

# **Biomarkers to predict advanced liver diseases.**

**Yusef Muhana Alenezi, MSc.**

Submitted in partial fulfilment of the requirements for the degree of  
Doctor of Philosophy

June 2024

Lifespan and Population Health

School of Medicine

University of Nottingham

## **Abstract**

**Background:** Chronic liver disease (CLD) is one of the leading causes of morbidity and mortality, accounting for more than two million deaths annually worldwide. In recent years, Non-alcoholic Fatty Liver Disease (NAFLD) has been a fast-emerging cause of CLD, which can progress to advanced conditions such as liver cirrhosis, hepatocellular carcinoma (HCC), and liver failure. While the condition of simple steatosis in NAFLD patients is usually benign, the risk of progression to advanced CLD underscores the importance of early detection, as late diagnosis frequently leads to poor patient outcomes. As a result, considerable effort is now being made to find non-invasive methods for the early detection of CLD.

This study assesses the feasibility of using non-invasive biomarkers to predict the progression to advanced CLD. In order to achieve this, it begins by assessing the prevalence of NAFLD in the United Kingdom (UK) and the Kingdom of Saudi Arabia (KSA), and then determines the availability of non-invasive biomarkers among UK primary care patients.

**Methodology:** Systematic review and meta-analysis were used to assess NAFLD prevalence among the general population in the UK and KSA. Then, three retrospective cohort studies were conducted using Clinical Practice Research Datalink (CPRD) data from 1 Jan 1990 – 31 Dec 2016. The first study determined the proportion of patients over 40 with data within their records to calculate normal or abnormal liver

fibrosis biomarkers. The second and third studies utilised CPRD data, comprising all cases eligible for linkage to the National Cancer Registration and Analysis Service (NCRAS) cancer registry database, ONS death registration, and Hospital Episode Statistics (HES) data to identify patients who have been diagnosed with advanced CLD and determine whether the AST/ALT ratio, APRI score, and FIB-4 index could be used to predict subsequent advanced CLD in those for whom they were assessed.

**Findings:** The systematic review studies showed an estimated pooled prevalence of NAFLD among disease-unrestricted populations to be 16.8% and 23.8% for KSA and the UK, respectively. There is good evidence that adults with type 2 diabetes mellitus (T2DM) have a higher prevalence of NAFLD in both countries. The CPRD findings showed that the most frequent calculable biomarker was the APRI score, followed by the AST/ALT, and the least calculable biomarker was the FIB-4. However, despite being the most calculable biomarker, the prevalence of abnormal APRI was lower than that of the other biomarkers.

The findings also demonstrated that the AST/ALT ratio, APRI score, and FIB-4 index allow for the appropriate identification of patients at a higher risk of developing advanced CLD or the outcomes of liver transplantation and, possibly, death. Notably, based on conventional cutoffs, abnormal APRI scores showed a stronger association with the development of the outcomes than the abnormal AST/ALT ratio and Fib4 index.

**Conclusion:** This thesis revealed a high prevalence of NAFLD among the general population in the UK and KSA. The studies in this thesis have clearly shown that abnormal biomarkers allow for the appropriate identification of patients at a higher risk of developing advanced CLD. The use of routinely collected data to measure the availability of these biomarkers reveals the potential and practicality of a large dataset for future biomarker studies.



## List of publications from the work in this thesis:

- **Published manuscript.**

1. Alenezi Y M, Harris R, Morling J, et al. (June 12, 2023) Prevalence of Non-alcoholic Fatty Liver Disease (NAFLD) in Saudi Arabia: Systematic Review and Meta-analysis. *Cureus* 15(6): e40308. doi:10.7759/cureus.40308.

- **Abstract presented at an international conference.**

1. Alenezi, Y., Card, T., Morling, J., & Harris, R. (2022). The prevalence of non-alcoholic fatty liver disease in the United Kingdom: a systematic review and meta-analysis. *Journal of Hepatology*, 77, S145.

## **Acknowledgements**

I thank Allah (God) for giving me the strength and patience to complete this thesis. I thank him for keeping me safe and alive to complete my studies in the UK and to pass through some hardships.

I express my deepest respect and gratitude to my supervisors, Dr. Joanne Morling and Dr. Tim Card, for their unlimited guidance and support throughout the past years. I highly appreciate that their suggestions and encouragement have not been interrupted, even with the impact of the COVID-19 pandemic. I thank them for taking me under their supervision and, most importantly, for believing in me. I have learned a lot from them, and their comments and feedback have always been constructive and invaluable. I am where I am because of them, and this thesis would not have been completed without their unlimited support and assistance.

My thanks extend to my sponsor (Northern Border University) and the Saudi-Arabian Cultural Bureau in London for giving me this opportunity to pursue my higher education in the UK.

My sincerest gratitude also extends to my parents for their endless prayers for my success. I thank them for not giving up on me during my worst, weakest, and lowest. I would also like to thank my family, including my brothers and sisters, for their support and concern.

My special thanks to my wife, Abrar, who has supported and encouraged me throughout this journey. I thank her for her sacrifices and

for looking after me during my study. The same also goes for my beloved sons, Nasser, Abdullah, and Tamim.

# Table of Contents

|                                                                                            |      |
|--------------------------------------------------------------------------------------------|------|
| Abstract .....                                                                             | i    |
| Acknowledgements.....                                                                      | v    |
| Table of Contents.....                                                                     | vii  |
| List of Tables .....                                                                       | xiv  |
| List of Figures .....                                                                      | xvii |
| List of Abbreviations.....                                                                 | xix  |
| Chapter 1 Background and thesis rationale .....                                            | 1    |
| 1.1 Introduction.....                                                                      | 2    |
| 1.2 What is chronic liver disease (CLD) and how does it develop? .....                     | 3    |
| 1.3 Epidemiology and public health impact of CLD .....                                     | 4    |
| 1.4 Aetiology of CLD.....                                                                  | 5    |
| 1.5 Liver diseases in the UK and KSA .....                                                 | 6    |
| 1.6 Non-alcoholic Fatty Liver Disease (NAFLD).....                                         | 8    |
| 1.6.1 NAFLD definitions and staging.....                                                   | 8    |
| 1.6.2 Global epidemiology .....                                                            | 11   |
| 1.6.3 Risk factors.....                                                                    | 12   |
| 1.6.4 Diagnosis.....                                                                       | 13   |
| 1.6.5 Clinical Presentation .....                                                          | 14   |
| 1.6.6 Management.....                                                                      | 14   |
| 1.6.7 Outcome/Prognosis .....                                                              | 14   |
| 1.7 Early detection of chronic liver disease .....                                         | 15   |
| 1.7.1 Recommendations for early CLD detection.....                                         | 15   |
| 1.7.2 Importance of early disease detection.....                                           | 18   |
| 1.7.3 Strategies for early detection of CLD .....                                          | 20   |
| 1.7.4 Characteristics of an ideal screening test for early advanced<br>CLD detection. .... | 23   |
| 1.8 Methods for early detection of CLD .....                                               | 26   |
| 1.8.1 Liver biopsy .....                                                                   | 26   |
| 1.8.2 Imaging modalities .....                                                             | 28   |

|           |                                                                                          |    |
|-----------|------------------------------------------------------------------------------------------|----|
| 1.8.3     | Direct serum markers.....                                                                | 30 |
| 1.8.4     | Indirect serum markers .....                                                             | 31 |
| 1.9       | Implications from the early disease detection using pragmatic case finding approach..... | 49 |
| 1.10      | Rationale, aims and objectives .....                                                     | 54 |
| 1.10.1    | Rationale .....                                                                          | 54 |
| 1.10.2    | Aims and objectives.....                                                                 | 56 |
| Chapter 2 | The prevalence of NAFLD in the UK: A Systematic Review and Meta-Analysis .....           | 58 |
| 2.1       | Introduction.....                                                                        | 59 |
| 2.2       | Methods.....                                                                             | 59 |
| 2.2.1     | Review question .....                                                                    | 59 |
| 2.2.2     | Study design and inclusion and exclusion criteria .....                                  | 59 |
| 2.2.3     | Search strategy.....                                                                     | 60 |
| 2.2.4     | Study selection and quality assessment .....                                             | 61 |
| 2.2.5     | Data management and analysis.....                                                        | 62 |
| 2.3       | Results .....                                                                            | 63 |
| 2.3.1     | Search results.....                                                                      | 63 |
| 2.3.2     | Quality assessment .....                                                                 | 71 |
| 2.3.3     | Heterogeneity and publication bias .....                                                 | 73 |
| 2.4       | Discussion .....                                                                         | 73 |
| 2.5       | Conclusion.....                                                                          | 82 |
| Chapter 3 | The prevalence of NAFLD in the KSA: A Systematic Review and Meta-Analysis. ....          | 83 |
| 3.1       | Introduction.....                                                                        | 84 |
| 3.2       | Methods.....                                                                             | 84 |
| 3.2.1     | Review question .....                                                                    | 84 |
| 3.2.2     | Study design and inclusion and exclusion criteria .....                                  | 84 |
| 3.2.3     | Search strategy.....                                                                     | 85 |
| 3.2.4     | Study selection and quality assessment .....                                             | 85 |
| 3.2.5     | Data management and analysis.....                                                        | 85 |

|           |                                                                                                    |     |
|-----------|----------------------------------------------------------------------------------------------------|-----|
| 3.3       | Results .....                                                                                      | 85  |
| 3.3.1     | Search results.....                                                                                | 85  |
| 3.3.2     | Quality assessment .....                                                                           | 91  |
| 3.3.3     | Heterogeneity and publication bias. ....                                                           | 93  |
| 3.4       | Discussion .....                                                                                   | 93  |
| 3.5       | Conclusion.....                                                                                    | 103 |
| Chapter 4 | Clinical Practice Research Datalink (CPRD) .....                                                   | 106 |
| 4.1       | Introduction.....                                                                                  | 107 |
| 4.2       | Development of CPRD.....                                                                           | 107 |
| 4.3       | Nature and structure of CPRD .....                                                                 | 108 |
| 4.4       | CPRD coverage.....                                                                                 | 112 |
| 4.5       | External linkage .....                                                                             | 112 |
| 4.6       | Diagnosis codes .....                                                                              | 113 |
| 4.7       | Strengths and limitations of CPRD.....                                                             | 113 |
| 4.7.1     | Size .....                                                                                         | 115 |
| 4.7.2     | Representative .....                                                                               | 115 |
| 4.7.3     | Prospectively collected .....                                                                      | 116 |
| 4.7.4     | Incomplete recording .....                                                                         | 117 |
| 4.7.5     | Duration of follow-up.....                                                                         | 117 |
| 4.7.6     | Reproducibility .....                                                                              | 118 |
| 4.8       | Ethical Consideration and Research Approval .....                                                  | 118 |
| 4.9       | Conclusion.....                                                                                    | 119 |
| Chapter 5 | Assessing the availability of non-invasive liver biomarkers<br>among UK primary care patients..... | 120 |
| 5.1       | Introduction.....                                                                                  | 121 |
| 5.2       | Aim .....                                                                                          | 121 |
| 5.2.1     | Objectives.....                                                                                    | 122 |
| 5.3       | Methods.....                                                                                       | 122 |
| 5.3.1     | Study design.....                                                                                  | 122 |
| 5.3.2     | Data source .....                                                                                  | 122 |

|           |                                                                                                                                 |     |
|-----------|---------------------------------------------------------------------------------------------------------------------------------|-----|
| 5.3.3     | Study population .....                                                                                                          | 123 |
| 5.3.4     | Data processing.....                                                                                                            | 124 |
| 5.3.5     | Biomarkers calculation.....                                                                                                     | 126 |
| 5.3.6     | Outcomes .....                                                                                                                  | 127 |
| 5.3.7     | Determining patients with normal and abnormal scores .....                                                                      | 127 |
| 5.3.8     | Statistical analysis .....                                                                                                      | 127 |
| 5.4       | Results .....                                                                                                                   | 130 |
| 5.4.1     | Characteristics of the participants .....                                                                                       | 130 |
| 5.4.2     | Results of data processing.....                                                                                                 | 132 |
| 5.4.3     | Time difference between the tests .....                                                                                         | 134 |
| 5.4.4     | Score results.....                                                                                                              | 135 |
| 5.5       | Discussion .....                                                                                                                | 143 |
| 5.5.1     | Availability of calculable non-invasive biomarkers among<br>general practice populations .....                                  | 143 |
| 5.5.2     | Variability in liver function testing practices in primary care;<br>implications for non-Invasive biomarker calculability ..... | 144 |
| 5.5.3     | Gender and age disparities in liver function testing .....                                                                      | 147 |
| 5.5.4     | Implications of biomarker utilization in case finding and risk<br>stratification.....                                           | 152 |
| 5.5.5     | Source of the data to calculate the Biomarkers. ....                                                                            | 154 |
| 5.6       | Conclusion.....                                                                                                                 | 158 |
| Chapter 6 | Identifying Advanced CLD outcomes among the Study<br>Population. 160                                                            |     |
| 6.1       | Introduction.....                                                                                                               | 161 |
| 6.2       | Study aim.....                                                                                                                  | 161 |
| 6.3       | Methods.....                                                                                                                    | 161 |
| 6.3.1     | Study design.....                                                                                                               | 161 |
| 6.3.2     | Study population.....                                                                                                           | 161 |
| 6.3.3     | Outcomes .....                                                                                                                  | 162 |
| 6.3.4     | Stratification of the cohort based on potential predictors. ....                                                                | 168 |

|           |                                                                               |     |
|-----------|-------------------------------------------------------------------------------|-----|
| 6.3.5     | Statistical analysis .....                                                    | 169 |
| 6.4       | Results .....                                                                 | 170 |
| 6.4.1     | Characteristics of patients with outcomes .....                               | 170 |
| 6.4.2     | Determination of new patient outcomes for each cohort.....                    | 174 |
| 6.5       | Discussion .....                                                              | 176 |
| 6.6       | Conclusion.....                                                               | 181 |
| Chapter 7 | Can Non-Invasive Biomarkers Be Used to Predict Advanced Liver Diseases? ..... | 182 |
| 7.1       | Study aim.....                                                                | 183 |
| 7.1.1     | Objectives.....                                                               | 183 |
| 7.2       | Methods.....                                                                  | 183 |
| 7.2.1     | Study design.....                                                             | 183 |
| 7.2.2     | Study population .....                                                        | 184 |
| 7.2.3     | Outcomes .....                                                                | 185 |
| 7.2.4     | Statistical analysis .....                                                    | 185 |
| 7.2.5     | Stratification analyses.....                                                  | 187 |
| 7.3       | Results .....                                                                 | 189 |
| 7.3.1     | Patients with biomarkers and outcomes.....                                    | 189 |
| 7.3.2     | Stratification analysis .....                                                 | 199 |
| 7.4       | Discussion .....                                                              | 207 |
| 7.4.1     | Interpreting the association between biomarkers and advanced CLD              | 207 |
| 7.4.2     | Stratification analyses.....                                                  | 217 |
| 7.5       | Conclusion.....                                                               | 223 |
| Chapter 8 | Sensitivity analyses .....                                                    | 226 |
| 8.1       | Objectives.....                                                               | 227 |
| 8.2       | Methods.....                                                                  | 227 |
| 8.2.1     | Starting follow-up for the exposed group at first abnormal test:              | 227 |
| 8.2.2     | Impact of follow-up duration on study outcomes: .....                         | 228 |



|                 |                                                                                                                     |     |
|-----------------|---------------------------------------------------------------------------------------------------------------------|-----|
| 8.2.3           | Assessment of biomarker diagnostic performance and hazard risk across varied cut-offs: .....                        | 229 |
| 8.2.4           | Analysing biomarker efficacy in a common cohort across three biomarkers subgroups: .....                            | 229 |
| 8.3             | Results .....                                                                                                       | 230 |
| 8.3.1           | changing the index date for the exposed group.....                                                                  | 230 |
| 8.3.2           | Minimum Five-Year of Follow-Up .....                                                                                | 232 |
| 8.3.3           | Univariate analysis of different cut-offs and biomarkers diagnostic performance.....                                | 234 |
| 8.3.4           | Cox regression and hazard ratio among people who had records in the three cohorts. ....                             | 242 |
| 8.4             | Discussion .....                                                                                                    | 244 |
| 8.4.1           | Changing the index date for the exposed group.....                                                                  | 244 |
| 8.4.2           | Five-year follow-up sensitivity analysis .....                                                                      | 246 |
| 8.4.3           | Cut-off variations with the sensitivity, specificity, positive predictive value and negative predictive value. .... | 249 |
| 8.4.4           | The efficacy of biomarkers within a common cohort across three subgroups.....                                       | 254 |
| 8.5             | Conclusion.....                                                                                                     | 254 |
| Chapter 9       | Overall Discussion and Conclusions .....                                                                            | 256 |
| 9.1             | Introduction.....                                                                                                   | 257 |
| 9.2             | Summary of main findings .....                                                                                      | 257 |
| 9.3             | Clinical sequences and implications.....                                                                            | 258 |
| 9.4             | Limitations of the study .....                                                                                      | 266 |
| 9.5             | Suggestions for future research .....                                                                               | 268 |
| 9.6             | Conclusion.....                                                                                                     | 271 |
| References      | .....                                                                                                               | 272 |
| Appendices      | .....                                                                                                               | 292 |
| Appendix A..... |                                                                                                                     | 292 |
| Appendix B..... |                                                                                                                     | 294 |
| Appendix C..... |                                                                                                                     | 309 |

|                 |     |
|-----------------|-----|
| Appendix D..... | 330 |
|-----------------|-----|

# List of Tables

|                                                                                                                                                                                               |            |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| <i>Table 1-1: Correspondence and comparison between Ishak, METAVIR and Batts-Ludwig scoring systems, adapted from reference (99). .....</i>                                                   | <i>27</i>  |
| <i>Table 2-1: Characteristics of the included studies for the prevalence of NAFLD in the UK: .....</i>                                                                                        | <i>65</i>  |
| <i>Table 2-2: Detailed characteristics of included studies in the UK systematic review: .....</i>                                                                                             | <i>66</i>  |
| <i>Table 2-3 Quality Assessment of the studies for the prevalence of NAFLD in the UK: .....</i>                                                                                               | <i>72</i>  |
| <i>Table 3-1: Characteristics of included studies for the prevalence of NAFLD in KSA: .....</i>                                                                                               | <i>88</i>  |
| <i>Table 3-2: Detailed characteristics of included studies for the prevalence of NAFLD in the KSA. ....</i>                                                                                   | <i>89</i>  |
| <i>Table 3-3: Quality assessment of included studies for NAFLD prevalence in KSA. ....</i>                                                                                                    | <i>92</i>  |
| <i>Table 5-1: Number of patients over a different time between AST and ALT date (to calculate ast/alt): .....</i>                                                                             | <i>134</i> |
| <i>Table 5-2: Number of patients over a different time between AST and platelet date (to calculate APRI): .....</i>                                                                           | <i>135</i> |
| <i>Table 5-3: Number of patients over a different time between AST and platelet date (when AST and ALT were done on the same day) to calculate Fib-4: .....</i>                               | <i>135</i> |
| <i>Table 5-4: Characteristics of participants (with at least 1 score measured) in the normal and abnormal groups stratified by sex and age group: .....</i>                                   | <i>137</i> |
| <i>Table 5-5: Annual distribution of AST, ALT, platelet, and calculable scores. ....</i>                                                                                                      | <i>139</i> |
| <i>Table 5-6: Regression analysis demonstrating the interaction effect of sex and age on the likelihood of being classified into the abnormal group among AST/ALT population. ....</i>        | <i>141</i> |
| <i>Table 5-7: Regression analysis demonstrating the interaction effect of sex and age on the likelihood of being classified into the abnormal group among the APRI score population....</i>   | <i>142</i> |
| <i>Table 5-8: Regression analysis demonstrating the interaction effect of sex and age on the likelihood of being classified into the abnormal group among the Fib4 index population. ....</i> | <i>143</i> |
| <i>Table 6-1: Cirrhosis medcodes.....</i>                                                                                                                                                     | <i>163</i> |
| <i>Table 6-2: cirrhosis ICD10 codes.....</i>                                                                                                                                                  | <i>165</i> |
| <i>Table 6-3: Medcodes and read terms for liver transplantation. ....</i>                                                                                                                     | <i>166</i> |
| <i>Table 6-4: ICD 10 codes for liver diseases. ....</i>                                                                                                                                       | <i>167</i> |
| <i>Table 6-7: Characteristics of patients with the Outcomes: .....</i>                                                                                                                        | <i>172</i> |
| <i>Table 6-8: Number of patients with the outcomes among each subpopulation: .....</i>                                                                                                        | <i>175</i> |
| <i>Table 6-9: Median follow-up time for each subpopulation and outcomes: .....</i>                                                                                                            | <i>176</i> |
| <i>Table 7-1: Univariate and multivariate cox regression model of the incidence of different outcomes for patients with normal and abnormal ast/alt ratio. ....</i>                           | <i>191</i> |
| <i>Table 7-2: Univariate and multivariate cox regression model of the incidence of different outcomes for patients with normal and abnormal APRI scores: .....</i>                            | <i>194</i> |

|                                                                                                                                                                                                                                                                                                                           |            |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| <i>Table 7-3: Univariate and multivariate cox regression model of the incidence of different outcomes for patients with normal and abnormal Fib4 index: .....</i>                                                                                                                                                         | <i>197</i> |
| <i>Table 7-4: Univariate Cox regression model of the incidence of the (combined outcome) for patients with normal and abnormal ast/alt ratio stratified by sex: .....</i>                                                                                                                                                 | <i>200</i> |
| <i>Table 7-5: Univariate Cox regression model of the incidence of the (combined outcome) for patients with normal and abnormal ast/alt ratio stratified by Age group: .....</i>                                                                                                                                           | <i>200</i> |
| <i>Table 7-6: Univariate Cox regression model of the incidence of the (combined outcome) for patients with normal and abnormal ast/alt ratio stratified by socioeconomic status: .....</i>                                                                                                                                | <i>200</i> |
| <i>Table 7-7: Univariate Cox regression model of the incidence of the (combined outcome) for patients with normal and abnormal APRI score stratified by sex: .....</i>                                                                                                                                                    | <i>201</i> |
| <i>Table 7-8: Univariate Cox regression model of the incidence of the (combined outcome) for patients with normal and abnormal APRI score stratified by Age group: .....</i>                                                                                                                                              | <i>202</i> |
| <i>Table 7-9: Univariate Cox regression model of the incidence of the (combined outcome) for patients with normal and abnormal APRI score stratified by socioeconomic status. ....</i>                                                                                                                                    | <i>202</i> |
| <i>Table 7-10: Univariate Cox regression model of the incidence of the (combined outcome) for patients with normal and abnormal FIB-4 index stratified by sex: .....</i>                                                                                                                                                  | <i>203</i> |
| <i>Table 7-11: Univariate Cox regression model of the incidence of the (combined outcome) for patients with normal and abnormal FIB-4 index stratified by Age group: .....</i>                                                                                                                                            | <i>204</i> |
| <i>Table 7-12: Univariate Cox regression model of the incidence of the (combined outcome) for patients with normal and abnormal FIB-4 index stratified by socioeconomic status: .....</i>                                                                                                                                 | <i>204</i> |
| <i>Table 7-13: Univariate Cox regression model of the incidence of the combined outcomes for patients with normal and abnormal AST/ALT ratio across different time periods. ....</i>                                                                                                                                      | <i>205</i> |
| <i>Table 7-14: Univariate Cox regression model of the incidence of the combined outcomes for patients with normal and abnormal APRI score across different time periods.....</i>                                                                                                                                          | <i>206</i> |
| <i>Table 7-15: Univariate Cox regression model of the incidence of the combined outcomes for patients with normal and abnormal Fib-4 index across different time periods. ....</i>                                                                                                                                        | <i>207</i> |
| <i>Table 8-1: Univariate Cox regression model of the incidence of the (combined outcome) for patients with normal and abnormal AST/ALT, APRI score, and Fib4 when the index date for the exposed group adjusted to be the date of the first abnormal score calculated compared to the original analysis results. ....</i> | <i>231</i> |
| <i>Table 8-2 : Univariate Cox regression model of the incidence of the (combined outcome) for patients with normal and abnormal AST/ALT, APRI score, and Fib4 restricted for patients with at least 5 years of follow-up.....</i>                                                                                         | <i>233</i> |
| <i>Table 8-3: Univariate Cox regression model of the incidence of the combined outcomes for patients with normal and abnormal AST/ALT ratio using different cut-offs. ....</i>                                                                                                                                            | <i>236</i> |
| <i>Table 8-4: Univariate Cox regression model of the incidence of the combined outcomes for patients with normal and abnormal APRI score using different cut-offs. ....</i>                                                                                                                                               | <i>238</i> |

|                                                                                                                                                                                                                                                        |            |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| <i>Table 8-5 Univariate Cox regression model of the incidence of the combined outcomes for patients with normal and abnormal Fib4 index using different cut-offs.....</i>                                                                              | <i>240</i> |
| <i>Table 8-6: Univariate Cox regression model of the incidence of the (combined outcome) for patients with normal and abnormal AST/ALT, APRI score, and Fib4 patients included in all 3 populations compared to the original analysis results.....</i> | <i>244</i> |

# List of Figures

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |     |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <i>Figure 1-1: General overview of the course of CLD, Adapted from reference (15).</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | 3   |
| <i>Figure 1-2: Progression of NAFLD to NASH with or without fibrosis, cirrhosis, and HCC, adapted from reference (50).</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | 10  |
| <i>Figure 2-1: Flow diagram of articles screening and selection process for studies of NAFLD in the UK.</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | 63  |
| <i>Figure 2-2: Forest plot of the Prevalence of NAFLD among disease-unrestricted studies in the UK.</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | 69  |
| <i>Figure 2-3: Forest plot of the Prevalence of NAFLD among disease-restricted studies in the UK.</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | 70  |
| <i>Figure 3-1: Flow diagram of articles screening and selection process for studies of NAFLD in the KSA.</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | 86  |
| <i>Figure 3-2: Forest plot of the Prevalence of NAFLD in KSA by subgroups:</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | 91  |
| <i>Figure 5-1: The process of extracting the study population from the source population (CPRD) and the number of included patients for the analysis.</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | 131 |
| <i>Figure 5-2: Results of data processing for AST from test files in CPRD.</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | 132 |
| <i>Figure 5-3: Results of data processing for ALT from test files in CPRD.</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | 133 |
| <i>Figure 5-4: Results of data processing for Platelet count from test files in CPRD.</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | 133 |
| <i>Figure 5-5: Annual Distribution of calculable scores.</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | 140 |
| <i>Figure 5-6: Annual Distribution of AST, ALT, and Platelet conducted for the study population.</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | 140 |
| <i>Figure 7-1: Kaplan-Meier plots comparing the incidence of combined outcomes for patients with normal ast/alt biomarkers and patients with abnormal ast/alt biomarkers. The Y-axis represents the probability of patients' free outcome. The X-axis represents the follow-up time (years) since enrolment in the study. The solid line represents subjects with normal biomarkers. The dashed line represents patients with abnormal biomarkers. The difference between the two curves is statistically significant (<math>P&lt;0.001</math>), indicating a clear disparity in survival probabilities based on AST/ALT ratio status.</i> | 192 |
| <i>Figure 7-2: The log-log plot analysis assesses the proportion hazards assumption by comparing the incidence of combined outcomes between patients with normal AST/ALT biomarkers and those with abnormal AST/ALT biomarkers. On this plot, the x-axis represents the logarithm of analysis time, and the y-axis displays the natural logarithm of the negative natural logarithm of survival probabilities.</i>                                                                                                                                                                                                                         | 192 |
| <i>Figure 7-3: Kaplan-Meier plots comparing the incidence of combined outcome for patients with normal APRI score and patients with abnormal APRI score. The Y-axis represents the probability of patients' free outcome. The X-axis represents the follow-up time (years) since</i>                                                                                                                                                                                                                                                                                                                                                       |     |

|                                                                                                                                                                                                                                                                                                                                                                                                                           |     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| enrolment in the study. The solid line represents subjects with normal biomarkers. The dashed line represents patients with abnormal biomarkers. ....                                                                                                                                                                                                                                                                     | 195 |
| Figure 7-4: The log-log plot analysis assesses the proportion hazards assumption by comparing the incidence of combined outcomes between patients with normal APRI biomarkers and those with abnormal APRI biomarkers. On this plot, the x-axis represents the logarithm of analysis time, and the y-axis displays the natural logarithm of the negative natural logarithm of survival probabilities.....                 | 195 |
| Figure 7-5: Kaplan-Meier plots comparing the incidence of combined outcome for patients with normal Fib-4 and patients with abnormal Fib-4. The Y-axis represents the probability of patients' free outcome. The X-axis represents the follow-up time (years) since enrolment in the study. The solid line represents subjects with normal biomarkers. The dashed line represents patients with abnormal biomarkers. .... | 198 |
| Figure 7-6: The log-log plot analysis assesses the proportion hazards assumption by comparing the incidence of combined outcomes between patients with normal Fib4 biomarkers and those with abnormal Fib4 biomarkers. On this plot, the x-axis represents the logarithm of analysis time, and the y-axis displays the natural logarithm of the negative natural logarithm of survival probabilities.....                 | 198 |
| Figure 8-1: Univariate Cox regression model of the incidence of combined outcomes for patients with normal and abnormal AST/ALT ratios using different cut-off. ....                                                                                                                                                                                                                                                      | 237 |
| Figure 8-2: Univariate Cox regression model of the incidence of combined outcomes for patients with normal and abnormal APRI score using different cut-offs. ....                                                                                                                                                                                                                                                         | 239 |
| Figure 8-3 Univariate Cox regression model of the incidence of combined outcomes for patients with normal and abnormal FIB4 index using different cut-offs .....                                                                                                                                                                                                                                                          | 242 |
| Figure 8-4 Venn diagram showing the overlap among patients in different cohorts. ....                                                                                                                                                                                                                                                                                                                                     | 243 |

## List of Abbreviations

|                |                                                           |
|----------------|-----------------------------------------------------------|
| <b>AASLD</b>   | The American Association for the Study of Liver Diseases  |
| <b>ALP</b>     | Alkaline phosphatase                                      |
| <b>ALT</b>     | Alanine Aminotransferase                                  |
| <b>AIH</b>     | Autoimmune hepatitis                                      |
| <b>APASL</b>   | Asian Pacific Association for the Study of the Liver      |
| <b>APRI</b>    | AST/Platelet Ratio Index                                  |
| <b>ARFI</b>    | Acoustic radiation force impulses                         |
| <b>ARLD</b>    | Alcohol-Related Liver Disease                             |
| <b>ASH</b>     | Alcoholic Steatohepatitis                                 |
| <b>AST/ALT</b> | Aspartate aminotransferase/alanine aminotransferase ratio |
| <b>AUROC</b>   | The area under receiver operation curve                   |
| <b>BMI</b>     | Body Mass Index                                           |
| <b>CI</b>      | Confidence interval                                       |
| <b>CLD</b>     | Chronic Liver Disease                                     |
| <b>CPRD</b>    | Clinical Practice Research Datalink                       |
| <b>CT</b>      | Computed tomography                                       |
| <b>CVD</b>     | Cardiovascular disease                                    |
| <b>EASD</b>    | European Associations for the Study of Diabetes           |
| <b>EASL</b>    | European Associations for the Study of the Liver          |
| <b>EASO</b>    | European Associations for the Study of Obesity            |
| <b>ECM</b>     | Extracellular Matrix                                      |
| <b>EHRs</b>    | Electronic Health Records                                 |
| <b>ELF</b>     | Enhanced liver fibrosis score                             |
| <b>FIB-4</b>   | Fibrosis 4 Index                                          |
| <b>GGT</b>     | gamma-glutamyl transferase                                |
| <b>GI</b>      | glycaemic index                                           |
| <b>GOT</b>     | Glutamic Oxaloacetic Transaminase                         |
| <b>GP</b>      | General practice                                          |
| <b>GPRD</b>    | General Practise Research Database                        |
| <b>GPT</b>     | Serum Glutamic Pyruvic Transaminase                       |
| <b>HA</b>      | Hyaluronic acid                                           |
| <b>HBV</b>     | Hepatitis B viral infection                               |
| <b>HCC</b>     | hepatocellular carcinoma                                  |
| <b>HCV</b>     | Hepatitis C viral infection                               |
| <b>HES</b>     | Hospital Episode Statistics                               |
| <b>HIV</b>     | Human Immunodeficiency Virus                              |
| <b>HR</b>      | Hazard Ratio                                              |
| <b>HSCs</b>    | Hepatic Stellate Cells                                    |
| <b>ICD10</b>   | International Classification of Diseases                  |
| <b>INR</b>     | The international normalized ratio                        |
| <b>ISAC</b>    | Independent Scientific Advisory Committee                 |
| <b>KM</b>      | Kaplan-Meier                                              |



|                       |                                                                |
|-----------------------|----------------------------------------------------------------|
| <b>KSA</b>            | Kingdom of Saudi Arabia                                        |
| <b>MASLD</b>          | Metabolic Dysfunction-Associated Steatotic Liver Disease       |
| <b>MELD</b>           | The Model for End-Stage Liver Disease                          |
| <b>MENA</b>           | Middle East and North Africa                                   |
| <b>MeSH<br/>Terms</b> | Medical Subject Headings                                       |
| <b>MHRA</b>           | The Medicines and Healthcare Products Regulatory Agency        |
| <b>MINAP</b>          | Myocardial Ischemia National Audit Project                     |
| <b>MRE</b>            | Magnetic Resonance Elastography                                |
| <b>MRI</b>            | Magnetic resonance imaging                                     |
| <b>MRI-<br/>PDFF</b>  | Magnetic Resonance Imaging-Derived Proton Density Fat Fraction |
| <b>MRS</b>            | Magnetic Resonance Spectroscopy                                |
| <b>NAFL</b>           | Non-Alcoholic Fatty Liver                                      |
| <b>NAFLD</b>          | Non-Alcoholic Fatty Liver Disease                              |
| <b>NAS</b>            | NAFLD Activity Score                                           |
| <b>NASH</b>           | Non-alcoholic steatohepatitis                                  |
| <b>NCAN</b>           | Neurocan                                                       |
| <b>NCRAS</b>          | National Cancer Registration and Analysis Service              |
| <b>NHS</b>            | National Health Service                                        |
| <b>NICE</b>           | The National Institute for Health and Care Excellence          |
| <b>NPV</b>            | Negative Predictive Value                                      |
| <b>ONS</b>            | Office of National Statistics                                  |
| <b>OR</b>             | Odd ratio                                                      |
| <b>OXMIS</b>          | Oxford Medical Information system                              |
| <b>PBC</b>            | Primary biliary cirrhosis                                      |
| <b>PIIINP</b>         | Type III Procollagen Peptide                                   |
| <b>PPV</b>            | Positive Predictive Value                                      |
| <b>PSC</b>            | Primary Sclerosing Cholangitis                                 |
| <b>PT</b>             | Prothrombin time                                               |
| <b>QCRI</b>           | Qatar Computing Research Institute                             |
| <b>QOF</b>            | The Quality and Outcomes Framework                             |
| <b>T1DM</b>           | Type 1 Diabetes Mellitus                                       |
| <b>T2DM</b>           | Type 2 Diabetes Mellitus                                       |
| <b>TACE</b>           | Transarterial chemoembolization                                |
| <b>TE</b>             | Transient Elastography                                         |
| <b>TIMP-1</b>         | Tissue inhibitors of metalloproteinase                         |
| <b>UK</b>             | United Kingdom                                                 |
| <b>US</b>             | Ultrasonography                                                |
| <b>UTS</b>            | Up to standard                                                 |
| <b>VAMP</b>           | Value-Added Medical Products                                   |
| <b>WHO</b>            | World Health Organization                                      |

# **Chapter 1      Background and thesis rationale**

## **1.1 Introduction**

Advanced liver diseases encompass a broad range of disorders that cause significant damage to the organ, including cirrhosis and liver failure (1). These diseases result from chronic inflammation, injury or scarring of the liver, which can be caused by factors such as Alcohol-related liver disease (ARLD), viral hepatitis, and, most commonly, Non-Alcoholic Fatty Liver Disease (NAFLD) (2-4).

Chronic liver disease (CLD) has become a significant global health issue that results in high mortality rates (5, 6); however, CLD often remains undiagnosed until it has reached an advanced stage (7). This has led to calls for healthcare providers to develop effective strategies for early detection and management of CLD, with a particular focus on those diseases linked to lifestyle issues (8), such as NAFLD, ARLD, and viral hepatitis (7).

The substantial healthcare burden associated with CLD emphasizes the urgent need for effective treatment and management strategies (9, 10), and current evidence demonstrates the significance of recognising high-risk patients in general practice (11). The World Health Organization (WHO) and several scientific organisations have advocated non-invasive biomarkers as a potential way to identify liver fibrosis as an alternative to liver biopsies; however, they have called for further research to test their efficacy and reliability (12). In response, this thesis uses routinely collected data to explore the potential of non-invasive biomarkers to predict advanced CLD among the general practice population in the UK.

## 1.2 What is chronic liver disease (CLD) and how does it develop?

CLD is a progressive deterioration of liver functions (2). The liver has numerous essential functions, including the production of coagulation factors and other proteins, the excretion of bile, and the detoxification of harmful products of metabolism (2). Chronic inflammation, destruction, and regeneration of liver parenchyma can lead to these functions being compromised. As shown in Figure 1-1, CLD can progress from early stages, such as fatty liver, to more advanced stages, such as fibrosis, cirrhosis, and HCC (13, 14).

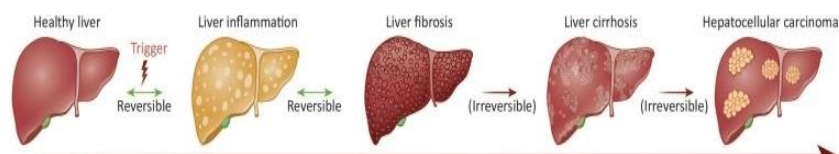


Figure 1-1: General overview of the course of CLD, Adapted from reference (15).

Liver fibrosis is a common feature of CLD. It is characterised by the accretion of extracellular matrix (ECM) proteins in the liver, including but not limited to collagens, especially collagen I and III, which can be induced due to various types of chronic liver injuries (16-18). Several events occur in the process of liver fibrogenesis, such as the activation of hepatic stellate cells (HSCs), bone marrow-derived cells and myofibroblasts (19).

The formation of a fibrous scar commonly deforms the typical architecture of the liver due to ECM protein accumulation. This deformation impairs hepatocellular function and enhances the intrahepatic opposition to blood flow, causing hepatic insufficiency and

portal hypertension, respectively and may cause liver failure and death (2).

Cirrhosis is the final stage of CLD and is defined as the irreversible deformation of liver architecture by fibrosis, scarring, and aberrant nodules (20). The formation of fibrous scar tissue can cause the liver to become hard and lumpy, and the liver tissue may be replaced by scar tissue and nodules (20).

Liver cirrhosis can cause portal hypertension since disruption of the normal liver architecture impairs blood flow through the liver creating a back pressure in the portal vein. This can lead to complications such as oesophageal varices and ascites (21). As cirrhosis progresses, the liver's ability to function normally is significantly impaired due to the diminishing number of functioning hepatocytes, which substantially increases the risk of liver failure (21).

### **1.3 Epidemiology and public health impact of CLD**

In 2017, there were 1.5 billion cases of CLD worldwide (22). The epidemiology of CLD varies widely between and even within countries and is subject to various factors, such as lifestyle behaviours, socioeconomic conditions, cultural influences, and historical events promoting each aetiology (23). Individuals of advanced age are more likely to have CLD than younger individuals (24). This is partially caused by ageing naturally, which can have an impact on the liver's capacity to maintain homeostasis and the accumulation of risk factors over time, including chronic alcohol use and viral hepatitis. There is an especially

notable increase in the prevalence of NAFLD in older age groups, particularly in type 2 diabetics (T2DM) (24).

Public health approaches require comprehensive efforts to tackle this problem. Promoting healthy lifestyles, reducing risk factors, and providing access to early detection and effective treatment options are essential to prevent the development of CLD and its severe complications (2, 25).

According to estimates from the WHO, cirrhosis and liver cancer are the eleventh and sixteenth leading causes of death, accounting for 1.16 million and 788,000 deaths annually, respectively (25, 26).

## **1.4 Aetiology of CLD**

Globally, the most prevalent underlying cause for CLD is NAFLD (60%), followed by Hepatitis B viral infection (HBV) (29%), Hepatitis C viral infection (HCV) (9%), and ARLD (2%) (27). However, the relative prevalence of these causes exhibits country-specific disparities, with certain regions witnessing higher rates of ARLD and viral hepatitis, while others confront a surge in NAFLD cases attributable to the escalating prevalence of obesity and metabolic syndrome (28).

Global variations in CLD incidence and prevalence stem from diverse factors, encompassing discrepancies in genetics, environmental risk factors, lifestyle choices, healthcare accessibility, and socioeconomic status. The prevalence of HBV and HCV further fluctuates across different regions worldwide due to variations in risk

factor prevalence, such as injection drug use, unsafe medical procedures, and mother-to-child transmission (29, 30).

Other Less common causes of CLD include (2):

- Genetic causes such as Alpha-1 antitrypsin deficiency, Hereditary hemochromatosis and Wilson disease.
- Autoimmune causes such as Primary biliary cirrhosis (PBC), Primary Sclerosing Cholangitis (PSC), and Autoimmune hepatitis (AIH).
- Drugs: methotrexate, isoniazid, Amiodarone, nitrofurantoin, phenytoin, and others.

## **1.5 Liver diseases in the UK and KSA**

In the UK, liver diseases are the 3<sup>rd</sup> most common cause of early death (8). The prevalence of liver fibrosis in the UK is difficult to determine as it often goes undiagnosed until it progresses to a more serious condition such as decompensated cirrhosis (31). However, an estimated one in five individuals in the UK has early-stage liver fibrosis (7, 32). This is often associated with risk factors such as obesity, T2DM, and excessive alcohol consumption. The recorded premature deaths due to liver diseases in the UK between 2015 and 2017 were 26,265 cases, with a rate of 18.5/100,000 population younger than 75 years old (33). Moreover, liver cirrhosis remains a significant risk factor for HCC, which ranks eighth in the list of leading causes of cancer deaths in the UK (34).

In the context of the UK, the future curve of NAFLD is of particular concern. The comprehensive study by Estes et al (35), which modelled

the NAFLD disease burden across various countries including the UK, provides critical insights into the expected trends and challenges for the upcoming years. The study indicates a significant and growing burden of NAFLD, with projections showing more substantial increase in advanced cases of the disease. This trend is particularly concerning in the context of the UK, where the already high levels of obesity and T2DM are key drivers of NAFLD prevalence. Despite any potential levelling off of obesity rates, the prevalence of NAFLD is expected to continue rising, largely due to the lag in the impact of obesity and T2DM on liver health.

As NAFLD has emerged as the predominant cause of CLD in developed nations, such as the UK, it is anticipated that rapidly developing countries, such as KSA, will experience a similar trend in the upcoming years (5, 36). Consequently, understanding and addressing NAFLD is crucial for the effective management of liver disease globally. This study focuses specifically on NAFLD within the broader spectrum of CLD to provide targeted insights and interventions that are directly relevant to current and emerging public health challenges.

The Kingdom of Saudi Arabia (KSA) is a rapidly developing country in the Middle East, with strong economic growth, and the accompanying changes in the prevalence of obesity and T2DM (37). Liver disease is now a significant health concern and HBV, HCV, NAFLD, and ARLD are recognised as significant contributors to CLD, resulting in varying prevalence rates (38, 39). In KSA context, NAFLD is becoming the leading cause of liver transplants, overtaking hepatitis-induced liver disease (40). Studies indicate that 22.5% of patients with



CLD have NAFLD (41). In 2017, KSA reported an incidence and mortality rate of 10.5 and 1.2 per 100,000 people, respectively, from cirrhosis caused by non-alcoholic steatohepatitis (NASH) (42). Additionally, there was a 25% increase in cirrhosis deaths between 2007 and 2017, with cirrhosis ranked 10th as a cause of premature death (43).

## **1.6 Non-alcoholic Fatty Liver Disease (NAFLD)**

### **1.6.1 NAFLD definitions and staging**

NAFLD encompasses a spectrum of liver conditions characterized by excessive hepatic fat build-up not attributed to alcohol consumption or other recognized primary or secondary causes of liver disease (44). This complex disease progresses through several stages, each with its distinct features. The initial stage of NAFLD is simple fatty liver, also known as Steatosis or Non-Alcoholic Fatty Liver (NAFL). It is defined by the presence of fat globules in more than 5% of hepatocytes without detectable inflammation (45). As the disease progresses to NASH, it is characterized by over 5% hepatic steatosis and inflammation, with hepatocyte ballooning, with or without fibrosis, causing liver swelling and tenderness (45).

The third stage, fibrosis, involves inflammation and scar tissue formation as the liver attempts to repair damage (46). The severity of fibrosis is crucial for prognosis, with various histopathological scoring systems used to measure its progression (47). Detailed scoring systems are discussed in section 1.8.1. The fourth stage, cirrhosis, is severe liver scarring leading to loss of liver function, taking decades to develop.

Cirrhosis takes a long time to develop and may even take decades; thus, therapies can be initiated early before this happens (46). Cirrhosis can lead to liver failure, cancer, and other complications (44, 48, 49). Figure 1-2 shows the progression of NAFLD into advanced liver disease and HCC. Additionally, the figure also includes factors that have the potential to influence the histological progression of NASH.

