the following conditions (multiple boxes can be crossed)?

☐ High blood pressure ☐ Diabetes ☐ Stroke
☐ Heart disease e.g. heart attack, angina ☐ Rheumatoid arthritis ☐ Osteoporosis
☐ High blood cholesterol ☐ Osteoarthritis
☐ I have not been treated for any of the above conditions

Since joining UKCTOCS have you had any of the following operations? ☐ Yes ☐ No

If "yes" please enter the details (multiple boxes can be filled; please fill even if you had them as part of the trial)

Operation to look at your ovaries -either by incision or keyhole (laparoscopy)

Year of operation: _____ Hospital No.: _____

Hospital at which operation took place: _____

_____

Name of Consultant: _____

Hysterectomy / Removal of womb

Year of operation: _____ Hospital No.: _____

Hospital at which operation took place: _____

_____

Name of Consultant: _____

Removal of ovaries (please tick one box) ☐ Right Ovary ☐ Left Ovary ☐ Both ovaries

Year of operation: _____ Hospital No.: _____

Hospital at which operation took place: _____

_____

Name of Consultant: _____

Hysteroscopy / D&C / Scrape of your womb / Operation to look at your womb

Year of operation: _____ Hospital No.: _____

Hospital at which operation took place: _____

_____

Name of Consultant: _____

Were there any complications resulting from any of the above procedures? ☐ Yes ☐ No

If yes, please give details using an extra sheet of paper if necessary.

_____

_____

Breast biopsy or other surgery involving the breast

Year of operation: _____ Hospital No.: _____

Hospital at which operation took place: _____

_____

_____

Name of Consultant: _____

Any other operation of any kind. Please describe:

Since joining UKCTOCS have you been diagnosed with any cancer? Please tell us about this(ese) cancer(s)

- ☐ Ovarian cancer
- ☐ Breast cancer
- ☐ Cervical cancer
- ☐ Endometrial/uterus/womb cancer
- ☐ Bowel/colorectal cancer
- ☐ Gastric/stomach cancer
- ☐ Pancreatic cancer
- ☐ Kidney cancer
- ☐ Lung cancer
- ☐ Vulval/vaginal cancer
- ☐ BCC/rodent/skin cancer
- ☐ Other cancer
- ☐ I have not been diagnosed with any cancers

Type of cancer: _____

Year of operation: _____ Hospital No.: _____

Hospital at which operation took place: _____

_____

_____

Name of Consultant: _____

Type of cancer: _____

Year of operation: _____ Hospital No.: _____

Hospital at which operation took place: _____

_____

_____

Name of Consultant: _____

Once you have completed this questionnaire please sign and date below, then return it in the Free Post envelope. Thanks you.

Name (please print): Date: ☐☐ / ☐☐ / ☐☐☐☐

Signature: (dd/mm/yyyy)

QUOX 1 ☐ 2 ☐ 3 ☐ **Office use only** **QU Consent** Yes ☐ No ☐

The following sections should ONLY be completed by those volunteers who are in the CONTROL group

Since joining UKCTOCS have you had an ultrasound scan of your ovaries? ☐ Yes ☐ No

If "yes" why was this performed?

- ☐ GP request
- ☐ Hospital Doctor request
- ☐ Your own request
- ☐ Other reason _____

Year of scan: _____ Hospital No.: _____

Hospital at which operation took place: _____

_____

_____

Name of Consultant: _____

Since joining UKCTOCS have you had a blood test for CA125? ☐ Yes ☐ No

(CA125 is a substance, which is released at higher levels into the blood in women with ovarian cancer. The test is carried out if doctors suspect that a woman may have ovarian cancer)

If "yes" why was this performed?

- ☐ GP request
- ☐ Hospital Doctor request
- ☐ Your own request
- ☐ Other reason _____

Year of CA125 test: _____ Hospital No.: _____

Hospital at which operation took place: _____

_____

_____

Name of Consultant: _____

Thank you for taking the time to complete this questionnaire. The information that you supply is of great importance to the success of the trial.

Regards

UKCTOCS Team

APPENDIX F

ICD-10 codes for studies cited in this thesis.
Those marked * were the codes used in my studies

ICD-10 code	Diagnosis
B18.0	Chronic viral hepatitis B with delta-agent
B18.1	Chronic viral hepatitis B without delta-agent
B18.2	Chronic viral hepatitis C
B18.9	Chronic viral hepatitis, unspecified
C22.0 *	Liver cell carcinoma
E83.1	Disorders of iron metabolism
E88.0	Disorders of plasma-protein metabolism, not elsewhere classified
I85	Oesophageal varices
I85.0	Oesophageal varices with bleeding
I85.9	Oesophageal varices without bleeding
I86.4	Gastric varices
I98.2	Oesophageal varices without bleeding in diseases classified elsewhere
K70 *	Alcoholic liver disease
K70.1	Alcoholic hepatitis
K70.2	Alcoholic fibrosis and sclerosis of liver
K70.3	Alcoholic cirrhosis of liver
K70.4	Alcoholic hepatic failure
K71	Toxic liver disease
K71.1	Toxic liver disease with fibrosis and cirrhosis of liver
K71.7	Toxic liver disease with fibrosis and cirrhosis of liver
K71.8	Toxic liver disease with other disorders of liver
K72	Hepatic failure, not elsewhere classified
K72.1	Chronic hepatic failure
K72.9	Hepatic failure, unspecified
K73 *	Chronic hepatitis, not elsewhere classified
K73.2	Chronic active hepatitis, not elsewhere classified
K73.9	Chronic hepatitis, unspecified
K74 *	Fibrosis and cirrhosis of liver
K74.3	Primary biliary cirrhosis
K74.4	Secondary biliary cirrhosis
K74.5	Biliary cirrhosis, unspecified
K74.6	Other and unspecified cirrhosis of liver
K75	Other inflammatory liver diseases
K75.8	Other specified inflammatory liver diseases
K75.9	Inflammatory liver disease, unspecified
K76 *	Other diseases of liver
K76.1	Chronic passive congestion of liver
K76.6	Portal hypertension
K76.7	Hepatorenal syndrome
K77	Liver disorders in diseases classified elsewhere
K83.0	Cholangitis
R18	Ascites
Z94.4 *	Liver transplant status

APPENDIX G

Comparisons between study cohort and cohort comprising remainder of participants in England who returned follow-up questionnaire

Characteristic	Participants in England *n* = 17,823	Study cohort *n* = 92,126	*p* value
Age in years, mean	62.2	60.9	<0.001[1]
BMI, mean	26.5	26.3	<0001[1]
Hypertension, %	33.2	31.6	<0.001[2]
Heart disease, %	6.8	5.5	<0.001[2]
Hypercholesterolaemia	26.8	23.7	<0.001[2]
Type 2 diabetes	6.4	5.0	<0.001[2]

1. Independent samples t-test; 2. Pearson Chi squared test

APPENDIX H
Permission to access UKCTOCS samples

25th March 2014

Dr Paul Trembling
Institute for Liver and Digestive Health,
University College London

Dear Paul

Re: Project entitled 'Performance of the Enhanced Liver Fibrosis (ELF) test in predicting clinical outcomes due to chronic liver disease'

The above project was reviewed at the UCL/Abcodia Joint Steering Committee on 25th February 2014 and I am pleased to inform you that the project was approved.

We look forward to hearing of the outcome of the project in due course

Julie Barnes, PhD
Chair of the Joint Steering Committee