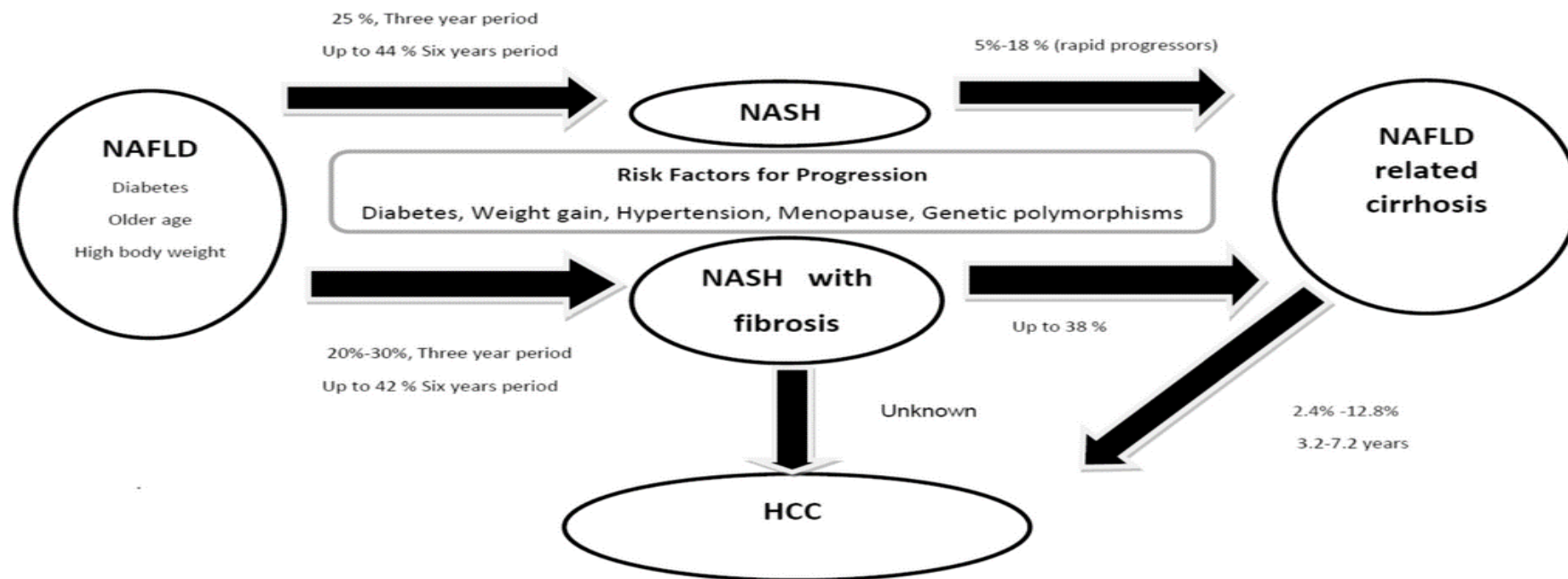


Figure 1-2: Progression of NAFLD to NASH with or without fibrosis, cirrhosis, and HCC, adapted from reference (50).

### **1.6.2 Global epidemiology**

In the past two decades, the prevalence of NAFLD has surpassed other causes of liver diseases, emerging as the most prevalent liver condition globally (51). However, the reliability of the reported prevalence of NAFLD can vary depending on the population under investigation and the diagnostic criteria utilized (51). A key study in assessing the growing worldwide problem of NAFLD is the systematic review and meta-analysis by Younossi et al. (52). This three-decade analysis carefully aggregates data from 1990 to 2019 to provide a broad picture of the epidemiological landscape of NAFLD. The authors report a global prevalence of NAFLD of about 30% based on 92 studies with a total sample size of more than 9 million people.

The results of Younossi et al. show notable regional differences in the prevalence of NAFLD, which is consistent with the larger body of literature. The regions with the highest reported prevalence were the Middle East and North Africa (MENA) regions at 42.6%, Latin America at 34.5%, and Asia at 30.9%. Notably, the analysis revealed a significant increase in prevalence rates over time, with rates rising from 24.4% during 1991-2006 to 38.0% during 2016-2020.

Therefore, while global prevalence data offer insights, caution should be exercised when extrapolating these findings to specific regions or populations without considering potential limitations and variations in diagnostic practices and coding.

Moreover, the prevalence and course of NAFLD are significantly influenced by ethnicity, indicating that genetic predispositions and distinct cultural and lifestyle factors have a major impact on disease development (53). A study by Al-Hazzaa et al. (54) underscores the importance of considering cultural, social, and environmental backgrounds when examining health behaviours. For example, the study found a significantly higher prevalence of overweight/obesity among Saudi adolescents (38.3%) compared to their British counterparts (24.1%). This difference could be attributed to varying lifestyle habits and dietary patterns influenced by cultural and environmental factors in each country.

So, addressing research gaps, creating focused prevention and treatment plans, and directing healthcare policy all depend on an understanding of these localised dynamics. As a result, the subsequent chapters employed a systematic review approach to explore the prevalence of NAFLD in the UK and KSA populations, ultimately contributing to the overarching goal of predicting advanced liver diseases.

### **1.6.3 Risk factors**

NAFLD is recognized as the hepatic manifestation of metabolic syndrome, characterized by interrelated metabolic abnormalities including obesity, dyslipidaemia, T2DM, and insulin resistance (48, 55). These factors are key risk factors for NAFLD due to their strong association with its incidence and severity (56).

The risk factors frequently overlap and interact; obesity often coincides with insulin resistance and dyslipidaemia, which are central to metabolic syndrome. Similarly, T2DM is closely related to both insulin resistance and obesity, emphasizing the interconnection among these risk factors (56). Patients exhibiting these risk factors are at a higher risk of progressing to advanced CLD. Early identification and management of these risk factors can help healthcare providers prevent the progression of NAFLD to more severe complications (see [1.7.2](#)).

#### **1.6.4 Diagnosis**

Diagnosing NAFLD involves a comprehensive approach that integrates various diagnostic procedures alongside medical history and physical examination findings. Patients with risk factors such as obesity, T2DM, high cholesterol, or elevated triglycerides are often prioritized for NAFLD evaluation. While LFTs are crucial for assessing liver health, they may not definitively confirm NAFLD, as the disease can progress without significant alterations in liver enzyme levels (57). Imaging modalities such as Ultrasonography (US), Computed tomography (CT), or Magnetic resonance imaging (MRI) are pivotal in detecting NAFLD. In certain cases, a liver biopsy may be necessary to accurately evaluate liver damage, particularly fibrosis. Furthermore, it's imperative to employ diagnostic criteria for the diagnosis of exclusion, ruling out other potential causes such as medication use or underlying conditions that may lead to secondary forms of the disease (58). This comprehensive diagnostic approach ensures a thorough assessment and accurate diagnosis of NAFLD, facilitating appropriate management strategies.

### **1.6.5 Clinical Presentation**

The asymptomatic nature of NAFLD in its early stages can make it challenging to identify the disease solely based on clinical symptoms or abnormal liver enzyme levels (59). The clinical manifestations of NAFLD may include pancreatic, hepatic, and cardiac endothelial cell dysfunction and illness (28).

### **1.6.6 Management**

Drug therapy for NAFLD aims to address fat accumulation, associated damage, and fibrosis (60, 61). Several medications, including pentoxifylline, UDCA, omega-3 fatty acids, and metformin, are under consideration but have not yet received approval, as clinical trial results have been inconsistent (62, 63)

Lifestyle interventions are crucial for managing NAFLD, focusing on dietary modifications, weight loss, and addressing metabolic risks (64). Effective dietary strategies involve achieving a negative calorie balance and substituting high-glycaemic index (GI) foods with low-GI, high-fiber options to improve insulin sensitivity and reduce T2DM and NAFLD risks (65, 66). Regular physical activity, essential for managing NAFLD and enhancing overall health, decreases the risk of metabolic syndrome, T2DM, and dyslipidaemia (67).

### **1.6.7 Outcome/Prognosis**

NAFLD is associated with a higher overall mortality rate compared to age and sex-matched controls, primarily due to vascular

and liver-related causes (68, 69). It is a significant factor in the progression to cirrhosis and potentially HCC, and by 2030, NAFLD is expected to become the leading indication for liver transplantation (70).

The prognosis for patients with NAFLD varies significantly based on the histological stage of the disease. Individuals with simple steatosis, characterized by excess fat in the liver without inflammation or fibrosis, face a relatively low risk of progressing to cirrhosis, estimated at 0% to 4% over 20 years (71). This understanding is critical for risk stratification and guiding treatment decisions, allowing for the identification of patients at risk of disease progression and helping to tailor management strategies to prevent overdiagnosis and unnecessary treatment in those with less severe forms of the disease.

However, the progression rates to cirrhosis in patients with NASH vary, underscoring the need for continued research to support clinical decision-making. Studies indicate that nearly 10% of individuals with NASH may develop decompensated liver disease over a 13-year period, and about 25% may progress to cirrhosis within nine years. Importantly, the presence and severity of fibrosis are major determinants of disease progression, with individuals without fibrosis facing significantly lower mortality risks than those with advanced fibrosis (50).

## **1.7 Early detection of chronic liver disease**

### **1.7.1 Recommendations for early CLD detection**

To address the growing health concern of NAFLD, guidelines from organizations such as the National Institute for Health and Care



Excellence (NICE), the American Association for the Study of Liver Diseases (AASLD), the European Association for the Study of the Liver (EASL), the European Association for the Study of Diabetes (EASD), and the European Association for the Study of Obesity (EASO) recommend implementing both opportunistic and population screening programs. These programs aim to facilitate evaluation and promote patient lifestyle modifications (72-74). Furthermore, early detection and intervention in NAFLD can significantly improve histological and metabolic outcomes, enhance overall health, and reduce the burden of advanced liver disease (75). However, there is a crucial need for less invasive methods to screen for NAFLD, evaluate disease severity, and monitor its progression (76).

In this thesis, the emphasis is placed on utilizing biomarkers to predict advanced CLD rather than solely focusing on NAFLD, for several reasons. Firstly, this approach is informed by significant data limitations encountered when utilizing electronic health records (EHRs) such as the Clinical Practice Research Datalink (CPRD) and Hospital Episode Statistics (HES). Specifically, the absence of specific diagnostic codes for NAFLD complicates the accurate identification and analysis of this patient subgroup. This coding limitations is coupled with the potential for misclassification between NAFLD and ARLD. In contrast, advanced CLD diagnoses are more consistently coded within these systems, enabling more reliable extraction and analysis of data related to liver disease progression. The challenges of accurately determining NAFLD prevalence in the general population using EHRs are well-documented

with studies such as Åström et al. (77), who note the lack of standardized diagnostic codes as a major barrier to identifying NAFLD cases accurately, which causes irregularities and errors in the identification and categorization of NAFLD cases.

The challenge of defining and coding the progression from NAFLD to more advanced stages, such as cirrhosis, introduces further complexity. Without validation studies for these codes, the reliability of EHR-derived data is questionable, underscoring the need for a broader focus on CLD in this research.

Furthermore, Interventions aimed at controlling or managing CLD could significantly mitigate the impact of NAFLD, which is projected to become the predominant cause of CLD in the future, both in KSA and the UK.

In addition, the recent shift in terminology from NAFLD to Metabolic Dysfunction-Associated Steatotic Liver Disease (MASLD) reflects a broader understanding of the disease as being related to metabolic dysfunctions. By studying CLD, I inherently include this important subset, allowing my research to stay relevant with current medical understanding and terminology. The concept of MASLD has evolved significantly from the traditional understanding of NAFLD. MASLD is recognized for its association with metabolic dysfunctions, encompassing a wider range of systemic disorders. The change from NAFLD to MASLD was proposed by multinational liver associations to better reflect the metabolic complexities underlying the disease. This

redefinition emphasizes the role of systemic insulin resistance and related metabolic dysfunctions in the development of liver-related morbidities and extrahepatic complications, including cardiovascular disease (CVD) and certain types of cancers (78).

### **1.7.2 Importance of early disease detection**

As liver disease prevalence is growing, healthcare providers in primary and secondary care must develop effective strategies to detect and manage CLD (51), focusing on addressing the disease's primary causes (see [1.4](#)). Fortunately, preventative measures can be taken to reduce the incidence of these causes, and HCV is now curable (7).

However, CLD is often diagnosed at an advanced stage (7). Three-quarters of individuals with cirrhosis are not diagnosed until they develop potentially life-threatening symptoms of advanced disease (8). In the early phases of cirrhosis, compensation occurs. At this stage, the majority of patients are asymptomatic, and cirrhosis is typically discovered during unrelated medical encounters (79). In patients with compensated cirrhosis, decompensation is typically defined as the onset of ascites, oesophageal variceal bleeding, hepatic encephalopathy, and, in some cases, an elevated bilirubin concentration (80, 81). Due to the nature of decompensation, these patients are quickly brought to medical attention; consequently, reports on the prevalence of decompensated cirrhosis are likely to be much more accurate than those of compensated cirrhosis (80).

Once decompensation occurs, mortality and morbidity due to cirrhosis increase dramatically, and depending on the source of decompensation, the 1-year case-fatality rate can reach up to 80% (20, 82). Patients with compensated cirrhosis have a life expectancy comparable to healthy adults if cirrhosis remains compensated (20). However, identifying these patients is crucial, as they are at risk of developing decompensated cirrhosis if they are not managed appropriately (80).

The prolonged natural history of the disease, characterised by progressive fibrosis leading to cirrhosis, in conjunction with well-established risk factors such as alcohol abuse, obesity, and metabolic syndrome, provides good opportunities for earlier diagnosis (7). In fact, up to 90% of CLD cases, including those related to ARLD, NAFLD, and viral hepatitis, are preventable through early detection and intervention (7).

When CLD is detected at an early stage, there is a better chance of implementing timely interventions and treatments such as weight loss programs, alcohol brief interventions, antiviral therapies, and screening for complications like varices and HCC before permanent damage occurs (83). This can prevent or slow down disease progression to more severe complications such as cirrhosis, liver failure, or HCC (83).

However, identifying at-risk patients and prioritising their care can be challenging for healthcare providers already struggling to meet the needs of patients with established CLD (7). In order to offer optimal care

to patients with CLD, the Lancet Commission on Liver Disease in 2013 made several recommendations. These included policy initiatives, such as strategies to address the obesity epidemic and minimum unit pricing for alcohol, and enhancing secondary care services by establishing liver centres in district general hospitals staffed by hepatologists (84).

Notably, more than seventy-five per cent of patients were not referred to liver services prior to their first hospitalisation for CLD decompensation (8). Therefore, a key recommendation was to strengthen early detection of liver disease, which would require collaboration between primary and secondary care providers.

Patients with early diagnosis have a higher chance of maintaining a better quality of life, experiencing reduced symptoms, and incurring lower healthcare costs. On the other hand, late diagnosis often means that the disease has advanced significantly, leading to limited treatment options, poorer prognosis, increased risk of severe complications, reduced quality of life, and higher healthcare expenses. Therefore, early detection of CLD is crucial to maximise the potential benefits for patients and minimise the disease burden on individuals and healthcare systems (85).

### **1.7.3 Strategies for early detection of CLD**

Early detection of CLD may be a key step in enhancing patient outcomes and effectively managing healthcare resources (72-74). There are various strategies for early detection, each with its own advantages

and challenges. The main methods include general population screening, opportunistic screening, and pragmatic case finding.

**General Population Screening:** This approach involves systematically testing a large portion or the entire population for liver disease, regardless of individual risk factors or symptoms. Its primary advantage lies in the potential for early disease detection across a wide population. However, this method is resource-intensive and may lead to over-diagnosis, particularly in low-risk individuals. General population screening could result in the inefficient use of healthcare resources, especially in regions where CLD prevalence is relatively low (86). An example of this approach would be a nationwide health initiative where all adults above a certain age undergo liver function tests, regardless of their individual risk factors or symptoms. While comprehensive, this approach aims to identify and address liver diseases at an early stage across the entire population, not solely in those already exhibiting symptoms or known to be at high risk.

**Opportunistic Screening:** This method refers to the practice of screening patients for CLD during routine healthcare visits, regardless of their reason for the visit. This approach takes advantage of the opportunity presented by the patient's presence in the healthcare setting to assess for CLD, even if the patient has not been pre-identified or specifically come in for liver disease screening. It offers a more targeted approach than general population screening but has limitations. It may overlook individuals who infrequently interact with the healthcare system or those who may be at risk but do not exhibit clear symptoms or risk

factors during their healthcare visits (87). An example of the opportunistic screening for CLD is when a patient goes for a regular health check-up or for unrelated medical issues, and the healthcare provider conducts liver function tests or checks for liver disease markers as part of the examination. This approach allows for the potential early detection of liver issues in patients who are not specifically there for liver disease screening and may not have obvious symptoms or known risk factors.

**Pragmatic Case Finding and risk stratification:** This strategy focuses on identifying individuals with a higher risk of developing CLD based on specific risk factors such as alcohol consumption, obesity, metabolic syndrome, or family history (85). Pragmatic Case Finding is more targeted than both general population and opportunistic screening. It concentrates resources on individuals most likely to benefit from early detection and intervention, enhancing healthcare efficiency and effectiveness. This method aligns with limited healthcare resources and focuses on high-risk groups, making it an efficient approach to manage and prevent the progression of CLD.

An example of the pragmatic case finding for CLD involves identifying and screening patients who are already known to have risk factors for liver disease. For instance, liver services like the Scarred Liver project in Nottingham have implemented strategies where individuals with documented risk factors in their medical records are specifically targeted for imaging techniques like fibroscan during their healthcare visits (88). In this approach, healthcare providers selectively test those

patients who have a higher likelihood of developing CLD based on their medical history, lifestyle factors, or family background, rather than screening every patient indiscriminately. This targeted approach aims to efficiently identify individuals at risk of CLD progression, allowing for timely interventions and management strategies.

In line with this focus, this thesis applies the pragmatic case finding approach to a specific subset of individuals who have undergone AST testing, identified within the dataset. AST tests, often indicative of liver health issues (see [1.8.4.1.1](#)), allow for the targeted examination of a group potentially more susceptible to CLD. The next section will focus on exploring the essential features of an ideal screening test for the early detection of advanced CLD.

#### **1.7.4 Characteristics of an ideal screening test for early advanced CLD detection.**

Clearly defining the different purposes of screening and diagnostic tests is essential. According to the 1998 UK National Screening Committee, screening tests are methodical applications of a test or inquiry intended to identify people who are at a high risk of developing a particular disorder to justify additional investigation or taking immediate preventive action. Usually, these people don't exhibit any symptoms of the illness. For instance, blood tests are not used for diagnosis in the context of cirrhosis; rather, they are used as screening tools to identify individuals at higher risk (89).



A diagnostic test is a crucial part of the first phase of clinical management which is employed to identify the illness or condition an individual may have, guided by their symptoms and signs. So, diagnostic tests are targeted and specific, used to confirm or rule out a condition in symptomatic individuals or in those who are screen positive, whereas screening tests serve a preventative purpose, applied broadly to asymptomatic populations (89). The primary goal of screening in this thesis is to risk stratify a population, identifying individuals who, despite not showing symptoms, may be at higher risk of developing advanced CLDs.

These characteristics, which are critical for determining the effectiveness of a screening test for early advanced CLD detection, will be assessed for the various screening methods for the non-invasive biomarkers for advanced CLD.

**General Requirements:** Cost-Effectiveness: The main factor to take into account is a screening test's economic viability. Economical solutions guarantee the effective distribution of healthcare resources, enabling a wider population to have access to CLD screening (90).

Acceptability: The screening test needs to be socially and culturally acceptable to the target population in order to be widely adopted. Convenience and comfort factors are important in encouraging test acceptability (91).

Availability: The successful implementation of CLD screening programmes depends on having easy access to testing facilities and resources (92).

Safety: This is still the top priority. In order to promote trust and engagement, the screening test should reduce any potential risks to participants' physical and mental health (93).

Pragmatics: It is imperative that the test be easily integrated into standard medical procedures. The test should be simple for medical professionals to administer, guaranteeing widespread adoption (94).

#### **Performance characteristics:**

Sensitivity: Accurately identifying people with CLD or its risk factors requires a high level of sensitivity. By reducing the possibility of false negative results, it guarantees that individuals in need of assistance are correctly identified (95). For a risk stratification test, high sensitivity specifically means the test effectively identifies all individuals who are at a higher risk of developing advanced CLD.

Specificity: A test with high specificity accurately identifies people without CLD or the risk of developing it. High specificity reduces the number of false positives, which in turn minimizes patient anxiety and avoids unnecessary follow-up testing (95). In the context of a risk stratification test, high specificity ensures the accurate identification of individuals who are not at risk of developing advanced CLD.

Positive Predictive Value (PPV): A high PPV minimizes overdiagnosis and unnecessary interventions by ensuring that a positive

test result reliably indicates the presence of CLD (95). A test with a high PPV means that individuals identified as at risk truly have a significant likelihood of disease progression.

Negative Predictive Value (NPV): A high NPV accurately rules out the presence of CLD and provides confidence in negative test results, which is essential for reassuring patients (95). For a risk stratification test, High NPV means that those identified as negative on the screen are highly unlikely to have Advanced CLD.

Reliability: It is crucial that the test delivers accurate results over time. Desirable characteristics in repeated testing are low variability and high reliability (96).

## **1.8 Methods for early detection of CLD**

### **1.8.1 Liver biopsy**

Liver biopsy is the conventional gold standard for assessing the necro-inflammatory grade and stages of liver fibrosis (97). The prognosis of liver fibrosis disease is mainly predicted by the precise staging of fibrosis in CLD (47). Three staging methods are most commonly used to categorise liver fibrosis: the Metavir score (F0: no detected fibrosis, F1: fibrosis with spread to portal zones, F2: affects the portal zones and slight bridging fibrosis, F3: affects the portal zones with marked bridging fibrosis, F4: development of cirrhosis), the Batt-Ludwig systems and the Ishak score, which is an adjustment of the Metavir score (98).

Table 1-1 below provides a comparison and correspondence between METAVIR, Batts-Ludwig and Ishak scoring systems. The

diagram suggests a potential correlation between the pathological stages in the three systems. While the METAVIR and Batt-Ludwig systems classify cirrhosis as stage 4, the Ishak system identifies it as stages 5 and 6.

Additionally, the Ishak staging system offers more extensive and detailed data on fibrosis than the METAVIR and Batt-Ludwig systems. One METAVIR fibrosis level or Batts-Ludwig stage comprises approximately two Ishak scores. Despite providing more information on fibrosis, the Ishak score's intricacy makes it less reproducible and impractical for clinical application.

*Table 1-1: Correspondence and comparison between Ishak, METAVIR and Batts-Ludwig scoring systems, adapted from reference (99).*

| <b>Ishak Score (Stage)</b>             | <b>METAVIR Fibrosis level</b>                                    | <b>Batts-Ludwig Stage</b>                                                |
|----------------------------------------|------------------------------------------------------------------|--------------------------------------------------------------------------|
| <b>0 (no fibrosis)</b>                 | <b>F0 (no fibrosis)</b>                                          | <b>Stage 0 (no fibrosis)</b>                                             |
| <b>1 (Mild fibrosis)</b>               | <b>F1 (Mild to moderate fibrosis)</b>                            | <b>Stage 1 (Mild fibrosis)</b>                                           |
| <b>2 (Mild to moderate fibrosis)</b>   | <b>Either or between F1 and F2</b>                               | <b>Either or between Stage 1 and Stage 2 (Mild to moderate fibrosis)</b> |
| <b>3 (Moderate fibrosis)</b>           | <b>F2 (Moderate fibrosis)</b>                                    | <b>Stage 2 (Moderate fibrosis)</b>                                       |
| <b>4 (Moderate to severe fibrosis)</b> | <b>F3 (Severe fibrosis)</b>                                      | <b>Stage 3 (Severe fibrosis)</b>                                         |
| <b>5 (Incomplete cirrhosis)</b>        | <b>Either or between F3 or F4 (Severe fibrosis or cirrhosis)</b> | <b>Either or between stage 3 or 4 (Severe fibrosis or cirrhosis)</b>     |
| <b>6 (Cirrhosis)</b>                   | <b>F4 (Cirrhosis)</b>                                            | <b>Stage 4 (Cirrhosis)</b>                                               |

However, liver biopsy is an invasive, hazardous, and painful procedure with numerous potential complications, including the risk of infection and bleeding (97). Sampling error, which occurs when a tissue sample does not accurately reflect the pathological condition, is a

concern in needle-based liver biopsies (76). Additionally, inter- and intra-observer variability can reduce the precision of histopathologic evaluations of CLD, thereby affecting physicians' clinical decision-making processes (100).

As a result, more effort has been made to find non-invasive measures to discover fibrosis (101). These methods not only assess the severity of liver fibrosis, aiding in treatment decisions and monitoring the condition's progression over time (102), but they are also increasingly recognized for their role in prognostication (103). Together, these non-invasive procedures offer safer and more convenient alternatives to liver biopsy (104, 105).

### **1.8.2 Imaging modalities**

Imaging modalities play a crucial role in detecting liver cirrhosis and fibrosis, yet each method has limitations. US is commonly used to detect liver cirrhosis in patients with portal hypertension, with a sensitivity of 84% and specificity of 93% (106). However, its sensitivity for early fibrosis detection is lower due to the subtle changes in liver texture that are hard to identify with conventional ultrasound (107).

CT scans can image the entire abdomen and identify small varices, which can improve the sensitivity of the modality (108). The use of ionizing radiation and intravenous contrast increases the procedure's cost and limits its practicality for early fibrosis detection (106, 107) .

MRI is employed to identify advanced liver disease through clear morphological changes and signs of portal hypertension, such as spleen

enlargement (109). However, like US and CT, MRI is generally limited in detecting early liver fibrosis and is more costly, less available, and requires high technical expertise (107, 110).

Transient Elastography (TE) measures liver stiffness to diagnose advanced liver disease. It is cost-effective, non-invasive, and provides quantitative measurements of liver stiffness. However, it sometimes fails to provide sufficient data for identifying significant fibrosis, and its findings require professional interpretation (46).

Acoustic radiation force impulses (ARFI) offer precision comparable to TE in diagnosing severe fibrosis and cirrhosis. Shear wave velocity can be determined using conventional ultrasound with ARFI (109). However, ARFI is highly operator-dependent and has a restricted range of values, affecting optimal cut-off determination (111, 112)

Magnetic resonance elastography (MRE) uses a transducer to transmit mechanical waves into the liver while concurrently obtaining shear elasticity and viscosity maps to estimate fibrosis over a larger area (113). Despite MRE's high accuracy and reliability in diagnosing cirrhosis, its use is limited due to expensive equipment, the need for considerable expertise, and potential issues with patient repositioning affecting diagnostic yield (114).

These limitations prompt us to reconsider the criteria outlined in section 1.7.4 for an ideal screening test for early CLD detection. Imaging modalities often lack sensitivity for early fibrosis detection, have high

costs, require significant technical expertise, and can be operator dependent. Consequently, serum biomarkers have emerged as a more promising alternative for screening CLD. These biomarkers offer potential advantages such as non-invasiveness, cost-effectiveness, and the ability to provide comprehensive information about liver function and fibrosis progression. With these advantages, serum biomarkers have gained momentum as a preferred choice for disease screening, surpassing the limitations associated with imaging modalities and offering a more practical and effective approach to identifying individuals at risk of CLD.

### **1.8.3 Direct serum markers**

Several direct serum biomarkers have been investigated as potential tools for assessing liver fibrosis and cirrhosis. These biomarkers are categorized based on their association with fibrogenesis and fibrinolysis, matrix deposition and degradation, and related cytokines and chemokines (115). Notable direct serum biomarkers include hyaluronic acid (HA), procollagen III peptides (PIIINP), tissue inhibitor of metalloproteinases-1 (TIMP-1), and YKL-40.

Although these direct serum markers have demonstrated promise as diagnostic instruments for liver fibrosis and cirrhosis, their performance characteristics may vary based on the disease's underlying aetiology. Certain biomarkers, such as HA and TIMP-1, may more effectively diagnose fibrosis in patients with viral hepatitis than those with NAFLD (115, 116). However, YKL-40 may be more useful in predicting

fibrosis progression in patients with NAFLD than those with viral hepatitis (115, 117). Research also suggests that certain biomarkers may be more specific to particular forms of liver damage. PIIINP, for example, has been demonstrated to be highly diagnostic for ARLD (116). These findings highlight the importance of considering the underlying aetiology of liver fibrosis when interpreting biomarker results.

Direct serum biomarkers have shown promise as non-invasive diagnostic approaches for liver fibrosis, with studies investigating their utility in predicting the severity and progression of the disease. However, their performance characteristics and clinical utility can vary, necessitating further research to establish their optimal use in clinical practice. Moreover, their routine use in healthcare settings may be difficult due to limited availability, a situation often exacerbated by the high costs associated with their implementation. Additionally, not all patients may be suitable candidates for these biomarkers due to factors like comorbidities or specific clinical conditions, so their use for longitudinal follow-up or monitoring of disease progression can be difficult. As a result, indirect serum biomarkers have also been investigated for their potential use in assessing liver fibrosis. These biomarkers are typically associated with liver function and injury, and their potential to predict CLD is explored in detail in the following section.

#### **1.8.4 Indirect serum markers**

Indirect serum biomarkers are a valuable tool for the non-invasive diagnosis and monitoring of liver fibrosis. These markers offer several



advantages over direct serum markers, including widespread availability, the ability to provide information about liver function, and usefulness for disease monitoring (103).

The AST/ALT ratio, AST to platelet ratio index (APRI) score, and The Fibrosis (FIB)-4 index are particularly favoured in clinical settings due to their simplicity and the minimal parameters required for their calculation. In contrast, other non-invasive biomarkers like the BARD score (118), Forns index (119), Fibro index (120), and NAFLD fibrosis score (121), necessitate more complex calculations and sometimes additional laboratory tests. This increased complexity can limit their practicality in routine clinical use, making them less convenient and more time-consuming compared to the straightforward nature of the AST/ALT ratio, APRI score, and Fib4 index.

Furthermore, in countries with limited healthcare resources, the challenge of accurately staging liver fibrosis necessitates reliable, accessible, and cost-effective methods. Among the various strategies for fibrosis assessment, the APRI and the FIB-4 index have emerged as prominent non-invasive tools recommended by leading health authorities, including the WHO and the Asian Pacific Association for the Study of the Liver (APASL) (122-124).

#### **1.8.4.1 Liver function tests (LFTs)**

Before investigating non-invasive biomarkers for liver disease, it's essential to first understand LFTs and their significance. Understanding

LFTs is necessary because they provide the baseline values required for the accurate calculation and interpretation of these biomarkers (125).

LFTs encompass a range of tests, including alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphatase (ALP), gamma-glutamyl transferase (GGT), and others like 5'nucleotidase, total and fractionated bilirubin, prothrombin time (PT), the international normalized ratio (INR), and various protein measures. These tests collectively offer insights into liver health, identifying potential areas of damage and assisting in developing a differential diagnosis based on specific patterns of enzyme elevation (126, 127).

In primary care, LFTs are essential tools for diagnosing and monitoring liver conditions. Moreover, it's crucial to emphasize that liver enzyme levels can vary, regardless of the stage of liver disease. It's also worth mentioning that normal liver biochemistry doesn't necessarily exclude the presence of liver disease (128, 129)

In the scope of this review, the focus will be narrowed down to specifically discuss AST and ALT. In addition to platelet count, these enzymes are critical components of LFTs and are instrumental in calculating the AST/ALT ratio, APRI score, and the Fib4 index.

AST and ALT are primarily liver enzymes, but they are also present in red blood cells, heart cells, muscle tissues, and other organs including the pancreas and kidneys. Previously known as serum glutamic oxaloacetic transaminase (GOT) and serum glutamic pyruvic transaminase (GPT) respectively, AST and ALT are crucial in diagnosing

liver diseases. While they are not exclusive to liver disorders, their levels, especially when assessed with other enzymes, are instrumental in monitoring various liver conditions (130).

It's important to recognize that elevated levels of AST and ALT are common, found in about 8% of the general population, and can often be transient, especially in individuals without symptoms. In fact, up to 30% of these elevations may resolve spontaneously within three weeks, highlighting the importance of interpreting these test results with caution to avoid unnecessary interventions (131).

#### **1.8.4.1.1 Indications for requesting LFTs**

Opportunistic evaluation of liver function is not recommended for individuals without symptoms and no known risk factors (132, 133). As noted above, it's important to recognize that healthy people might occasionally show slightly abnormal LFT results. Conducting tests for non-specific, transient symptoms like short-term fatigue might lead to unnecessary anxiety or further medical actions if it results in false positive outcomes. So, LFTs, including AST and ALT, are most effectively used when requested based on specific clinical indications (134). Liver function testing is principally recommended in the following situations (132-134):

**Presence of risk factors for viral hepatitis.** Includes HCV, such as injecting drug users and children of mothers with HCV, and HBV, including individuals engaging in high-risk sexual practices, especially men who have sex with men. Hepatitis A risk factors include recent travel

to countries with poor sanitation or a lack of safe water within the last two months. Other rarer causes of viral hepatitis include Epstein-Barr virus and travel-related viruses.

**Presence of risk factors for MASLD**, previously known as NAFLD. Examples include obesity, T2DM, dyslipidaemia, and hypertension.

**Presence of risk factors for ARLD**, notably excessive alcohol consumption.

**Presence of risk factors for immune-mediated liver diseases** such as PBC, PSC, AIH, or other pre-existing autoimmune disorders or inflammatory bowel disease. Examples for the risk factors environmental triggers span viral infections (e.g., hepatitis, Epstein-Barr), low vitamin D levels, altered gut microbiota, exposure to sex hormones, and urinary tract infections (135).

**Monitoring the impact of medications that can influence liver function or are known to be hepatotoxic**, for instance, methotrexate, sodium valproate.

Additionally, conditions affecting other organs, especially muscles, can lead to increased aminotransferase levels. Thus, a thorough clinical evaluation, including history, physical examination, and appropriate laboratory testing, is essential to determine the cause of LFT abnormalities (136).

#### **1.8.4.1.2 LFTs in the CPRD**

CPRD is a large, primary care database in the UK, collecting anonymized patient data from general practices for public health research and evidence-based policy making; full details of it are provided in Chapter 4. In CPRD, A study by Scutt et al (137), focused on the utilization of AST and ALT in general practice for liver disease management. The study found that while the use of various liver blood markers has increased over the last decade, their application varied significantly across different population subgroups.

A striking observation was the sharp decline in the use of AST, with its measurement falling to just 2% per annum among general practice users. This decline poses a significant challenge to the implementation of current diagnostic guidelines, which often incorporate AST as a critical component, such as in the FIB-4 index for liver fibrosis assessment (12, 70, 129). In addition, the study highlighted a notable gap in practice: in cases of NAFLD, only 11% of patients in the UK had the necessary tests to assess FIB-4 (138).

According to Scutt et al study, the preference for ALT testing over AST might be attributed to a focus on efficiency, with ALT being deemed more liver specific. However, the study suggests that AST could be a more sensitive indicator of chronic liver injury, especially when used in conjunction with ALT. This finding raises questions about current testing practices and underscores the need for a more balanced approach that includes both AST and ALT measurements, particularly in the assessment of liver diseases (137).

#### **1.8.4.2 AST/ALT ratio**

The AST/ALT ratio is a non-invasive blood test that measures the ratio of two liver enzymes, AST and ALT (139).

A cut-off of 1 for the AST/ALT ratio is widely recognized as the conventional threshold in the diagnosis of liver diseases. This cut-off has been validated by numerous studies, which demonstrate its utility in distinguishing between various liver conditions, particularly in identifying the presence of cirrhosis among patients with CLD. The study by Williams and Hoofnagle (139), demonstrated that an AST/ALT ratio of 1 or more is strongly suggestive of cirrhosis in patients with NAFLD. This finding has been pivotal in guiding clinical practice, emphasizing the importance of this ratio in the diagnostic process.

The short-term and long-term prognostic utility of the AST/ALT ratio was assessed in a study by Giannini et al. (140), who showed a significant correlation between the ratio and The Model for End-Stage Liver Disease (MELD) scores in patients with viral liver cirrhosis. Specifically, an AST/ALT ratio cutoff of 1.17 demonstrated notable sensitivity and specificity in predicting mortality, suggesting that this ratio, alongside the MELD score, could be a valuable prognostic tool in this patient population. However, the study also highlighted that while the AST/ALT ratio and MELD score each provide prognostic insights independently, their combined assessment significantly improves prognostic accuracy. This finding indicates the importance of a multifaceted approach to prognosis, where the AST/ALT ratio is one of several tools used to evaluate outcomes in patients with viral liver

diseases. The study underscores the need for broader prognostic assessments, recognizing the AST/ALT ratio's value within a comprehensive evaluation framework to enhance predictive accuracy for patient outcomes.

In addition, in their retrospective cohort study, Lai et al. (141) analysed data from 1754 patients with chronic HBV, aiming to evaluate the AST/ALT ratio's predictive capacity for cirrhosis development. Employing Cox proportional hazards analysis showed an association between elevated AST/ALT ratio levels and an increased risk of cirrhosis. Particularly, they reported that patients with higher AST/ALT ratio levels (more than 0.7) presented a 2.77 to 2.95-fold increased risk of developing cirrhosis compared to those with lower AST/ALT ratio levels.

The study conducted by Giannini et al. (142), on the validity and clinical utility of the AST/ALT in assessing disease severity and prognosis in patients with HCV-related CLD has shown that the AST/ALT ratio is a valuable non-invasive tool for assessing the severity of CLD in patients with HCV. This comprehensive research retrospectively evaluated 252 patients to explore the correlation between the AST/ALT ratio and the degree of liver fibrosis, as well as its association with the Child-Pugh score in cirrhotic patients. The research demonstrated that a ratio of 1 or higher is significantly more prevalent in cirrhotic patients, establishing a strong statistical correlation with severe liver disease. The study also assessed the prognostic capability of the AST/ALT ratio in predicting mortality among cirrhotic patients followed up for at least one

year and demonstrated that it could identify cirrhotic patients who died within one year of follow-up, with an 81.3% sensitivity and 55.3% specificity.

Furthermore, a study conducted by Wang et al. (143), explored the prognostic significance of the AST/ALT ratio in patients with HCC undergoing thermal ablation combined with trans arterial chemoembolization (TACE). Analysing data from 117 patients, the researchers aimed to establish the pre-treatment AST/ALT ratio as a predictive marker for overall survival. They determined an optimal cut-off value for the AST/ALT ratio at 0.89, finding that patients with a ratio above this threshold had notably poorer survival outcomes. The AST/ALT ratio, alongside alpha-fetoprotein levels and tumour count, emerged as independent prognostic indicators for overall survival.

The study by Amernia et al. (144), sheds light on the diagnostic accuracy of the AST/ALT ratio, alongside FIB-4 and APRI, in assessing hepatic fibrosis among patients with NAFLD or NASH, comparing their effectiveness against FibroScan results. This research critically evaluates the AST/ALT ratio's ability to distinguish between early and advanced stages of liver fibrosis, setting a cut-off at 0.94 for differentiating F1/F2 from F3/F4 fibrosis stages. Despite finding a significant correlation between FibroScan scores and the AST/ALT ratio, albeit weaker than with FIB-4 and APRI, the study concludes APRI as the most reliable substitute for FibroScan in detecting significant fibrosis. This conclusion is particularly insightful for clinical settings lacking advanced diagnostic tools like FibroScan, positioning the AST/ALT ratio,



FIB-4, and APRI as valuable, non-invasive markers in managing NAFLD. Moreover, the research underlines the necessity for further exploration into non-invasive diagnostic methods, offering a foundation for future studies aimed at refining the assessment and management of liver fibrosis.

Angulo et al.'s (145), study conducted on 144 patients with NASH aimed to identify independent predictors of severe hepatic fibrosis. Their analysis, through liver biopsies and clinical data, found that an AST/ALT ratio greater than one, along with factors like older age, obesity, and diabetes mellitus, were significant independent predictors of advanced liver fibrosis in NASH patients. This groundbreaking work emphasized the AST/ALT ratio's role in non-invasively assessing liver health, suggesting that it could be instrumental in identifying patients at higher risk of severe liver disease without resorting to invasive procedures like biopsies.

Building on this, McPherson et al. (146), evaluated the utility of various non-invasive scoring systems, including the AST/ALT ratio, in diagnosing advanced fibrosis among 145 patients with biopsy-proven NAFLD. The study underscored AST/ALT ratio alone could effectively rule out advanced fibrosis in patients with NAFLD, offering a valuable tool for primary care settings. By reducing the reliance on liver biopsies, McPherson et al.'s work significantly contributes to the non-invasive management strategies for NAFLD, offering a practical approach to identifying patients who require more intensive evaluation or treatment.

Park et al. (147), focused on HCV infection, assessing whether the AST/ALT ratio could predict cirrhosis in this patient group. Through a retrospective analysis of 153 patients, they discovered that an AST/ALT ratio  $\geq 1$  offered a 95.9% specificity and 73.7% PPV in distinguishing cirrhotic from non-cirrhotic patients, with a 46.7% sensitivity and 88.1% NPV. This ratio also correlated well with the degree of liver fibrosis, providing a non-invasive marker that could help distinguish between cirrhotic and non-cirrhotic patients within the chronic HCV population. Park et al.'s findings reinforce the AST/ALT ratio's utility across different liver diseases, suggesting it as a beneficial tool in the clinical evaluation of patients with chronic HCV, further broadening the scope of non-invasive liver disease assessment.

#### **1.8.4.3 AST to Platelet Ratio Index (APRI) score**

APRI score is another widely used non-invasive biomarker for assessing liver fibrosis. It provides a rapid estimation for cirrhosis or severe fibrosis based on the ratio of AST level to platelet count in the blood. The study by Wai and colleagues aimed to develop a simple model using routine laboratory data to predict significant fibrosis and cirrhosis among patients with HCV. This endeavour led to the creation of APRI, which uses the opposing effects of liver fibrosis on AST levels and platelet counts to predict liver fibrosis and cirrhosis. The study's findings revealed that the APRI, utilizing optimized cut-off values, could accurately predict significant fibrosis in 51% of patients and cirrhosis in 81% of patients, demonstrating a high degree of accuracy (148).

WHO recommends a cutoff of 1 for the APRI as a threshold for identifying patients at a higher risk of liver fibrosis and cirrhosis (149, 150). A study by Zhu et al. (151), conducted to evaluate the diagnostic performance of various non-invasive methods for detecting liver fibrosis in patients with HBV infection. Involving a cohort of 178 Chinese HBV patients, the research aimed to compare the utility of FibroScan, APRI, and FIB-4 against liver biopsy. The findings revealed that APRI, with a cutoff value greater than 1, achieved an AUROC of 0.83. This indicates a strong capacity to differentiate between patients with and without cirrhosis, evidenced by a sensitivity of 75.9% and a specificity of 69.2%.

APRI was also extensively validated in a meta-analysis of 18 studies (152) for HCV patients, in which an APRI score greater than 2 was shown to exhibit a high specificity of 94% for cirrhosis diagnosis, albeit with a modest sensitivity of 48%. When the APRI threshold was set above 0.5, it identified fibrosis with a sensitivity of 81% and specificity of 55%, achieving an AUROC of 0.77.

In the context of NAFLD, a systematic review involving 242 patients identified that an APRI threshold of 0.54 is indicative of liver fibrosis, with an AUROC of 0.75. This APRI cut-off demonstrated a sensitivity of 77.3% and a specificity of 70.9% for the diagnosis of liver fibrosis in NAFLD patients (153).

Lin et al.'s (154) updated meta-analysis significantly enhances our knowledge of the APRI's diagnostic utility. Their comprehensive evaluation across various APRI cutoffs, such as 0.3, 0.5, 0.7, 1, 1.2, 1.5,

and 2, provides valuable insights into the test's performance in diagnosing liver conditions. At a lower threshold of 0.5, APRI shows a high sensitivity of 74% but a moderate specificity of 49%, indicating its strong ability to detect liver fibrosis. Conversely, a higher threshold of 1.5 yields a much lower sensitivity of 37% but significantly increases specificity to 93%, reducing false positives but risking missing true cases. The crucial trade-off between sensitivity and specificity when utilising APRI for liver fibrosis staging is illustrated by this comparison. The PPV and NPV at these thresholds, for instance, differ significantly with a fibrosis prevalence of 46%: 55% PPV and 69% NPV at the 0.5 cutoff, compared to 82% PPV and 63% NPV at the 1.5 cutoff.

The study conducted by Oliveira et al. (155), aimed to evaluate the efficacy of the APRI and FIB-4 tests as non-invasive markers for staging fibrosis in patients with HCV. The research focused on a cohort of nonelected outpatients from a referral centre in São Paulo, Brazil, highlighting the practical application of these tests in a real-world clinical setting. The study encompassed 798 adult patients, analysing medical records retrospectively. The criteria for the APRI and FIB-4 calculations were based on original descriptions, comparing these markers against the degree of liver injury determined through liver biopsy. Key findings were the cutoff values for APRI and FIB-4 that are crucial in classifying patients into different fibrosis stages, essentially providing a non-invasive means of assessing liver damage which traditionally requires a biopsy. For APRI, cutoff values of  $\leq 0.5$  suggested a lower probability of significant fibrosis, while values  $\geq 1.5$  indicated a higher likelihood. For

FIB-4, the cutoffs were  $\leq 1.45$  for lower and  $\geq 3.25$  for higher chances of advanced fibrosis or cirrhosis. The effectiveness of the APRI test for detecting liver cirrhosis varies, requiring careful adjustment of its thresholds to balance detecting true cases against avoiding false positives. Its performance differs across liver diseases, highlighting the need for tailored application and further research to improve its accuracy in diagnosing and predicting cirrhosis.

#### **1.8.4.4 Fibrosis (FIB)-4 index**

The Fib-4 is a non-invasive biomarker commonly used to assess liver fibrosis. It is a simple, rapid, and cost-effective test that provides immediate results (156). Fib4 is calculated using a simple formula that considers four clinical parameters: age, AST levels, ALT levels, and platelet count (157).

The conventional cutoff for the FIB-4 indexes, commonly adopted in clinical practice for the assessment of liver fibrosis, is 1.3. This threshold is recognized for its utility in distinguishing patients with minimal or no fibrosis from those likely to have significant fibrosis. In the study conducted by Shah et al.(158), the objective was to assess the FIB-4 index's effectiveness as a non-invasive marker for evaluating fibrosis in NAFLD patients and compare its performance against six other non-invasive fibrosis markers. Utilizing data from a national database of 541 adults diagnosed with NAFLD, the study meticulously calculated the AUROCs for FIB-4 and the other markers, aiming to identify those with advanced fibrosis accurately.

The standout finding of this analysis was the FIB-4 index's AUROC, which was 0.8, surpassing that of the NAFLD fibrosis score and significantly outperforming other evaluated markers such as the AST/ALT ratio. This superiority indicates FIB-4's enhanced ability to discern patients with significant fibrosis, potentially mitigating the need for invasive liver biopsies.

Specifically, the study illuminated FIB-4's practical thresholds for identifying advanced fibrosis, with a score above 2.7 showcasing an 80% PPV for advanced fibrosis, thereby highlighting individuals at greater risk. Conversely, a score below 1.3 was associated with a 90% NPV, effectively identifying patients with a low probability of advanced fibrosis.

In the study of Sterling et al. (156), a comprehensive analysis of FIB-4 index cutoffs was conducted, providing valuable insights into their diagnostic accuracy for liver fibrosis. This study examined a wide range of cutoffs, including 0.3, 0.5, 0.7, 1, 1.3, 1.5, 1.7, 2, 2.3, 2.5, 2.8, 3, 3.3, and 3.5, assessing their sensitivity and specificity for significant fibrosis, severe fibrosis, and cirrhosis. For example, at a lower cutoff of  $\leq 0.6$ , the study demonstrated a high sensitivity of 71% in the total cohort (830 patients), indicating a strong ability to detect liver fibrosis. However, the specificity at this cutoff was 43%, suggesting a moderate rate of false positives. In contrast, a higher cutoff of  $\geq 3.25$  showed a substantially increased specificity of 96.6% in the total cohort, significantly reducing the likelihood of false positives. However, the sensitivity at this threshold was only 23%, indicating a higher chance of missing true cases of fibrosis.

In the study conducted by Vallet-Pichard and colleagues (159), the effectiveness of the FIB-4 index as a non-invasive marker for assessing liver fibrosis in HCV infection was rigorously evaluated. The investigation sought to validate FIB-4 against 847 liver biopsies in HCV patients.

The research findings revealed that the FIB-4 index proficiently identified patients with severe fibrosis (F3-F4) and cirrhosis with AUROCs of 0.85 and 0.91, respectively. Notably, a FIB-4 value below 1.45 was able to exclude severe fibrosis with a 94.7% NPV and a sensitivity of 74.3%. Conversely, a FIB-4 value above 3.25 affirmed the existence of significant fibrosis with an 82.1% PPV and a specificity of 98.2%. By applying these thresholds, the study accurately classified 72.8% of the liver biopsies, underscoring the utility of FIB-4 in diagnosing liver fibrosis levels.

Moreover, the study emphasized the concordance between FIB-4 and FibroTest results, particularly for values outside the 1.45-3.25 range. A substantial agreement was observed in cases with FIB-4 values less than 1.45 or greater than 3.25, correlating well with FibroTest outcomes in 92.1% and 76% of instances, respectively. This analysis not only highlighted the FIB-4 index's capability to mirror FibroTest results but also its potential to serve as a simple, cost-effective, and immediately available diagnostic tool for liver fibrosis assessment.

In the meta-analysis conducted by Poynard et al.(160), the focus was on evaluating the prognostic value of various liver fibrosis

biomarkers, including FibroTest, APRI, FIB-4, and FibroScan, compared directly against liver biopsy for diagnosing fibrosis in patients with CLD. Their meta-analysis aimed to assess these biomarkers' 5-year prognostic values, particularly highlighting FibroTest's comparable prognostic utility to liver biopsy and its superiority over APRI and FIB-4 in predicting survival without liver-related death. This comprehensive analysis underscores the potential of non-invasive biomarkers as valuable tools in the prognostication of liver diseases, offering insights into their clinical relevance and applicability.

The findings of the meta-analysis showed that APRI was analysed across five studies involving 2,422 patients with various liver diseases, including HCV, HBV, and ARLD. It yielded an AUROC of 0.66, indicating its capability to predict survival without liver-related death. However, when compared directly to liver biopsy, APRI's prognostic value was significantly lower, with a mean difference in AUROCs of -0.21, suggesting it may not fully substitute the prognostic accuracy provided by histological evaluation.

Similarly, FIB-4's utility was examined in three studies with 1,184 patients. This biomarker achieved an AUROC of 0.73 for predicting survival without liver-related death. Like APRI, FIB-4's performance was found to be inferior to that of liver biopsy, indicating a reduced prognostic effectiveness with a mean difference in AUROCs of -0.21 compared to biopsy.



Despite the lower prognostic value of APRI and FIB-4 when directly compared to liver biopsy, their significance in clinical practice cannot be understated. These biomarkers offer a non-invasive, accessible means for prognostication in CLD, presenting a valuable tool for clinicians in the assessment and management of patients. The findings underscore the importance of adopting a comprehensive approach in prognostication, incorporating non-invasive biomarkers alongside other diagnostic tools as necessary. This multifaceted strategy is crucial for facilitating informed clinical decision-making and ensuring optimal patient care in the management of CLDs.

So, the review of existing literature, as conducted in the preceding sections, underscores the significance of indirect serum biomarkers such as AST/ALT ratios, APRI, and FIB-4 scores. These biomarkers are heralded for their critical role in the early detection of liver diseases, primarily due to three key advantages they offer over other diagnostic methods.

Firstly, the non-invasive nature of these biomarkers greatly enhances patient safety, eliminating the risks associated with more invasive procedures such as liver biopsies. Secondly, these tests are accessible: they can be readily performed in most healthcare settings without the need for specialized equipment or highly trained personnel. This sets them apart from methods such as FibroScan or liver biopsy, which require specific technology and expertise and is particularly significant in resource-limited settings where advanced diagnostic tools might not be available. Thirdly, the cost-effectiveness of AST/ALT, APRI,

and FIB-4 scores cannot be overstated. These tests utilize routine blood work, making them significantly less expensive than imaging studies or liver biopsies, both of which involve higher costs from the procedure itself and the necessary professional expertise. This aspect is especially important in managing chronic liver diseases, where ongoing monitoring of liver health is required, potentially imposing a substantial financial burden on healthcare systems and patients. These biomarkers form the basis for the investigations reported in Chapters 5, 6, 7 and 8, aligning with the requirements outlined in section 1.7.4.

### **1.9 Implications from the early disease detection using pragmatic case finding approach.**

The importance of early identification has been consistently emphasized (as discussed in section 1.7.2). However, to provide a more comprehensive understanding of the implications of widespread testing and the populations deemed 'at risk,' it is imperative to delve deeper into the expectations and potential consequences of such an approach.

As mentioned earlier (see section 1.7.3), the pragmatic case finding approach is especially applicable and useful for the 'blood test' study. In the context of this thesis, the term "blood test" specifically refers to the subset of individuals who underwent an AST test within the dataset. AST is frequently used as a marker for underlying liver diseases compared to the broader application of LFTs which may be conducted opportunistically. This approach is based on the understanding that while LFTs are commonly used in a wide range of contexts, AST testing specifically is more likely to be intentionally directed towards individuals

suspected of having liver health issues, although the precise rationale for this selective targeting is not fully understood.

The 'blood test' cohort may have a higher prevalence of CLD than the general population and may also have different risk factors. Thus, in terms of higher risk and susceptibility to CLD, the study's population differs from the general population. This distinction emphasises that using pragmatic case finding in this situation is appropriate. Focusing on a population that has already been identified by AST testing effectively directs resources to those who stand to gain the most from early detection and treatment of CLD. In contrast to the general population, it also enables deeper comprehension and management of the progression of CLD in a high-risk group. This focused strategy maximises the efficacy of healthcare interventions while also being in line with the constrained resources available.

Recent liver guidelines from organizations such as EASL-EASO and NICE consistently recommend screening for advanced fibrosis in individuals diagnosed with NAFLD (72-74). Patients with NAFLD often rely heavily on secondary care liver services due to inconsistent and irregular management in primary care, despite the strong links between obesity, diabetes, and most NAFLD cases. The lack of established diagnostic pathways makes it challenging to identify those at risk of liver disease progression in primary care settings (161).

NAFLD, prevalent in the community but less severe with only 5% having advanced fibrosis, shows that non-invasive tools can effectively

exclude advanced disease with a NPV exceeding 98% in these settings (161). However, LFTs in NAFLD patients often fail to reflect disease severity until advanced stages, presenting with liver failure signs. Despite this, primary care practitioners continue to use elevated liver enzymes from LFTs to decide referrals to specialist liver services (162), leading to unnecessary secondary care referrals and missing advanced but asymptomatic cases in the community.

Srivastava and colleagues conducted a large prospective study in the UK primary care setting, involving 1,452 patients with NAFLD (163). Their goal was to establish a referral pathway designed to identify advanced fibrosis or cirrhosis within the community. This innovative approach was implemented in 2013 and aligns with research proposals outlined in the EASL-EASD-EASO guidelines. The NAFLD pathway consisted of a 2-step process, initially focusing on patients with elevated ALT levels, negative alcohol/liver screen results, and/or fatty liver detected on ultrasound. Subsequently, all patients underwent a FIB-4 score assessment, and those with an indeterminate FIB-4 score (ranging from 1.3 to 3.25) received a sequential ELF test. As a result, 81% of patients were identified as having a low risk of advanced fibrosis (FIB-4 <1.3 or ELF <9.5) and remained in the community, while the remaining 19% were recommended for referral to liver services.

This study represents a significant advancement in the field of NAFLD diagnosis and management, and the implementation of this approach offers several potential benefits. It has the potential to result in significant cost savings by reducing unnecessary referrals to specialist

liver services. It can increase the capacity of specialist clinics, ensuring that patients with advanced disease receive the attention they need, and help to alleviate unnecessary patient anxiety and potential harm from additional tests associated with referrals to hospital specialists. Beyond these immediate advantages, identifying patients with advanced disease at an earlier stage enables targeted interventions, including HCC/variceal surveillance, intensive lifestyle/nutritional management, and participation in clinical trials. Over time, it is anticipated that these measures may contribute positively to reducing liver-related morbidity and mortality rates. However, it is essential to emphasize that these potential benefits require further evaluation and long-term follow-up to be substantiated conclusively.

The use of non-invasive markers originally developed and validated in secondary care settings in primary care presents its own set of challenges. For example, FIB-4 for patients over the age of 65 would have provided false reassurance in 27% of cases with advanced fibrosis (163). These findings underscore the importance of validating specific cut-off values tailored to primary care populations. It highlights the need to strike the right balance between minimizing the referral burden and ensuring accurate disease exclusion in primary care settings.

Importantly, it is widely recognized that many patients with NAFLD, including those with cirrhosis, may have normal LFTs (162). However, the current proposed NAFLD pathway relies solely on the patient having an abnormal ALT measurement to initiate a FIB-4 calculation (164). So, a substantial number of individuals with NAFLD

and advanced fibrosis may go undiagnosed, leading to misallocation of healthcare resources.

Instead of relying only on the abnormal ALT to trigger the calculation of Fib4 in order to refer patients for further analysis for advanced liver disease, this thesis will assess whether the routinely collected data to calculate AST/ALT ratios, APRI scores, and FIB-4 indexes can be used in pragmatic case finding and risk stratification for patients with advanced CLD. This research may provide valuable insights that could lead to the integration of these biomarkers into routine clinical visits for improved patient care and risk assessment.

Pryke and Guha (165), provide compelling evidence for the integration of non-invasive biomarkers in the primary care setting as a significant advancement in the management of CLD, driven by the growing prevalence of lifestyle-related metabolic risk factors. They identify a gap in primary care's confidence and tools for detecting and managing CLD, emphasizing the potential of non-invasive biomarkers, such as the FIB-4 index, to address this gap effectively. Their review highlights the underutilization of LFT for specific liver disease risk assessment and outlines educational steps to promote the use of pragmatic algorithms for identifying fibrosis risk.

Pryke and Guha argue for aligning liver disease management with the care pathways of other metabolic disorders, suggesting that this approach can reduce unnecessary testing and improve referral quality to secondary care. They also discuss the benefits of integrating liver

disease pathways with community lifestyle support services, positing that such integration can enhance the patient journey through long-term multimorbidity care and have wide-ranging benefits across co-existing metabolic disorders.

Critically, Pryke and Guha's findings underscore the feasibility of integrating non-invasive biomarkers into primary care without significantly increasing the workload of healthcare professionals. This approach not only aims to improve the accuracy of CLD diagnosis and management but also to streamline the patient management process by leveraging existing frameworks for managing long-term conditions. Their work suggests that the adoption of non-invasive biomarkers in primary care could significantly enhance patient care by facilitating early detection, enabling appropriate referrals, and fostering an integrated care model that addresses the multifaceted needs of patients with CLD.

In light of these findings, integrating non-invasive biomarkers into primary care represents a forward-thinking approach to CLD management. This integration holds the promise of transforming patient outcomes by enabling early intervention, personalized care strategies, and the efficient allocation of healthcare resources.

## **1.10 Rationale, aims and objectives**

### **1.10.1 Rationale**

Effective screening and early treatment strategies are key for preventing the progression of CLD and improving patient outcomes (2, 25); however, there are a significant gaps in our knowledge of the

diagnostic pathway, specifically in relation to diagnostic criteria and the cost-effectiveness of screening and early diagnosis of CLD (166, 167). Also of concern are the potential harms (e.g. patient concerns relating to false positive results, insufficient resources to meet demand) that could result from the implementation of poorly planned pragmatic case finding programs.

The widespread use of non-invasive biomarkers has been identified as a valuable diagnostic tool in respect of a range of conditions; however, in the primary care setting, there is a need to investigate whether these biomarkers can be applied effectively in pragmatic case finding and risk stratification for patients at a higher risk of developing advanced CLD. As a result, this study draws on routinely collected data to conduct cohort studies to assess how common non-invasive biomarker recordings are among UK primary care patients. These focus on identifying abnormal AST/ALT, APRI, and FIB-4 scores and determining their ability to predict subsequent advanced CLD. As CLD is typically undiagnosed until it reaches an advanced stage, demonstrating that routinely collected patient data could be used to identify those at greater risk would enable interventions to be introduced at a much earlier stage, potentially limiting the progression of the disease and reducing the associated healthcare burden.

In addition, this thesis estimates the prevalence of NAFLD in the UK and KSA. As a clinician with experience in the KSA, I am particularly interested in affecting positive change in the country, and it is anticipated that the findings of this thesis have the potential to be transferred to the



KSA, where they could have significant implications for the diagnosis of NAFLD and CLD and the further development of diagnostic services within the country's healthcare system.

While previous studies may provide some insights into NAFLD prevalence in the UK (168-171) and KSA (172-175), they may lack comprehensive methodologies, sample representativeness, or standardization, making their findings less reliable for informing healthcare policies and interventions. These observations underscore the value of conducting a systematic review to rigorously evaluate the existing evidence, ensuring a thorough understanding of the published data of NAFLD's epidemiology across both countries. A detailed examination of these and other pertinent studies will be carried out in the systematic reviews detailed in the subsequent chapters of this thesis. Through this meticulous scrutiny, the reviews seek to identify gaps in the current literature, address the heterogeneity in study designs and outcomes, and ultimately contribute to a more nuanced understanding of NAFLD epidemiology.

### **1.10.2 Aims and objectives**

As indicated above, this thesis primarily aims to assess the feasibility of employing non-invasive biomarkers (AST/ALT ratio, APRI score, and FIB-4 index) for pragmatic case finding of patients at higher risk of developing advanced liver diseases among individuals attending primary care centres in the UK. The secondary aim of estimating the

prevalence of NAFLD among the UK and KSA populations was also established.

In order to achieve these aims, the following objectives were identified:

- To estimate the prevalence of NAFLD in the UK and KSA using systematic reviews.
- To calculate the AST/ALT ratio, APRI score, and FIB-4 index from existing data within patient records where available for patients older than 40 years who did not have a previous diagnosis of CLD.
- To estimate the proportion of patients with (normal or abnormal) AST/ALT ratios, APRI scores, and FIB-4 indexes based on conventional cut-offs.
- To identify all patients aged 40 years or older with at least one year of up-to-standard (UTS) data and with data allowing the calculation of at least one of AST/ALT ratio, APRI and FIB4-index who have been diagnosed with liver cirrhosis, HCC, liver transplantation, and liver death using CPRD and linked data.
- To determine whether the AST/ALT ratio, APRI score, and FIB-4 index can be used to predict subsequent Advanced CLD in those for whom these calculations can be performed in UK primary care centres.

## **Chapter 2      The prevalence of NAFLD in the UK: A Systematic Review and Meta-Analysis**

## **2.1 Introduction**

The number of cases of NAFLD varies greatly worldwide, with an estimated overall prevalence of 30% (52). However, the prevalence of NAFLD in the general population of the UK and KSA is not well established. Consequently, one of the objectives of this thesis is to address this knowledge gap through two systematic reviews, which aim to determine the prevalence of NAFLD in both countries. This research will contribute to assessing the associated health needs more accurately.

## **2.2 Methods**

This section describes the methodology adopted for both systematic reviews. In the second review, a number of minor changes were made (for example, references to the 'UK' were replaced with 'Saudi Arabia'), and these are detailed in the Methods section of the Saudi study.

### **2.2.1 Review question**

What is the prevalence of NAFLD among the UK population?

### **2.2.2 Study design and inclusion and exclusion criteria**

A systematic review and meta-analysis was conducted to estimate the prevalence of NAFLD among the UK population. This included studies investigating the prevalence of NAFLD in the UK diagnosed by any method previously validated for a diagnosis of NAFLD, such as liver biopsy, US, MRI, other liver scans, liver enzymes, and blood-based biomarkers, after excluding other liver diseases (176, 177).

We included both community and hospital-based studies and those in other specific populations. Studies were excluded if the diagnosis of NAFLD was self-reported. We excluded studies of other liver diseases that cause fatty liver, including excessive alcohol consumption, side effects of certain medications (e.g., glucocorticoids, methotrexate, chemotherapy, tamoxifen, etc.), HCV, particular endocrine conditions, as well as many other conditions (e.g., jejunal diverticulitis with bacterial overgrowth, inflammatory bowel disease, occupational exposure) (178). We included all studies that met the inclusion criteria without language or time restrictions.

### **2.2.3 Search strategy**

We identified the studies by conducting a systematic search of electronic databases in PubMed, EMBASE via OVID, CINAHL, Web of Science, and Google Scholar utilizing Medical Subject Headings (PubMed) and other search words pertaining to “the prevalence of NAFLD in the UK” as well as Boolean operators “AND” and “OR” for articles from inception to January 2021. Grey literature was also searched, including annual and government reports, newsletters, reference lists, and discussion papers. We included all studies that met the inclusion criteria without restriction on language or time.

The search terms were: ((Non-alcoholic fatty liver) OR (NAFLD) OR (Steatosis) OR (fatty liver) OR (Non-alcoholic steatosis) OR (Non-alcoholic steatohepatitis) OR (hepatic steatosis) OR (Hepatic fat)) AND

((Prevalence) OR (Prevalence rate) OR (trend)) OR ((UK) OR (United Kingdom))

#### **2.2.4 Study selection and quality assessment**

Articles identified by the search were imported to the EndNote citation manager, allowing duplicates to be removed. Two reviewers (Yusef Alenezi (YA) and Tim Card (TC)) independently conducted title and then abstract screening and assessment, excluding non-relevant items. Full texts of the remaining papers were then assessed (by YA and Joanne Morling (JM)). Two reviewers (YA and Rebecca Harris (RH)) independently assessed the quality of the included studies and extracted data into a standardized form. Disagreements at all stages were resolved by consensus.

All relevant data were extracted, including:

- The study setting was classified as a hospital (inpatient or outpatient) or community.
- The study population was defined as the general population when sampled from the entire population of the UK; disease-unrestricted populations when composed of people selected due to hospital/clinic attendance but not limited to a specific disease; and disease-specific populations when the cohort was selected due to the presence of a specific co-morbidity (e.g., T2DM).

Diagnostic methods for NAFLD were classified as blood tests only, US, CT, MRI, or other.

The quality of the studies was assessed using the Joanna Briggs Institute's critical appraisal tool for prevalence studies ([Appendix A](#)) in order to assess the quality of methodology and the possibility of bias. Each domain was evaluated as Yes, No, Unclear, or Non-applicable. All included studies were subject to critical appraisal, and disagreements were resolved by consensus. Data were extracted on an Excel sheet. Studies were rated "low risk" when at least 50% of the JBI domains were replied to with "Yes."

### **2.2.5 Data management and analysis**

Rayyan-Qatar Computing Research Institute (QCRI) (179) was used for article management, and statistical analyses were done using STATA version 17 for meta-analysis. Prevalence meta-analysis was performed both on all the included articles and on subgroups on the basis of population type (disease-unrestricted population estimates coming from studies whose inclusion criteria did not specify any pathological conditions or diseases). A random-effects model was utilized in order to perform the meta-analysis using a weighted inverse variance model (DerSimonian and Laird) (180). We used Higgin's  $I^2$  test to evaluate the percentage of variability with inter-studies heterogeneity's cut-off point at  $I^2 > 50\%$ . We used a funnel plot and Egger's test to assess publication bias with a cut-off point with a P-value of 0.05 or less.

## 2.3 Results

### 2.3.1 Search results

We identified a total of 328 studies for title screening, of which 243 were retained for abstract screening. We did not find more eligible studies on searching grey literature. A total of 211 studies were excluded for not showing NAFLD prevalence. The remaining 32 articles were included for full-paper assessment, and 23 of them were excluded due to inappropriate analysis or self-reported NAFLD diagnosis. Finally, nine studies were enrolled for data extraction. The study selection process is illustrated in Figure 2-1.

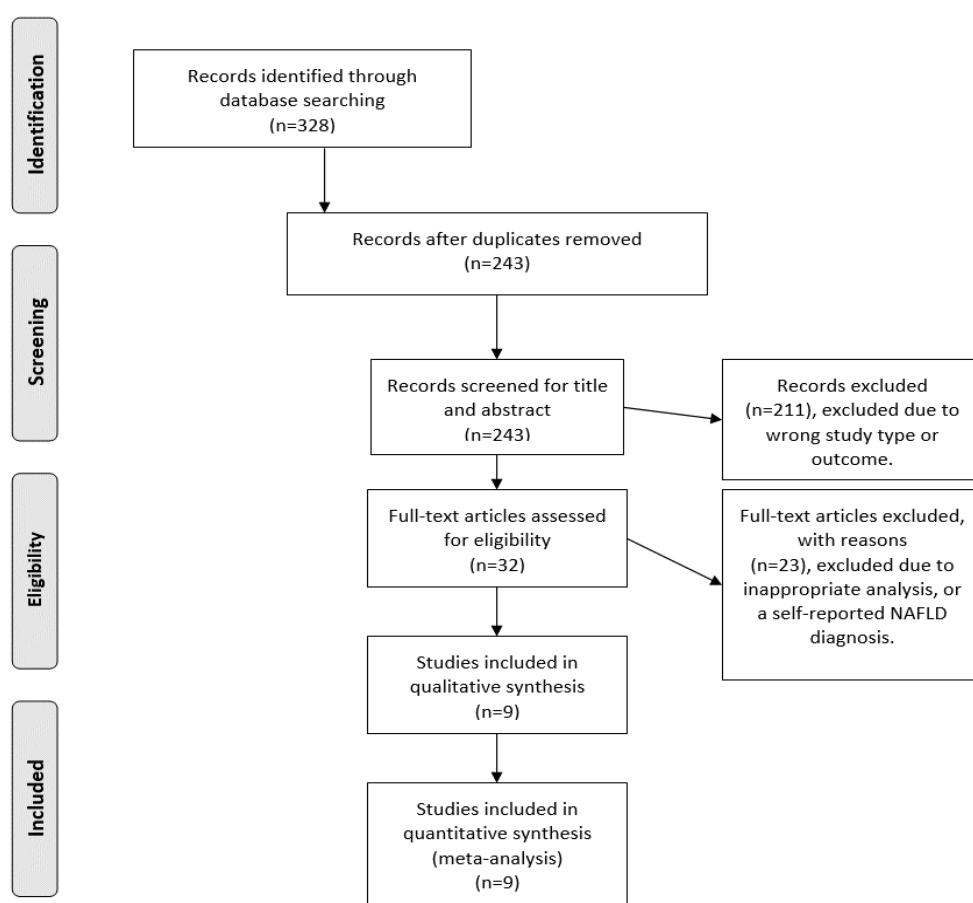


Figure 2-1: Flow diagram of articles screening and selection process for studies of NAFLD in the UK.



### **2.3.1.1 NAFLD prevalence in the UK**

Our study included a total of 25,439 participants from 9 studies, 18,880 of which were from 4 disease-unrestricted populations. The disease-restricted populations included 1,306 from three T2DM studies, 2818 obese subjects (170), 1,118 with abnormal LFTs (181), 910 patients with liver-related conditions who were referred from a general practitioner, found to be high risk (e.g. T2DM or alcohol misuse), and screened for excess alcohol (182), 156 subjects with HIV (183), and 251 with T1DM (184). Characteristics of each study are to be found in Table 2-1 and Table 2-2.

Table 2-1: Characteristics of the included studies for the prevalence of NAFLD in the UK:

| Author                                   | Year      | Study type                                       | Study setting   | Population                 | Participants | Males (%) | Age range | Age (Mean±SD) | Diagnostic method                                        | Prevalence (95% CI)       |
|------------------------------------------|-----------|--------------------------------------------------|-----------------|----------------------------|--------------|-----------|-----------|---------------|----------------------------------------------------------|---------------------------|
| <b>Khalatbari-Soltani S et al. (169)</b> | 2005-2015 | Cross-sectional                                  | Community-based | Disease-unrestricted       | 9645         | 44.9%     |           | 48.9±7.4      | US, Fatty Liver Index (FLI) & NAFLD score                | 23.9% (95% CI:23.1-24.8)  |
| <b>Armstrong M et al. (181)</b>          | 2006-2008 | Cross-sectional                                  | Hospital-based  | Abnormal LFTs Patients     | 1118         | 56.0%     | 48-70     | 60            | US                                                       | 26.4% (95% CI:23.9-29.1)  |
| <b>Neukam K et al. (168)</b>             | 2010-2012 | Cross-sectional                                  | Community-based | Disease-unrestricted       | 597          | 48.7%     | 40-55     |               | FLI & BARD                                               | 30.8% (95% CI:27.3-34.6)  |
| <b>Williamson RM et al. (185)</b>        | 2010      | Cross-sectional                                  | Community-based | T2DM Patients              | 939          | 52.0%     | 61-76     | 68.9±4.2      | US                                                       | 42.6% (95% CI:39.5-45.8)  |
| <b>El-Gohary M et al.(182)</b>           | 2014-2016 | Prospective cluster randomized feasibility trial | Hospital-based  | High-risk liver admissions | 910          | 56.3%     | 31-58     | 47            | Southampton Traffic Light Test & Transient Elastography  | 49.5% (95% CI:46.2-52.7)  |
| <b>Lombardi R et al. (183)</b>           | 2014      | Cross-sectional                                  | Hospital-based  | HIV patients               | 156          | 91.7%     |           | 47.5±8.5      | US                                                       | 71.2% (95% CI:63.6-77.7)  |
| <b>Marjot T et al. (184)</b>             | 2015      | Prospective                                      | Hospital-based  | T2DM Patients              | 141          | 61.0%     |           | 53.5±15.4     | LFTs                                                     | 13.1% (95% CI:8.5-19.7)   |
|                                          |           |                                                  |                 | T1DM Patients              | 251          |           |           |               |                                                          | 2.0% (95% CI: 0.0-4.0)    |
| <b>Wilman HR et al. (170)</b>            | 2015      | Prospective                                      | Community-based | Disease-unrestricted       | 4617         | 47.3%     | 45-73     |               | MRI-Derived Proton Density Fat Fraction (MRI PDFF) ≥5.5% | 19.9% (95% CI:18.8-21)    |
|                                          |           |                                                  |                 | BMI≥25                     | 2818         |           |           |               |                                                          | 29.6% (95% CI: 27.9-31.3) |
|                                          |           |                                                  |                 | T2DM Patients              | 226          |           |           |               |                                                          | 50% (95% CI: 43.5-56.5)   |
| <b>Abeysekera KWM et al.(171)</b>        | 2015-2017 | Cross-sectional                                  | Community-based | Disease-unrestricted       | 4021         | 37.5%     | 22-26     | 24            | ALT, AST, GGT & Lipid Profiles                           | 20.7% (95% CI:19.5-21.9)  |

Table 2-2: Detailed characteristics of included studies in the UK systematic review:

| Study name                             | Study purpose/aim                                                                                                                                                   | Target population, sampling frame, power calculation and response rate                                                                                                                                                                                                      | Definition of NAFLD (including diagnostic criteria)                                                                                                                                                                                                                                                                                                                                                                          |
|----------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Khalatbari-Soltani et al. (169)</b> | To estimate the association between adherence to the Mediterranean diet and hepatic steatosis among healthy middle-aged adults in two cohorts (Fenland and CoLaus). | <ul style="list-style-type: none"> <li>The target population is all healthy middle-aged adults in these two countries.</li> <li>The sample was derived from general practice lists in Cambridgeshire in the UK.</li> <li>No power of calculation was estimated.</li> </ul>  | NAFLD was defined as a steatosis score >4 using USS or FLI>60 with the exclusion of patients with alcohol consumption.                                                                                                                                                                                                                                                                                                       |
| <b>Armstrong et al. (181).</b>         | To determine the prevalence and disease severity of silent NAFLD in a primary care setting                                                                          | <ul style="list-style-type: none"> <li>The target population would be all patients visiting primary care units.</li> <li>The sample was obtained from eight primary care practices within the Birmingham region.</li> <li>No power of calculation was estimated.</li> </ul> | NAFLD was diagnosed depending on the sonographic diagnosis of fatty liver, defined as diffusely increased liver echogenicity (>right renal parenchyma) with vascular blurring, exclusion of history of alcohol consumption exceeding current UK health guidelines, and exclusion of liver disease of other aetiology, including drug-induced, autoimmune, viral hepatitis, cholestatic, metabolic and genetic liver disease. |
| <b>Neukam et al. (168).</b>            | Assess the prevalence of NAFLD among a cohort of first- and second-generation British Indians of Gujarati origin living in North London.                            | <ul style="list-style-type: none"> <li>The target population is all Gujarati Indians living in North London.</li> <li>The sample was taken, however, from two Hindu Temples situated in London.</li> <li>No power of calculation was estimated.</li> </ul>                  | FLI (Fatty Liver Index) was used to determine the presence of NAFLD, the calculation of which depended on BMI, plasma gamma-glutamyl transferase (gGT), triglyceride levels, and waist circumference measurements. An FLI of more than 60 rules in NAFLD.                                                                                                                                                                    |
| <b>Williamson et al.(185)</b>          | To assess the prevalence of NAFLD in older patients with type 2 diabetes mellitus.                                                                                  | <ul style="list-style-type: none"> <li>The target population is all older patients with type 2 diabetes mellitus.</li> <li>The sample was retrieved from patients registered with the Lothian Diabetes Register.</li> </ul>                                                 | NAFLD was defined as the presence of hepatic steatosis on an ultrasound scan in the absence of a secondary cause for hepatic steatosis. Secondary causes included alcohol consumption, use of hepatotoxic medication within the six months prior to the year one clinic, positive                                                                                                                                            |

|                                 |                                                                                                                                                                |                                                                                                                                                                                                                                                                                                                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|---------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                 |                                                                                                                                                                | <ul style="list-style-type: none"> <li>The initial power calculation estimated the need for 800 patients. Ultimately, 1066 patients were successfully enrolled, exceeding the target and further securing the study's capacity to accurately assess the intended outcomes.</li> </ul>                                           | hepatitis B or C serology, and ferritin concentration $\geq 1,000$ $\mu\text{g/L}$ .                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| <b>El-Gohary et al. (182).</b>  | To test the effectiveness of using nurse-led primary care-based liver clinics utilising additional non-invasive testing to diagnose liver diseases.            | <ul style="list-style-type: none"> <li>The target population is all patients who visit nurse-led clinics in Southampton, UK.</li> <li>The sample was derived from patients registered with intervention practice.</li> <li>No power of calculation was estimated.</li> </ul>                                                    | Blood-based biomarkers were used to diagnose NAFLD, including FBC, LFTs including AST and GGT, and were obtained from blood samples in addition to the serum fibrosis markers HA and P3NP. NO cut-offs were mentioned. TE (used to measure a physical parameter in liver stiffness) was pragmatically graded as follows: Median value of $<6\text{kPa}$ = no fibrosis, $6\text{--}8\text{kPa}$ = 'liver warning', $8\text{--}12.9\text{ kPa}$ = 'progressive fibrosis', $\geq 13\text{kPa}$ = 'probable cirrhosis'; the $13\text{kPa}$ cut off being almost conservative and patients did not have a liver biopsy. |
| <b>Lombardi R et al. (183).</b> | To assess the prevalence of NAFLD and predictors in HIV-mono-infected patients with persistently elevated transaminases.                                       | <ul style="list-style-type: none"> <li>The aim implies that the target population is all HIV patients with elevated transaminases.</li> <li>The sample was driven from HIV patients who attended Royal Free Hospital, London, UK, between January and December 2014.</li> <li>No power of calculation was estimated.</li> </ul> | The presence of fatty liver upon abdominal US scans and TE with FibroScan. APRI (It has dual cut-offs for ruling out ( $<0.5$ ) or diagnosing ( $>1.5$ ) significant fibrosis) and FIB4 scores (It also has dual cut-offs for ruling out ( $<1.45$ ) or diagnosing ( $>3.25$ ) advanced fibrosis) were used to evaluate the presence of significant or advanced liver fibrosis.                                                                                                                                                                                                                                    |
| <b>Marjot et al. (184).</b>     | To define the attitudes and current clinical practice of diabetes specialists towards NAFLD across the UK<br>To assess the prevalence of NAFLD in all patients | <ul style="list-style-type: none"> <li>The target population is all patients attending outpatient diabetes clinics in the UK.</li> <li>The sample was obtained from all diabetic patients attending outpatient diabetes clinics within the Oxford Centre for Diabetes, Endocrinology and Metabolism (OCDEM).</li> </ul>         | NAFLD was diagnosed using blood biomarkers, such as raised ALT and $\text{Fib4} > 1.3$ , at risk of advanced fibrosis. $\text{TE} > 7.9\text{kPa}$ confirmed dx. In-text, it states that 13.1% had advanced fibrotic NAFLD, but this is based upon an assumption of those who did not attend TE.                                                                                                                                                                                                                                                                                                                   |

|                             |                                                                                                                                                                                                                                   |                                                                                                                                                                                                                                                                                                    |                                                                                                   |
|-----------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------|
|                             | attending diabetes outpatient clinics at a large teaching hospital/NHS Trust and to estimate the proportions of people with Type 1 and Type 2 diabetes attending outpatient clinics who are at risk of advanced hepatic fibrosis. | <ul style="list-style-type: none"> <li>No power of calculation was estimated.</li> </ul>                                                                                                                                                                                                           |                                                                                                   |
| <b>Wilman et al. (170).</b> | To investigate the prevalence of NAFLD in the UK Biobank cohort and to assess the potential impact of age, gender and BMI on hepatic fat deposition.                                                                              | <ul style="list-style-type: none"> <li>The target population implied by the aim is all patients signed in the UK Biobank database.</li> <li>The sample was drawn from all patients signed in the UK Biobank database from 2014 to 2015.</li> <li>No power of calculation was estimated.</li> </ul> | The diagnosis was dependent on the presence of fat upon the performance of MRI with a PDFF >5.5%. |

We found a pooled prevalence of 23.8% in the disease-unrestricted population (95% CI: 20.7%-26.8%) (Figure 2-2). Of the disease-restricted populations, a single study of HIV patients reported a prevalence of 71.2% (95% CI: 59.7% - 82.7%), two studies of liver-related hospital admissions yielded a prevalence of 37.90% (95% CI: 15.3% - 60.5%), a single study BMI  $\geq 25$  (29.6%) (95% CI: 27.9% - 31.3%) and T1DM patients (1.8%) (95% CI: 0.0% - 3.6%).

Our results also included three studies of T2DM patients with an overall prevalence of 38.6% (95% CI: 25.5% - 51.7%). The overall prevalence in the disease-restricted studies was 36.4% (95% CI: 23.5 – 49.2) (Figure 2-3).

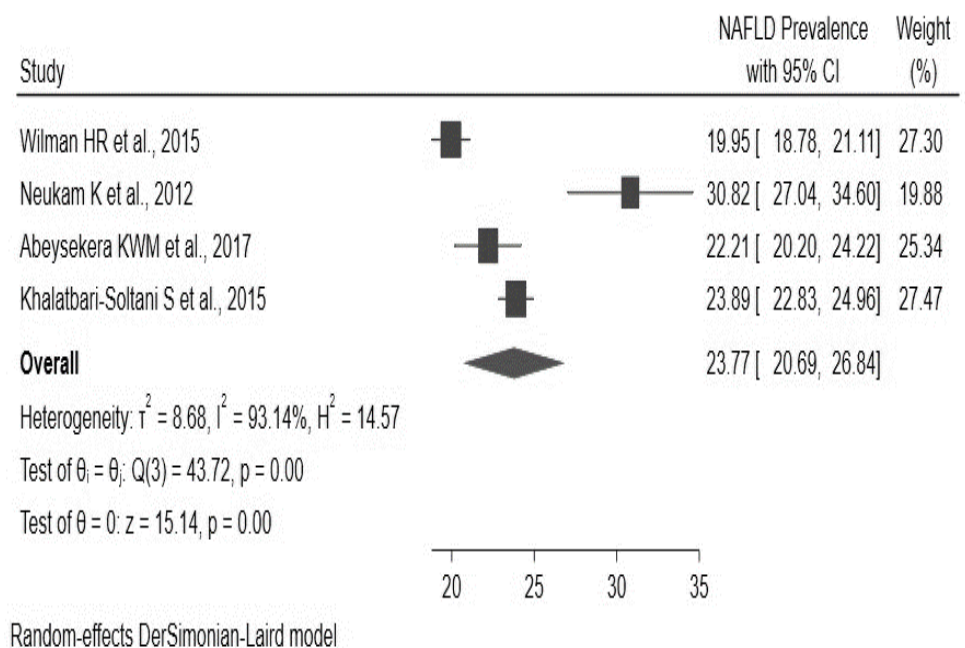


Figure 2-2: Forest plot of the Prevalence of NAFLD among disease-unrestricted studies in the UK.

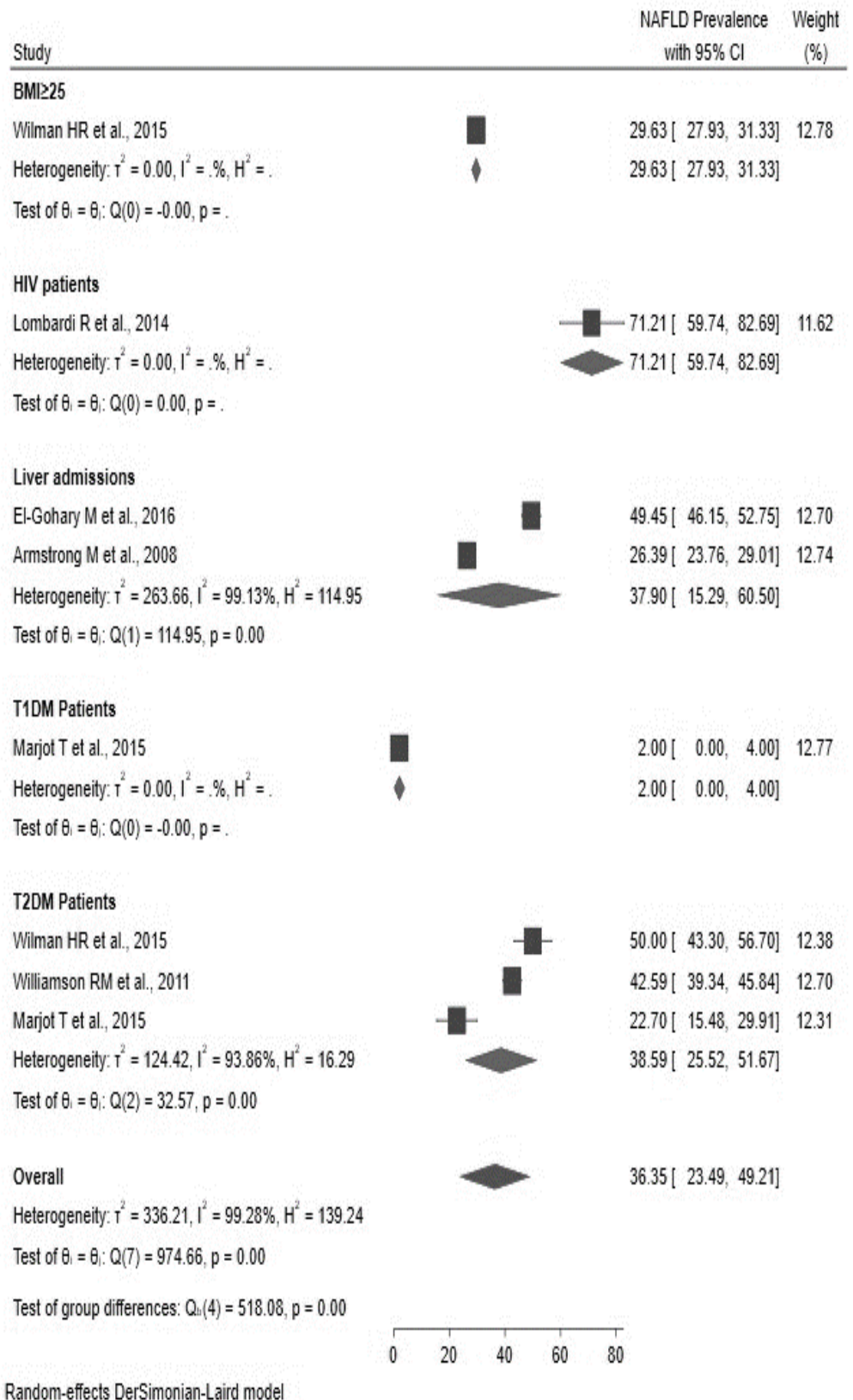


Figure 2-3: Forest plot of the Prevalence of NAFLD among disease-restricted studies in the UK.

### **2.3.2 Quality assessment**

A summary of the quality assessment of the included articles is presented in Table 2-2 and Table 2-3. All studies were positively rated on all quality criteria except for the study by Wilman et al. (170), in which a detailed description of the study participants was rated “unclear,” and that by Lombardi R et al. (183), where ‘Adequacy of response rate and well-handled lack of data’ was also unclear.



Table 2-3 Quality Assessment of the studies for the prevalence of NAFLD in the UK:

| Author                                   | Appropriate sample frame to reach the target population | Appropriate way of sampling participants | Was the sample size adequate? | Detailed description of study subjects | Presence of sufficient coverage of the identified sample in the data analysis | Valid methods were used for the identification of the condition | Standard and reliable way was used to measure the condition for all participants | Was there an appropriate statistical analysis? | Adequacy of response rate, and well-handled lack of data | Inclusion in meta-analysis |
|------------------------------------------|---------------------------------------------------------|------------------------------------------|-------------------------------|----------------------------------------|-------------------------------------------------------------------------------|-----------------------------------------------------------------|----------------------------------------------------------------------------------|------------------------------------------------|----------------------------------------------------------|----------------------------|
| Khalatbari-Soltani S <i>et al.</i> (169) | Yes                                                     | Yes                                      | Yes                           | Yes                                    | Yes                                                                           | Yes                                                             | Yes                                                                              | Yes                                            | Yes                                                      | Include                    |
| Armstrong M <i>et al.</i> (181)          | Yes                                                     | Yes                                      | Yes                           | Yes                                    | Yes                                                                           | Yes                                                             | Yes                                                                              | Yes                                            | Yes                                                      | Include                    |
| Neukam K <i>et al.</i> (168)             | Yes                                                     | Yes                                      | Yes                           | Yes                                    | Yes                                                                           | Yes                                                             | Yes                                                                              | Yes                                            | Yes                                                      | Include                    |
| Williamson RM <i>et al.</i> (185)        | Yes                                                     | Yes                                      | Yes                           | Yes                                    | Yes                                                                           | Yes                                                             | Yes                                                                              | Yes                                            | Yes                                                      | Include                    |
| El-Gohary M <i>et al.</i> (182)          | Yes                                                     | Yes                                      | Yes                           | Yes                                    | Yes                                                                           | Yes                                                             | Yes                                                                              | Yes                                            | Yes                                                      | Include                    |
| Lombardi R <i>et al.</i> (183)           | Yes                                                     | Yes                                      | Yes                           | Yes                                    | Yes                                                                           | Yes                                                             | Yes                                                                              | Yes                                            | Unclear                                                  | Include                    |
| Marjot T <i>et al.</i> (184)             | Yes                                                     | Yes                                      | Yes                           | Yes                                    | Yes                                                                           | Yes                                                             | Yes                                                                              | Yes                                            | Yes                                                      | Include                    |
| Wilman HR <i>et al.</i> (170)            | Yes                                                     | Yes                                      | Unclear                       | Unclear                                | Yes                                                                           | Yes                                                             | Yes                                                                              | Yes                                            | Yes                                                      | Include                    |
| Abeysekera KWM <i>et al.</i> (171)       | Yes                                                     | Yes                                      | Yes                           | Yes                                    | Yes                                                                           | Yes                                                             | Yes                                                                              | Yes                                            | Yes                                                      | Include                    |

### **2.3.3 Heterogeneity and publication bias**

The heterogeneity among the studies was high, with an  $I^2$  of 93.1% in the disease non-restricted populations. An  $I^2$  of 99.1% was recorded in liver admissions and 93.9% in T2DM studies, as shown in Figure 2-2 and Figure 2-3.

## **2.4 Discussion**

In our systematic review and meta-analysis, we aimed to comprehensively collect and analyse studies that would enable us to estimate the prevalence of NAFLD within the general population of the UK. This encompassed studies from a broad spectrum of cohorts, including the general population, disease-unrestricted populations attending hospitals or clinics without specific disease focus, and disease-specific populations with conditions such as T2DM. However, it is imperative to clarify that the notable finding of a pooled NAFLD prevalence of 23.8% emerges specifically from the disease-unrestricted group, rather than the general UK population.

Moreover, our analysis included specific subgroups, such as T2DM patients and individuals admitted to hospitals for liver-related issues, which exhibited higher NAFLD prevalence rates of 38.6% and 37.9%, respectively. These results point to significant variations in NAFLD prevalence across different population segments.

Amongst our disease-unrestricted studies, the study of Neukam et al. (168) was an outlier. It found a prevalence of 30.8%, rather higher than the findings of other studies within our sample ranging between

19.9% and 23.9%. This study sample, however, though not restricted by disease, was restricted to Gujarati Indians living in North London, which may suggest a greater predisposition to NAFLD in this group for either genetic or environmental reasons. A study conducted on 1,002 healthy Indian blood donors using US found a prevalence of about 53.5%, which indicates a very high prevalence within India. However, another study conducted by Amarapurkar et al. found a relatively low prevalence among adult Indians living in Mumbai, 16.6% upon US examination (186).

This discrepancy in results can be explained by several factors. The variations in geographical distribution within India could have influenced the findings. Local dietary habits, which can vary significantly across different regions, have a strong impact on the development of NAFLD. Therefore, these regional differences can significantly affect the resulting prevalence estimates of NAFLD. Also, genetic variability might be of great weight in this difference (187).

Furthermore, Khalatbari-Soltani et al.'s study (169), focusing on adults in Cambridgeshire, excludes certain health conditions, which might lead to a healthier population subset, not reflective of the general UK demographic. The study's limited age range and geographic focus also restrict its generalizability. Similarly, the study on Gujarati Indians in North London primarily examines a specific ethnic group within a religious setting, which may not mirror the wider UK population's lifestyle, dietary habits, and health practices. This raises concerns about the

study's applicability in understanding NAFLD prevalence on a national scale.

The Wilman et al.'s study (170), due to its volunteer-based nature, potentially includes participants who are healthier and of higher socio-economic status than the general UK population. Its age range and lower representation of ethnic minorities further limit its generalizability.

Lastly, Abeysekera et al.'s study (171) on young adults in Bristol, which focused on a specific birth cohort and geographic area, may not accurately represent the diversity in health, lifestyle, and socio-economic conditions found across young adults in the UK.

The heterogeneity observed among the studies included in the systematic review is indeed high, raising questions about the appropriateness of combining them in a single meta-analysis. Diverse methodologies, populations, and contexts contribute to this heterogeneity, affecting the robustness of the synthesized prevalence estimates. However, in the absence of more comprehensive population studies, these selected studies represent the best available data to estimate NAFLD prevalence in the UK. It is crucial to interpret the combined results with caution, acknowledging the inherent variability and the limitations it imposes on the generalizability of the findings.

While these studies provide valuable data, their generalizability to the entire UK population is limited. Specific cohorts and contexts may not fully capture the diversity of the UK population, potentially overlooking regional variations and socio-economic factors affecting NAFLD prevalence. This limitation raises the question of whether the

true prevalence of NAFLD might be higher or lower than the estimates derived from the meta-analysis.

Considering potential biases, it is plausible that the true prevalence of NAFLD in the UK is likely higher than the estimates derived from the meta-analysis. This is because the disease unrestricted cohorts, such as biobank studies, may have recruited individuals who are more health-conscious and therefore at a potentially lower risk of NAFLD. Additionally, these cohorts might not adequately represent higher-risk populations who may have different health-seeking behaviours. Consequently, the estimates might not fully capture the true burden of NAFLD in the broader UK population.

Moreover, determining how much heterogeneity is too much is complex and context dependent. High heterogeneity does not automatically invalidate a meta-analysis but necessitates a cautious interpretation of results (188). Statistical measures such as the  $I^2$  statistic can indicate the degree of heterogeneity, with values above 50% suggesting substantial heterogeneity. Exploring sources of heterogeneity through subgroup and sensitivity analyses can help understand how different study characteristics influence results. Despite high heterogeneity, combining these studies may still be appropriate to provide a broad estimate of NAFLD prevalence, with the understanding that significant uncertainties and variations exist.

Our results are very similar to those reported in the meta-analysis conducted by Younossi et al., where the global estimated prevalence of

NAFLD was found to be 24% (189). However, the results around the world have also varied greatly. A systematic review in Asia between 1999 and 2019 estimated the prevalence of NAFLD at 29.6%, noting an increase over time of about 50.9 cases per 1000 persons/year (190).

The prevalence of NAFLD exhibits significant geographic variation. In Asia, for instance, the numbers vary considerably; a pooled prevalence of 30.3% was reported in South Korea (191), whereas it ranged from 5% in Singapore to 30% in Indonesia (192). Similarly, the prevalence is relatively high in the United States, particularly among the elderly, where it ranges between 39.2% and 40.3% (193). In contrast, a meta-analysis pooling data from 17 studies encompassing 85,203 individuals, estimated the NAFLD prevalence in Europe to be 26.9% (194).

The variation in NAFLD prevalence globally can be attributed to several key factors. Genetics can make certain populations more prone to NAFLD (195). Dietary habits, especially those high in saturated fats and sugars, also play a significant role. Lifestyle elements, such as physical activity levels (196) and the prevalence of comorbid conditions like obesity and T2DM (189, 197, 198), also influence NAFLD rates.

The role of T2DM as a risk factor for NAFLD was also apparent in our results, where there was a higher prevalence than in the general population- this is an important feature that has already been widely recognized (199, 200).

Our study has considerable strengths. It is the only meta-analysis of UK NAFLD prevalence to date and has been conducted in a rigorous manner according to standard systematic review methodology. This, therefore, will mean that it is likely to have produced the least biased and most generalizable estimate of NAFLD prevalence available for the UK.

It has, however, also considerable weaknesses. These derive from the limited quantity of data available to answer our question and the heterogeneity of results. Heterogeneity may be rooted in the variation in the study locations, samples, and time. Variations in study locations can reflect differences in population genetics, diet, lifestyle, and healthcare practices. Differences in study samples may influence results depending on the demographics, comorbid conditions, and specific risk factors of the sampled population. Temporal changes can also contribute to heterogeneity, as diagnostic practices, awareness, and reporting of NAFLD have evolved over time.

The heterogeneity was also present within the subset of studies examining NAFLD in T2DM patients. Marjot et al. (184) found a prevalence far lower than the other two results: 22.7% compared to 50% and 42.6%. This variation may be attributed to the difference in diagnostic methodology since Marjot et al. used blood analysis to diagnose NAFLD in their sample (184).

The observed heterogeneity in NAFLD prevalence can be partly attributed to the varied methodologies employed in different studies, as included in this systematic review. From conventional techniques such

as US and LFTs to more specific ones like the Fatty Liver Index, BARD score, Southampton Traffic Light Test, and TE, there is a wide array of diagnostic methods in use. Some studies even incorporate advanced diagnostics, like MRI PDFF (170).

It is also critical to note that each methodology possesses unique sensitivity and specificity for detecting NAFLD, subsequently influencing the reported prevalence rates. For instance, a study in India associated ALT and GGT levels with NAFLD, contributing to their prevalence rates (201). Conversely, a study in West Virginia focused on plasma biomarkers like IL-6, Leptin, and FABP-1, which could influence the reported prevalence there (202). The uncertainty in the ALT level cut-off points for diagnosing NAFLD or NASH further adds to this variability (203). Therefore, the reported NAFLD prevalence might reflect the specific methodologies used in each study, adding to the global variations in reported figures.

Another weakness of the available data is that it is inadequate to study trends over time, and it will be important to continue to monitor the prevalence. Although our systematic review and meta-analysis included nine studies, four were not restricted to particular diseases or presentations, and none can be expected to reflect the general population of the UK accurately. As might be expected, the prevalence seems to be higher in T2DM, but there is limited data for other subgroups.



One limitation of the current systematic review is the variability in the time points at which the included studies measured the prevalence of NAFLD. Each study's prevalence estimate corresponds to the specific period during which the study was conducted, leading to potential discrepancies when these estimates are synthesized. This temporal variability must be considered when interpreting the overall prevalence rates derived from our meta-analysis. To approximate changes in NAFLD prevalence over time, we examined the publication years and study periods of the included studies. Although this method provides some insights, it is not a substitute for longitudinal analyses. The studies included in this review primarily offer cross-sectional data, capturing snapshots of NAFLD prevalence at different times and in different populations. Consequently, while we can hypothesize about potential trends, our ability to draw definitive conclusions about temporal changes in NAFLD prevalence is limited.

To examine changes in NAFLD prevalence over time, one would ideally conduct longitudinal studies that follow specific cohorts over extended periods. This approach allows for the observation of NAFLD development and progression within the same population, providing clearer insights into temporal trends. Alternatively, repeated cross-sectional studies conducted at regular intervals could offer valuable data on how NAFLD prevalence shifts at the population level over time.

Alternatives may include focusing on targeted cohort studies within existing healthcare frameworks, utilizing data from existing national health surveys and databases, and engaging in collaborative

research efforts that pool resources and data from multiple centres. These approaches offer practical pathways to gather meaningful insights into NAFLD prevalence and its risk factors, allowing for targeted interventions within the constraints of current research environments. Additionally, standardized diagnostic criteria and imaging modalities should be used to accurately diagnose NAFLD across study sites. Longitudinal follow-up would enable the assessment of disease progression and associated risk factors over time.

Collaborative efforts among researchers, healthcare providers, and policymakers are essential to facilitate the implementation of such a study. By adhering to these recommendations, future research can generate robust evidence on the epidemiology of NAFLD in the UK, informing targeted prevention and intervention strategies.

The systematic reviews are specifically reporting on NAFLD. This confidence stems from the stringent inclusion criteria we employed, which required explicit diagnosis of NAFLD based on established clinical guidelines such as imaging studies, liver biopsy results, or biochemical markers. This rigorous selection process ensures that our prevalence estimates accurately reflect the burden of NAFLD.

The decision to focus exclusively on NAFLD, rather than conducting a wider systematic review of all CLD forms, is rooted in several key considerations. NAFLD has rapidly become the leading cause of CLD in developed countries and is expected to follow similar trends in developing nations. Its association with metabolic syndrome,

obesity, and T2DM makes it a critical public health issue. By concentrating on NAFLD, we aim to address the specific risk factors and pathophysiology unique to this condition, enabling the development of targeted interventions and management strategies. This focus also allows for a more efficient allocation of research resources, ensuring that our efforts directly address the most pressing liver health challenge.

## **2.5 Conclusion**

From our analysis, the current best estimate of NAFLD prevalence in the UK disease un-restricted population stands at 23.7%, with a higher prevalence noted in high-risk subgroups like those with T2DM. Given the high degree of heterogeneity, further studies in the UK are warranted to increase precision and confidence in these prevalence estimates. However, the call for additional research is not merely to refine this number but to deepen our understanding of the evolving nature of NAFLD. A more precise estimate could help us track trends over time, identify potential factors influencing these trends, and monitor the impact of public health interventions on NAFLD prevalence.

Having identified the current best estimate for the prevalence of NAFLD in the UK, the next chapter adopts the same methodology to estimate its prevalence in the KSA. Brief comparisons between the findings in both countries are also drawn.

### **Chapter 3      The prevalence of NAFLD in the KSA: A Systematic Review and Meta-Analysis.**

### **3.1 Introduction**

Although the worldwide prevalence of NAFLD is thought to be around 30%, research suggests that it may be even higher in the Middle East (26). This is due to rapidly increasing rates of obesity, metabolic syndrome, and T2DM, notably in the Gulf states (204), of which KSA is the largest by far. Therefore, this chapter presents a second systematic review which aims to estimate the prevalence of NAFLD in KSA to aid in the assessment of related health needs.

### **3.2 Methods**

The same methods were used for both the UK and Saudi studies; in order to avoid repetition, this section simply identifies areas where changes were made to reflect the different study locations.

#### **3.2.1 Review question**

What is the prevalence of NAFLD among the Saudi population?

#### **3.2.2 Study design and inclusion and exclusion criteria**

A systematic review and meta-analysis was conducted to estimate the prevalence of NAFLD among the Saudi population. We included studies investigating the prevalence of NAFLD in KSA diagnosed by any method previously validated for a diagnosis of NAFLD, such as liver biopsy, US, MRI, other liver scans, liver enzymes, and blood-based biomarkers, after excluding other liver diseases (176, 177). Other criteria were the same as in the UK review.

### **3.2.3 Search strategy**

The same search strategy was applied; however, In place of the UK, in this case the search terms were combined with ((Saudi Arabia) or (Saudia) or (Kingdom of Saudi Arabia)).

### **3.2.4 Study selection and quality assessment**

This followed the same process as the UK review; however, the study population was defined as the general population when sampled from the entire population of KSA.

### **3.2.5 Data management and analysis**

The process adopted was the same as the UK systematic review with no variation.

## **3.3 Results**

### **3.3.1 Search results**

The search returned a total of 168 publications that were included for title screening. 46 of them were included for abstract screening, which led to the exclusion of 22 articles. The remaining 24 publications' full texts were reviewed, leading to the inclusion of eight for final data extraction (172-175, 205-208). None of the studies estimated the prevalence in the true general population in the KSA, and the study most nearly approximating such an approach estimated the prevalence among an age-restricted (20-30 years old) sample of Saudi women, producing a prevalence of 20.4%. The detailed screening process is presented in Figure 3-1.

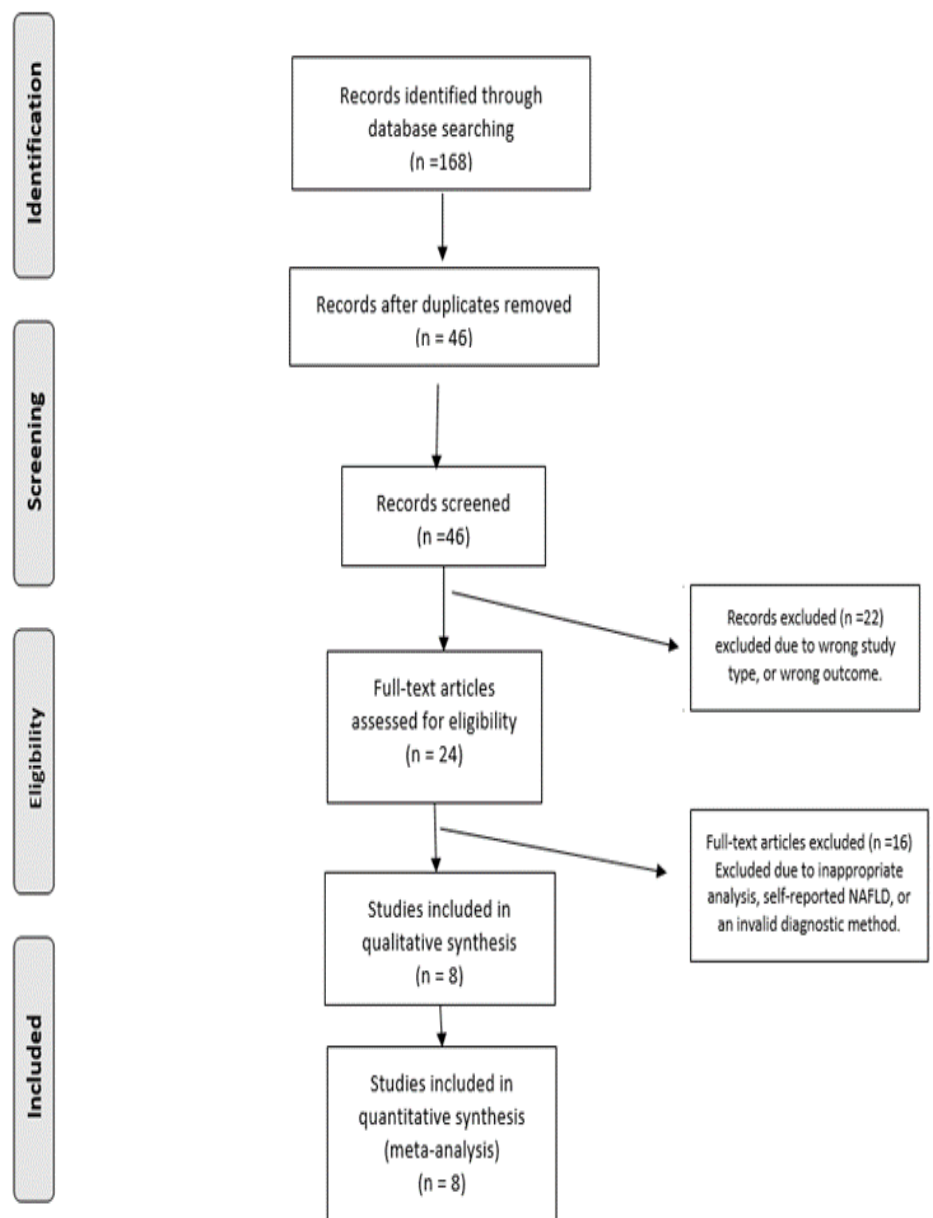


Figure 3-1: Flow diagram of articles screening and selection process for studies of NAFLD in the KSA.

### **3.3.1.1 NAFLD prevalence in KSA**

A total of 4045 participants from eight studies were included in the meta-analysis, of which 3405 were sampled from disease-unrestricted populations (four studies) and 640 from patients diagnosed with T2DM (four studies). Characteristics of the studies are in Table 3-1 and Table 3-2. Across the studies, there was a wide range of disease prevalence, from 9.7% (95% CI: 8.2-11.2) among a disease-unrestricted population in 1989 to 72.8% (95% CI: 67.1-78.6) among T2DM patients in 2017.

Across all studies, the pooled prevalence was 37.2% (95%CI: 25.0-49.6). The pooled NAFLD prevalence in the disease-unrestricted populations was 16.8% (95% CI: 11.1-22.5  $I^2=94.6\%$ ), whereas the pooled prevalence of NAFLD among T2DM patients was 58.0% (95%CI: 45.0-70.9  $I^2=91.0\%$ ). Forrest plots for these analyses are shown in Figure 3-2.



Table 3-1: Characteristics of included studies for the prevalence of NAFLD in KSA:

| Author                         | Year      | Study type      | Study setting   | Population                      | Participants | Males | Age, years | Diagnostic method | Prevalence (95% CI)      |
|--------------------------------|-----------|-----------------|-----------------|---------------------------------|--------------|-------|------------|-------------------|--------------------------|
| <b>El-Hassan et al. (172)</b>  | 1985-1989 | Cohort study    | Hospital-based  | Disease-unrestricted population | 1425         | 57.0% | 45.9±15.7  | CT (HU < -10)     | 9.7% (95% CI:8.3-11.3)   |
| <b>Akbar et al. (205)</b>      | 2003      | Cross-sectional | Hospital-based  | T2DM Patients                   | 116          | 28.0% | 54±12.8    | US                | 55% (95% CI:46.1-63.9)   |
| <b>Fallatah et al. (206)</b>   | 2008-2009 | Cohort study    | Hospital-based  | T2DM Patients                   | 72           | 41.7% | 58.5       | US                | 55.6% (95% CI:44.1-66.5) |
| <b>Al-hamoudi et al. (173)</b> | 2009      | Cross-sectional | Hospital-based  | Disease-unrestricted population | 1312         | 51.0% | 44.7±11.5  | US                | 16.6% (95% CI:14.7-18.7) |
| <b>Alshumrani et al. (174)</b> | 2012      | Cohort study    | Hospital-based  | Disease-unrestricted population | 100          | 76.0% | 41.7±11.2  | CT (HU < 40)      | 23% (95% CI:15.8-32.2)   |
| <b>Elmakki et al(207)</b>      | 2013      | Cross-sectional | Hospital-based  | T2DM Patients                   | 207          | 54.1% | -          | US                | 47.8% (95% CI:41.1-54.6) |
| <b>Alsabaani et al. (208)</b>  | 2016      | Cross-sectional | Community-based | T2DM Patients                   | 245          | 66.1% | 57.1±13.5  | US                | 72.8% (95% CI:66.6-78.1) |
| <b>Alazzeah et al. (175)</b>   | 2017      | Cross-sectional | Community-based | Disease-unrestricted population | 568          | 0.0%  | 22.6±4.7   | US                | 20.4% (95% CI:17.3-23.9) |

Table 3-2: Detailed characteristics of included studies for the prevalence of NAFLD in the KSA.

| Study name                           | Study purpose/aim                                                                                                                                                                  | Target population, sampling frame, power calculation and response rate                                                                                                                                                                                                                                                                                                                                                                                                           | Definition of NAFLD (including diagnostic criteria)                                                                                                                                                                                                                           |
|--------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b><i>El-Hassan et al(172)</i></b>   | To better examine the intricacy of cases and diagnostic signs and to cover inadequate data concerning clinical and radiological features of fatty infiltration of the liver (FIL). | <ul style="list-style-type: none"> <li>All patients suffering from fatty infiltration of liver FIL attended King Fahd University Hospital.</li> <li>Two independent reviewers reviewed the files of all patients over a five-year period (1985-1989).</li> <li>The sample includes all patients with positive CT results on the presence of FIL.</li> <li>The power of calculation was not mentioned.</li> </ul>                                                                 | The presence of unequivocal radiological evidence, positive liver biopsy results or/and appropriate radiological follow-up was performed, defined as diffuse or focal FIL. The existence of a non-spherical focal area with no mass effect and low density defined local FIL. |
| <b><i>Akbar et al (205)</i></b>      | Assessing prevalence and characteristics of NAFLD in patients with type 2 diabetes attending King Abdulaziz University Hospital.                                                   | <ul style="list-style-type: none"> <li>All type-2 diabetic patients in King Abdulaziz University Hospital.</li> <li>The sample was drawn from type-2 diabetic patients attending King Abdulaziz University Hospital. 116 patients were randomly selected. Abdominal ultrasound and liver measurement were performed for the entire sample.</li> <li>No power of calculation was mentioned.</li> </ul>                                                                            | An increase in the echogenicity of the liver parenchyma in sonography was diagnostic of NAFLD, which is determined by the lack of hyperechoic walls of the portal vein radicals and an interface between the diaphragm and liver.                                             |
| <b><i>Fallatah et al(206)</i></b>    | Comparison between Fibroscan, FIB-4 scores, APRI scores, and AST/ALT ratios in the assessment of liver fibrosis in patients with NAFLD.                                            | <ul style="list-style-type: none"> <li>All patients with NAFLD attending King Abdulaziz University Hospital (KAUH).</li> <li>The sample was extracted from patients with NAFLD based on abdominal ultrasound and serum hepatic enzyme testing. All patients who attended the diabetic clinic were included in the study (122 patients).</li> <li>No power of calculation was indicated.</li> </ul>                                                                               | Those patients with abnormal findings of abdominal ultrasound (showed fatty change), serum hepatic enzyme testing, Fibroscan, FIB-4 scores and APRI scores.                                                                                                                   |
| <b><i>Al-hamoudi et al (173)</i></b> | To assess characteristics of Saudi NAFLD patients attending university hospitals and other factors that affect ALT levels.                                                         | <ul style="list-style-type: none"> <li>All patients undergo ultrasound regardless of the primary ultrasound indication. At King Khalid University Hospital, they were screened for NAFLD. A total of 218 patients were labelled as having NAFLD, leaving 202 after applying the remaining exclusion criteria (if they had other liver diseases such as viral liver disease, consumed alcohol, and were on medications).</li> <li>Power calculation was not mentioned.</li> </ul> | Existence of fatty liver upon the performance of abdominal ultrasound without the presence of coexisting liver problems or alcohol abuse.                                                                                                                                     |

|                               |                                                                                                                                                                                                                                           |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                                              |
|-------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Alshumrani et al (174)</b> | Using unenhanced CT to assess the prevalence of hepatic steatosis among adults in southwest Saudi Arabia.                                                                                                                                 | <ul style="list-style-type: none"> <li>All adults with hepatic steatosis based on CT results in southwest Saudi Arabia.</li> <li>The sample was extracted from patients attending a tertiary hospital in southwest Saudi Arabia after unenhanced CT testing to evaluate the urinary tract stones.</li> <li>No power of calculation was mentioned.</li> </ul>                                                                                                                                       | The definition of NAFLD was found in abnormal CT findings depending on different cut-offs of fatty liver disease. Five cut-offs depend on mean liver density MLD and mean spleen density MSD. $MLD \leq 40$ HU, $MLD \leq (MSD-10)$ HU, $MLD \leq MSD 36$ , $MLD \leq (MSD+5)$ HU $MLD/MSD \leq 1.1$<br>HU: Hounsfield units |
| <b>Elmakki et al(207)</b>     | Assessing the Prevalence of NAFLD among type-2 diabetic patients in the Jazan region, Saudi Arabia and explore the most important associated factors.                                                                                     | <ul style="list-style-type: none"> <li>Type-2 diabetic patients in the Jazan region with a diagnosis of NAFLD.</li> <li>The sample was extracted from patients attending Jazan General Hospital. Patients were chosen through systematic random sampling.</li> <li>A representative sample of 230 participants was calculated, estimating a non-response rate of 15%.</li> </ul>                                                                                                                   | Abnormal Ultrasound results are indicative of the presence of fatty liver-defined patients with NAFLD.                                                                                                                                                                                                                       |
| <b>Alsabaani et al (208)</b>  | Determining the Prevalence of NAFLD among diabetic patients in Abha region, Saudi Arabia, and the associated factors.                                                                                                                     | <ul style="list-style-type: none"> <li>All patients diagnosed with diabetes in Abha region, Saudi Arabia.</li> <li>The sample was taken from patients attending diabetes clinics in Abha region using systematic random sampling.</li> <li>A sample size of 267 was calculated.</li> </ul>                                                                                                                                                                                                         | Presence of fatty liver upon performing an ultrasound with the absence of coexisting liver disease, alcohol or steatogenic drugs consumption.                                                                                                                                                                                |
| <b>Alazzeah et al (175)</b>   | To assess the Prevalence of NAFLD among female university students in the Hail region, Saudi Arabia, examine the associated risk factors and try to validate different simple, non-invasive indices, which can be of clinical importance. | <ul style="list-style-type: none"> <li>All female university students attending university in Hail region.</li> <li>The sample was retrieved from females attending the University of Hail. Data were collected by recruiting 568 female students to study the Prevalence of NAFLD among female students.</li> <li>A cross-sectional study was initially performed, followed by a case-control study to assess the different risk factors.</li> <li>No power calculation was estimated.</li> </ul> | Positive findings using ultrasound with the exception of coexistence of CLD, hepatitis B virus surface antigen or the presence of hepatitis C virus antibodies, abuse of alcohol or steatogenic drugs.                                                                                                                       |

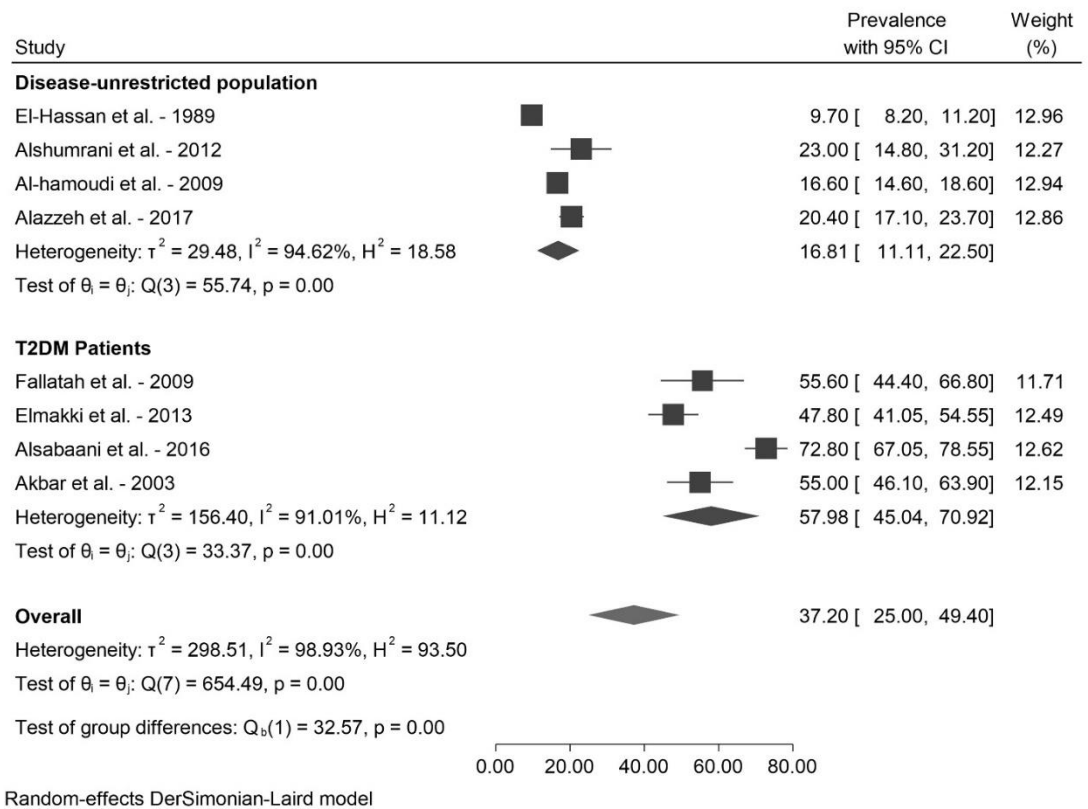


Figure 3-2: Forest plot of the Prevalence of NAFLD in KSA by subgroups:

### 3.3.2 Quality assessment

A summary of the quality assessment of the included articles is presented in Table 3-3. All articles were of adequate quality, meeting our definition of low risk of bias. Five studies were unclear about the adequacy of sample size (172, 173, 175, 205, 206), sampling was not clearly explained in two studies (175, 205), one did not properly describe subjects and settings (174), and one study did not mention the response rate (205).

Table 3-3: Quality assessment of included studies for NAFLD prevalence in KSA.

| Author                  | Was the sample frame appropriate to address the target population? | Were study participants sampled in an appropriate way? | Was the sample size adequate? | Were the study subjects and the setting described in detail? | Was the data analysis conducted with sufficient coverage of the identified sample? | Were valid methods used for the identification of the condition? | Was the condition measured in a standard, reliable way for all participants? | Was there an appropriate statistical analysis? | Was the response rate adequate, and if not, was the low response rate managed appropriately? |
|-------------------------|--------------------------------------------------------------------|--------------------------------------------------------|-------------------------------|--------------------------------------------------------------|------------------------------------------------------------------------------------|------------------------------------------------------------------|------------------------------------------------------------------------------|------------------------------------------------|----------------------------------------------------------------------------------------------|
| El-Hassan et al. (172)  | Yes                                                                | Yes                                                    | Unclear                       | Yes                                                          | Yes                                                                                | Yes                                                              | Yes                                                                          | Yes                                            | Yes                                                                                          |
| Akbar et al. (205)      | Yes                                                                | Unclear                                                | Unclear                       | Yes                                                          | Yes                                                                                | Yes                                                              | Yes                                                                          | Yes                                            | Unclear                                                                                      |
| Fallatah et al. (206)   | Yes                                                                | Yes                                                    | Unclear                       | Yes                                                          | Yes                                                                                | Yes                                                              | No                                                                           | Yes                                            | Yes                                                                                          |
| Al-hamoudi et al. (173) | Yes                                                                | Yes                                                    | Unclear                       | Yes                                                          | Yes                                                                                | Yes                                                              | Yes                                                                          | Yes                                            | Yes                                                                                          |
| Alshumrani et al. (174) | Yes                                                                | Yes                                                    | Yes                           | Unclear                                                      | Yes                                                                                | Yes                                                              | No                                                                           | Yes                                            | Yes                                                                                          |
| Elmakki et al(207)      | Yes                                                                | Yes                                                    | Yes                           | Yes                                                          | Yes                                                                                | Yes                                                              | Yes                                                                          | Yes                                            | Yes                                                                                          |
| Alsabaani et al. (208)  | Yes                                                                | Yes                                                    | Yes                           | Yes                                                          | Yes                                                                                | Yes                                                              | Yes                                                                          | Yes                                            | Yes                                                                                          |
| Alazzeah et al. (175)   | Yes                                                                | Unclear                                                | Unclear                       | Yes                                                          | Yes                                                                                | Yes                                                              | Yes                                                                          | Yes                                            | Yes                                                                                          |

### **3.3.3 Heterogeneity and publication bias.**

The between-study variability was extremely high, with  $I^2$  for the overall analysis being 98.9%. There was high heterogeneity within population subgroups, with  $I^2=94.6\%$  among the disease-unrestricted population and  $I^2=91.0\%$  among T2DM patients.

## **3.4 Discussion**

We have shown that the available literature to assess the prevalence of NAFLD in KSA is limited. Our review included eight studies of adequate but variable quality, none of which addressed a true general population. In this study, we have found the pooled prevalence of NAFLD in KSA to be 37.2%. However, we were unable to identify any general population studies.

Amongst people attending hospitals/community healthcare for any reason (disease unrestricted), the prevalence was 16.8%, and the only disease-specific subgroup was T2DM, with a prevalence of 58.0%. The closest study to a general population was (175), which found a prevalence of 20.4% in a group of young (20-30 years) women.

No previous systematic reviews have estimated the prevalence of NAFLD in the KSA. We, therefore, provide the best estimates to date. NAFLD is often a diagnosis of exclusion,

identified in a broad range of specialties, and requires the exclusion of several other conditions. As a result, there is no single diagnostic test, leading to inconsistencies in reporting. This complex diagnostic process makes conducting truly population-based studies challenging compared to conditions where a single, inexpensive diagnostic test is available. Consequently, few such studies are conducted.

Four of the studies we included could be considered to approximate a general population study (172-175). Three of these studies included people attending for imaging (either CT or US) for reasons unrelated to NAFLD (172-174), and the other was a cross-sectional study of female University students (175).

The study of El-Hassan et al (172), has several critiques regarding its representativeness of the general population in KSA. The study's patient cohort was derived from adults undergoing CT studies at a single university hospital over a six-year period. This selection criterion inherently introduces a bias towards individuals with underlying health conditions severe enough to warrant a CT scan. Therefore, the study population may have a higher prevalence of health issues compared to the general population. Moreover, as it was conducted in Dammam in the Eastern Province of KSA, the study's findings may not be generalizable across the entire country due to regional variations in lifestyle and

diet. The study's demographic details also reveal a predominance of middle-aged males (mean age of 45.9 years and 57% male), which may not accurately represent the age and gender distribution of the entire population.

The study by Al-hamoudi et al (173). conducted at King Khalid University Hospital in Riyadh, Saudi Arabia, aimed to determine the characteristics of Saudi patients with NAFLD. While the study provides important insights into the prevalence and characteristics of NAFLD among patients in a hospital setting, several critiques regarding its representativeness of the general Saudi population. The study's hospital-based design means that the participants were patients referred for US at a university hospital, likely due to existing health concerns or symptoms suggestive of liver disease. This selection bias limits the study's ability to capture the true prevalence and spectrum of NAFLD in the general population, as it excludes healthy individuals or those without indications for an ultrasound who might still have NAFLD. In addition, as the study was conducted in Riyadh, the capital and one of the largest cities in KSA, the study's findings may not be generalizable to the entire country. Furthermore, the study reports high prevalences of obesity and diabetes among the participants, which are major risk factors for NAFLD. While these conditions are indeed prevalent in KSA, the



concentration of patients with these risk factors in a hospital setting might inflate the association between these conditions and NAFLD compared to a community-based sample. This could lead to an overestimation of NAFLD prevalence linked to obesity and diabetes, overlooking other potential risk factors or causes of the disease in the general population.

The study by Alshumrani et al (174). assesses the distribution of hepatic steatosis among adults in Southwestern Saudi Arabia using unenhanced CT scans. While the study makes an important contribution to understanding the prevalence of fatty liver disease in this region, there are several critiques concerning its representativeness of the general population in Saudi Arabia.

The study's sample consisted of 100 consecutive adult patients who presented with suspected symptoms of renal colic to the Outpatient Clinic or Emergency Department of a tertiary care hospital. This sampling method results in a population that may not be representative of the general community, as it is biased towards individuals with specific symptoms leading to the performance of unenhanced CT scans primarily for detecting urinary stones, not liver disease. The gender distribution within the study was notably skewed, with 76% of the participants being men and only 24% women. This imbalance could affect the

generalizability of the findings, especially considering that fatty liver disease might present differently between genders. Additionally, while the study mentions nationality, it does not elaborate on the proportions of Saudi nationals versus expatriates, which could also influence the representativeness of the findings. And, finally, the focus in the study on the Southwestern region of Saudi Arabia may mean the study's findings are not generalizable to the entire country.

The study by Alazzeah et al (175). focuses on identifying NAFLD in Saudi females and validating non-invasive indices. While this study contributes important insights into NAFLD among Saudi females and explores non-invasive diagnostic tools, there are several considerations regarding its generalizability and representativeness. By exclusively studying Saudi females, the research provides valuable data on this demographic. However, the findings cannot be directly extrapolated to the entire population, including males, due to physiological, metabolic, and potentially lifestyle differences between genders that can influence NAFLD prevalence and characteristics. While focusing on females addresses an important gap in the literature, additional research is necessary to understand NAFLD's full impact across all demographics within KSA. Furthermore, the study was conducted only in Hail, so the findings may reflect the

prevalence and characteristics of NAFLD in this specific region. However, KSA is a country with diverse populations and varying dietary, lifestyle, and genetic factors across its regions, so the results from one region might not accurately represent the situation in other areas, especially urban versus rural settings, which can have different risk factor profiles for NAFLD.

Although not true general population studies, it is reasonable to consider these as providing more realistic reflections of the general population risk than is available in studies selected for reasons related to liver disease, NAFLD or its risk factors. It's evident that while individually, each study may have limitations in fully estimating the true general prevalence of NAFLD across the KSA, collectively, they provide a valuable estimation. Each study, with its unique focus ranging from specific populations, such as Saudi females, to different geographic regions within the country, and employing various methodologies contributes to the larger of understanding of NAFLD prevalence in Saudi Arabia.

Collectively, these studies underscore the multifaceted nature of NAFLD in Saudi Arabia. While no single study captures the entire scope of NAFLD prevalence across the Kingdom, together, they form a comprehensive picture that highlights the significant health burden posed by NAFLD.

Moreover, these studies collectively call attention to the urgent need for nationwide, representative epidemiological research on NAFLD. Such research would ideally integrate diverse methodologies and populations to accurately assess the prevalence, risk factors, and regional variations of NAFLD across Saudi Arabia. By building on the existing research foundation, future studies can aim to provide a more definitive understanding of NAFLD's impact, guiding public health initiatives and healthcare resource allocation effectively.

It is striking that the prevalence from the three disease-unrestricted imaging-based studies is close to that from the cross-sectional survey and also that these estimates are between those provided by cross-sectional surveys from Iran and Israel, which were included in the meta-analysis by Younossi et al. studying the global epidemiology of NAFLD (209, 210). The overall estimate from that study of Middle Eastern NAFLD prevalence of 32% was greatly raised by including a study from Turkey conducted on gastroenterological and general internal medical outpatients, among whom 55.1% were found to have NAFLD (211).

Compared to the Younossi review, this study's estimate falls between the population-based estimates from Iran and Israel and is lower than the estimate from Turkey, which was specific to gastroenterology outpatients. Iran and Israel's estimates provide

a broader representation of the disease burden in the general population, while Turkey's estimate may overestimate the disease burden due to the selection of a specific patient population.

In addition to being broadly similar to other Middle Eastern estimates of NAFLD prevalence, our results also mirror those elsewhere, showing that NAFLD is more common in people with diabetes. We found a pooled prevalence in studies of T2DM populations of 58.0%, and this greatly elevated prevalence is reflected in studies from China, India, Africa and Europe (212-214).

Looking forward, the one-quarter of the adult population of KSA affected by T2DM is predicted to more than double by 2030 (215). Similarly, the worrying levels of obesity in the population are projected to rise further (216). Reflecting this, modelling estimates NAFLD prevalence of 25.7% and with likely trends in risk factors, this has been estimated to rise by 2030 to reach 31.7% among the Saudi general population (43).

Despite the limitations of the data available, it could be argued that this rising trend is already clear within the studies we have included based on disease unrestricted imaging series. The study by El-Hassan (172) and colleagues utilising CT scans from 1985-89 found a NAFLD prevalence of 9.7%, which rose to 16.6% in the US-based study of Al-hamoudi (173) and colleagues in

2009 and to 23% in the CT based study of Alshumrani (174) in 2012.

This rising tide of NAFLD is likely to have a profound impact on the people and healthcare systems of KSA. NASH has already overtaken HCV hepatitis as the most common cause of liver transplant in KSA (40). However, NAFLD progresses slowly, so these transplants are likely to reflect NAFLD, which developed many years ago when the prevalence was lower (217).

Hence, the current levels of clinical liver disease due to NAFLD are likely to be just the beginning of far larger impacts in future. It is not yet known what the rate of progression of screen detected NAFLD will be. We do, though, know that about 20% of T2DM patients have elevated liver stiffness internationally (suggesting the presence of significant fibrosis), that NAFLD patients with fibrosis have a high risk of progressing to cirrhosis (218) and that the rate of T2DM in KSA is high and rising (215). It is reasonable, therefore, to assume that there will be high levels of cirrhosis, decompensated cirrhosis and HCC in the coming years.

The systematic methodology of this study reduces the risk of selection bias in study selection and ensures that our results are as authoritative as possible. However, as for any systematic review, the quality of the output is dependent upon the available

studies. In this case, the studies have high heterogeneity due to the variety of study designs and diagnostic methods utilised. In addition, and arguably a more serious issue, a paucity of studies limits the scope and precision of the results. None of the studies addressed a true general population prevalence of NAFLD or reported on high-risk subgroups other than diabetics.

Our study sheds light on the prevalence of NAFLD in the KSA, providing insights into the epidemiological landscape of this condition within the country. The wide range of prevalence rates observed across the included studies underscores the heterogeneity in NAFLD prevalence estimates, reflecting variations in population demographics, diagnostic criteria, and study methodologies. The substantial burden of NAFLD among individuals with T2DM is particularly noteworthy, emphasizing the intricate interplay between metabolic disorders and hepatic steatosis. These findings underscore the urgent need for targeted interventions to address the growing public health challenge posed by NAFLD in the KSA.

The findings of our study have significant clinical and public health implications for the management of NAFLD in the KSA. Given the high prevalence of NAFLD observed among individuals with T2DM and other metabolic disorders, healthcare providers must prioritize the early detection and management of

NAFLD to mitigate the risk of disease progression and associated complications. The clinical implications of our findings underscore the necessity for a multidisciplinary approach involving collaboration between healthcare professionals, policymakers, and community stakeholders are essential to develop and implement comprehensive strategies for the prevention, diagnosis, and management of NAFLD in the KSA. Moreover, public health initiatives aimed at promoting healthy lifestyle behaviours, including regular physical activity and a balanced diet, are crucial for reducing the burden of NAFLD and its associated comorbidities in the population.

### **3.5 Conclusion**

The best available estimates of NAFLD prevalence in KSA show that it is common and has similar risk factors to those elsewhere, as far as can currently be assessed, and it is reasonable to assume that it will become an ever-greater disease burden without action being taken. Without better estimates of the prevalence in both the general population and at-risk groups in KSA, precise estimates of the size of this problem are made more difficult; however, based on the available knowledge, interventions targeting obesity and T2DM are likely to be appropriate measures to limit the spread of this disease in the KSA.



Overall, the findings of these systematic reviews indicate that, in general, the burden of NAFLD in KSA and the UK closely aligns with global estimates for the general population and for specific groups, such as individuals with T2DM. This suggests that risk factors for NAFLD may be similar in the two nations and that hence the broad strategy of targeting those risk factors to reduce the prevalence of NAFLD may be generalisable from the UK to KSA.

However, it is essential to consider the different factors driving NAFLD in these regions. In the UK, lifestyle factors such as high rates of obesity and sedentary behaviour are significant contributors (219). In contrast, in KSA, dietary patterns high in fats and sugars, along with genetic predispositions, play a more prominent role (43). Given these differences, public health measures that are effective in the UK may need to be adapted to suit the unique context of KSA.

For example, interventions focusing on dietary modifications might need to account for traditional dietary practices in KSA (220). Similarly, strategies to promote physical activity may require different approaches tailored to cultural norms and available infrastructure (221). This implies that while the core components of public health measures may be similar, their

implementation needs to be context-specific, respecting the cultural, social, and environmental factors unique to each region.

As non-invasive biomarkers, notably the AST/ALT ratio, APRI score, and FIB-4 index, have already demonstrated their diagnostic accuracy in respect of NAFLD (109), the next part of this study establishes whether these biomarkers can also serve as predictive tools for CLD in order to facilitate the introduction of interventions at an earlier and more effective stage.

## **Chapter 4      Clinical Practice Research Datalink (CPRD)**

## **4.1 Introduction**

The CPRD is one of the world's largest electronic health research databases based on general practice records. General practices (GPs) are the gatekeepers of the National Health Service (NHS) in the UK, which serves roughly 98% of the total population. The CPRD collects anonymised electronic health record data from participating practices on a regular basis (222), and these provide the basis for the studies reported in Chapters 6 and 7.

This chapter, therefore, provides an introduction to the CPRD, describing its nature and structure and identifying its strengths and limitations. The methodology portion of the subsequent chapters contains further study-specific information on the actual data used. The ethical approval process involved in accessing the CPRD for the purposes of this research is also described.

## **4.2 Development of CPRD**

The database was founded in 1987 in London as the Value Added Medical Products (VAMP) computer research database as a business endeavour (223). However, VAMP Ltd was acquired by Reuters in 1993, which in turn donated the database to the Department of Health, who renamed it the General Practice Research Database (GPRD) (224). Administration of the GPRD

became the responsibility of the Medicines and Healthcare Products Regulatory Agency (MHRA) (225) in 2003, at which point the use of the database expanded considerably. It was renamed the Clinical Practice Research Datalink (CPRD) in 2012, with the aim of integrating a broader spectrum of health data and extending the scope of research possible with the database (225).

### **4.3 Nature and structure of CPRD**

CPRD data is collated into two distinct formats, CPRD GOLD and CPRD Aurum, each dependent on the software employed in the general practice. GPs routinely record patient information during their clinical practice using their personal desktop computers with either Vision software for CPRD Gold or EMIS software for CPRD Aurum. The fundamental differences between these two databases stem from the unique structures and coding methods used by each software system. Following the recording process, this data, initially generated for clinical use, is then repurposed to meet CPRD recording requirements, which allows for its further utilization in research endeavours.

In all of these systems, which use fundamentally similar data models for primary care data, single healthcare events are kept as date-stamped coded records. The way this data is organised varies due to differences in system design, which slightly influences recording methods. However, it's crucial to note that

despite these variations, the baseline data remains consistent (225). The following information must be filled in for all active patients in practice:

- Birth date, sex, date of registration with the practice, and current registration status are all included in the registration details.
- Morbidity events include those that end in hospitalisation or referral, events that result in a medicine prescription, and events that require many consultations.
- Symptoms, investigations, procedures, and diagnosis.
- Details of the death, including the date and reason, if known.
- Pregnancy outcomes.
- Prescriptions' information.

The practice must provide at least 95% of this "Research Information" after being accepted into the CPRD Gold in order to be declared UTS (226).

In CPRD, data is structured into two primary categories: practice data and patient data, each distinguished by pseudonymized IDs. The patients' data is subdivided into more categories: clinical, test and therapy data, immunisation, and referral. Patients' demographics, clinical data, and medical staff and practices information are included (227). Appendix B (CPRD

GOLD Data Specification) contains descriptions of the CPRD's organizational structure.

This thesis exclusively utilizes data input using Vision software for CPRD GOLD. The selection of CPRD Gold over CPRD Aurum for my study lies in the commencement of data coverage. Although both databases provide similar linkages to HES data, Death Registration, Cancer data, and the Index of Multiple Deprivation, CPRD Gold's initiation from approximately 1988 seamlessly aligns with the start of our study in 1990. Additionally, CPRD GOLD provides access to a time when AST testing was more prevalent, which is critical for this study. This prevalence offers a more comprehensive view of the period when AST testing was commonly conducted.

While CPRD GOLD provides a longer historical coverage starting from 1988, which aligns with the commencement of our study in 1990, there is a recognized limitation regarding the linkage to HES data. The linkage with HES data did not start until 1997, which means that for the initial years of the study, the benefits of this extended historical coverage are not fully realized in terms of hospital episode statistics. This poses a potential weakness in the rationale for exclusively using CPRD GOLD over CPRD Aurum, which could have provided similar linkages starting from a later period when HES data were available.

In contrast, although CPRD Aurum boasts a higher coverage population, with 19 million patients from 738 practices, its inclusion of records predating 1995 introduces an arbitrary cut-off that may compromise the completeness and reliability of recorded information following this point (228). Therefore, while CPRD Aurum offers a broader dataset, the historical completeness and the prevalence of AST testing available in CPRD GOLD were deemed more critical for the purposes of this study. Additionally, the study was limited by the number of cases that could be drawn from CPRD, meaning that there was no real power benefit from using CPRD Aurum, as the maximum number of cases allowed had already been reached with CPRD GOLD.

To address the potential weaknesses and assess the impact of these temporal data variations, chapter 8 of this thesis performs a sensitivity analysis stratifying by three distinct time periods. This analysis examines the effect of changes in available data throughout the study, ensuring that the findings are robust and account for the varying data coverage over time.

Vision software is an electronic patient management system for doctors that records information mainly as "events." Patient information is kept on file from their first interaction with the participating clinic until their last. Data from individual practices are uploaded to CPRD on a monthly basis. The obtained data is



subjected to a quality check to guarantee complete longitudinal records, adherence to CPRD recording recommendations, practice-level data, and internal consistency of patient data.

#### **4.4 CPRD coverage**

By December 2017, the CPRD GOLD database contained information on nearly 17 million patients or 12 per cent of the population of the UK. Nearly 5 million of these were active registered patients (222). Data is recorded in practice data according to the 13 geographical regions on nearly 735 practices that exist in the UK, with ten representing England and one representing each of the three other countries: Northern Ireland, Scotland, and Wales.

However, the level of participation in practices from different regions is unequal, with some regions contributing much more than others. The greatest number of patients were from London (13.6% of the total) and the least in the East Midlands (0.7%), according to a publication published in 2013 (222).

#### **4.5 External linkage**

In addition to CPRD, data from other databases are linked to provide a richer health record. These include the HES, offering details about hospital admissions, outpatient appointments, and

A&E attendances (229); the Office of National Statistics (ONS) death registration data with primary and secondary causes of death (230); and the National Cancer Registration and Analysis Service (NCRAS) cancer registry data on malignant and pre-malignant neoplasms (231).

Additional linked databases include the Myocardial Ischemia National Audit Project (MINAP), the mother-baby link, the national joint registry, and data on socioeconomic deprivation. This linkage enhances the information available for approximately 75% of English practices within the CPRD, providing a more comprehensive view of patients' health trajectories (222)

#### **4.6 Diagnosis codes**

Diagnoses were coded using the Oxford Medical Information System (OXMIS) during the early years of the CPRD, which was eventually replaced by Read codes in 1995. Read codes are available for CPRD gold. HES maintains information about hospital admissions, but instead of using the Read coding system, it utilises the International Classification of Diseases (ICD-10) coding system for diagnosis.

#### **4.7 Strengths and limitations of CPRD**

CPRD is one of the largest primary care databases available, containing information on lifestyle and morbidity

determinants, as well as links to secondary care and mortality statistics (222).

The large number of patients in the database and the longitudinal data make it a valuable resource for researchers looking into illness links and outcomes, among other things. The CPRD broadly represents the UK's general population, making population-based studies feasible in this country (222, 228).

The Quality and Outcomes Framework (QOF), which was first implemented in 2004, is an optional annual incentive programme requiring general practice services to track numerous quality-of-care parameters to obtain monetary rewards. QOF provides a set of indicators that can be used to assess the overall quality of services offered at the practice level (232). The QOF enhanced the completeness of many variables recorded (222).

While the early years of CPRD might have seen less comprehensive recording of lifestyle determinants, the advent of QOF has substantially improved these records, making them invaluable for present-day research. However, despite the lack of liver-specific indicators in the QOF, its implementation has led to more consistent and complete patient data recording, indirectly benefiting liver-related research. Several QOF indicators, such as alcohol consumption, obesity, diabetes, and cholesterol

management, are relevant to liver disease risk factors and comorbidities, especially NAFLD.

Other advantages include the availability of laboratory and examination results data, many external linkages, and a high proportion of secondary care events recorded in general practice records. Investigators may acquire access to anonymised extracts of original medical records and contact primary care physicians to conduct supplementary questionnaires dependent upon approvals and payment of fees (225).

#### **4.7.1 Size**

Because the CPRD is such a big longitudinal primary care database, it is especially promising for study into relatively uncommon illnesses like cirrhosis. Using a big database like the CPRD will enable more exact estimations of the prevalence of cirrhosis and the consequences associated with this chronic disease than smaller, clinically oriented databases could previously provide.

#### **4.7.2 Representative**

Although the practices included in the CPRD were initially self-selected, they have since been found to be generally representative of the UK's general population in terms of morbidity and mortality estimates and terms of age, sex, ethnicity

and BMI. Thus, it would appear reasonable to generalise data from the CPRD to the entire UK population (222). The CPRD provides a good opportunity for various academic research such as clinical epidemiology, health service research, drug safety, health outcomes research, disease management, and pharmacoeconomics (222).

Nonetheless, it's essential to recognize that our study's 'blood test' population, consisting of individuals with AST measurements, represents a specific subset within the CPRD database. This subset may have distinct characteristics and risk factors compared to the entire UK population. As a result, the generalizability of our study's findings to the broader population should be considered with caution.

#### **4.7.3 Prospectively collected**

The majority of data in the CPRD are acquired prospectively as part of each patient's routine medical care. As a result, there is significantly less potential for bias in recording individual diagnoses, symptoms, or, more specifically, exposures than there may be in smaller studies meant to address a specific research question. Indeed, there is no possibility of bias in reporting primary care prescriptions, as the database's prescription information is generated electronically concurrently with the prescription itself.

However, because some of the data in the CPRD are based on letters from secondary care, some occurrences may be misclassified in terms of their precise dates of occurrence.

#### **4.7.4 Incomplete recording**

As data in the CPRD is collected on a regular basis as part of the usual course of care, certain information that would be useful for research purposes, particularly information about previous exposures, may be missing. Furthermore, because the data collection is based on what the GPs consider to be significant for the patient's ongoing care, there is likely to be a bias in the information obtained.

Medication that is either prescribed in secondary care or consumed as over-the-counter medication without a prescription is another potential gap in the information captured in the CPRD.

#### **4.7.5 Duration of follow-up**

Although the CPRD is the largest longitudinal database of primary care data, it is limited in terms of the duration of follow-up time that may be stored on each patient. Patients are free to change general practice whenever they like, and there has been turnover in participating practices, so many individuals will contribute a limited follow up period. Though the database's power for estimating incidence will be high, the power for following

up on these event cohorts will be poorer due to the short period of individual person time.

#### **4.7.6 Reproducibility**

The CPRD is an ongoing database that allows for the use of the same methods repeatedly to evaluate trends over time.

### **4.8 Ethical Consideration and Research Approval**

The ethical approval of CPRD provides assurance that the service is committed to adhering to strict ethical standards in research. The approval process involves a thorough review of the research proposal to ensure that it meets ethical standards and that the benefits of the research outweigh any potential risks.

Permission from the Independent Scientific Advisory Committee (ISAC) is required to undertake research that can be published. This committee was formed in February 2006 by The Secretary of State for Health to examine research proposals using the Yellow Card Scheme and MHRA's CPRD datasets (233). The ISAC committee evaluates the proposed research's quality, feasibility, and usefulness to public health using anonymized patient-level CPRD data. It assesses planned studies' medical, epidemiological, and methodological aspects and offers high-quality peer assessments on the precise validity of submitted study ideas.

After completing and successfully reviewing the ISAC protocol form (Appendix C), Protocol number 17\_235 was approved, and the data used in this thesis was provided according to this protocol. The details of the study population are presented in the following chapters.

## **4.9 Conclusion**

This chapter has introduced CPRD which forms the primary data source for the following studies, describing its development, nature and structure, and considering its strengths and limitations. For the purposes of this research study, it provides an established source of routinely collected patient data, meaning that no additional patient data would need to be collected if the procedures described in the next chapters were to be adopted as part of wider early identification processes, thereby reducing the time and costs required.



## **Chapter 5      Assessing the availability of non-invasive liver biomarkers among UK primary care patients**

## **5.1 Introduction**

As previously discussed in (section [1.7.2](#)), CLD is often diagnosed at a late stage, typically when severe and irreversible liver damage, such as liver cirrhosis or HCC, has already occurred. While liver biopsy is the gold standard procedure for diagnosing and assessing the degree of liver inflammation and fibrosis in patients (see Section [1.8.1](#)), the procedure has some limitations, including the risk of sampling error, infection, and bleeding (76, 234), and there has been a significant push towards identifying non-invasive measures for detecting fibrosis (145, 235).

Before examining the utility of biomarkers such as the AST/ALT ratio, APRI score, and FIB-4 index, it is essential to establish their availability within the general practice population in the UK. This is fundamental as it provides insights into the retrospective application of these biomarkers, enabling their use without necessitating general practitioners to conduct additional blood tests, and it forms the basis for understanding the practicality and feasibility of using these biomarkers in the real-world context of primary care.

## **5.2 Aim**

This study therefore aims to estimate the proportion of patients over 40 years of age, registered with a GP on CPRD and

free of known advanced CLD, who had data within their records to calculate liver fibrosis biomarkers. A further aim is to establish the proportions of these patients with normal or abnormal biomarkers based on conventional cut-offs.

### **5.2.1 Objectives**

- To calculate the AST/ALT ratio, APRI score, and FIB-4 index from existing data within patient records where available for patients older than 40 years who did not have a previous diagnosis of advanced CLD.
- To estimate the proportion of patients with (normal or abnormal) AST/ALT ratios, APRI scores, and FIB-4 indexes based on conventional cut-offs.

## **5.3 Methods**

### **5.3.1 Study design**

A retrospective cohort study was conducted using CPRD data.

### **5.3.2 Data source**

CPRD data were prepared for use in this study and approved by ISAC under protocol number 17\_235. The data source for this study has been fully described in Chapter 4.

The population from the CPRD GOLD database (October 2018) comprised all cases eligible for data linkage to NCRAS,

ONS, and HES. In defining our cohort within NCRAS, Patients are identified in the CPRD GOLD database during the coverage period from 1 Jan 1990 – 31 Dec 2016. Additionally, I obtained data from the ONS's registered deaths from 2 Jan 1998 to 31 Dec 2016 and from CPRD practices in England linked to HES from 1 Jan 1997 to 31 Dec 2016.

### **5.3.3 Study population**

The study population are patients aged 40 years or older during the study period (1 Jan 1990 – 31 Dec 2016) and registered for at least one year on CPRD after the latest of either the date of registration or the date the practice was labelled UTS. The included patients are those among this population who have an AST measurement whilst over age 40 and did not have known advanced CLD, as defined in the next chapter, prior to the first AST measurement.

The sample was limited to those over age 40 primarily because it is unlikely enough time has passed for cirrhosis to have developed in younger patients, making this age group more relevant for the study. Additionally, the limits upon the number of cases that could be drawn from CPRD prevented all cases from being utilized. Age over 40 is the age group within which most outcomes occur, and hence this choice maximized power for chapters 6, 7, and 8 within the permitted data limits.

The study population's characteristics (birth, sex, UTS date, and death date 'if applicable') were extracted using patients' files from CPRD GOLD.

#### **5.3.4 Data processing**

I processed the data to isolate reliable and accurate blood marker values for AST, ALT, and Platelets. I applied a thorough data cleaning process to the records extracted from the entity files, to enhance the integrity and usability of this data, Data cleaning, which is also referred to as data cleansing or scrubbing, entails rigorously identifying and correcting or removing any inaccuracies and inconsistencies (236).

This involved extracting all relevant test results recorded during the study period from general practice sources, as documented in the CPRD GOLD's 'test' files. Specifically, these included results for ALT (entity file code 155), AST (entity file code 156), and platelet count (entity file code 189).

The raw data from the CPRD GOLD database, which is a large collection of test results, often needs careful cleaning and checking to make it ready for detailed analysis. These included the removal of records with missing or incomplete values and the identification and exclusion of potential outliers or anomalous reading. This step is important to ensure the data's quality and

reliability for analysis. Detailed steps and methods used in the data cleaning process are outlined in the following sections.

#### **5.3.4.1 AST and ALT**

AST and ALT had the same data cleaning steps. Firstly, all records under 155 and 156 entities were extracted from the patient test files. Then, any missed results with no values were removed from the recorded observations.

Subsequently, further cleaning was carried out to ensure the accuracy of the data. This involved removing all results that had decimals (because AST and ALT are expressed as integers without decimal fractions) and all readings that were higher than 1500 U/L. The value of 1500 U/L was more than four times the level of 99% of the data and, therefore, likely an erroneous value.

#### **5.3.4.2 Platelets**

Firstly, all records under entity 189, which included both platelet counts and platelet volumes, were extracted. Following this, I used medcodes 7 (platelet count), 4415 (platelet count abnormal), 26926 (platelet count NOS), and 26927 (platelet count normal) to extract only platelet count records within entity 189. Then, all observations with missed values (i.e., records with no value) and those containing decimals were removed (because

platelet count is expressed as an integer without a decimal fraction).

Based on data observation, some platelet counts had decimal values, which is unusual and likely indicates overlap with platelet volume measurements, possibly because platelet count and platelet volume are stored under the same entity number. So, further data cleaning was needed. As a result, values of less than  $15 \times 10^9/L$  were removed during cleaning to prevent overlap with platelet volume measurements. The value of 15 was chosen because it represents the highest level of normal platelet volume. Values greater than  $2000 \times 10^9/L$  were also removed during the cleaning process as they were deemed implausible and likely erroneous. This decision was based on the fact that test results higher than  $2000 \times 10^9/L$  were four times higher than 99% of the data, making them potentially indicative of measurement or data entry errors.

### 5.3.5 Biomarkers calculation

- **AST/ALT** was calculated as  $AST(IU/L)/ALT(IU/L)$ ,
- **APRI** was calculated as  $[[AST(IU/L)/upper\ limit\ normal] / platelets \times (10^9/L)] \times 100$ , and
- **FIB-4** was calculated as  $[age(years) \times AST(IU/L)] / [platelets \times (10^9/L) \times \sqrt{ALT(IU/L)}]$ .

### **5.3.6 Outcomes**

Each biomarker was assessed in a separate cohort study. The normal and abnormal groups were defined using cut-offs for each index, which were chosen based on existing literature. A cut-off of one was used for the AST/ALT ratio (139), a cut-off of one for the APRI score (151), and a cut-off of 1.3 for the Fib4 index (158). More detail on the validity of different cut-offs has been discussed in the introduction chapter (section [1.8.4](#)).

### **5.3.7 Determining patients with normal and abnormal scores**

Patients were included in the normal group when they had no history of abnormal scores. In contrast, patients were included in the abnormal group when they had at least one abnormal score.

### **5.3.8 Statistical analysis**

#### **5.3.8.1 Determining the acceptable period between the tests**

To determine the most appropriate interval between the biomarker components, all the results of the tests undertaken for each patient were combined, and the durations between the tests were determined. Ideally, all components of a calculated score would be measured at the same time. If the time interval between



the biomarker components had been extended, it could have potentially allowed for more individuals to be included in the study. On the other hand, reducing the time between tests will likely improve the validity of the biomarkers used in the study.

I divided the time between tests into three categories. The first category included all possible combinations without a time limit. The second category was from zero to ten days, and the third was from zero to one day.

#### **5.3.8.2 Proportion calculations**

The frequency of liver biomarkers was presented for each of the three biomarkers as follows:

- The proportion of patients with at least one calculable biomarker (for each biomarker) out of the total eligible population (Unique patients aged 40 years or older during the study period (01/01/1990–31/12/2016)) over the study period was calculated as follows:

$$\frac{\text{All patients with a score measured}}{\text{Total eligible population over the study period}}$$

- The proportion of patients with an abnormal value out of the total eligible population over the study period for each biomarker was calculated as follows:

$$\frac{\text{All patients with an abnormal score measured}}{\text{Total eligible population over the study period for each biomarker}}$$

Analyses were repeated with the patients stratified by sex (male, female) and age (at the date of biomarker measurement) based on ten-year age bands (i.e., 40–49 years, 50–59, 60–69, and more than 70 years old).

#### **5.3.8.3 Annual distribution of calculable scores.**

To determine the annual availability of calculable scores for my study population throughout the study period, I calculated the annual distribution of scores for the AST/ALT ratio, APRI score, and FIB-4 index. Additionally, I documented the annual frequency of conducted AST, ALT, and platelet tests during the study period.

#### **5.3.8.4 Age sex interaction**

I conducted a logistic regression analysis to assess the impact of age and gender on the likelihood of being classified into the normal and abnormal groups. The model incorporated age as a continuous variable, sex as a categorical variable, and an interaction term between age and gender to explore potential differences in the age effect on being classified into the normal and abnormal groups status by sex.

The model's fit was evaluated using the likelihood ratio chi-square test. The significance of the predictors, including the interaction term, was determined through Wald z-tests. Odds ratios (ORs) with 95% (CIs) were computed to elucidate the

influence of age, gender, and their interaction on the odds of exposure. A P-value less than 0.05 was deemed to indicate statistical significance.

## **5.4 Results**

### **5.4.1 Characteristics of the participants**

Figure 5-1 shows the process of extracting the study population from the source population (CPRD) and the number of included patients for the analysis. From the source population, 15,081,817 patients were registered between 1 Jan 1990 – 31 Dec 2016, and 11,083,422 patients had at least one year of UTS follow-up.

During the study period, 3,513,680 (23.3%) unique patients were aged 40 years or older at the end of the study period. This does include anyone who died or transferred before 2016, provided that they were at least 40 years old at the time of death or transfer and were eligible to be included in the study. A total of 2,946,250 (19.5%) patients were excluded because they had no AST measured in their records or they had advanced CLD before the date of the first AST measurement. The data provided from the CPRD combines the number of patients excluded for not having an AST measurement and those with pre-existing advanced CLD into a single figure.

Consequently, out of the 3,513,680 eligible patients, 567,430 (approximately 16.1%) were ultimately included in the study.

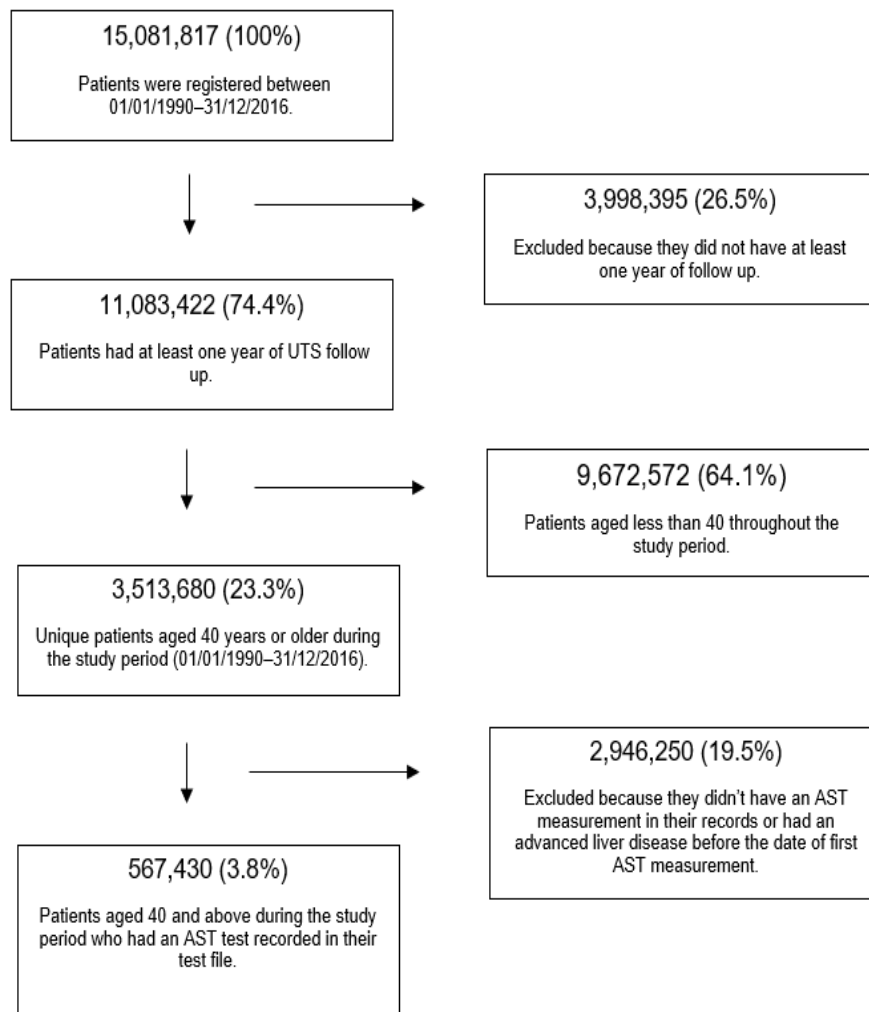


*Figure 5-1: The process of extracting the study population from the source population (CPRD) and the number of included patients for the analysis.*

### 5.4.2 Results of data processing

Platelet count was the most frequently measured blood marker, with 4,383,974 observations for 546,862 patients, followed by 2,702,579 ALT observations for 393,317 patients. For AST, in the initial dataset, there were a total of 567,430 patients over 40 years old with recorded AST measurements. However, following rigorous data processing and cleaning procedures, which included removing inaccuracies and inconsistencies, the final analysis was conducted on 564,367 patients with a total of 2,402,106 AST observations. The details of AST, ALT, and Platelet data processing results are shown in Figure 5-2, Figure 5-3, and Figure 5-4.

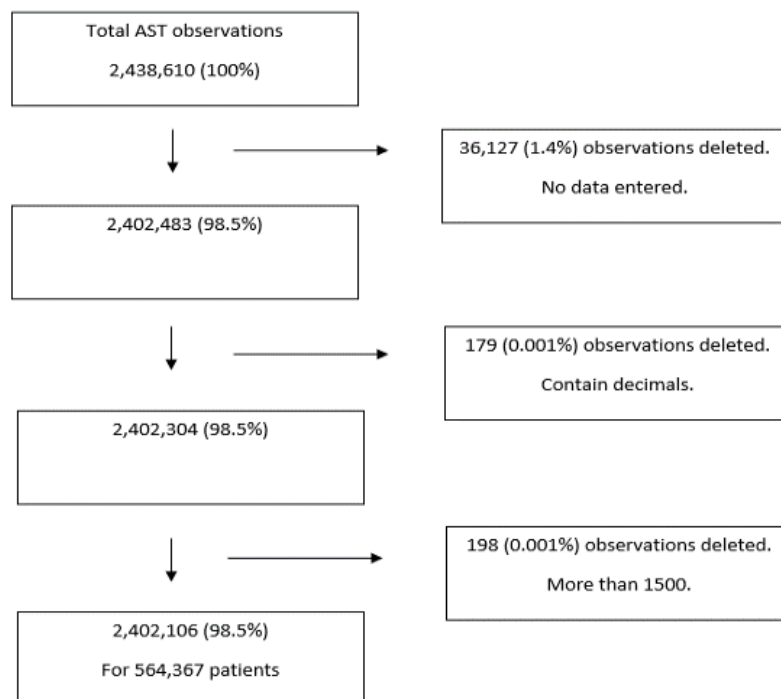


Figure 5-2: Results of data processing for AST from test files in CPRD.

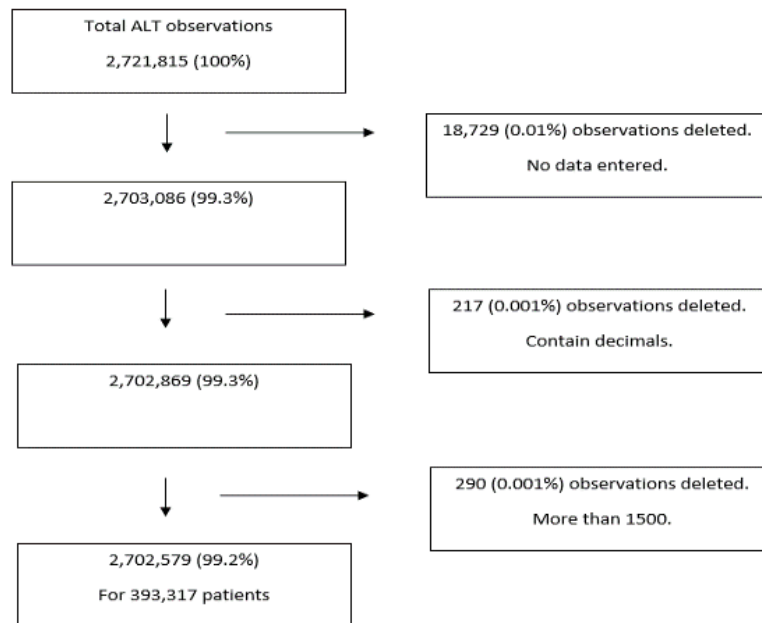


Figure 5-3: Results of data processing for ALT from test files in CPRD.

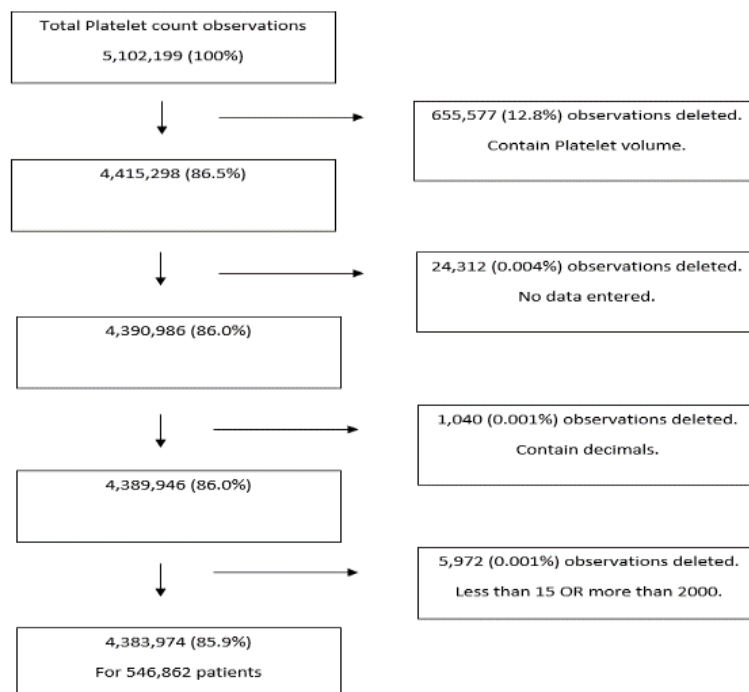


Figure 5-4: Results of data processing for Platelet count from test files in CPRD.

### 5.4.3 Time difference between the tests

A time difference between tests of one day covered 47.2%, 88.2%, and 87.4% of the all possible combinations for AST/ALT, APRI, and FIB-4 measurements, respectively. Extending the interval for including test results from 1 day to 10 days resulted in a marginal increase in the study population. Specifically, the initial 1-day interval captured 98% of the observations that would have been included with a 10-day separation. This minor addition to the study population did not justify the potential reduction in study validity. Instead, maintaining a 1-day interval significantly enhanced the validity of my findings. Accordingly, the maximum time between the tests included in the calculation of the scores was one day. More details are shown in Table 5-1, Table 5-2, and Table 5-3.

*Table 5-1: Number of patients over a different time between AST and ALT date (to calculate ast/alt):*

|                               |   | <b>All possible combinations</b> |  | <b>0-10 days</b> | <b>0-1 day</b> |
|-------------------------------|---|----------------------------------|--|------------------|----------------|
| <b>Number of observations</b> | n | 22,276,329                       |  | 865,847          | 821,572        |
|                               | % | 100%                             |  | 3.8%             | 3.6%           |
| <b>Number patients</b>        | n | 394,414                          |  | 190,600          | 186,504        |
|                               | % | 100%                             |  | 48.3%            | 47.2%          |

Table 5-2: Number of patients over a different time between AST and platelet date (to calculate APRI):

|                        |    |  | All possible combinations | 0-10 days | 0-1 day   |
|------------------------|----|--|---------------------------|-----------|-----------|
| Number of observations | n. |  | 42,283,729                | 1,847,382 | 1,700,555 |
|                        | %  |  | 100%                      | 4.3%      | 4.0%      |
| Number of patients     | n. |  | 547,304                   | 493,571   | 483,097   |
|                        | %  |  | 100%                      | 90.1%     | 88.2%     |

Table 5-3: Number of patients over a different time between AST and platelet date (when AST and ALT were done on the same day) to calculate Fib-4:

|                        |    |  | All possible combinations | 0-10 days | 0-1 day |
|------------------------|----|--|---------------------------|-----------|---------|
| Number of observations | n. |  | 15,524,407                | 654,528   | 601,267 |
|                        | %  |  | 100%                      | 4.2%      | 3.8%    |
| Number of patients     | n. |  | 180,517                   | 161,400   | 157,804 |
|                        | %  |  | 100%                      | 89.4%     | 87.4%   |

## 5.4.4 Score results

### 5.4.4.1 Patients with normal and abnormal AST/ALT ratios

After excluding those with advanced CLD at the time of the biomarker measurement, there were 821,572 calculable AST/ALT ratio scores from 186,504 patients. The proportion of all patients with at least one AST/ALT ratio among the study population (i.e., those aged 40 years or more, 3,513,680 patients) was 5.3%.



Of those with a valid AST/ALT ratio, 63.3% (n=118,039) had at least one abnormal score. Females were more likely than males to have abnormal scores (72.3 vs. 53.4%, respectively). The likelihood of being in the abnormal group increased with age, with the highest percentage among patients over 70 years old (84% abnormal vs. 48% abnormal in those aged 40-49 years). Full details of the numbers of patients in the normal and abnormal groups stratified by sex and age are shown in Table 5-4.

Table 5-4: Characteristics of participants (with at least 1 score measured) in the normal and abnormal groups stratified by sex and age group:

|                                              |   | AST/ALT |          | APRI    |          | FIB-4   |          |
|----------------------------------------------|---|---------|----------|---------|----------|---------|----------|
| n. of patients                               |   | 186,504 |          | 483,097 |          | 157,804 |          |
|                                              |   | Normal  | Abnormal | Normal  | Abnormal | Normal  | Abnormal |
| n. of patients in normal and abnormal groups | N | 68,465  | 118,039  | 464,204 | 18,893   | 77,590  | 80,214   |
|                                              | % | 36.7 %  | 63.3%    | 96.1%   | 3.9%     | 49.2%   | 50.8%    |
| <b>Sex</b>                                   |   |         |          |         |          |         |          |
| <b>Male</b>                                  | N | 41,169  | 47,135   | 209,498 | 10,810   | 33,712  | 39,177   |
|                                              | % | 46.6 %  | 53.4%    | 95.1%   | 4.9%     | 46.3%   | 53.7%    |
| <b>Female</b>                                | N | 27,296  | 70,904   | 254,706 | 8,083    | 43,878  | 41,037   |
|                                              | % | 27.7%   | 72.3%    | 96.9%   | 3.1%     | 51.7%   | 48.3%    |
| <b>Age group</b>                             |   |         |          |         |          |         |          |
| <b>40-49</b>                                 | N | 26,339  | 24,334   | 131,369 | 4,114    | 35,579  | 6,093    |
|                                              | % | 52.0%   | 48.0%    | 97.0%   | 3.0%     | 85.4%   | 14.6%    |
| <b>50-59</b>                                 | N | 21,321  | 25,246   | 113,898 | 4,792    | 23,345  | 15,094   |
|                                              | % | 45.8%   | 54.2%    | 95.9%   | 4.1%     | 60.7%   | 39.3%    |
| <b>60-69</b>                                 | N | 12,908  | 27,885   | 96,581  | 4,216    | 11,849  | 22,566   |
|                                              | % | 31.6%   | 68.4%    | 95.8%   | 4.2%     | 34.5%   | 65.5%    |
| <b>70 years and older</b>                    | N | 7,897   | 40,574   | 122,356 | 5,771    | 6,817   | 36,461   |
|                                              | % | 16.3%   | 83.7%    | 95.5%   | 4.5%     | 15.7%   | 84.3%    |

#### 5.4.4.2 Patients with normal and abnormal APRI scores

From a total of 483,097 patients, 1,700,555 APRI scores were analysed, accounting for 13.7% of the study cohort. A subset (3.9%) had high-risk values, with males showing a slightly higher prevalence (4.9% vs. 3.1% in females). Older age groups had a higher proportion of abnormal results. Details are available in Table 5-4.

#### **5.4.4.3 Patients with normal and abnormal FIB-4 indexes:**

I analysed 601,267 FIB-4 scores from 157,804 patients, which is 4.5% of our study population. Of these, 50.8% had abnormal results. There was a higher incidence of abnormal scores in males (53.7%) compared to females (48.3%). Abnormalities were notably more common in older age groups, with 84.3% of those over 70 years displaying abnormal scores, versus 14.6% in the 40–49-year range. Details are available in Table 5-4.

#### **5.4.4.4 Annual distribution of calculable scores.**

My study documented a consistent trend in the annual calculation of AST scores, AST/ALT ratios, APRI, and FIB-4 from 1990 to 2016, as shown in Figure 5-5 and Figure 5-6. Initially, AST scores were low, rising to peak in 2004, then declining notably. A similar trend was observed in AST/ALT ratios, APRI, and FIB-4, all peaking around 2004 and declining by 2010. This pattern underscores the pivotal role of AST in liver health assessments and indicates that increases in AST testing paralleled the availability of calculable liver health scores. Detailed annual distributions can be found in Table 5-5.

Table 5-5: Annual distribution of AST, ALT, platelet, and calculable scores.

|              | <b>AST</b> | <b>ALT</b> | <b>Platelet</b> | <b>AST/ALT</b> | <b>APRI</b> | <b>FIB4</b> |
|--------------|------------|------------|-----------------|----------------|-------------|-------------|
| <b>1990</b>  | 230        | 92         | 722             | 130            | 93          | 71          |
| <b>1991</b>  | 682        | 246        | 2088            | 347            | 370         | 101         |
| <b>1992</b>  | 2211       | 963        | 5824            | 768            | 1248        | 264         |
| <b>1993</b>  | 3464       | 1535       | 8429            | 1164           | 1945        | 452         |
| <b>1994</b>  | 3274       | 1587       | 8077            | 1063           | 2018        | 474         |
| <b>1995</b>  | 5620       | 2225       | 11998           | 1196           | 3666        | 569         |
| <b>1996</b>  | 11501      | 4347       | 21431           | 2613           | 7818        | 1589        |
| <b>1997</b>  | 22362      | 10457      | 38446           | 6159           | 14249       | 3830        |
| <b>1998</b>  | 36866      | 17513      | 58468           | 10981          | 25189       | 7816        |
| <b>1999</b>  | 64244      | 25411      | 90440           | 15664          | 45489       | 11357       |
| <b>2000</b>  | 92816      | 38961      | 136465          | 22376          | 63111       | 14950       |
| <b>2001</b>  | 116454     | 64510      | 167364          | 34030          | 76850       | 22562       |
| <b>2002</b>  | 164354     | 89019      | 212703          | 41724          | 110577      | 26412       |
| <b>2003</b>  | 205244     | 124854     | 257713          | 60528          | 133619      | 38875       |
| <b>2004</b>  | 203990     | 154821     | 275359          | 67542          | 132877      | 44301       |
| <b>2005</b>  | 200222     | 179377     | 293331          | 69294          | 131001      | 46560       |
| <b>2006</b>  | 173780     | 209427     | 302770          | 64799          | 117433      | 43869       |
| <b>2007</b>  | 177764     | 217437     | 325339          | 67423          | 125185      | 48329       |
| <b>2008</b>  | 179,487    | 219,196    | 335,143         | 70,792         | 131,359     | 51,951      |
| <b>2009</b>  | 164865     | 219263     | 324846          | 66634          | 124394      | 56755       |
| <b>2010</b>  | 156611     | 200380     | 311123          | 54232          | 121561      | 42770       |
| <b>2011</b>  | 132847     | 207609     | 308663          | 50305          | 106606      | 41320       |
| <b>2012</b>  | 100673     | 200063     | 276400          | 39001          | 78776       | 32907       |
| <b>2013</b>  | 78537      | 165248     | 227369          | 30328          | 61562       | 25396       |
| <b>2014</b>  | 60218      | 141373     | 184862          | 26675          | 48502       | 23001       |
| <b>2015</b>  | 29330      | 94833      | 116844          | 12343          | 23922       | 11174       |
| <b>2016</b>  | 14460      | 71445      | 81756           | 3461           | 11135       | 3312        |
| <b>Total</b> | 2,402,106  | 2,702,579  | 4,383,974       | 821,572        | 1,700,555   | 601,267     |

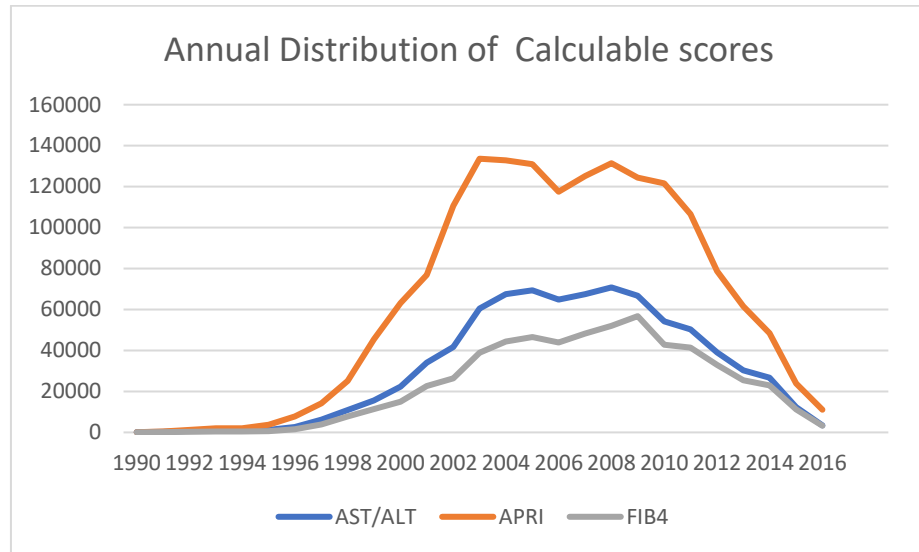


Figure 5-5: Annual Distribution of calculable scores.

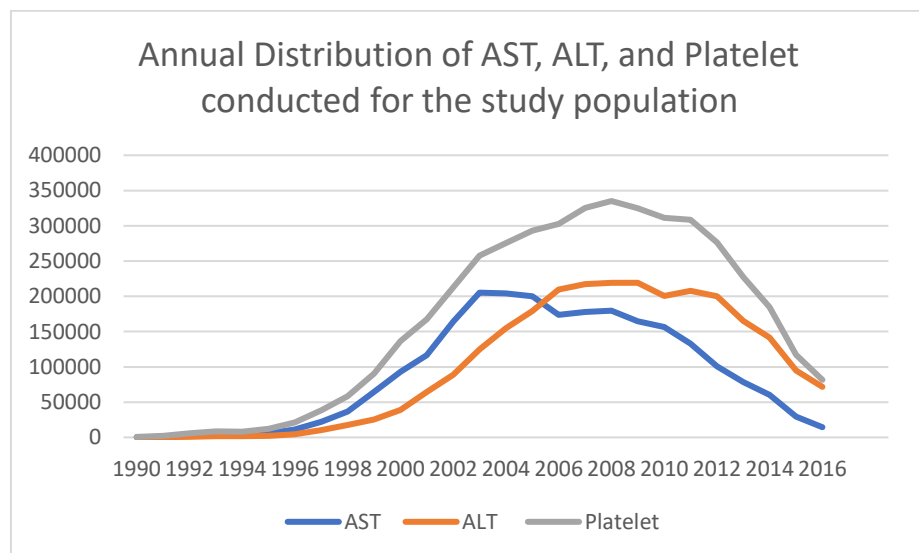


Figure 5-6: Annual Distribution of AST, ALT, and Platelet conducted for the study population.

#### 5.4.4.5 Age sex interaction

##### 5.4.4.5.1 AST/ALT ratio analysis

My analysis revealed a significant correlation between age and an increased risk of being classified into the abnormal group, with the log odds of exposure rising by approximately 0.06 for every additional year of age ( $p < 0.001$ ). Gender also played a critical role, with females showing significantly higher log odds of being classified into the abnormal group (an increase of 2.2) compared to males ( $p < 0.001$ ). However, this increased risk was moderated by the interaction between age and female gender, resulting in a decrease in log odds by 0.023 for each additional year ( $p < 0.001$ ), suggesting that the increase in the risk of being classified into the abnormal group with age is less pronounced in females. The results of the regression analysis are presented in Table 5-6.

*Table 5-6: Regression analysis demonstrating the interaction effect of sex and age on the likelihood of being classified into the abnormal group among AST/ALT population.*

| Variables        | log odds | Std. err | z<br>score | P> z  | (95% conf.<br>interval) |
|------------------|----------|----------|------------|-------|-------------------------|
| age              | 0.06     | 0.001    | 201.34     | 0.001 | (0.06, 0.06)            |
| Sex-female       | 2.2      | 0.026    | 84.93      | 0.001 | (2.2, 2.3)              |
| <b>Sex * age</b> |          |          |            |       |                         |
| Female           | -0.023   | 0.0004   | -59.84     | 0.001 | (-0.024-, -0.023)       |

#### 5.4.4.5.2 APRI score analysis

Age plays a significant role in the APRI score analysis, with each additional year increasing the log odds of being classified into the abnormal group by 0.01 ( $p < 0.001$ ). Female gender was associated with a decrease in the log odds by 1.2. The interaction between sex and age further impacted this trend, with an increase in the log odds by 0.008 per year for females, indicating that the risk associated with age is slightly more pronounced in females (Table 5-7).

*Table 5-7: Regression analysis demonstrating the interaction effect of sex and age on the likelihood of being classified into the abnormal group among the APRI score population.*

| Variables  | log odds | Std. err | z<br>score | P> z  | (95% conf.<br>interval) |
|------------|----------|----------|------------|-------|-------------------------|
| age        | 0.01     | 0.001    | -26.80     | 0.001 | (-.016 -.014)           |
| Sex-female | -1.15    | 0.05     | -21.23     | 0.001 | (-1.2 -1.0)             |
| Sex * age  |          |          |            |       |                         |
| Female     | 0.008    | 0.001    | 10.49      | 0.001 | (0.007 0.01)            |

#### 5.4.4.5.3 FIB-4 index analysis

Age was significantly correlated with an increased risk of being classified into the abnormal group, raising log odds by 0.1 per year ( $p < 0.001$ ). Female gender slightly reduced the risk, decreasing log odds by 0.05. The age-sex interaction led to a further reduction of 0.007 in log odds per year (Table 5-8).

Table 5-8: Regression analysis demonstrating the interaction effect of sex and age on the likelihood of being classified into the abnormal group among the Fib4 index population.

| Variables  | log odds | Std. err | z<br>score | P> z  | (95% conf.<br>interval) |
|------------|----------|----------|------------|-------|-------------------------|
| age        | 0.1      | 0.001    | 233.91     | 0.001 | (0.1, 0.1)              |
| Sex-female | -0.05    | 0.03     | -1.31      | 0.001 | (-0.12, 0.02)           |
| Sex * age  |          |          |            |       |                         |
| Female     | -0.007   | 0.001    | -13.87     | 0.001 | (-0.008, -<br>0.006)    |

## 5.5 Discussion

### 5.5.1 Availability of calculable non-invasive biomarkers among general practice populations

This study shows that, of those assessed, the most commonly calculable liver marker was APRI score in approximately 14% of those aged over 40 years attending UK general practice. The least calculable biomarker was the FIB-4 index (4.5%). Using the conventional cutoffs, despite APRI being the most readily calculable marker, it was found to yield the lowest proportion of abnormal scores. There was wide variation in the proportion of abnormal scores across the three biomarkers: 63.3% for the AST/ALT ratio, 50.8% for the FIB-4 index, and 3.9% for the APRI score.

This thesis explores the feasibility of using non-invasive biomarkers like the AST/ALT ratio, APRI score, and FIB-4 index for identifying patients at higher risk of advanced liver diseases.



The findings of this study are enlightening as they demonstrate that these biomarkers can be calculated from existing patient records in GP, emphasizing their accessibility and practical utility in clinical settings.

The lower calculability of the FIB-4 index along with the AST/ALT ratio in compare to APRI score, which may be the result of less frequent recording of the required parameters in routine patient evaluations, raises concerns about potential issues with its routine application.

#### **5.5.2 Variability in liver function testing practices in primary care; implications for non-Invasive biomarker calculability**

The inequality in the scores' measurements can be attributed to the variable approaches in conducting LFTs across different primary care settings.

The current study found that a significant variation was observed in the use of ALT and AST tests when analysing LFT practices in primary care in the UK. After thorough data processing and analysis, my study findings revealed that a total of 2,402,106 AST tests were given to a larger group of 564,367 patients. However, within this group, a specific subset of 393,317 (69.7%) patients not only had AST tests but also underwent ALT

testing. In this particular subset, a substantial number of ALT tests, totalling 2,702,579, were conducted.

So, as I am describing routine data, the observations reflect variations in practice and not just in the performance of the biomarkers. APRI's higher calculability, requiring only AST, contrasts with other indices that need both AST and ALT. This finding aligns with Scutt et al. (137), who noted an increase in most liver indicators over the last decade but a significant drop in AST testing in primary care—only about 2% of patients receive annual AST tests. Since AST is essential for calculating AST/ALT ratios, APRI scores, and FIB-4 indices, reduced testing impacts these biomarker measurements significantly.

It's important to recognize that simply increasing AST testing doesn't automatically enhance biomarker calculability, given that the AST/ALT ratio and FIB-4 index also rely on ALT values, and both the APRI and FIB-4 indices require platelet counts. Integrating these tests could greatly reduce the current variability in calculable scores caused by diverse testing practices. This comprehensive approach would provide a more accurate assessment of liver health and align with our study's goal to effectively use non-invasive biomarkers in primary care for liver disease detection.

Furthermore, the observed differences in AST and ALT testing frequencies in this study are not contradictory to the literature but rather indicative of the study's focus on patients with measured AST levels. While ALT is more specific to hepatic conditions (237), AST could potentially be a more sensitive marker for chronic liver damage (238, 239). Consequently, the study's results may be more representative of this specific patient subgroup rather than the broader primary care population.

The annual distribution of calculable scores observed in my study lends support to the findings presented by Scutt et al. (137), as the reduction in AST calculation directly impacts the number of calculable scores, corroborating my observations that the ability to calculate the AST/ALT ratio, APRI, and FIB-4 has diminished over time.

It is important to note that the plots display raw counts for time trends rather than rates, as we did not have access to the denominator population for each year of the study. While I do have the denominator for the overall study population, the lack of annual denominators means that changes between years might reflect shifts in the underlying population over time. This means that changes between years might reflect shifts in the underlying population over time, such as an initial increase followed by a decrease as practices left the CPRD Gold database. Despite this

limitation, comparisons between the tests within the same year remain valid because they were conducted within the same populations. So, these raw counts should be interpreted with caution, recognizing that they do not account for population changes over time. Future studies should aim to access denominator data to calculate rates and provide a more accurate representation of time trends.

### **5.5.3 Gender and age disparities in liver function testing**

Another area in which findings may reflect clinical practice rather than the performance of the tests is in the variations between ages and sexes regarding the proportions of patients tested. Females predominated over males in numbers with scores calculable for each of the three indices may reflect the fact that females are known to consult their general practice services more than males and hence have an increased opportunity for testing (240, 241).

This is also supported by the finding from Scutt et al. (137), study who showed that females generally undergo AST and ALT testing more frequently than males. It is possible that this is enough to counter the knowledge that liver disease is more

common in males than females (242, 243) since otherwise targeted testing of males may be expected.

Scutt et al. (137) reported that females are more likely to have normal AST and ALT levels compared to males, which can be attributed to physiological differences such as hormonal variations, lower average muscle mass, and different fat distributions in females. These factors might influence baseline ALT levels. Additionally, since AST is present in various tissues beyond the liver, including muscle and heart, this finding suggests gender-specific differences in factors like muscle mass, cardiac health, and other systemic conditions affecting AST levels (244, 245).

Contrary to this, my study findings show that females more frequently have abnormal scores in the AST/ALT ratio, while males are more likely to exhibit abnormal scores in the APRI and FIB-4. This discrepancy adds complexity to the interpretation of liver enzyme tests and biomarkers, highlighting that while females generally present with normal AST and ALT levels, they might still have abnormalities in specific liver function ratios and scores. This underlines the importance of a nuanced approach in assessing liver health, considering gender-specific physiological and systemic factors.

A similar phenomenon may be occurring in the elderly. Since about 90% of liver disease deaths in the UK occur under age 70 (246), one might expect that general practitioners would conduct fewer tests of liver function in this age group. That similar numbers of scores are calculable in the over 70s may reflect the fact that LFTs are not measured purely to assess the liver and that older individuals tend to have more healthcare contact (247, 248).

My study identified an increased probability of presenting abnormal liver biomarker levels in patients over 70 years old, diverging from the pattern observed by Scutt et al. (137), who found that normal ALT and AST levels generally rise with age, with the proportion of the population undergoing testing peaking at ages 70-79 before slightly declining. Notably the peak age for abnormal test results was identified as 50-59. This difference underlines a crucial aspect: the calculated biomarkers, rather than AST levels, guide the selection of individuals for further investigation. This distinction supports the subsequent discussion that liver enzymes and calculated scores may not identify the same individuals for additional diagnostic procedures, highlighting the importance of context in interpreting these biomarkers.

As individuals age, the metabolic demands on the liver may decrease, accompanied by a reduction in cellular turnover. These changes could contribute to the stabilization of liver enzyme levels, resulting in more frequent 'normal' ALT and AST levels in older populations (249, 250). However, it's important to interpret these 'normal' levels cautiously. In older adults, even 'normal' enzyme levels may not accurately reflect the liver's functional capacity, as aging is associated with a general decline in organ efficiency and regenerative ability (250). Therefore, the increased likelihood of abnormal biomarker results in older patients in our study might actually reflect a more accurate indication of liver health in this age group.

#### **5.5.3.1 Age sex interaction.**

Age-sex interaction refers to the differential impact that age has on a health outcome based on sex, and vice versa. Given the variability in liver disease epidemiology across ages and sexes, examining this interaction allows for a more refined understanding of disease risk, potentially leading to improved case findings and the personalized treatment approaches. My analysis revealed a significant correlation between age and an increased risk of liver disease, as indicated by all three biomarkers. However, this risk was modulated by sex, with females experiencing a different trajectory of risk increase with

age compared to males. Specifically, while females showed a baseline higher risk in some biomarkers, the age-sex interaction indicated that the rate of risk increase with age is less pronounced for females.

These findings align with the literature, notably the comprehensive review by Cooper et al.(251), which underscores sex differences in chronic liver disease and the protective effects of oestrogens against hepatic fibrosis. Their review elaborates on how sex hormones contribute to the pathophysiology of liver diseases, with oestrogen having a direct inhibitory effect on HSCs and thereby modulating fibrosis. This is consistent with my observation that the increase in liver disease risk with age is moderated in females, likely due to oestrogen's protective roles. Cooper et al. also discuss the increased risk of fibrosis in post-menopausal women, aligning with my findings that suggest an increased risk of being in the abnormal group with age among females. This comparison emphasizes the critical need for considering age-sex interactions in liver disease research and clinical practice, advocating for personalized approaches to liver health assessment and management.



#### **5.5.4 Implications of biomarker utilization in case finding and risk stratification.**

The EASL-EASD-EASO clinical practice guidelines propose using the FIB-4 score as part of the diagnostic protocol for ruling out advanced fibrosis (252). The recommendations also suggest using FIB-4 as a prognostic indicator to exclude progression to severe liver disease and death. If we were to implement widespread FIB-4 testing among the UK population aged over 40 years, this could have a substantial impact on general practice, laboratory and hepatology services.

Furthermore, though some measures of diagnostic accuracy, such as specificity and sensitivity, are generally thought of as being characteristics purely of a test, others, such as PPV, are also affected by the population tested. In a low-risk population, the PPV of a FIB-4 of 1.3, may be as low as 5% (253). This means that out of the 17 million abnormal results, only 5% would be true positives. Consequently, with a PPV of 5%, we would expect only 850,000 of these 17 million abnormal results to accurately indicate the presence of the condition, leading to a large number of false positives and potentially unnecessary further investigations. Therefore, any approach needs to target those at the highest risk of disease without compromising the potential to detect cases or intervene to improve outcomes.

The extrapolation of abnormal FIB-4 tests results from a study population primarily consisting of high-risk individuals to the general UK population over 40 years of age requires careful consideration. The study cohort we have used may not be representative of the broader population since general practitioners are likely to have to some degree targeted the testing of blood for AST and other LFTs. Hence the proportion of the general population over 40 who would have abnormal results is likely to be lower.

This difference in test outcomes between high-risk individuals and the general population highlights the importance of targeted testing. Screening the entire population over 40 years might not be as effective or efficient as continuing the current practice of opportunistic case finding approach. Such a targeted approach ensures that resources are efficiently used, focusing on those most likely to benefit from early detection and intervention, thereby optimizing healthcare outcomes without overwhelming the system with low-yield screenings.

Another consideration is the use of such biomarkers for surveillance rather than diagnosis. Recent studies have shown that serial measurements of FIB-4 over time can be useful in monitoring disease progression and identifying patients at higher risk of developing liver complications such as cirrhosis or HCC

(129). By tracking changes in FIB-4 over time, clinicians can identify patients who may require closer surveillance or more aggressive treatment to prevent or delay the development of liver complications. This approach could help to improve risk stratification and guide personalised management strategies for patients with chronic viral hepatitis (254).

So, the surveillance potential of biomarkers, particularly FIB-4, enhances the scope of our study. While our main focus is on the feasibility of using non-invasive biomarkers for identifying advanced liver diseases in primary care, understanding their role in ongoing disease monitoring offers valuable insights. This broader view of biomarker applications, beyond initial diagnosis to include disease progression surveillance, aligns with the overall objective of improving liver disease management in primary care settings.

#### **5.5.5 Source of the data to calculate the Biomarkers.**

Several studies have looked at the role primary care plays in managing liver disease, emphasising how important it is for early disease intervention and management (88, 163) as well as its cost-effectiveness (255). However, Jarvis et al. (256), have emphasized that despite the established advantages, routine primary care practices do not fully integrate liver disease pathways.

The conduct of my study in primary care data allows me to examine further the potential for such interventions, focusing on a patient population at a stage where liver disease might not yet be evident, but where early detection can have a significant impact.

Moreover, Pryke and Guha, in their article (165), highlighted the benefits of integrating non-invasive biomarkers into primary care for liver disease management, suggesting their significant role in improving patient care and managing associated conditions. These biomarkers, as they note, enhance patient care in primary care settings without substantially increasing the workload of healthcare professionals, demonstrating their efficiency and effectiveness.

My research offers an innovative viewpoint by proving that a significant amount of data is available in primary care to calculate non-invasive liver biomarkers even in the absence of an active case-finding strategy. This is a noteworthy finding because it implies that primary care settings in the UK already have the data needed to use non-invasive biomarkers for liver disease in an efficient manner. However, it is important to note that this data is primarily available for a subset of the population who have had blood tests, specifically AST tests, as part of their clinical care.

The population undergoing AST/ALT testing is likely to differ from the general UK population. Typically, these individuals are those at higher risk of liver disease. General practitioners GPs are more likely to check LFTs in patients in whom they either suspect liver disease or in whom clinical guidelines suggest testing. This includes patients with conditions such as obesity, diabetes, alcohol use disorder, or other metabolic conditions. Additionally, patients presenting with symptoms indicative of liver dysfunction or those being monitored for known liver conditions are more likely to undergo these tests.

Demographically, the population getting AST/ALT testing may have a higher prevalence of CLD and associated risk factors compared to the general population. They may also differ in age, with older individuals more likely to be tested, and could have a higher representation of males, as men are often at greater risk for certain liver conditions. Socioeconomic status may also play a role, as those with better access to healthcare services are more likely to receive regular testing.

Therefore, while primary care settings do have the necessary data to use non-invasive biomarkers, this data primarily reflects a high-risk group rather than the general UK population. This distinction is crucial, as it suggests that while the findings demonstrate the availability of data within the CPRD,

they may not be entirely representative of the broader population's health status.

However, it is worth noting that the results of my study, which primarily came from a primary care setting, may have limited relevance in management of the advanced CLD within secondary care settings. As the disease progresses and requires more complex diagnostic and management strategies, the usefulness of these biomarkers may not be as significant. In advanced stages, which are characterized by various clinical symptoms and complications, a shift towards more comprehensive diagnostic approaches like liver biopsy becomes necessary. Grattagliano et al. (257), emphasize that the progression of liver disease and the increasing need for specialized care in advanced stages, my findings contribute a complementary perspective by demonstrating the availability and potential utility of non-invasive biomarkers in primary care. This is particularly significant given the rising prevalence of liver diseases and the need for early intervention strategies.

However, the integration of non-invasive biomarkers into referral pathways from primary to secondary care aligns with the main aim of my thesis. By effectively categorizing patients in primary care, these biomarkers have the potential to rationalize the referral process, thereby optimizing healthcare resources and

enhancing patient management in the continuum of care from primary to secondary settings.

As discussed in the introduction chapter (Section [1.9](#)), non-invasive biomarkers can be used in the referral pathways from primary to secondary care. The study by Srivastava et al (164) presented that the integration of non-invasive biomarkers into referral systems for liver diseases can bring about wide-ranging advantages. These advantages include financial savings and improved patient treatment, which align with the overall goals of many healthcare systems to provide efficient and patient-focused care. By accurately categorizing patients in primary care and minimizing unnecessary referrals to specialists, this approach highlights the potential of non-invasive biomarkers to optimize healthcare resources and enhance the management of liver diseases.

## **5.6 Conclusion**

The study presented in this chapter has shown that a considerable number of patients in UK general practice have calculable AST/ALT ratios, APRI scores, and FIB-4 indexes without a targeted case-finding approach. The results suggest that, although the APRI score was the most calculated among the study population, it had the lowest proportion of abnormal scores at conventional cut-offs.

The utilisation of data within general practice to explore the potential of such risk scores could be particularly advantageous for enhancing early case identification and improving disease management.



## **Chapter 6      Identifying Advanced CLD outcomes among the Study Population.**

## **6.1 Introduction**

The introduction chapter of this thesis explored the silent progression of NAFLD and its significant contribution to CLD. Given the difficulty of reliably identifying NAFLD in CPRD as discussed in (section [1.7.1](#)), this chapter focuses on the advanced CLD allowing for a comprehensive approach, targeting a range of liver illnesses and ensuring impactful research on detection and management of advanced liver diseases.

## **6.2 Study aim**

This study aims to identify all patients aged 40 years or older with at least one year of UTS data and with data allowing the calculation of at least one of AST/ALT ratio, APRI and FIB4-index who have been diagnosed with liver cirrhosis, HCC, liver transplantation, and liver death using CPRD and linked data.

## **6.3 Methods**

### **6.3.1 Study design**

Cohort study was conducted using prospectively collected clinical data.

### **6.3.2 Study population**

This study utilises the population described in Chapter 5 (section [5.3.3](#)). Within this population, I utilised the groups

identified in Chapter 5 who had data available to calculate AST/ALT ratios, APRI scores, and FIB-4 indexes. To this data, I added the linked NCRAS, HES, and ONS records available for the patient cohort.

### **6.3.3 Outcomes**

In this study, the primary advanced CLD outcome was liver cirrhosis. The secondary outcomes were HCC, Liver transplantation, Liver disease-related death, and any advanced CLD (all primary and secondary outcomes combined).

The number of patients exhibiting new outcomes was determined for each of the three cohorts under study. The following steps were undertaken to determine the primary and secondary outcomes from the data:

#### **6.3.3.1 Liver cirrhosis**

Cirrhosis was defined based on the presence of cirrhosis codes from CPRD, HES, and the death registry. The ICD codes used to identify cirrhosis cases in HES and the death registry were previously validated by Shearer et al. (254). A set of codes was mapped to these from the Medcodes available in CPRD. Additionally, more cirrhosis codes were ascertained according to the codes list approved in the ISAC form ([Appendix C](#)). All of

these codes are separately presented in ([Appendix D](#)). My final codes list is presented in Table 6-1 and Table 6-2.

As the data in CPRD were linked to the data in HES and the death registry, I combined and included the results from these three sources. The date of cirrhosis diagnosis was considered the date upon which the diagnosis was first coded.

*Table 6-1: Cirrhosis medcodes*

| <b>MedCode</b> | <b>Description</b>                                      |
|----------------|---------------------------------------------------------|
| <b>6863</b>    | Cirrhosis and chronic liver disease                     |
| <b>1638</b>    | Cirrhosis of liver NOS                                  |
| <b>16725</b>   | Cirrhosis - non alcoholic                               |
| <b>69204</b>   | Multilobular portal cirrhosis                           |
| <b>40567</b>   | Capsular portal cirrhosis                               |
| <b>26319</b>   | Oesophageal varices in cirrhosis of the liver           |
| <b>58630</b>   | Biliary cirrhosis NOS                                   |
| <b>68376</b>   | Florid cirrhosis                                        |
| <b>6015</b>    | [X]Other and unspecified cirrhosis of liver             |
| <b>55454</b>   | Portal cirrhosis unspecified                            |
| <b>47257</b>   | Portal cirrhosis                                        |
| <b>15424</b>   | Secondary biliary cirrhosis                             |
| <b>22841</b>   | Macronodular cirrhosis of liver                         |
| <b>44676</b>   | Fatty portal cirrhosis                                  |
| <b>8363</b>    | Oesophageal varices in alcoholic cirrhosis of the liver |
| <b>5638</b>    | Primary biliary cirrhosis                               |
| <b>92909</b>   | Hypertrophic portal cirrhosis                           |
| <b>4743</b>    | Alcoholic cirrhosis of liver                            |
| <b>16455</b>   | Non-alcoholic cirrhosis NOS                             |
| <b>3450</b>    | Diffuse nodular cirrhosis                               |
| <b>18739</b>   | Cryptogenic cirrhosis of liver                          |
| <b>9494</b>    | Biliary cirrhosis                                       |
| <b>24220</b>   | Open operations on oesophageal varices                  |
| <b>73139</b>   | Oesophageal varices without bleeding in diseases EC     |
| <b>10797</b>   | Oesophageal varices NOS                                 |
| <b>107975</b>  | Other specified open operation on oesophageal varices   |

|               |                                                               |
|---------------|---------------------------------------------------------------|
| <b>20912</b>  | Open injection sclerotherapy to oesophageal varices           |
| <b>20233</b>  | Open operation on oesophageal varices NOS                     |
| <b>30655</b>  | Oesophageal varices without bleeding                          |
| <b>11960</b>  | Fibreoptic endoscopic injection sclerotherapy oesoph varices  |
| <b>46647</b>  | Rigid oesophagoscopy banding of oesophageal varices           |
| <b>1641</b>   | Oesophageal varices                                           |
| <b>11972</b>  | Varices - other                                               |
| <b>89717</b>  | Fibreoptic endoscop rubber band ligation of upper GIT varices |
| <b>28929</b>  | Gastric varices                                               |
| <b>24989</b>  | Oesophageal varices with bleeding                             |
| <b>47214</b>  | Rigid oesophagoscopy injection sclerotherapy oesoph varices   |
| <b>105611</b> | [X]Oesophageal varices in diseases classified elsewhere       |
| <b>43404</b>  | Local ligation of oesophageal varices                         |
| <b>16759</b>  | Fibreoptic endoscopic banding of oesophageal varices          |
| <b>96756</b>  | Oesophageal varices with bleeding in diseases EC              |
| <b>44424</b>  | Oesophageal varices in diseases EC                            |
| <b>22411</b>  | Encephalopathy - hepatic                                      |
| <b>1508</b>   | [D]Ascites                                                    |
| <b>58736</b>  | O/E - ascites NOS                                             |
| <b>48138</b>  | Insertion of peritoneal to venous shunt for ascites           |
| <b>3503</b>   | Drainage of ascites NEC                                       |
| <b>44222</b>  | Peritoneal to venous drainage for ascites                     |
| <b>37930</b>  | [D]Ascites NOS                                                |
| <b>16113</b>  | O/E - ascites                                                 |
| <b>19438</b>  | Paracentesis abdominis for ascites                            |
| <b>52838</b>  | O/E -ascites-shifting dullness                                |
| <b>73305</b>  | O/E-ascites-fluid thrill shown                                |
| <b>8206</b>   | Pigmentary cirrhosis of liver                                 |
| <b>19512</b>  | Glycogenosis with hepatic cirrhosis                           |
| <b>27438</b>  | Cardiac portal cirrhosis                                      |
| <b>40963</b>  | Portal fibrosis without cirrhosis                             |
| <b>44120</b>  | Toxic liver disease with fibrosis and cirrhosis of liver      |
| <b>48928</b>  | Infectious cirrhosis NOS                                      |
| <b>55333</b>  | Sengstaken intubation of oesophagus                           |
| <b>62582</b>  | Oesophageal varices in diseases EC NOS                        |

Table 6-2: cirrhosis ICD10 codes

| ICD Code     | Description                                                           |
|--------------|-----------------------------------------------------------------------|
| <b>K70.3</b> | Alcoholic cirrhosis of liver                                          |
| <b>K74</b>   | Fibrosis and cirrhosis of liver                                       |
| <b>K74.6</b> | Other and unspecified cirrhosis of liver                              |
| <b>K76.6</b> | Portal hypertension                                                   |
| <b>K76.7</b> | Hepatorenal syndrome                                                  |
| <b>C22.0</b> | Liver cell carcinoma                                                  |
| <b>I85</b>   | Oesophageal varices                                                   |
| <b>I85.0</b> | Oesophageal varices with bleeding                                     |
| <b>I85.9</b> | Oesophageal varices without bleeding                                  |
| <b>I98</b>   | Oesophageal varices in diseases classified elsewhere                  |
| <b>I98.2</b> | Oesophageal varices without bleeding in diseases classified elsewhere |
| <b>I98.3</b> | Oesophageal varices with bleeding in diseases classified elsewhere    |
| <b>K72.9</b> | Hepatic failure, unspecified                                          |
| <b>R18</b>   | Ascites                                                               |

### 6.3.3.2 HCC

HCC was determined based on the presence of the HCC **C22.0** code from the cancer registry data. The NCRAS is recognized as a pivotal source of cancer data in the UK, meticulously collecting and analysing information on all diagnosed cancer cases (231). The diagnosis date for HCC was exclusively obtained from the cancer registry data.

### 6.3.3.3 Liver transplantation

Liver transplantation was included based on the presence of liver transplantation codes from CPRD, and liver transplantation **Z94.4** code from HES, and the death registry.

In CPRD, a liver transplantation case was defined as any patient with a Read code indicating liver transplantation in the clinical files. The list of Read codes used to identify liver transplantation was previously validated by Muller et al. (258), and the specific codes are provided in (Table 6-3).

As the data in CPRD were linked to the data in HES and the death registry, I combined and included the results from these three sources. The included date of liver transplantation was considered the date when the condition was first coded.

*Table 6-3: Medcodes and read terms for liver transplantation.*

| <b>medcode</b> | <b>Read Term</b>                          |
|----------------|-------------------------------------------|
| <b>4405</b>    | Transplantation of liver                  |
| <b>100073</b>  | Piggy back liver transplant               |
| <b>71422</b>   | Heterotopic transplantation of liver      |
| <b>32025</b>   | Orthotopic transplantation of liver       |
| <b>27319</b>   | Transplantation of liver                  |
| <b>89445</b>   | Auxillary liver transplant                |
| <b>105506</b>  | Orthotopic transplantation of whole liver |
| <b>97157</b>   | Orthotopic transplantation of liver       |
| <b>69194</b>   | Replacement of previous liver transplant  |
| <b>47861</b>   | Exploration of liver transplant           |
| <b>99250</b>   | Other specified transplantation of liver  |

#### **6.3.3.4 Liver death**

Liver death was included based on the presence of deaths where liver disease was coded as the underlying cause of death. The code list utilized to identify liver diseases is detailed in (Table 6-4) and was previously validated by Ratib et al. (259).

Table 6-4: ICD 10 codes for liver diseases.

| ICD 10 code | Term                                                         |
|-------------|--------------------------------------------------------------|
| K70.0       | Alcoholic fatty liver                                        |
| K70.1       | Alcoholic hepatitis                                          |
| K70.2       | Alcoholic fibrosis and sclerosis of liver                    |
| K70.3       | Alcoholic cirrhosis of liver                                 |
| K70.4       | Alcoholic hepatic failure                                    |
| K70.9       | Alcoholic liver disease, unspecified                         |
| K71.0       | Toxic liver disease with cholestasis                         |
| K71.1       | Toxic liver disease with hepatic necrosis                    |
| K71.2       | Toxic liver disease with acute hepatitis                     |
| K71.3       | Toxic liver disease with chronic persistent hepatitis        |
| K71.4       | Toxic liver disease with chronic lobular hepatitis           |
| K71.5       | Toxic liver disease with chronic active hepatitis            |
| K71.6       | Toxic liver disease with hepatitis, not elsewhere classified |
| K71.7       | Toxic liver disease with fibrosis and cirrhosis of liver     |
| K71.8       | Toxic liver disease with disorders of liver                  |
| K71.9       | Toxic liver disease with disorders of liver                  |
| K72.0       | Acute and subacute hepatic failure                           |
| K72.1       | Chronic hepatic failure                                      |
| K72.9       | Hepatic failure, unspecified                                 |
| K73         | Chronic hepatitis, not elsewhere classified                  |
| K74.0       | Hepatic fibrosis                                             |
| K74.1       | Hepatic sclerosis                                            |
| K74.2       | Hepatic fibrosis with hepatic sclerosis                      |
| 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                            |
| K76.0       | Other diseases of liver                                      |
| K76.1       | Chronic passive congestion of liver                          |
| K76.2       | Central haemorrhagic necrosis of liver                       |
| K76.3       | Infarction of liver                                          |
| K76.4       | Peliosis hepatis                                             |
| K76.5       | Hepatic veno-occlusive disease                               |
| K76.6       | Portal hypertension                                          |
| K76.7       | Hepatorenal syndrome                                         |
| K76.8       | Other specified diseases of liver                            |
| K76.9       | Liver disease, unspecified                                   |
| K77         | Liver disorders in diseases classified elsewhere             |



|              |                                                      |
|--------------|------------------------------------------------------|
| <b>I85.0</b> | Oesophageal varices with bleeding                    |
| <b>I85.9</b> | Oesophageal varices without bleeding                 |
| <b>I86.4</b> | Gastric varices                                      |
| <b>I98.2</b> | Oesophageal varices in diseases classified elsewhere |

#### **6.3.3.5 Any liver event (combined outcome)**

All primary and secondary outcomes mentioned earlier were combined into one (combined outcome). In the data analysis, the outcome date for each patient was determined as the date of the first occurrence of any individual outcomes.

#### **6.3.4 Stratification of the cohort based on potential predictors.**

At baseline, several key independent explanatory predictors (covariates) were identified to guide the stratification of participants. Participants were stratified first, by their sex, categorizing them as either male or female. Second, by age, with ages classified based on ten-year age bands taken at the diagnosis date, specifically 40–49 years, 50–59 years, 60–69 years, and 70 years and older.

Lastly, participants' socioeconomic status was determined by linking their data from the CPRD to the Index of Multiple Deprivation (IMD). The IMD is a composite measure derived from various indicators covering aspects of material deprivation, calculated at the Lower Layer Super Output Area (LSOA) level (260). The IMD decile rankings of participants' LSOA of residence

were used to categorize socioeconomic status in this study. Participants were divided into four groups reflecting their deprivation level, based on these deciles: 1-2 (least deprived), 3-5, 6-8, and 9-10 (most deprived). The decision to group IMD deciles into four broader categories was made for reasons similar to my age classification approach. This categorisation allowed retention of the ability to examine differences and trends across these factors whilst maintaining adequate sample size for rare outcomes.

### **6.3.5 Statistical analysis**

In chapter 5: Patients were classified into two groups for each subpopulation: those with abnormal scores (exposed group) and those with normal scores (control group). Subsequently, I identified all patients with the primary and secondary outcomes within the study population. Among these patients, P-values, calculated using the Pearson chi-square test, are used to assess the statistical significance of differences in the occurrence of the primary and secondary outcomes between each demographic group within the diagnosed group. The primary variables tested were gender (male vs. female), age groups, and socioeconomic status. A p-value < 0.001 indicates significant difference.

I calculated for each patient with an outcome their age at the diagnosis of this (as Outcome date – Date of birth in years)

and then calculated the mean of these ages. Finally, I calculated the median follow up time from when the abnormal biomarkers were calculated to the observed outcomes was determined for all biomarkers across all outcomes, including the calculation of the interquartile range (IQR).

## **6.4 Results**

### **6.4.1 Characteristics of patients with outcomes**

#### **6.4.1.1 Liver cirrhosis**

Of the study population, 2,135 patients were diagnosed with liver cirrhosis in CPRD. A total of 5,949 patients were coded with cirrhosis within HES data, and 838 patients had liver cirrhosis registered as their primary cause of death.

Among patients who were diagnosed with liver cirrhosis in the CPRD, 74.9% also had corresponding records in the HES database. Conversely, among those diagnosed with liver cirrhosis in the HES, 26.8% had documented records in the CPRD.

The total number of unique patients diagnosed with cirrhosis across all files was 6,676, comprising 4,102 men and 2,574 women, with men being significantly more affected than women. The mean age at diagnosis was 64.9 years (SD  $\pm$  12.1 years). Among patients stratified by socioeconomic status, the results showed that cirrhosis was more common in the most

deprived population. The characteristics of the patients diagnosed with cirrhosis are given in Table 6-5.

Table 6-5: Characteristics of patients with the Outcomes:

|                             |                                        | <b>Cirrhosis</b> | <b>HCC</b>     | <b>Liver transplantation</b> | <b>Liver death</b> | <b>Combined outcomes</b> | <b>*P value</b> |
|-----------------------------|----------------------------------------|------------------|----------------|------------------------------|--------------------|--------------------------|-----------------|
| <b>Number of patients</b>   |                                        | 6,676            | 476            | 188                          | 1,922              | 7,546                    |                 |
| <b>sex</b>                  | Males (n= 262,956)                     | 4,102<br>61.4 %  | 347<br>72.8%   | 118<br>62.8%                 | 1,185<br>61.6%     | 4,661<br>61.7%           | < 0.0001        |
|                             | Females (n= 301,411)                   | 2,574<br>38.6%   | 129<br>27.2%   | 70<br>37.2%                  | 737<br>38.4%       | 2885<br>38.3%            |                 |
| <b>Age at diagnosis</b>     | Mean age years (Standard Deviation SD) | 64.9<br>(12.1)   | 70.2<br>(10.7) | 58.3<br>(9.6)                | 63.4<br>(11.3)     | 65.1<br>(12.1)           | < 0.0001        |
|                             | 40-49                                  | 674<br>10.1 %    | 15<br>3.0%     | 27<br>14.4%                  | 201<br>10.5%       | 769<br>10.2%             |                 |
|                             | 50-59                                  | 1,756<br>26.3 %  | 64<br>13.5%    | 81<br>43.1%                  | 590<br>30.7%       | 1,931<br>25.6%           |                 |
|                             | 60-69                                  | 1,849<br>27.7 %  | 132<br>27.8%   | 61<br>32.4%                  | 576<br>30.0%       | 2,075<br>27.5%           |                 |
|                             | More than 70                           | 2,397<br>35.9%   | 265<br>55.7%   | 19<br>10.1%                  | 555<br>28.7%       | 2,771<br>36.7%           |                 |
|                             |                                        |                  |                |                              |                    |                          |                 |
|                             |                                        |                  |                |                              |                    |                          |                 |
| <b>Socioeconomic status</b> | 1 (least deprived)<br>(n= 140,396)     | 1,656<br>24.8%   | 131<br>27.6%   | 61<br>32.4%                  | 501<br>26.1%       | 2,173<br>28.8%           | < 0.0001        |
|                             | 2<br>(n= 171,605)                      | 1,569<br>23.5%   | 146<br>30.7%   | 48<br>25.5%                  | 563<br>29.3%       | 2,127<br>28.2%           |                 |
|                             | 3<br>(n= 150,964)                      | 1,702<br>25.3%   | 106<br>22.3%   | 43<br>22.9%                  | 395<br>20.6%       | 1,592<br>21.1%           |                 |
|                             | 4 (most deprived)<br>(n= 101,402)      | 1,749<br>26.4%   | 93<br>19.4%    | 36<br>19.1%                  | 463<br>23.8%       | 1,654<br>22.0%           |                 |
|                             |                                        |                  |                |                              |                    |                          |                 |

\* P-values, calculated using the Pearson chi-square test, are used to assess the statistical significance of differences in the incidence of outcomes (Cirrhosis, HCC, Liver Transplantation, Liver Death, Combined Outcomes) between each demographic group within the diagnosed group. A p-value < 0.0001 indicates significant differences.

#### **6.4.1.2 HCC**

The results from the cancer registry showed that there were 476 patients diagnosed with HCC. Of these patients, 347 were men and 129 were women, with men being significantly more affected than women. The mean age at diagnosis was 70.24 (SD  $\pm$  10.7 years). Among patients stratified by socioeconomic status, HCC was more common in patients in the second least deprived population. The characteristics of patients diagnosed with HCC are shown in Table 6-5.

#### **6.4.1.3 Liver transplantation**

Liver transplantation was coded in the general practice records of 63 patients and the HES records of 183. No patients had a recorded death due to liver transplantation.

The total number of unique patients who underwent liver transplantation was 188, with 118 being men and 70 women, with men being significantly more affected than women. Liver transplants generally had a younger age distribution than other outcomes, with a mean age at diagnosis of 58.30 years (SD  $\pm$  9.6 years). The characteristics of the patients recorded as having liver transplantation are presented in Table 6-5.

#### **6.4.1.4 Liver death**

The total number of deaths reported primarily due to liver diseases was 1,922 patients. Of these, 1,185 were men, and 737 were women, with men being significantly more affected than women. The mean age at diagnosis was 63.39 years (SD  $\pm$  11.3 years). The characteristics of the patients coded with liver death are given in Table 6-5.

#### **6.4.1.5 Combined outcomes**

Of the study population, the number of unique patients with at least one of the outcomes was 7,546. Among this cohort, 4,661 were males (61.7%) and 2,885 were females (38.3%).

The mean age at diagnosis was 65.1 years (SD  $\pm$  12.1 years). Among patients stratified by socioeconomic status, the results showed that liver-related combined outcomes were more common in the least deprived population. Detailed characteristics of these patients, including age, sex, and socioeconomic status, are outlined in Table 6-5.

#### **6.4.2 Determination of new patient outcomes for each cohort**

Different numbers of patients with the outcomes were available for each biomarker. Among the patients with AST/ALT scores, 2,424 had cirrhosis. There were 5,372 patients with

cirrhosis among the APRI population, and 1,988 among the FIB-4 population. The populations available for cohort studies with each biomarker are shown in Table 6-6.

*Table 6-6: Number of patients with the outcomes among each subpopulation:*

|                                                                             | <b>Cirrhosis</b> | <b>HCC</b>     | <b>Liver transplant</b> | <b>Liver death</b> | <b>Combined outcome</b> |
|-----------------------------------------------------------------------------|------------------|----------------|-------------------------|--------------------|-------------------------|
| <b>Patients with the outcomes among the study population (n =3,513,680)</b> | 6,676<br>(0.2%)  | 476<br>(0.01%) | 188<br>(0.01%)          | 1,922<br>(0.05%)   | 7,546<br>(0.2%)         |
| <b>Patients with the outcomes in AST/ALT subpopulation (n= 186,504)</b>     | 2,424<br>(1.3%)  | 189<br>(0.1%)  | 60<br>(0.03%)           | 743<br>(0.4%)      | 2,733<br>(1.4%)         |
| <b>Patients with the outcomes in APRI subpopulation (n= 483,097)</b>        | 5,372<br>(1.1%)  | 398<br>(0.1%)  | 118<br>(0.02%)          | 1,635<br>(0.3%)    | 6,025<br>(1.2%)         |
| <b>Patients with the outcomes in Fib4 subpopulation (n=157,804)</b>         | 1,988<br>(1.2%)  | 101<br>(0.1%)  | 54<br>(0.03%)           | 623<br>(0.3%)      | 2,235<br>(1.4%)         |

#### 6.4.2.1 Follow-up time

The follow-up times were different for each subpopulation. The median follow-up time for the combined outcome was 5.1 years (IQR 2.1, 9.4), 5.0 years (IQR 1.8, 9.1), and 4.4 years (IQR 1.6, 8.0) for the AST/ALT ratio, APRI score, and FIB-4 subpopulations, respectively. The median follow-up time for each outcome by population is presented in Table 6-7.



Table 6-7: Median follow-up time for each subpopulation and outcomes:

|                                                                          | <b>Cirrhosis</b> | <b>HCC</b>      | <b>Liver transplant</b> | <b>Liver death</b> | <b>Combined outcome</b> |
|--------------------------------------------------------------------------|------------------|-----------------|-------------------------|--------------------|-------------------------|
| <b>The median follow-up time for AST/ALT and outcomes by years (IQR)</b> | 4.7 (1.8, 8.7)   | 5.9 (2.5, 9.8)  | 5.5 (2.1, 8.6)          | 5.2 (2.5, 8.7)     | 5.1 (2.1, 9.4)          |
| <b>The median follow-up time for APRI and outcomes by years (IQR)</b>    | 5.0 (1.8, 9.1)   | 6.3 (2.1, 10.2) | 5.6 (2.2, 10.2)         | 5.1 (2.2, 9.0)     | 5.0 (1.8, 9.1)          |
| <b>The median follow-up time for Fib4 and outcomes by years (IQR)</b>    | 4.4 (1.6, 8.2)   | 5.6 (1.9, 8.9)  | 5.2 (2.1, 9.8)          | 4.9 (2.3, 7.9)     | 4.4 (1.6, 8.0)          |

## 6.5 Discussion

The findings of our study highlight liver cirrhosis as a significant outcome of advanced CLD observed across healthcare settings. While cirrhosis is expected to be more common than HCC, transplant, or liver-related deaths, which are often complications of cirrhosis, it is important to note that our focus was solely on liver-related outcomes. Therefore, other potential outcomes and causes of death were not considered in this analysis. Notably, the analysis of HES diagnosis files revealed a higher incidence of liver cirrhosis diagnoses in hospitals compared to primary care centres.

A significant disparity in the documentation of liver cirrhosis cases between these settings is evident: while 74.9% of patients with liver cirrhosis identified in the CPRD have

corresponding records in the HES, only a fraction (26.8%) of cases diagnosed in HES are documented in the CPRD.

The diagnosis of liver cirrhosis in primary and secondary care in the UK has been previously examined in a study by Ratib et al (261). This study found that of the 2,975 people identified as incident cirrhosis cases in primary care records during the study period, 2,230 (75%) had a hospital admission related to liver disease. Furthermore, of the 2,282 patients who were diagnosed with cirrhosis in secondary care only during the study period, and who had no cirrhosis diagnostic Read code in primary care, 2,062 (90.3%) had either death or additional evidence related to liver disease in their records, or a confirmation of a cirrhosis diagnosis in their free text.

The observed discrepancies between my study and Ratib's in the diagnosis of liver cirrhosis in primary and secondary care in the UK can largely be attributed to a key methodological difference in case definition. My study was specifically focused on confirmed cases of cirrhosis, as defined by my codes lists in both primary and secondary care files. In contrast, Ratib's study adopted a broader approach, encompassing any diagnosis or procedure related to liver disease as indicated by a more extensive code list used in their analysis. This wider scope in Ratib's research would naturally include a more extensive array

of cases, capturing some that may not meet the specific definition of confirmed cirrhosis used in my study. However, we need to consider that their inclusion of procedures as part of the cirrhosis definition would likely increase the proportion of hospital-diagnosed cases even further.

The higher incidence of liver cirrhosis coding in HES compared to primary care can be attributed to several factors. First, cirrhosis is not typically diagnosed in primary care because it often requires specialized testing and imaging that are available only in secondary care settings. This means that primary care physicians may not have the necessary resources to diagnose cirrhosis definitively, leading to more diagnoses being recorded in hospital records. Second, there is a lack of secondary care outpatient coding for cirrhosis diagnoses, which could be the primary place of diagnosis. This absence contributes to the underrepresentation of cirrhosis diagnoses in primary care records.

Additionally, differential recording practices and delays in communication between secondary and primary care settings further contribute to the disparity. Poor communication from secondary to primary care means that diagnoses made in hospital settings are not always promptly or accurately recorded in primary care records. This variation is supported by Saine et al., who

demonstrated that not all cases noted in the HES find a corresponding entry in the CPRD (262). This is further supported by McDonald et al (263), who highlighted that up to 10% of hospitalizations noted in the HES are only documented in the CPRD within free text notes. Another potential factor is that delays in communication between secondary and primary care may result in some patients with cirrhosis dying before their diagnosis is recorded in the CPRD, leading to their inclusion in death records without a corresponding primary care entry.

My study's reliance on Read codes to identify cirrhosis cases within the CPRD likely led to the exclusion of these cases from my analysis, potentially underestimating the true prevalence of cirrhosis in primary care records. Moreover, the use of Read codes may not capture all relevant cases if the coding is inconsistent or incomplete, further contributing to the discrepancy between primary and secondary care records.

My study's methodology, which includes data linkage between the CPRD and HES, mitigates this limitation. By incorporating patients with a confirmed diagnosis of cirrhosis in HES, my approach ensures the inclusion of cases that might be missing from the CPRD due to the absence of Read codes. This linkage provides a more accurate reflection of cirrhosis incidence and management across the healthcare system.

In addition, I explained in the methodology section (see [6.3.3.2](#)) that the NCRAS is the gold standard for cancer data recording in the UK, emphasizing its unparalleled reliability and comprehensiveness in capturing cancer cases.

Although my study population was predominantly female, as detailed in Table 5-4, my findings revealed that advanced CLDs were more common in men. This observation aligns with the broader findings reported in the global literature. A recent global update on the burden of liver disease in 2023 has underscored that approximately two-thirds of all liver-related deaths occur in men, primarily due to complications associated with cirrhosis and HCC (264).

Additionally, a population-based cohort study in Sweden has provided detailed insights into the incidence rates of cirrhosis, delineating a higher crude annual incidence in men compared to women. Specifically, the study estimates that the overall crude annual incidence of cirrhosis among adults for the period 2011–2018 was markedly higher in men than in women (265).

The study findings indicate that liver cirrhosis has a higher reported incidence compared to other advanced CLD outcomes. This observation is in line with the natural history of liver disease progression as liver cirrhosis is the culmination of a chronic

inflammatory process and the beginning of a series of potentially fatal complications, including the development of HCC (266).

## **6.6 Conclusion**

This chapter showed that cirrhosis was the primary outcome observed in the majority of patients with the outcomes. Furthermore, the analysis of HES diagnosis files revealed a higher incidence of liver cirrhosis diagnoses in hospitals compared to primary care centres. Additionally, my findings revealed that advanced CLDs were more common in men.

## **Chapter 7      Can Non-Invasive Biomarkers Be Used to Predict Advanced Liver Diseases?**

## **7.1 Study aim**

This study aims to determine whether the AST/ALT ratio, APRI score, and FIB-4 index can be used within primary care settings to predict subsequent advanced liver disease in those for whom these calculations can be performed in the primary care in the UK.

### **7.1.1 Objectives**

- To estimate the crude hazard ratio (HR) of the first recorded diagnosis of each outcome in those with abnormal markers compared to those with normal markers.
- To assess the performance of different biomarkers in predicting liver diseases and detecting outcomes among different strata of age, sex, and socioeconomic status and among different time periods.

## **7.2 Methods**

### **7.2.1 Study design**

Three cohort studies were conducted using prospectively collected clinical data.



## **7.2.2 Study population**

The study population was previously described in section (6.3.2).

### **7.2.2.1 Selection of exposed and non-exposed groups**

I undertook a series of three subpopulation cohort studies (one for each marker). From the three subpopulations identified in Chapter 5, for each biomarker, patients with at least one abnormal score calculable were termed the exposed group. The start of the follow-up period (index date) for these patients was set at the date of their first recorded test result, irrespective of whether this result was normal or abnormal. The control group consisted of patients with only normal individual markers. The entry date to the study for the control group was defined as the first occasion on which the normal marker was calculable (index date). The characteristics of patients with normal and abnormal scores are described in Chapter 5 (See Table 5-4).

By choosing to start follow up for exposed subjects from before they were actually exposed (that is before they had an abnormal risk score), I inevitably impart a bias to the analysis. I chose to do this because the bias here is likely to be conservative (that is it will underestimate the utility of the scores) whereas starting follow up from the first abnormal score for the exposed

and the first score for the unexposed could produce the opposite bias. A sensitivity analysis adopting the alternative approach is presented in chapter 8.

### **7.2.3 Outcomes**

This study utilised the primary and secondary outcomes identified in section 6.3.3.

### **7.2.4 Statistical analysis**

In chapter 5: Patients were classified into two groups for each subpopulation: those with abnormal scores (exposed group) and those with normal scores (control group). In chapter 6, I identified all patients with the primary and secondary outcomes within the study population

In this study, longitudinal analyses were conducted separately for each marker. Cox proportion hazard models were used to estimate the crude HR with a 95% (CI) of the first recorded diagnosis of each outcome in those with abnormal markers compared to those with normal markers. The results of the Cox regression models show the HR with a 95% CI as a measure of the variation in risk between the groups and a p-value of less than 0.05 was considered statistically significant.

I calculated the time of observation as person-time, starting at the index date and ending when the outcome of interest occurred (i.e., liver cirrhosis, HCC, liver transplantation, liver death, and combined outcomes), subjects left the study, subjects died, or at the end of the study period (December 31, 2016), whichever came first. Person-time was calculated separately for the exposed and unexposed groups.

Multivariate Cox regression models, adjusted for gender, age at the diagnosis, and socioeconomic status, were employed to examine the association between explanatory independent predictors and the predictive ability of the biomarkers for the outcomes. Hazard ratios from these models are presented with 95% confidence intervals. To assess the proportional hazards assumption in my study, I implemented a combination of Kaplan-Meier (KM) curves and log-log plot analysis. As a further test of the significance in the difference in the survival function between the exposed and the unexposed cohorts I present the results of the logrank test implemented in the KM analysis.

## **7.2.5 Stratification analyses**

### **7.2.5.1 Performance of the biomarkers among the different strata**

As mentioned earlier; to assess the performance of the biomarkers among different strata, I began with a multivariate Cox proportional hazards model that included relevant explanatory variables. Then, to examine potential effect modification by the explanatory variables, interaction terms between the biomarkers and these variables were included in the model. This allowed me to investigate whether the relationship between biomarkers and outcomes differed across different strata.

The likelihood ratio test was then utilized to assess the statistical significance of the interaction terms. A p-value below 0.05 indicated significant effect modification, suggesting that the performance of the biomarkers varied across different strata.

Given significant effect modification, stratified univariate Cox models were subsequently used to evaluate the performance of the biomarkers within each stratum separately. This analysis aimed to determine which strata the biomarkers performed better in predicting the incidence of outcomes among the exposed (abnormal markers) and non-exposed (normal markers) groups. Finally, the performance of each biomarker was compared across the different strata, providing detailed insights into how these

biomarkers function within various demographic and clinical contexts.

#### **7.2.5.2 Evaluating biomarker-associated risk over three specific time periods:**

To examine any variation in the performance of the biomarkers over time, I divided the follow-up period into three distinct intervals (January 1, 1990- December 31, 1999, January 1, 2000- December 31, 2009, and January 1, 2010- December 31, 2016). The selection of the time periods closely matches the trends on the application of biomarkers that are emphasised in Chapter 5 and shown in Figure 5-5. Additionally, this stratification considers the variation in the availability of linked data, as HES follow-up started later, from January 2, 1998, to December 31, 2016. This approach allows us to assess both the impact of the reduction in AST testing and the change in the availability of linked data for outcomes. By examining these time periods separately, I can better understand how these factors influence my study results and ensure that my findings are robust across different data availability scenarios.

The analysis involved calculating the time of observation as person-time, starting at the index date and ending when the outcome of interest occurred, subjects left the study, subjects died, or at the end of the study period, whichever came first.

To account for temporal variations in biomarker performance, I included an interaction term between the biomarkers and the time periods in the multivariate Cox proportional hazards model. This approach allowed me to evaluate the influence of time stratification on the predictive power of the biomarkers. To assess the statistical significance of these interactions, I employed the likelihood ratio test. A p-value less than 0.05 would indicate a statistically significant variation in the effectiveness of biomarkers across the defined time intervals, underscoring the temporal dynamics in their predictive utility.

## **7.3 Results**

### **7.3.1 Patients with biomarkers and outcomes**

#### **7.3.1.1 AST/ALT ratio Cox regression and hazard ratio analysis**

In the analysis presented in Table 7-1, Cox regression analysis was conducted to evaluate the risk of various liver-related outcomes among patients. The analysis revealed that patients with abnormal biomarker scores demonstrated a higher risk of developing liver cirrhosis, HCC, undergoing liver transplantation, liver-related death, and the combined outcome compared to those with normal scores. The unadjusted HR for these outcomes were as follows: 1.7 (95% CI: 1.6 to 2.0) for liver cirrhosis, 1.1 (1.0 to 1.5) for HCC, 1.3 (0.7 to 2.6) for liver

transplantation, 3.4 (2.8 to 4.2) for liver death, and 1.8 (1.6 to 2.0) for the combined outcomes.

In the adjusted analysis, which accounted for potential covariates including gender, socioeconomic status, and age at diagnosis, the hazard ratios between the exposed and unexposed were increased. This was true for all outcomes assessed with adjusted hazard ratios ranging from 1.3 (1.0-1.7) for HCC to 4.9 (4.0-6.1) for liver death. Full results are presented in Table 7-1.

The total person-time at risk for the combined outcome, was 874,422 person-years for the exposed group and 520,028.3 person-years for the non-exposed group. The person-time for the other outcomes among the exposed and non-exposed groups is presented in Table 7-1. Examination of the proportional hazards assumption by examination of both K-M and log log curves (Figure 7-1 and Figure 7-2 ) showed no gross violation of this assumption.

Table 7-1: Univariate and multivariate cox regression model of the incidence of different outcomes for patients with normal and abnormal ast/alt ratio.

|                                                                 | <b>Cirrhosis</b> | <b>HCC</b>    | <b>Liver transplantation</b> | <b>Liver death</b> | <b>Combined outcomes</b> |
|-----------------------------------------------------------------|------------------|---------------|------------------------------|--------------------|--------------------------|
| <b>Patients with outcomes</b>                                   | 2,424            | 189           | 60                           | 743                | 2,733                    |
| <b>Person years among the exposed group</b>                     | 874671.84        | 879243.39     | 879178.1                     | 879290.95          | 874422                   |
| <b>Person years among the non-exposed group</b>                 | 520177.23        | 521195.94     | 521218.62                    | 521281.18          | 520028.3                 |
| <b>Univariate analysis</b>                                      |                  |               |                              |                    |                          |
| <b>HRs for Subjects with normal AST/ALT (n= 68,465)</b>         | 1                | 1             | 1                            | 1                  | 1                        |
| <b>HRs (CI) for Subjects with abnormal AST/ALT (n= 118,039)</b> | 1.7 (1.6-2.0)    | 1.1 (1.0-1.5) | 1.3 (0.7-2.6)                | 3.4 (2.8-4.2)      | 1.8 (1.6-2.0)            |
| <b>Multivariate analysis*</b>                                   |                  |               |                              |                    |                          |
| <b>Subjects with normal AST/ALT (n= 68,465)</b>                 | 1                | 1             | 1                            | 1                  | 1                        |
| <b>Subjects with abnormal AST/ALT (n= 118,039)</b>              | 2.2 (2.06-2.5)   | 1.3 (1.0-1.7) | 2.8(1.5-5.0)                 | 4.9(4.0-6.1)       | 2.2(2.05-2.4)            |

Abbreviations: CI: confidence interval; HR: hazard ratio.

\*The cox regression model was adjusted for gender, age at diagnosis, and socioeconomic status.



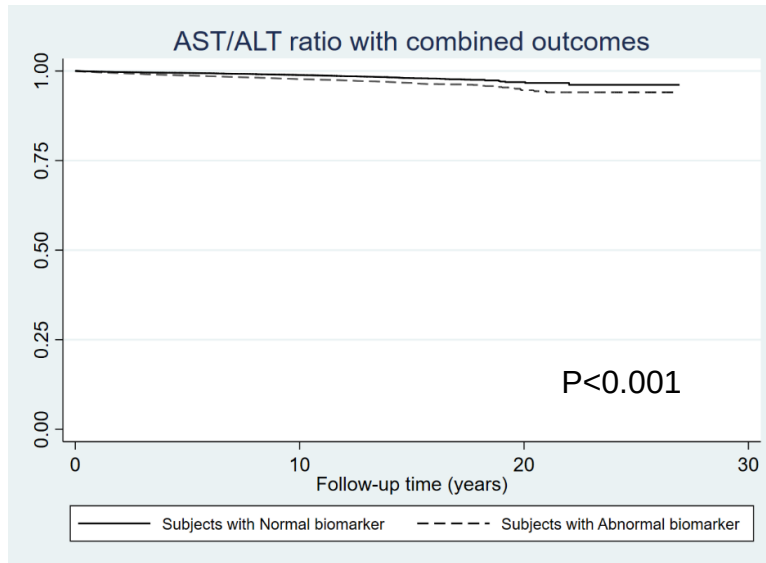


Figure 7-1: Kaplan-Meier plots comparing the incidence of combined outcomes for patients with normal ast/alt biomarkers and patients with abnormal ast/alt biomarkers. The Y-axis represents the probability of patients' free outcome. The X-axis represents the follow-up time (years) since enrolment in the study. The solid line represents subjects with normal biomarkers. The dashed line represents patients with abnormal biomarkers. The difference between the two curves is statistically significant ( $P<0.001$ ), indicating a clear disparity in survival probabilities based on AST/ALT ratio status.

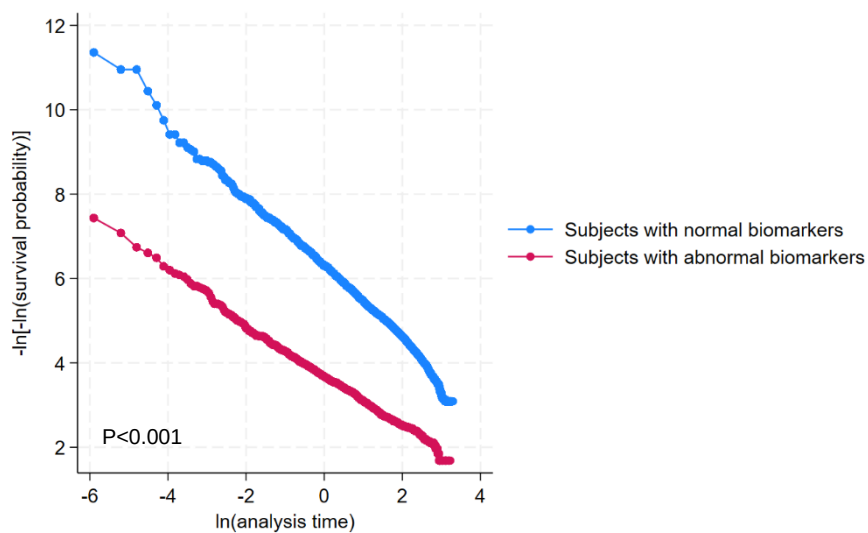


Figure 7-2: The log-log plot analysis assesses the proportion hazards assumption by comparing the incidence of combined outcomes between patients with normal AST/ALT biomarkers and those with abnormal AST/ALT biomarkers. On this plot, the x-axis represents the logarithm of analysis time, and the y-axis displays the natural logarithm of the negative natural logarithm of survival probabilities.

#### **7.3.1.2 APRI score Cox regression and hazard ratio analysis.**

When I repeated the above analyses with APRI score population, the findings demonstrated far larger hazard ratios among those with abnormal scores, with an abnormal APRI score carrying a more than 30-folds increased hazard of cirrhosis compared to those with normal scores and an almost 60-folds increased hazard of liver death. On this occasion adjusting for the effects of age, sex and socioeconomic deprivation attenuated the hazards slightly rather than accentuating them. Full details are given in Table 7-2. Again K-M and log-log curves (Figure 7-3 and Figure 7-4) showed no gross violation of the proportional hazards assumption.

Table 7-2: Univariate and multivariate cox regression model of the incidence of different outcomes for patients with normal and abnormal APRI scores:

|                                                                                      | <b>Cirrhosis</b> | <b>HCC</b>       | <b>Liver<br/>transplantation</b> | <b>Liver death</b> | <b>Combined<br/>outcomes</b> |
|--------------------------------------------------------------------------------------|------------------|------------------|----------------------------------|--------------------|------------------------------|
| <b>Patients with outcomes</b>                                                        | 5,372            | 398              | 118                              | 1,635              | 6,025                        |
| <b>Person year Exposed</b>                                                           | 85461.51         | 91634.16         | 91667.32                         | 91785.62           | 85793.83                     |
| <b>Person year Nonexposed</b>                                                        | 3848476          | 3854303          | 3854292                          | 3854532            | 3848495                      |
| <b>Univariate analysis</b>                                                           |                  |                  |                                  |                    |                              |
| <b>HR for Subjects with normal APRI score (n= 464,204)</b>                           | 1                | 1                | 1                                | 1                  | 1                            |
| <b>HRs (CI) Subjects with abnormal APRI score (n= 18,893)</b>                        | 31.7 (30.1-33.6) | 37.9 (32.3-44.4) | 53.2 (37.1-76.4)                 | 58.6 (53.1-64.7)   | 30.8 (29.3-32.4)             |
| <b>Multivariate analysis*</b>                                                        |                  |                  |                                  |                    |                              |
| <b>Subjects with normal APRI<br/>Subjects with normal APRI score (n= 464,204)</b>    | 1                | 1                | 1                                | 1                  | 1                            |
| <b>Subjects with abnormal APRI<br/>Subjects with abnormal APRI score (n= 18,893)</b> | 27.8(26.3-29.3)  | 30.9(26.3-36.4)  | 46.5 (32.0-67.48                 | 53.2(48.12-58.9)   | 26.7 (25.4-28.1)             |

Abbreviations: CI: confidence interval; HR: hazard ratio.

\*The cox regression model was adjusted for gender, age at diagnosis, and socioeconomic status.

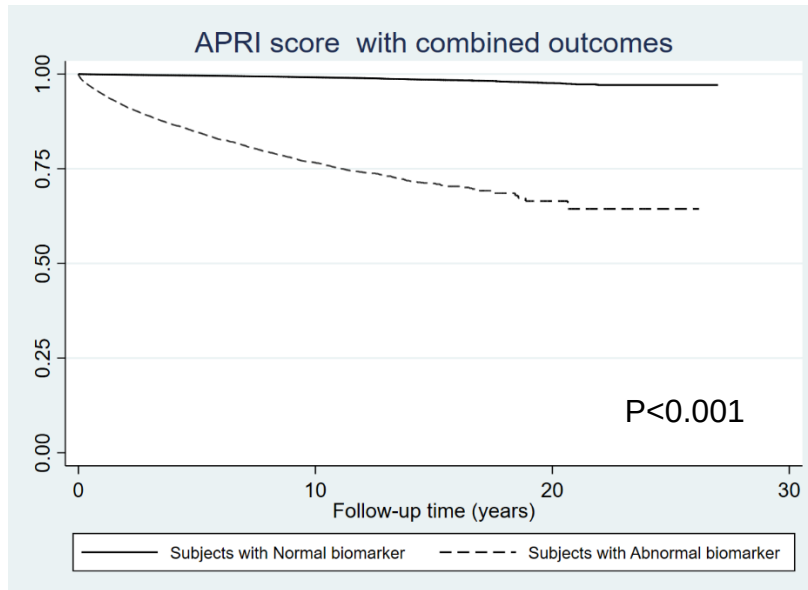


Figure 7-3: Kaplan-Meier plots comparing the incidence of combined outcome for patients with normal APRI score and patients with abnormal APRI score. The Y-axis represents the probability of patients' free outcome. The X-axis represents the follow-up time (years) since enrolment in the study. The solid line represents subjects with normal biomarkers. The dashed line represents patients with abnormal biomarkers.

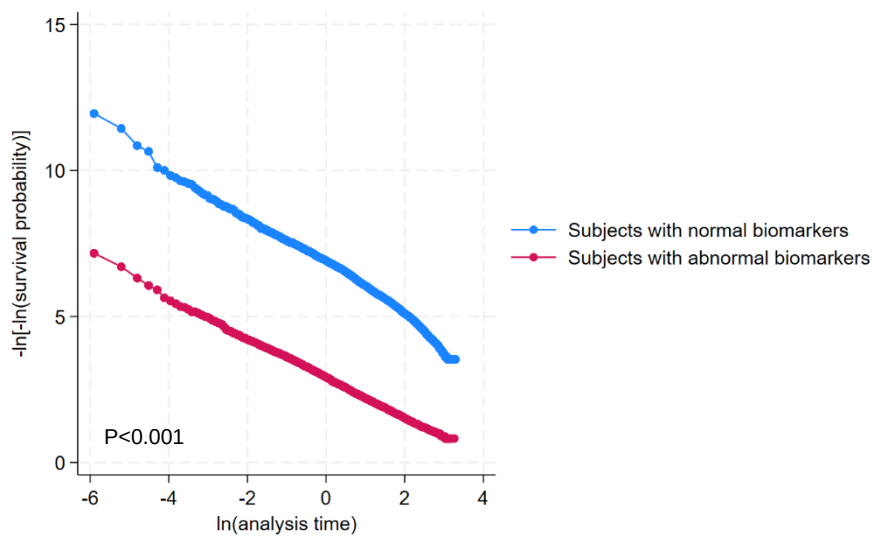


Figure 7-4: The log-log plot analysis assesses the proportion hazards assumption by comparing the incidence of combined outcomes between patients with normal APRI biomarkers and those with abnormal APRI biomarkers. On this plot, the x-axis represents the logarithm of analysis time, and the y-axis displays the natural logarithm of the negative natural logarithm of survival probabilities.

#### **7.3.1.3 Fib4 index Cox regression and hazard ratio analysis.**

When the analysis was conducted using the FIB-4 score, the results demonstrated significantly increased hazard ratios among those with abnormal scores. Specifically, an abnormal FIB-4 score was associated with a 5.7-folds increase in the hazard of cirrhosis and a 9.1-folds increase in the hazard of liver death, compared to those with normal scores. The risk was notably higher for all outcomes studied, with adjusted hazard ratios showing substantial increases: 13.1 for HCC and 12.4 for liver transplantation. Adjusting for age, gender, and socioeconomic status resulted in an increase in the hazard ratios, rather than a reduction. Detailed results of these analyses are presented in Table 7-3.

Kaplan-Meier plots and log-log curves (Figure 7-5 and Figure 7-6) demonstrated no significant violation of the proportional hazards assumption. The plots indicated that the time to the first recorded incidence of the outcomes was considerably shorter for patients with abnormal FIB-4 scores compared to those with normal scores.

Table 7-3: Univariate and multivariate cox regression model of the incidence of different outcomes for patients with normal and abnormal Fib4 index:

|                                                               | <b>Cirrhosis</b> | <b>HCC</b>      | <b>Liver transplantation</b> | <b>Liver death</b> | <b>Combined outcomes</b> |
|---------------------------------------------------------------|------------------|-----------------|------------------------------|--------------------|--------------------------|
| <b>Patients with outcomes</b>                                 | 1,988            | 101             | 54                           | 623                | 2,235                    |
| <b>Person years Exposed</b>                                   | 540665.6         | 544922.9        | 544883.9                     | 545031.1           | 540400.1                 |
| <b>Person years non-exposed</b>                               | 585590.9         | 586081.6        | 586056.8                     | 586078.4           | 585520.3                 |
| <b>Univariate analysis</b>                                    |                  |                 |                              |                    |                          |
| <b>HR for Subjects with normal Fib4 index (n= 77,590)</b>     | 1                | 1               | 1                            | 1                  | 1                        |
| <b>HRs (CI) Subjects with abnormal Fib4 index (n= 80,214)</b> | 5.7 (5.1-6.5)    | 11.4 (7.2-18.0) | 4.7 (2.4-9.3)                | 9.1 (7.0-11.8)     | 5.6 (5.0-6.3)            |
| <b>Multivariate analysis*</b>                                 |                  |                 |                              |                    |                          |
| <b>Subjects with normal Fib 4 (n= 77,590)</b>                 | 1                | 1               | 1                            | 1                  | 1                        |
| <b>Subjects with abnormal Fib 4 (n= 80,214)</b>               | 9.1(8.06-10.4)   | 13.1(8.1-21.1)  | 12.4(5.9-25.9)               | 16.9(13.1-21.9)    | 8.8 (7.8-9.9)            |

Abbreviations: CI: confidence interval; HR: hazard ratio.

\*The cox regression model was adjusted for gender, age at diagnosis, and socioeconomic status.

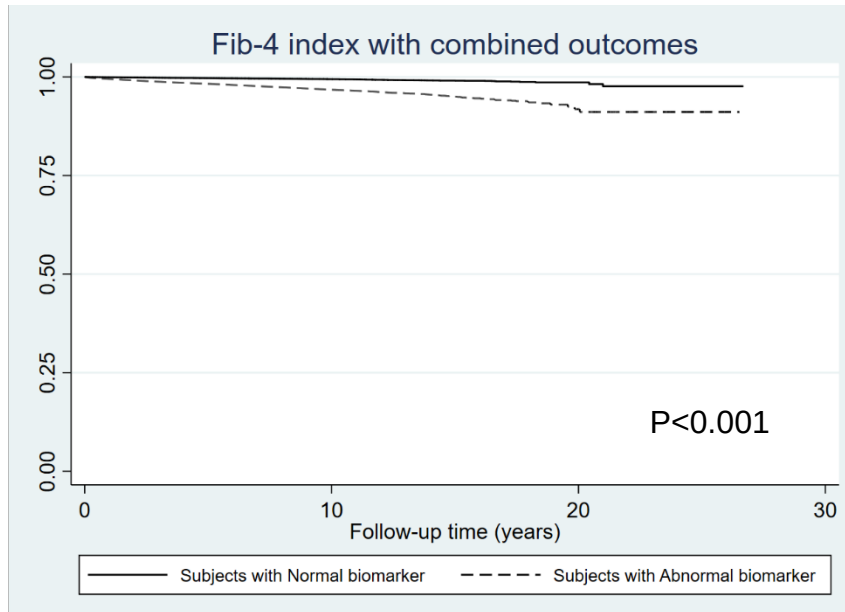


Figure 7-5: Kaplan-Meier plots comparing the incidence of combined outcome for patients with normal Fib-4 and patients with abnormal Fib-4. The Y-axis represents the probability of patients' free outcome. The X-axis represents the follow-up time (years) since enrolment in the study. The solid line represents subjects with normal biomarkers. The dashed line represents patients with abnormal biomarkers.

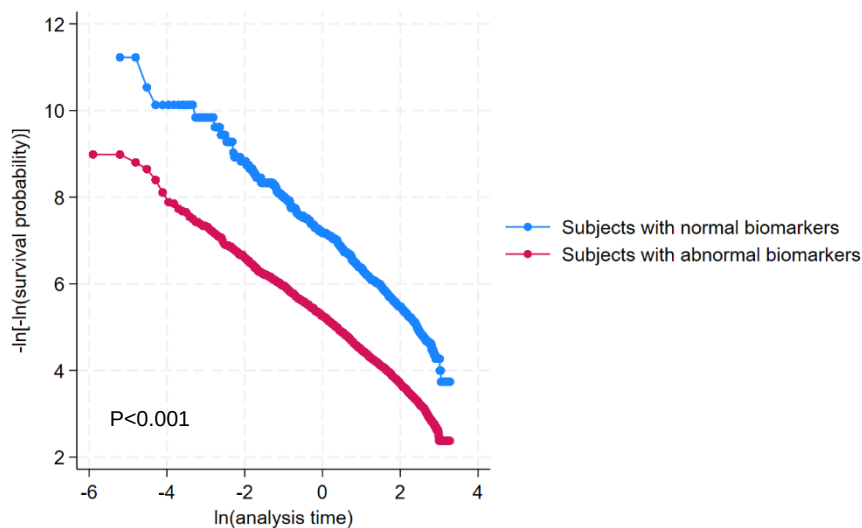


Figure 7-6: The log-log plot analysis assesses the proportion hazards assumption by comparing the incidence of combined outcomes between patients with normal Fib4 biomarkers and those with abnormal Fib4 biomarkers. On this plot, the x-axis represents the logarithm of analysis time, and the y-axis displays the natural logarithm of the negative natural logarithm of survival probabilities.

## **7.3.2 Stratification analysis**

### **7.3.2.1 Performance of the biomarkers among the different strata**

#### **7.3.2.1.1 Patients with normal and abnormal AST/ALT ratios**

Fitting interaction terms for age, sex, and socioeconomic status showed clear effect modification by these factors, with the significance of these modifications confirmed by the likelihood ratio test (p-value < 0.001).

The analysis showed that an abnormal AST/ALT ratio is associated with a significantly greater increase in the hazard of adverse liver outcomes in males than in females (Table 7-4). For age, younger patients exhibited a stronger association with adverse outcomes compared to older patients, with the interaction terms highlighting a significant modification of risk by age (Table 7-5).

Socioeconomic status significantly modified the relationship between abnormal AST/ALT ratios and liver outcomes, with the most deprived group showing a slightly lower hazard ratio despite high hazard ratios across most other deprivation quintiles (Table 7-6).



Table 7-4: Univariate Cox regression model of the incidence of the (combined outcome) for patients with normal and abnormal ast/alt ratio stratified by sex:

| sex            | Subjects with a normal and abnormal score        | Hazard risk (CI) |
|----------------|--------------------------------------------------|------------------|
| <b>Males</b>   | Subjects with normal AST/ALT ratio (n= 41,169)   | 1                |
|                | Subjects with abnormal AST/ALT ratio (n= 47,135) | 2.6 (2.3-2.8)    |
| <b>Females</b> | Subjects with normal AST/ALT ratio (n= 27,296)   | 1                |
|                | Subjects with abnormal AST/ALT ratio (n= 70,904) | 1.7 (1.4-2.0)    |

Table 7-5: Univariate Cox regression model of the incidence of the (combined outcome) for patients with normal and abnormal ast/alt ratio stratified by Age group:

| Age group                 | Subjects with a normal and abnormal score        | Hazard risk (CI) |
|---------------------------|--------------------------------------------------|------------------|
| <b>40-49</b>              | Subjects with normal AST/ALT ratio (n= 26,339)   | 1                |
|                           | Subjects with abnormal AST/ALT ratio (n= 24,334) | 2.6 (2.2-3.0)    |
| <b>50-59</b>              | Subjects with normal AST/ALT ratio (n= 21,321)   | 1                |
|                           | Subjects with abnormal AST/ALT ratio (n= 25,246) | 2.6 (2.1-3.0)    |
| <b>60-69</b>              | Subjects with normal AST/ALT ratio (n= 12,908)   | 1                |
|                           | Subjects with abnormal AST/ALT ratio (n= 27,885) | 1.3 (1.1-1.5)    |
| <b>70 years and older</b> | Subjects with normal AST/ALT ratio (n= 7,897)    | 1                |
|                           | Subjects with abnormal AST/ALT ratio (n= 40,574) | 1.1 (0.9-1.5)    |

Table 7-6: Univariate Cox regression model of the incidence of the (combined outcome) for patients with normal and abnormal ast/alt ratio stratified by socioeconomic status:

| Socioeconomic               | Subjects with a normal and abnormal score        | Hazard risk (CI) |
|-----------------------------|--------------------------------------------------|------------------|
| <b>1-2 (least deprived)</b> | Subjects with normal AST/ALT ratio (n= 22,830)   | 1                |
|                             | Subjects with abnormal AST/ALT ratio (n= 30,807) | 2.0 (1.2-3.4)    |
| <b>3-5</b>                  | Subjects with normal AST/ALT ratio (n= 18,539)   | 1                |
|                             | Subjects with abnormal AST/ALT ratio (n= 32,096) | 1.4 (0.8-2.7)    |
| <b>6-8</b>                  | Subjects with normal AST/ALT ratio (n= 14,954)   | 1                |
|                             | Subjects with abnormal AST/ALT ratio (n= 30,218) | 1.6 (0.7-3.4)    |
| <b>9-10 (most deprived)</b> | Subjects with normal AST/ALT ratio (n= 12,142)   | 1                |
|                             | Subjects with abnormal AST/ALT ratio (n= 24,918) | 1.2 (0.5-2.6)    |

#### 7.3.2.1.2 Patients with normal and abnormal APRI scores

The stratified analyses indicated that an increased APRI score was a predictor of adverse liver outcomes in both males and females. The HRs were higher in female patients compared to male patients (Table 7-7).

An abnormal APRI score was more closely associated with adverse outcomes in younger patients than in older patients, indicating that age significantly modifies the risk associated with APRI scores (Table 7-8). Socioeconomic status also influenced the risk associated with abnormal APRI scores. While high HRs were observed across most deprivation quintiles, the least deprived group exhibited a slightly higher HR (Table 7-9).

The findings of the likelihood ratio test, marked by p-values lower than 0.001 for each stratification, confirm that the impact of abnormal APRI scores on liver-related outcomes varies significantly based on gender, age, and socioeconomic status.

*Table 7-7: Univariate Cox regression model of the incidence of the (combined outcome) for patients with normal and abnormal APRI score stratified by sex:*

| <b>sex</b>     | <b>Subjects with normal and abnormal score</b> | <b>Hazard risk (CI)</b> |
|----------------|------------------------------------------------|-------------------------|
| <b>Males</b>   | Subjects with normal APRI score (n= 209,498)   | 1                       |
|                | Subjects with abnormal APRI score (n= 10,810)  | 26.6 (24.9-28.5)        |
| <b>females</b> | Subjects with normal APRI score (n= 254,706)   | 1                       |
|                | Subjects with abnormal APRI score (n=8,083)    | 33.9 (31.2-36.9)        |

Table 7-8: *Univariate Cox regression model of the incidence of the (combined outcome) for patients with normal and abnormal APRI score stratified by Age group:*

| <b>Age group</b>          | <b>Subjects with normal and abnormal score</b> | <b>Hazard risk (CI)</b> |
|---------------------------|------------------------------------------------|-------------------------|
| <b>40-49</b>              | Subjects with normal APRI score (n= 131,369)   | 1                       |
|                           | Subjects with abnormal APRI score (n= 4,114)   | 42.9(38.8-47.3)         |
| <b>50-59</b>              | Subjects with normal APRI score (n= 113,898)   | 1                       |
|                           | Subjects with abnormal APRI score (n=4,792)    | 36.9 (33.5-40.5)        |
| <b>60-69</b>              | Subjects with normal APRI score (n= 96,581)    | 1                       |
|                           | Subjects with abnormal APRI score (n= 4,216)   | 26.6 (23.9-29.6)        |
| <b>70 years and older</b> | Subjects with normal APRI score (n= 122,356)   | 1                       |
|                           | Subjects with abnormal APRI score (n= 5,771)   | 17.4 (15.4-19.7)        |

Table 7-9: *Univariate Cox regression model of the incidence of the (combined outcome) for patients with normal and abnormal APRI score stratified by socioeconomic status.*

| <b>Socioeconomic</b>        | <b>Subjects with normal and abnormal score</b> | <b>Hazard risk (CI)</b> |
|-----------------------------|------------------------------------------------|-------------------------|
| <b>1-2 (least deprived)</b> | Subjects with normal APRI score (n= 138,127)   | 1                       |
|                             | Subjects with abnormal APRI score (n= 4,531)   | 33.7 (30.2-37.6)        |
| <b>3-5</b>                  | Subjects with normal APRI score (n= 131,136)   | 1                       |
|                             | Subjects with abnormal APRI score (n=5,300)    | 28.7 (25.9-31.9)        |
| <b>6-8</b>                  | Subjects with normal APRI score (n= 104,331)   | 1                       |
|                             | Subjects with abnormal APRI score (n= 4,615)   | 29.9 (26.9-33.2)        |
| <b>9-10 (most deprived)</b> | Subjects with normal APRI score (n= 90,610)    | 1                       |
|                             | Subjects with abnormal APRI score (n= 4,447)   | 29.7 (26.9-32.8)        |

### 7.3.2.1.3 Patients with normal and abnormal FIB-4 indexes

Stratification for gender in the FIB-4 analysis produced similar HRs in both males and females (Table 7-10). The likelihood ratio tests for interaction confirmed the lack of statistically significant interaction (p-value > 0.05). Age-stratified analysis demonstrated that the FIB-4 index was more strongly associated with adverse liver outcomes in younger patients compared to older patients. The likelihood ratio test for this stratification was significant (p-value < 0.001), indicating a substantial age-related variation in the risk associated with the FIB-4 index (Table 7-11).

When stratifying by socioeconomic status, similar HRs were observed across most deprivation quintiles, except for the most deprived group, which exhibited a slightly lower HR. The likelihood ratio test for socioeconomic status stratification also showed significance (p-value < 0.001) (Table 7-12).

*Table 7-10: Univariate Cox regression model of the incidence of the (combined outcome) for patients with normal and abnormal FIB-4 index stratified by sex:*

| <b>sex</b>     | <b>Subjects with a normal and abnormal score</b> | <b>Hazard risk (CI)</b> |
|----------------|--------------------------------------------------|-------------------------|
| <b>Males</b>   | Subjects with normal Fib4 index (n= 33,712)      | 1                       |
|                | Subjects with abnormal Fib4 index (n= 39,177)    | 5.3 (4.5-6.1)           |
| <b>females</b> | Subjects with normal Fib4 index (n= 43,878)      | 1                       |
|                | Subjects with abnormal Fib4 index (n= 41,037)    | 5.7 (4.8-6.8)           |

*Table 7-11: Univariate Cox regression model of the incidence of the (combined outcome) for patients with normal and abnormal FIB-4 index stratified by Age group:*

| <b>Age group</b>          | <b>Subjects with a normal and abnormal score</b> | <b>Hazard risk (CI)</b> |
|---------------------------|--------------------------------------------------|-------------------------|
| <b>40-49</b>              | Subjects with normal Fib4 index (n= 35,579)      | 1                       |
|                           | Subjects with abnormal Fib4 index (n= 6,093)     | 17.4(14.5-20.9)         |
| <b>50-59</b>              | Subjects with normal Fib4 index (n= 23,345)      | 1                       |
|                           | Subjects with abnormal Fib4 index (n= 15,094)    | 7.7 (6.3-9.3)           |
| <b>60-69</b>              | Subjects with normal Fib4 index (n= 11,849)      | 1                       |
|                           | Subjects with abnormal Fib4 index (n= 22,566)    | 4.4 (3.3-5.9)           |
| <b>70 years and older</b> | Subjects with normal Fib4 index (n= 6,817)       | 1                       |
|                           | Subjects with abnormal Fib4 index (n= 36,461)    | 1.9 (1.33-2.7)          |

*Table 7-12: Univariate Cox regression model of the incidence of the (combined outcome) for patients with normal and abnormal FIB-4 index stratified by socioeconomic status:*

| <b>Socioeconomic</b>        | <b>Subjects with a normal and abnormal score</b> | <b>Hazard risk (CI)</b> |
|-----------------------------|--------------------------------------------------|-------------------------|
| <b>1-2 (least deprived)</b> | Subjects with normal Fib4 index (n= 22,508)      | 1                       |
|                             | Subjects with abnormal Fib4 index (n= 22,279)    | 6.0 (4.7-7.7)           |
| <b>3-5</b>                  | Subjects with normal Fib4 index (n= 20,864)      | 1                       |
|                             | Subjects with abnormal Fib4 index (n=21,633)     | 6.9 (5.3-9.0)           |
| <b>6-8</b>                  | Subjects with normal Fib4 index (n= 18,620)      | 1                       |
|                             | Subjects with abnormal Fib4 index (n= 19,986)    | 6.0 (4.8-7.8)           |
| <b>9-10 (most deprived)</b> | Subjects with normal Fib4 index (n= 15,598)      | 1                       |
|                             | Subjects with abnormal Fib4 index (n= 16,334)    | 4.4 (3.6-5.3)           |

### **7.3.2.2 Temporal analysis of biomarker-associated risk across distinct time intervals**

#### **7.3.2.2.1 Patients with normal and abnormal AST/ALT ratios**

A Cox model of the relationship between AST/ALT ratio and the occurrence of adverse liver outcomes was constructed,

including an interaction term to evaluate the effect of different time periods. This model was significantly better at describing the data compared to one without the interaction term ( $P < 0.0001$  by likelihood ratio test).

As shown in Table 7-13, the HR between abnormal and normal AST/ALT ratios was marginally higher in later periods, indicating that the risk associated with abnormal AST/ALT ratios increased over time. Specifically, the HR increased from 1.7 (95% CI: 1.3-2.2) in the earliest period (1990-1999) to 2.0 (95% CI: 1.6-2.4) in the most recent period (2010-2016). These results suggest that the predictive power of the AST/ALT ratio for adverse liver outcomes has strengthened over time.

*Table 7-13: Univariate Cox regression model of the incidence of the combined outcomes for patients with normal and abnormal AST/ALT ratio across different time periods.*

| <b>AST/ALT</b>                                         | <b>combined outcome</b> |
|--------------------------------------------------------|-------------------------|
| <b>Window 1 (1990-1999)</b>                            |                         |
| Person year Exposed                                    | 102023.2                |
| Person year Non-exposed                                | 57263.25                |
| HR for Subjects with normal AST/ALT (n= 5,598)         | 1                       |
| HR (CI) for Subjects with abnormal AST/ALT (n= 8,241)  | 1.7 (1.3-2.2)           |
| <b>Window 2 (2000-2009)</b>                            |                         |
| Person year Exposed                                    | 967204.7                |
| Person year Non-exposed                                | 520028.3                |
| HR for Subjects with normal AST/ALT (n= 44,948)        | 1                       |
| HR (CI) for Subjects with abnormal AST/ALT (n= 87,743) | 1.8 (1.6-2.0)           |
| <b>Window 3 (2010-2016)</b>                            |                         |
| Person year Exposed                                    | 775611.1                |
| Person year Non-exposed                                | 395096.4                |
| HR for Subjects with normal AST/ALT (n= 16,918)        | 1                       |
| HR (CI) for Subjects with abnormal AST/ALT (n= 22,055) | 2.0 (1.6-2.4)           |

Abbreviations: CI: confidence interval; HR: hazard ratio.

### 7.3.2.2.2 Patients with normal and abnormal APRI scores

A Cox model of the relationship between APRI scores and adverse liver outcomes, including an interaction term, was significantly better than one without ( $P < 0.0001$  by likelihood ratio test). During Window 1, the HR was 24.0 (95% CI: 20.0-28.0), while Window 3 showed a marked increase in risk with an HR of 41.0 (95% CI: 35.6-47.2). More details are provided in Table 7-14.

*Table 7-14: Univariate Cox regression model of the incidence of the combined outcomes for patients with normal and abnormal APRI score across different time periods.*

| <b>APRI score</b>                                         | <b>combined outcome</b> |
|-----------------------------------------------------------|-------------------------|
| <b>Window 1 (1990-1999)</b>                               |                         |
| Person year Exposed                                       | 8874.681                |
| Person year non-exposed                                   | 395539.4                |
| HR for Subjects with normal APRI score (n= 32,424)        | 1                       |
| HR (CI) for Subjects with abnormal APRI score (n= 2,010)  | 24.0 (20.0-28.0)        |
| <b>Window 2 (2000-2009)</b>                               |                         |
| Person year Exposed                                       | 63331.39                |
| Person year non-exposed                                   | 3121338                 |
| HR for Subjects with normal APRI score (n= 350,607)       | 1                       |
| HR (CI) for Subjects with abnormal APRI score (n= 11,575) | 29.9 (28.1-31.7)        |
| <b>Window 3 (2010-2016)</b>                               |                         |
| Person year Exposed                                       | 13064.88                |
| Person year non-exposed                                   | 330999.1                |
| HR for Subjects with normal APRI score (n= 81,173)        | 1                       |
| HR (CI) for Subjects with abnormal APRI score (n= 5,308)  | 41.0 (35.6-47.2)        |

Abbreviations: CI: confidence interval; HR: hazard ratio.

### 7.3.2.2.3 Patients with normal and abnormal FIB-4 indexes

A Cox model of the relationship between FIB-4 index and the occurrence of adverse liver outcomes with an interaction term was significantly better at describing the data than one without ( $P < 0.0001$  by likelihood ratio test). In Window 1, the HR was 5.2 (95% CI: 3.4-8.0). Window 2 revealed a slight peak in risk with an HR of 5.8 (95% CI: 5.0-6.7). Window 3 maintained a similar level

of risk, with an HR of 5.7 (95% CI: 4.3-6.7). This pattern indicates a relative stability in the risk associated with abnormal FIB-4 scores over time, closely mirroring the original full period analysis. More details can be found in Table 7-15.

*Table 7-15: Univariate Cox regression model of the incidence of the combined outcomes for patients with normal and abnormal Fib-4 index across different time periods.*

| <b>Fib-4 index</b>                                         | <b>combined outcome</b> |
|------------------------------------------------------------|-------------------------|
| <b>Window 1 (1990-1999)</b>                                |                         |
| Person year Exposed                                        | 36283.15                |
| Person year non-exposed                                    | 47241.57                |
| HR for Subjects with normal Fib-4 index (n= 4,758)         | 1                       |
| HR (CI) for Subjects with abnormal Fib-4 index (n= 3,615)  | 5.2 (3.4-8.0)           |
| <b>Window 2 (2000-2009)</b>                                |                         |
| Person year Exposed                                        | 87962.61                |
| Person year non-exposed                                    | 82026.34                |
| HR for Subjects with normal Fib-4 index (n= 52,624)        | 1                       |
| HR (CI) for Subjects with abnormal Fib-4 index (n= 53,982) | 5.8 (5.0-6.7)           |
| <b>Window 3 (2010-2016)</b>                                |                         |
| Person year Exposed                                        | 416154.4                |
| Person year non-exposed                                    | 456252.4                |
| HR for Subjects with normal Fib-4 index (n= 21,208)        | 1                       |
| HR (CI) for Subjects with abnormal Fib-4 index (n= 22,617) | 5.7 (4.3-6.7)           |

Abbreviations: CI: confidence interval; HR: hazard ratio.

## 7.4 Discussion

### 7.4.1 Interpreting the association between biomarkers and advanced CLD

The main findings of this study show that the risk of developing any advanced liver outcome was higher amongst those with abnormal markers (any) than with normal markers. Moreover, when using conventional cut-offs, the study revealed that abnormal APRI scores showed a stronger association with



the development of the outcomes compared to abnormal AST/ALT ratio and Fib4 index.

This study has several strengths. It is the first to examine the efficacy of biomarkers in predicting advanced liver diseases among the primary care population in the UK. The use of routine data ensures real-world applicability, and linking multiple databases (CPRD, HES, cancer, and death registries) enhance outcome ascertainment, capturing cases not recorded in primary care data.

The adoption of a consensus code list facilitates global collaboration, allowing comparisons of cirrhosis patient groups across different regions, enhancing the generalizability of the results. The extended follow-up period, spanning three decades, is particularly advantageous for capturing the slow progression of liver diseases, providing rich longitudinal data.

However, some limitations should be acknowledged. The findings may have limited generalizability beyond the UK context, especially concerning the IMD scores. Additionally, potential bias may arise from the inclusion of a sample based on general practitioners' decisions to order tests. This selection could be influenced by various factors such as clinical symptoms, patient history, or the practitioner's biases, affecting the generalizability of the findings. While there is potential for ascertainment bias due

to GPs recognizing abnormal biomarker scores, this bias is unlikely to fully explain the increased rates of severe outcomes like transplantation and death, as these conditions are typically well-documented and diagnosed by the time they reach critical stages. This supports the argument that the observed increased risk of adverse liver outcomes reflects a genuine heightened risk.

It is important to note that the study did not investigate how many patients received secondary care consultations for their liver conditions before experiencing adverse outcomes. The reliability of coding in outpatient clinics by specialty presents challenges in accurately tracking such consultations. Additionally, the potential to prevent these adverse outcomes with earlier medical intervention remains unproven within the context of my research. These considerations highlight significant avenues for future investigations, emphasizing the need for detailed studies to understand the full spectrum of clinical management and intervention possibilities for patients identified with abnormal liver biomarkers.

In the context of primary care, the utilization of non-invasive biomarkers such as AST/ALT, APRI, and FIB-4 represents a transformative advance for predicting and managing advanced liver disease outcomes. The extensive application of non-invasive biomarkers has been recognized as a diagnostic

and prognostic tool for advanced liver diseases (See [1.8.4](#)). However, there is still a lack of thorough understanding about how to best utilize these markers in primary care clinical practice for pragmatic case identification of patients at higher risk of developing advanced CLD.

The AST/ALT ratio, extensively studied in primary care settings by Giannini et al. (140) and Wang et al. (143), demonstrates its effectiveness as a diagnostic tool for liver disease. Similarly, the APRI score, validated by Wai et al. (148) and further supported by Zhu et al. (151), is reliable in predicting fibrosis and cirrhosis and can distinguish between cirrhotic and non-cirrhotic states, particularly in HBV patients. The FIB-4 index, originally validated in HIV/HCV co-infected patients, has shown high accuracy in diagnosing cirrhosis, as evidenced by subsequent studies (156).

The biomarkers relevance in primary care were underscored by Oliveira et al.'s (155), study in a Brazilian cohort, highlighting its practicality in routine clinical practice and potential to reduce the need for liver biopsies. Poynard et al.'s (160) comprehensive analysis supports this shift. Their findings validate the significant prognostic value of APRI score and Fib4 index and endorse their integration into clinical pathways, thus enabling

early intervention and more nuanced patient management strategies.

The observation that patients with elevated biomarkers have an increased risk of adverse liver outcomes underscores the robustness of these markers. Despite potential biases, the higher rates of severe outcomes like transplantation and death confirm a genuine heightened risk, highlighting the importance of early identification of high-risk patients for timely intervention. Notably, the stronger association of APRI scores with adverse outcomes emphasizes its potential as a frontline screening tool in primary care.

Furthermore, this study demonstrates that an abnormal AST/ALT ratio, APRI score, and Fib4 index are linked to an increased risk across various liver disease outcomes. Existing literature, including key studies, has shown the effectiveness of these biomarkers in predicting cirrhosis in patients with specific liver conditions. Typically, these biomarkers have been applied to patients with confirmed diagnoses. In contrast, my research highlights the utility of these biomarkers for pragmatic case finding of patients at higher risk of developing advanced CLD within a broader primary care context.

For instance, Angulo et al. (145) identified independent predictors of severe hepatic fibrosis in individuals with NASH, finding an association between an AST/ALT ratio greater than one and severe liver fibrosis. Similarly, McPherson et al. (146) evaluated the diagnostic effectiveness of non-invasive tests such as the FIB-4 score and AST/ALT ratio for NAFLD patients, demonstrating the utility of the AST/ALT ratio in ruling out advanced fibrosis. Park et al. (147) focused on the AST/ALT ratio's specificity in diagnosing cirrhosis among HCV patients, supporting my findings regarding the biomarker's significance. My study builds on these insights by exploring the broader application of these biomarkers in predicting advanced liver diseases in a general primary care population, illustrating their wider relevance in liver disease management. By integrating these findings, my study highlights the potential of these biomarkers not only in diagnosis but also in the early prognostic assessment of liver fibrosis in a primary care setting, demonstrating their adaptability in different healthcare environments.

Building on the evidence of elevated biomarkers predicting adverse liver outcomes, Lee et al. (267) demonstrated a strong association between elevated serum aminotransferase levels and increased mortality from various causes, including CVD and liver disease, among Korean patients. This study underscores the

potential of serum aminotransferase levels as predictive markers for mortality risk and emphasizes their importance in clinical assessments. Notably, patients with sustained elevation over two consecutive tests had a significantly higher risk of CVD mortality compared to those with normal levels. Furthermore, elevated AST levels were specifically linked to increased liver disease mortality, reinforcing the need for continuous monitoring of liver enzyme levels, particularly in patients at risk for CLD, early HCC, and NASH. These findings align with my study's conclusion that non-invasive biomarkers are crucial for early identification and management of high-risk patients in primary care settings.

Further supporting my findings, APRI and FIB-4 have been recommended by numerous guidelines, including those of the WHO, to determine the stage of fibrosis in resource-limited settings (122-124). Implementing these scores in pragmatic case finding means integrating these tests into routine general practice to opportunistically identify patients at elevated risk of advanced liver disease. This approach would involve initiating further interventions, follow-up, and case management for patients identified as high-risk during routine visits, thereby enabling early and targeted interventions. For example, patients who have elevated biomarker scores during routine blood tests could be flagged for additional testing or referred to specialists for further

evaluation and management. This proactive approach could lead to earlier detection of liver disease, potentially improving patient outcomes through timely intervention.

Although the FIB-4 index shows a lower association with advanced liver disease compared to the APRI score, it still correlates more strongly than the AST/ALT ratio.

These findings are critical in light of the EASL-Lancet Liver Commission's emphasis on shifting from a diagnostic-centric approach to one that proactively identifies individuals at heightened risk of severe disease progression (268). The current approach, which is predominantly diagnostic, falls short of this objective (129, 269).

Continuous observation of patients, even after their APRI or FIB-4 scores normalize, is necessary. This implies advocating for serial testing, especially in populations with a history of abnormal scores, to ensure that intermittent improvements do not mask underlying chronic liver conditions. Regular monitoring can help detect any relapse or progression of liver disease early. For instance, patients who initially present with abnormal biomarker scores but later show normal results should still undergo periodic re-testing to ensure that any potential recurrence or worsening of the condition is promptly identified and managed. However, it is

important to acknowledge that continuous follow-up and serial testing can lead to increased healthcare costs due to the additional tests required. Each follow-up visit involves costs related to laboratory tests and clinical evaluations. This can place a significant financial burden on healthcare systems, especially in resource-limited settings.

These results underscore the importance of adopting non-invasive biomarkers in primary care to enhance early diagnosis and management strategies, aligning with global health guidelines and supporting a proactive approach to liver disease management

Liver disease, particularly in its early stages, can exhibit considerable unpredictability and may not adhere to a linear progression (270, 271). A patient who initially manifests an abnormal score suggestive of potential liver fibrosis or injury might subsequently demonstrate normal outcomes in subsequent examinations due to possible factors such as temporary enhancements in hepatic function, modifications in lifestyle or pharmacotherapy, or even the inherent variability of the diagnostic tests (270, 272). Nevertheless, this restoration of normality does not necessarily ensure a lasting reversal of hepatic impairment or a substantial reduction in the risk of fibrosis.



Initial univariate analysis revealed a significant association between being in the abnormal scores' groups and advanced liver disease outcomes. However, to examine whether these associations were influenced by factors such as age, sex, and socioeconomic status, a multivariate Cox regression model with interaction terms was employed. This adjusted analysis demonstrated that the elevated risk associated with these liver markers persisted even when accounting for these potential effect modifiers. The final model, incorporating interaction terms, confirmed that the identified markers are robust predictors of disease progression.

The justification for including these covariates stems from their well-documented influence on the progression of liver disease. Gender differences have been observed in the epidemiology and natural history of liver diseases, with some studies suggesting variations in the progression rates and outcomes between males and females (264, 265).

Age at diagnosis is a critical factor, as the risk and severity of liver diseases, including cirrhosis, can increase with age due to cumulative exposure to risk factors and a possible decline in liver regenerative capacity (24, 273, 274). Socioeconomic status is included as a covariate due to its potential impact on health outcomes, where lower socioeconomic status may be associated

with barriers to healthcare access, differences in lifestyle factors, and increased exposure to environmental or occupational hazards that can influence liver disease progression.

To further elucidate the effect of these covariates on the association between liver markers and advanced liver outcomes, a stratified analysis was conducted.

#### **7.4.2 Stratification analyses.**

##### **7.4.2.1 Evaluating biomarkers' performance across varied strata.**

My analysis highlighted significant variations in the performance of the AST/ALT ratio, APRI score, and FIB-4 index across different strata of sex, age, and socioeconomic status, demonstrating statistical significance ( $p < 0.05$ ). These findings underscore the importance of considering individual patient characteristics in the risk stratification and management of advanced CLD within primary care settings.

The study identified sex-specific risks associated with abnormal biomarker scores, with the AST/ALT ratio posing a higher risk for males and APRI for females, while the FIB-4 index showed similar risks for both genders. This distinction was not observed in the FIB-4 index, where both genders demonstrated

similar risks. There is a notable gap in the existing literature regarding sex-stratified univariate analysis of these markers.

Several factors may explain these differences. Women tend to have lower protein mass and liver enzyme levels, which can lead to less variation in biomarkers such as the AST/ALT ratio and APRI. This lower baseline variability might make it easier to detect abnormalities in males compared to females. Additionally, hormonal differences between men and women could influence liver enzyme levels and their response to liver damage (275).

A correlation was found between normal enzyme levels predominantly in females, as reported by Scutt et al. (137), and the increased incidence of abnormalities in certain biomarkers identified in my analysis. Furthermore, the study indicated a notable decrease in the excess hazard for developing the studied outcomes among individuals with abnormal biomarkers as age increases, observed across all age-stratified subpopulations. This pattern suggests that age modulates the relationship between biomarker abnormalities and the risk of developing advanced liver diseases.

The decrease in biomarker effectiveness with age could be due to age-related physiological changes in the liver, such as reduced liver mass and function, which impact the reliability of these biomarkers. Additionally, older individuals are more likely to

have multiple comorbidities, which can interfere with the specificity of liver biomarkers for diagnosing liver disease (276). Competing risks, such as cardiovascular diseases and cancer, may also overshadow liver-related outcomes in older populations, making it more challenging to identify liver-specific abnormalities.

The observed decrease in excess hazard can be attributed to age-related physiological changes in the liver, including a decrease in liver mass and function, impacting the reliability of these biomarkers. Reduced liver function can lead to alterations in biomarker levels or impaired clearance, affecting their sensitivity and specificity. Competing risks in older populations, who are more susceptible to various health complications or diseases, can skew the results towards a null association of abnormal biomarkers with liver death, overshadowing liver disease progression.

It is also important to consider the testing patterns among different demographics. Younger men (e.g., a 30-year-old man) are less likely to have LFTs unless there are specific indications, whereas younger women (e.g., a 30-year-old woman) are more likely to be tested due to factors such as pregnancy, contraceptive use, and more frequent healthcare visits (277). This increased testing frequency in women might result in higher detection rates of abnormalities. In contrast, older men (e.g., an 85-year-old man)

are very likely to have had their LFTs checked due to increased healthcare interactions and comorbidities typical of advanced age (278). However, older men are generally sicker and die younger, which might explain why abnormal biomarker scores are more predictive in this age group (279).

The population differences in testing patterns could affect the observed performance of the biomarkers. For instance, the biomarkers might appear to work better in populations that are tested less frequently, such as younger men and the least deprived groups, because these groups are less likely to have competing health issues that could confound the test results. Recognizing these population differences is crucial for understanding the varying performance of the tests and adjusting future risk thresholds accordingly.

These findings suggest that future risk thresholds might need to be different according to age and sex. For example, lower cut-off values might be more appropriate for older adults and women to account for their lower baseline enzyme levels and reduced liver function. This approach could improve the sensitivity and specificity of biomarkers in these subgroups, enhancing their predictive value for advanced liver disease.

Additionally, competing risks in older populations, who are more susceptible to various health complications or diseases, can

skew the results towards a null association of abnormal biomarkers with liver death, overshadowing liver disease progression. The identification of sex-specific risks complements the age-sex interaction analysis presented in Chapter 5. While females are more likely to present with abnormalities in the AST/ALT ratio and FIB-4, indicating underlying physiological and possibly pathological distinctions, males show a distinct pattern of risk associated with the APRI score.

Furthermore, the slight increase in risks among patients with the least deprivation could be attributed to the idea that individuals with higher socioeconomic status generally have better access to healthcare facilities, including regular check-ups, preventive screenings, and early detection of diseases (280). However, even among these patients, there may still be some individuals who face barriers to accessing healthcare services (281).

#### **7.4.2.2 Temporal analysis of biomarker-associated risk across distinct time intervals**

My stratification analysis revealed statistically significant changes over time in the risk associated with liver biomarkers, validated by the likelihood ratio test ( $p < 0.05$  for all). The AST/ALT ratio showed a gradual increase in HR, peaking between 2010 and 2016. The APRI score's HR spiked notably in

recent years, highlighting its growing importance in detecting liver disease risk. Meanwhile, the FIB-4 index demonstrated stable HRs over time, suggesting consistent prognostic value. This evolving pattern of biomarker efficacy underscores the dynamic nature of biomarker performance over different periods and supports the central thesis objective to assess the practical utility of these non-invasive tools in primary care settings for liver disease prediction.

The decline in AST testing frequency, as noted in Scutt et al.'s UK study (137), is particularly relevant to my thesis. The increasing HR for APRI scores suggests that APRI may be a more targeted indicator of liver disease risk in primary care. This trend highlights the potential of APRI in the early identification of patients needing further investigation or intervention for advanced liver diseases, underscoring the importance of considering temporal changes in biomarker performance when developing clinical guidelines and strategies for liver disease management.

Furthermore, the second and third time periods (2000-2009 and 2010-2016) have full linkage to HES data, whereas the first period (1990-1999) only includes two years of such linkage. This increased availability of linked outcome data in the later periods likely enhances the robustness of the study findings. The sensitivity analysis stratifying by the three time periods examines

the effect of the change in available data through the study. Additionally, it considers the variation in the availability of linked data, as HES follow-up started later, from January 2, 1998, to December 31, 2016. This approach allows us to assess both the impact of the reduction in AST testing and the change in the availability of linked data for outcomes.

The results suggest that the comprehensive data available in the later periods strengthens the observed associations and provides a more detailed understanding of biomarker performance. When interpreting the temporal trends observed in liver disease prevalence and associated outcomes, it is important to note that where there is no significant change between periods, it suggests stability in the prevalence and outcomes of liver disease over time. This stability could indicate that despite variations in testing practices and data availability, the underlying prevalence and burden of liver disease remained relatively constant. These factors must be considered to ensure a balanced understanding of the temporal trends and their implications.

## **7.5 Conclusion**

The EASL-Lancet Liver Commission emphasises the significance of refocusing CLD diagnostic efforts on those at the highest risk of developing severe disease (268). However, the current paradigm, which focuses exclusively on a diagnostic



perspective rather than a prognostic one (129, 269), is inadequate for achieving this objective.

The analysis conducted suggests that abnormal markers are linked to a higher risk of developing advanced liver outcomes compared to normal markers. Among these, abnormal APRI scores demonstrated a stronger association with these outcomes, in contrast to the abnormal AST/ALT ratio and FIB-4 index.

While abnormal APRI scores demonstrated a stronger association with these outcomes, it is important to consider that the thresholds used for APRI, FIB-4, and AST/ALT were not directly comparable, which influenced the results. Specifically, the higher threshold set for APRI resulted in fewer patients being classified as high-risk compared to the thresholds for FIB-4 and AST/ALT. This discrepancy in threshold settings means that APRI's apparent stronger association may be partly due to its more stringent criteria, rather than an inherent superiority of the biomarker itself.

Therefore, it is essential to interpret these findings with caution and to recognize the potential impact of differing threshold levels on the observed outcomes. This disparity highlights the need for a balanced interpretation of biomarker performance, taking into account the specific cut-offs used for each test. Given

these considerations, the conclusion that APRI is superior to FIB-4 and AST/ALT in terms of prevalence and association should be revised. It is more accurate to state that the thresholds selected for APRI resulted in a more stringent classification of high-risk patients, which contributed to its stronger association in the initial analysis. However, this does not necessarily indicate that APRI is inherently a better biomarker.

Ultimately, the findings underscore the importance of using appropriate and comparable thresholds when evaluating the performance of different biomarkers. This balanced interpretation ensures a more accurate assessment of their utility in predicting liver disease outcomes.

The study underscores that patients with elevated biomarkers experience a rapid progression of liver diseases. This finding supports the use of these biomarkers not just for risk assessment but also for early referral and treatment, potentially improving outcomes for patients with advanced liver disease risks.

## **Chapter 8      Sensitivity analyses**

## 8.1 Objectives

- To evaluate the impact of initiating the follow-up period from the first recorded abnormal test result for the exposed group.
- To assess the impact of follow-up duration on study outcomes by evaluating participants with a minimum of five years of follow-up.
- To assess the diagnostic performance and hazard risk of different biomarkers' cut-offs in predicting liver diseases.
- To analyse the biomarker efficacy in a common cohort across the three biomarkers populations.

## 8.2 Methods

### 8.2.1 Starting follow-up for the exposed group at first abnormal test:

As mentioned in section [7.2.2.1](#), I carried out a sensitivity analysis examining whether moving the start date for the exposed group to the date of their first abnormal marker altered the estimation of the performance of the indices. This analysis in addition to the previously mentioned rationale had the added benefit of mirroring how the indices would be used in clinical practise more closely. (Clearly those who develop abnormal markers will not be referred for assessment or intervention before they do so).

### **8.2.2 Impact of follow-up duration on study outcomes:**

In this sensitivity analysis, I evaluated participants with a minimum of five years of follow-up post-index date. This approach recognizes that the duration of follow-up can significantly affect study outcomes. Specifically, participants who joined the cohort in the final year of the study period and had less than a full year of follow-up might not fully represent the disease's natural progression or the biomarkers' true efficacy. These individuals might not have had enough time for the relevant outcomes to develop or be accurately captured.

Additionally, the performance of the biomarkers could decline over longer follow-up periods if they are detecting existing disease rather than predicting future disease. Therefore, while biomarkers might show strong predictive power in the short term, their effectiveness may diminish over extended periods. This potential decline highlights the need for ongoing evaluation and adjustment of follow-up strategies to ensure accurate long-term predictions.

By restricting the analysis to participants with at least five years of follow-up post-index date, I aimed to mitigate the impact of varying follow-up durations and provide a more accurate assessment of the biomarkers' long-term predictive capabilities. This approach ensures that the study outcomes reflect a more

reliable picture of disease progression and the biomarkers' efficacy over time.

### **8.2.3 Assessment of biomarker diagnostic performance and hazard risk across varied cut-offs:**

Univariate Cox modelling of the combined outcome was repeated varying the cut offs of the biomarkers. Sensitivity, specificity, NPV and PPV were also calculated at each cut off. Cut offs used were 0.3, 0.5, 0.7, 1, 1.2, 1.5, and 2 for AST/ALT ratios and APRI scores, and 0.3, 0.5, 0.7, 1, 1.3, 1.5, 1.7, 2, 2.3, 2.5, 2.8, 3, 3.3, and 3.5 for FIB-4. (Different biomarkers cutoffs has been discussed in section 1.8.4).

### **8.2.4 Analysing biomarker efficacy in a common cohort across three biomarkers subgroups:**

To address potential biases from analysing different subpopulations for each biomarker's test performance, a common cohort was identified. This cohort included patients who had the combined outcome across all three biomarker subgroups (AST/ALT ratio, APRI score, and FIB-4 index). A Venn diagram was created to visually represent the overlap among these subgroups.

## **8.3 Results**

### **8.3.1 changing the index date for the exposed group.**

As can be seen in Table 8-1, adjusting the start time of the group with abnormal biomarkers so that their follow up started only when the first abnormal biomarker was calculable made little difference to the performance of the biomarkers. The HR for AST/ALT rose slightly from 1.8 (95%CI: 1.6-2.0) to 1.9 (95% CI: 1.7-2.1). APRI score showed a marginally reduced risk to 29.0 (95% CI: 28.5 to 30.5). FIB-4 index indicated a minor decrease in risk from 5.6 (95% CI: 5.0 to 6.3), to 5.4 (95% CI: 4.8 to 6.1).

Table 8-1: Univariate Cox regression model of the incidence of the (combined outcome) for patients with normal and abnormal AST/ALT, APRI score, and Fib4 when the index date for the exposed group adjusted to be the date of the first abnormal score calculated compared to the original analysis results.

| Number of subjects in normal and abnormal groups         | combined outcome<br>(original analysis) | combined outcome<br>(sensitivity analysis) |
|----------------------------------------------------------|-----------------------------------------|--------------------------------------------|
| <b>AST/ALT</b>                                           |                                         |                                            |
| Person year Exposed                                      | 874422                                  | 867204.7                                   |
| Person year Non-exposed                                  | 520028.3                                | 520028.3                                   |
| HRs for Subjects with normal AST/ALT (n= 68,465)         | 1                                       | 1                                          |
| HRs (CI) for Subjects with abnormal AST/ALT (n= 118,039) | 1.8 (1.6-2.0)                           | 1.9 (1.7-2.1)                              |
| <b>APRI</b>                                              |                                         |                                            |
| Person year Exposed                                      | 85793.83                                | 85270.95                                   |
| Person year non-exposed                                  | 3847876                                 | 3847876                                    |
| HR for Subjects with normal APRI score (n= 464,204)      | 1                                       | 1                                          |
| HRs (CI) Subjects with abnormal APRI score (n= 18,893)   | 30.8 (29.3-32.4)                        | 29.0 (95% CI: 28.5 to 30.5)                |
| <b>Fib4</b>                                              |                                         |                                            |
| Person year Exposed                                      | 540400.1                                | 540400.1                                   |
| Person year Non-exposed                                  | 585520.3                                | 585520.3                                   |
| HR for Subjects with normal<br>Fib4 index (n= 77,590)    | 1                                       | 1                                          |
| HRs (CI) Subjects with abnormal Fib4 index (n= 80,214)   | 5.6 (5.0-6.3)                           | 5.4 (95% CI: 4.8 to 6.1)                   |

Abbreviations: CI: confidence interval; HR: hazard ratio.



### **8.3.2 Minimum Five-Year of Follow-Up**

As can be seen in Table 8-2, the sensitivity analysis, which focused on patients with at least five years of follow-up made some differences to the performance of the biomarkers

The HR for AST/ALT slightly decreased from 1.8 (95%CI: 1.6-2.0) to 1.6 (95% CI: 1.4-1.8). APRI score showed a notably decrease in the risk from 30.8 (95% CI: 29.3 to 32.4) to 18.6 (95% CI: 17.0-20.4). FIB-4 index showed slightly decreased risk from 5.6 (95% CI: 5.0 to 6.3) to 5.0 (95% CI: 5.2-6.0).

Table 8-2 : Univariate Cox regression model of the incidence of the (combined outcome) for patients with normal and abnormal AST/ALT, APRI score, and Fib4 restricted for patients with at least 5 years of follow-up.

| Number of subjects in normal and abnormal groups | combined outcome<br>(original analysis) | combined outcome<br>(sensitivity analysis) |
|--------------------------------------------------|-----------------------------------------|--------------------------------------------|
| <b>AST/ALT</b>                                   |                                         |                                            |
| Person year exposed                              | 874422                                  | 876745.3                                   |
| Person year non exposed                          | 520028.3                                | 467253.8                                   |
| HR for Subjects with normal AST/ALT              | 1                                       | 1                                          |
| HR (CI) for Subjects with abnormal AST/ALT       | 1.8 (1.6-2.0)                           | 1.6 (1.4- 1.8)                             |
| <b>APRI</b>                                      |                                         |                                            |
| Person year Exposed                              | 85793.83                                | 70173.45                                   |
| Person year non exposed                          | 3847876                                 | 3497408                                    |
| HR for Subjects with normal APRI                 | 1                                       | 1                                          |
| HR (CI) for Subjects with abnormal APRI          | 30.8 (29.3-32.4)                        | 18.6 (17.0 – 20.4)                         |
| <b>Fib4</b>                                      |                                         |                                            |
| Person year Exposed                              | 540400.1                                | 462700.6                                   |
| Person year non exposed                          | 585520.3                                | 524053.4                                   |
| HR for Subjects with normal Fib4                 | 1                                       | 1                                          |
| HR (CI) for Subjects with abnormal Fib4          | 5.6 (5.0-6.3)                           | 5.0 (5.2- 6.0)                             |

Abbreviations: CI: confidence interval; HR: hazard ratio.

### **8.3.3 Univariate analysis of different cut-offs and biomarkers diagnostic performance**

#### **8.3.3.1 AST/ALT ratio univariate analysis of different cut-offs.**

The analysis of AST/ALT ratio at various cut-offs, as detailed in Table 8-3 and illustrated in Figure 8-1, shows that the exposed as defined at almost all cut-offs have greater risk of outcomes than the unexposed, but that the magnitude of this effect was greater for higher cut-offs. The exceptions were the lowest 2 cut-offs with the lowest of 0.3 showing a HR of 0.8 (95% CI: 0.4-1.5), which escalated to 7.7 (95% CI: 7.0-8.4) at the highest cut-off of 2.0.

As might be expected I found that there was an inverse relationship between the changes in sensitivity and specificity with varying cut-offs, with rising cut-off being associated with increasing specificity and PPV but falling sensitivity. The sensitivity was at its peak of 99.7% at the 0.3 cut-off but significantly reduced to 24.9% at the 2.0 cut-off. At the 0.3 cut-off, specificity was minimal at 0.2%, but it markedly increased to 93.8% at the 2.0 cut-off. This improvement implies that the AST/ALT ratio becomes more proficient in correctly identifying

individuals without the disease as the cut-off value rises. At a sensitivity of about 75-80% PPV for AST/ALT is 1.8%.

Table 8-3: Univariate Cox regression model of the incidence of the combined outcomes for patients with normal and abnormal AST/ALT ratio using different cut-offs.

| cut-off    | Subjects with normal and abnormal score     | Hazard risk (CI) | Sensitivity | Specificity | PPV  | NPV   |
|------------|---------------------------------------------|------------------|-------------|-------------|------|-------|
| <b>0.3</b> | Subjects with normal AST/ALT (n= 388)       | 1                | 99.7%       | 0.2%        | 1.5% | 98.4% |
|            | Subjects with abnormal AST/ALT (n= 183,063) | 0.8 (0.4-1.5)    |             |             |      |       |
| <b>0.5</b> | Subjects with normal AST/ALT (n= 2,485)     | 1                | 98.8%       | 1.4%        | 1.5% | 98.6% |
|            | Subjects with abnormal AST/ALT (n=180,959)  | 1.0 (0.7-1.4)    |             |             |      |       |
| <b>0.7</b> | Subjects with normal AST/ALT (n= 15,222)    | 1                | 93.9        | 8.3         | 1.5% | 98.9% |
|            | Subjects with abnormal AST/ALT (n=168,163)  | 1.3 (1.1-1.5)    |             |             |      |       |
| <b>1*</b>  | Subjects with normal AST/ALT (n=68,465)     | 1                | 76.5%       | 36.0%       | 1.8% | 99.0% |
|            | Subjects with abnormal AST/ALT (n= 118,127) | 1.8 (1.6-2.0)    |             |             |      |       |
| <b>1.2</b> | Subjects with normal AST/ALT (n= 96,570)    | 1                | 63.9%       | 52.9%       | 2.0% | 99.0% |
|            | Subjects with abnormal AST/ALT (n= 86,627)  | 2.3 (2.0-2.4)    |             |             |      |       |
| <b>1.5</b> | Subjects with normal AST/ALT (n= 141,280)   | 1                | 45.3%       | 77.4%       | 2.9% | 99.0% |
|            | Subjects with abnormal AST/ALT (n=41,935)   | 3.6 (3.4-3.9)    |             |             |      |       |
| <b>2</b>   | Subjects with normal AST/ALT (n= 171,440)   | 1                | 24.9%       | 93.8%       | 5.6% | 99.0% |
|            | Subjects with abnormal AST/ALT (n= 11,871)  | 7.7 (7.0-8.4)    |             |             |      |       |

\*Conventional cut-off used in the sex, age, and socioeconomic stratification.

abbreviations: CI: confidence interval, PPV: positive predictive value, NPV: negative predictive value.

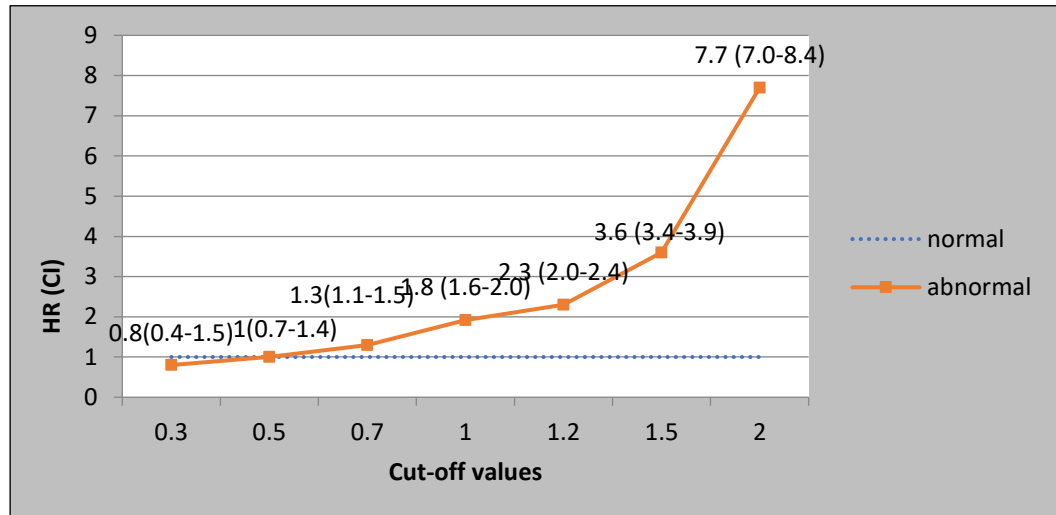


Figure 8-1: Univariate Cox regression model of the incidence of combined outcomes for patients with normal and abnormal AST/ALT ratios using different cut-off.

### 8.3.3.2 APRI score univariate analysis of different cut-offs.

Table 8-4 and Figure 8-2 revealed the same pattern of abnormal markers having a higher HR of clinical outcomes than that of the normal population. At the lower cut-off of 0.3, the HR was observed at 7.0 (95% CI: 6.6-7.4), escalating sharply to 35.0 (95% CI: 32.9-37.3) at the higher cut-off of 2.0.

As might be expected, there was a significant decrease in sensitivity alongside the rising cut-offs. In contrast, specificity exhibited an opposite pattern. Starting at a modest 63.3% at the 0.3 cut-off, it impressively rose to 99.0% at the 2.0 cut-off. PPV showed an increasing trend with higher cut-offs. At a sensitivity of about 75-80% PPV for APRI is 2.7%.

Table 8-4: Univariate Cox regression model of the incidence of the combined outcomes for patients with normal and abnormal APRI score using different cut-offs.

| cut-off    | Subjects with normal and abnormal score        | Hazard risk (CI) | Sensitivity | Specificity | PPV  | NPV   |
|------------|------------------------------------------------|------------------|-------------|-------------|------|-------|
| <b>0.3</b> | Subjects with normal APRI score (n= 300,726)   | 1                | 79.7%       | 63.3%       | 2.7% | 99.6% |
|            | Subjects with abnormal APRI score (n= 178,786) | 7.0 (6.6-7.4)    |             |             |      |       |
| <b>0.5</b> | Subjects with normal APRI score (n= 426,423)   | 1                | 61.6%       | 89.5%       | 7.0% | 99.4% |
|            | Subjects with abnormal APRI score (n=53,199)   | 16.0 (15.3-16.9) |             |             |      |       |
| <b>0.7</b> | Subjects with normal APRI score (n= 452,401)   | 1                | 51.4%       | 94.8%       | 11.4 | 99.3% |
|            | Subjects with abnormal APRI score (n= 27,298)  | 24.3 (23.0-25.5) |             |             |      |       |
| <b>1*</b>  | Subjects with normal APRI score (n= 464,204)   | 1                | 41.1%       | 97.2%       | 15.9 | 99.2% |
|            | Subjects with abnormal APRI score (n= 18,893)  | 30.8 (29.3-32.4) |             |             |      |       |
| <b>1.2</b> | Subjects with normal APRI score (n= 467,742)   | 1                | 35.7%       | 97.9%       | 17.8 | 99.1% |
|            | Subjects with abnormal APRI score (n= 12,006)  | 33.2 (31.4-35.0) |             |             |      |       |
| <b>1.5</b> | Subjects with normal APRI score (n= 470,991)   | 1                | 28.7%       | 98.5%       | 19.7 | 99.0% |
|            | Subjects with abnormal APRI score (n= 8,785)   | 34.3 (32.4-36.3) |             |             |      |       |
| <b>2</b>   | Subjects with normal APRI score (n= 473,972)   | 1                | 21.0%       | 99.0%       | 21.7 | 99.0% |
|            | Subjects with abnormal APRI score (n= 5,820)   | 35.0 (32.9-37.3) |             |             |      |       |

\*Conventional cut-off used in the sex, age, and socioeconomic stratification.

abbreviations: CI: confidence interval, PPV: positive predictive value, NPV: negative predictive value.

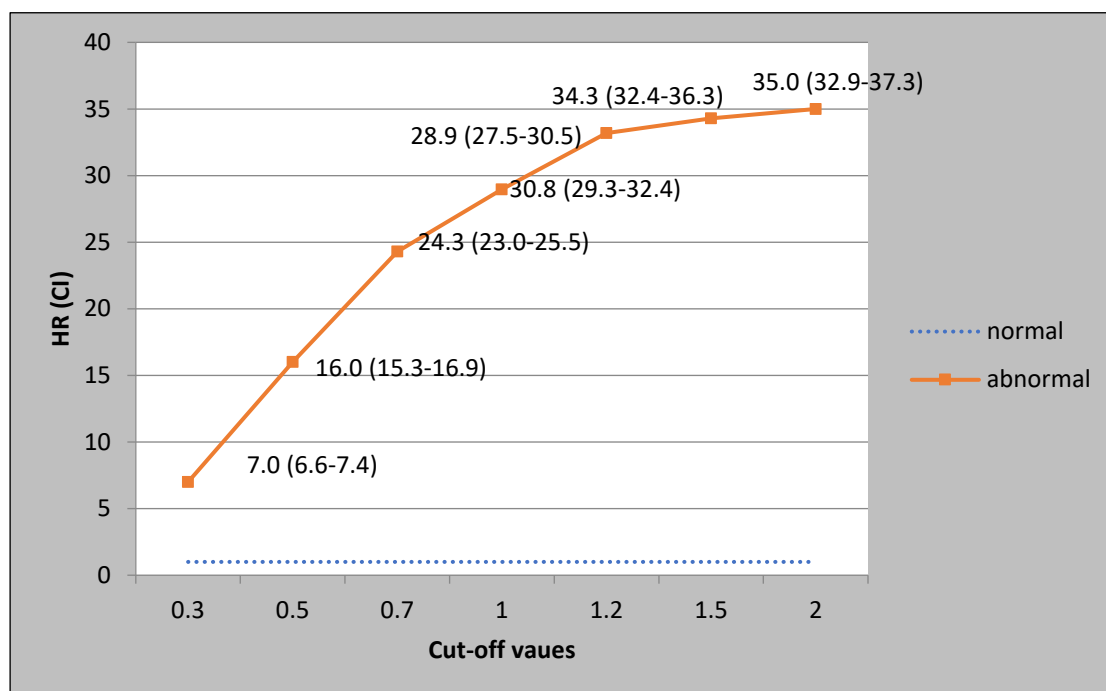


Figure 8-2: Univariate Cox regression model of the incidence of combined outcomes for patients with normal and abnormal APRI score using different cut-offs.

### 8.3.3.3 Fib-4 index univariate analysis of different cut-offs.

Table 8-5 and Figure 8-3 revealed the same pattern of abnormal markers having a higher HR of clinical outcomes than that of the normal population. At the lower cut-off of 0.3, the HR was 1.1 (95% CI: 0.3-3.4), which surged dramatically to 27.4 (95% CI: 25.1-30.0) at the higher cut-off of 3.5 and there was a notable decrease in sensitivity as the cut-offs increased. The PPV demonstrated an increasing trend with higher cut-offs. At a sensitivity of about 75-80% PPV for FIB-4 is 2.7%



Table 8-5 Univariate Cox regression model of the incidence of the combined outcomes for patients with normal and abnormal Fib4 index using different cut-offs.

| cut-off     | Subjects with normal and abnormal score         | Hazard risk (CI) | Sensitivity | Specificity | PPV  | NPV   |
|-------------|-------------------------------------------------|------------------|-------------|-------------|------|-------|
| <b>0.3</b>  | Subjects with normal Fib 4 score (n= 195)       | 1                | 99.9%       | 0.1%        | 1.4% | 98.5% |
|             | Subjects with abnormal Fib 4 score (n= 154,988) | 1.1 (0.3-3.4)    |             |             |      |       |
| <b>0.5</b>  | Subjects with normal Fib 4 score (n= 2,306)     | 1                | 99.3%       | 1.5%        | 1.4% | 99.3% |
|             | Subjects with abnormal Fib 4 score (n= 152,872) | 2.2 (1.3-3.8)    |             |             |      |       |
| <b>0.7</b>  | Subjects with normal Fib 4 score (n= 12,692)    | 1                | 97.2%       | 8.2%        | 1.5% | 99.5% |
|             | Subjects with abnormal Fib 4 score (n= 142,447) | 3.0 (2.4-3.9)    |             |             |      |       |
| <b>1</b>    | Subjects with normal Fib 4 score (n= 43,338)    | 1                | 91.4%       | 28.2%       | 1.8% | 99.5% |
|             | Subjects with abnormal Fib 4 score (n= 111,695) | 4.3 (3.7-5.0)    |             |             |      |       |
| <b>1.3*</b> | Subjects with normal Fib 4 score (n= 77,590)    | 1                | 84.0%       | 49.0%       | 2.3% | 99.5% |
|             | Subjects with abnormal Fib 4 score (n= 80,214)  | 5.6 (5.0-6.3)    |             |             |      |       |
| <b>1.5</b>  | Subjects with normal Fib 4 score (n= 92,511)    | 1                | 78.3%       | 60.2%       | 2.7% | 99.4% |
|             | Subjects with abnormal Fib 4 score (n= 62,507)  | 6.5 (5.8-7.1)    |             |             |      |       |
| <b>1.7</b>  | Subjects with normal Fib 4 score (n= 106,458)   | 1                | 73.2%       | 69.2%       | 3.3% | 99.4% |
|             | Subjects with abnormal Fib 4 score (n= 48,545)  | 7.7 (6.9-8.4)    |             |             |      |       |
| <b>2</b>    | Subjects with normal Fib 4 score (n=121,590)    | 1                | 66.7%       | 79.0%       | 4.4% | 99.4% |

|            |                                                |                  |       |       |      |       |
|------------|------------------------------------------------|------------------|-------|-------|------|-------|
|            | Subjects with abnormal Fib 4 score (n= 33,426) | 10.0 (9.2-11.0)  |       |       |      |       |
| <b>2.3</b> | Subjects with normal Fib 4 score (n= 131,587)  | 1                | 60.4% | 85.5% | 5.6% | 99.3% |
|            | Subjects with abnormal Fib 4 score (n= 23,439) | 12.6 (11.6-13.8) |       |       |      |       |
| <b>2.5</b> | Subjects with normal Fib 4 score (n= 136,368)  | 1                | 57.0% | 88.6% | 6.7% | 99.3% |
|            | Subjects with abnormal Fib 4 score (n= 18,680) | 14.9 (13.6-16.2) |       |       |      |       |
| <b>2.8</b> | Subjects with normal Fib 4 score (n= 141,401)  | 1                | 52.1% | 91.8% | 7.6% | 99.2% |
|            | Subjects with abnormal Fib 4 score (n= 13,653) | 18.1 (16.7-19.8) |       |       |      |       |
| <b>3</b>   | Subjects with normal Fib 4 score (n= 143,800)  | 1                | 49.1% | 93.3% | 8.7% | 99.2% |
|            | Subjects with abnormal Fib 4 score (n= 11,268) | 20.7 (19.0-22.5) |       |       |      |       |
| <b>3.3</b> | Subjects with normal Fib 4 score (n= 146,508)  | 1                | 45.9% | 95.1% | 11.7 | 99.2% |
|            | Subjects with abnormal Fib 4 score (n= 8,566)  | 25.3 (23.2-27.5) |       |       | %    |       |
| <b>3.5</b> | Subjects with normal Fib 4 score (n= 147,840)  | 1                | 43.1% | 95.9% | 13.0 | 99.2% |
|            | Subjects with abnormal Fib 4 score (n= 7,236)  | 27.4 (25.1-30.0) |       |       | %    |       |

\*Conventional cut-off used in the sex, age, and socioeconomic stratification.

abbreviations: CI: confidence interval, PPV : positive predictive value, NPV : negative predictive value.

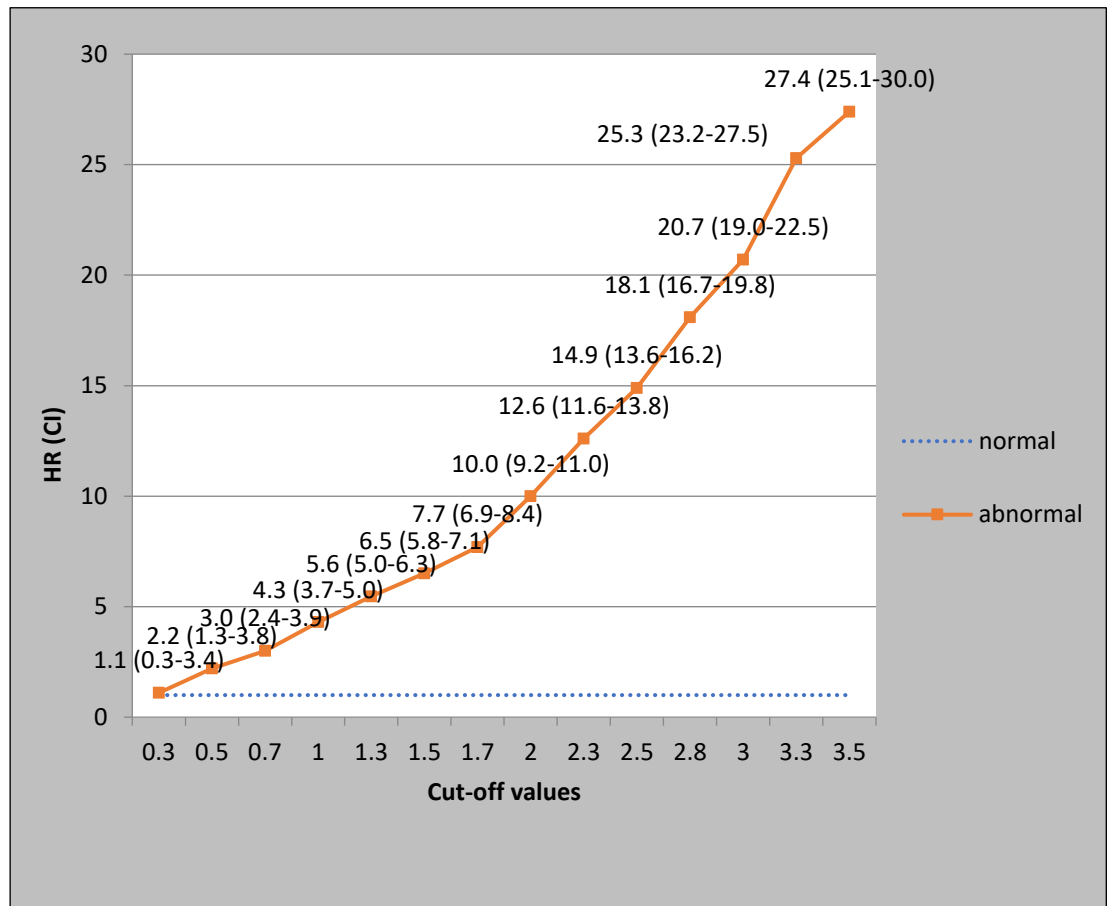


Figure 8-3 Univariate Cox regression model of the incidence of combined outcomes for patients with normal and abnormal FIB4 index using different cut-offs

### 8.3.4 Cox regression and hazard ratio among people who had records in the three cohorts.

Because the populations studied for each biomarker were slightly different as illustrated in Figure 8-4, I carried out an analysis limited to the 157,757 (31%) patients included in all 3 populations. This was done to check whether the differing populations substantially biased the results. The results within this subpopulation were very similar to those of the primary

analysis as shown in Table 8-6. AST/ALT subpopulation within the common cohort presented a slight increased risk from 1.8 (95% CI: 1.6 to 2.0) reported in the original group to 1.9 (1.7-2.1). APRI subpopulation showed a slight increased risk from 30.8 (95% CI: 29.3 to 32.4) to 33.4 (30.7-36.3), and Fib4 showed a slight decreased risk from 5.6 (95% CI: 5.0 to 6.3) to 5.5 (4.9-6.2).

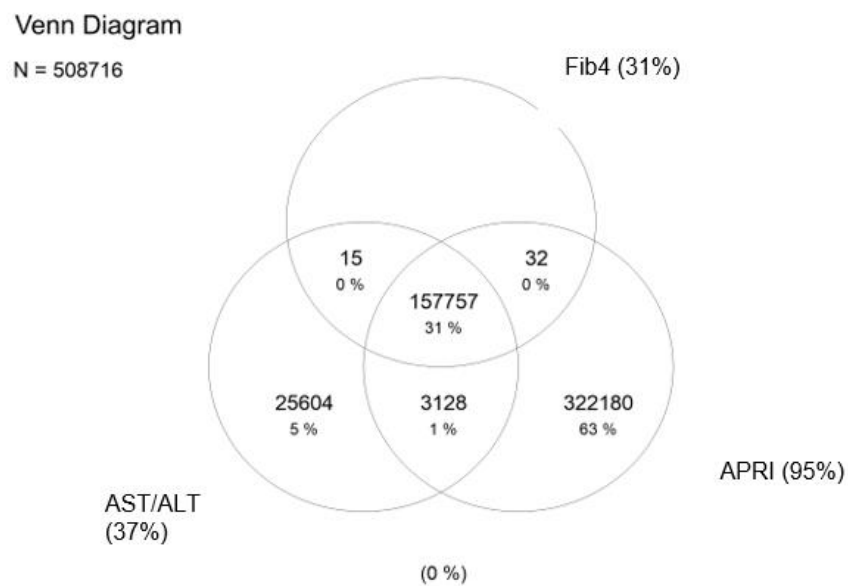


Figure 8-4 Venn diagram showing the overlap among patients in different cohorts.

Table 8-6: Univariate Cox regression model of the incidence of the (combined outcome) for patients with normal and abnormal AST/ALT, APRI score, and Fib4 patients included in all 3 populations compared to the original analysis results.

| Number of subjects in normal and abnormal groups        | combined outcome (original analysis) | combined outcome (sensitivity analysis) |
|---------------------------------------------------------|--------------------------------------|-----------------------------------------|
| <b>AST/ALT</b>                                          |                                      |                                         |
| Person year exposed                                     | 874422                               | 865397.3                                |
| Person year non exposed                                 | 520028.3                             | 410472.2                                |
| HR for Subjects with normal AST/ALT (n= 53,997)         | 1                                    | 1                                       |
| HR (CI) for Subjects with abnormal AST/ALT (n= 103,760) | 1.8 (1.6-2.0)                        | 1.9 (1.7-2.1)                           |
| <b>APRI</b>                                             |                                      |                                         |
| Person year Exposed                                     | 85793.83                             | 32954.04                                |
| Person year non exposed                                 | 3847876                              | 1200314                                 |
| HR for Subjects with normal APRI (n= 151,508)           | 1                                    | 1                                       |
| HR (CI) for Subjects with abnormal APRI (n= 6,249)      | 30.8 (29.3-32.4)                     | 33.4 (30.7-36.3)                        |
| <b>Fib4</b>                                             |                                      |                                         |
| Person year Exposed                                     | 540400.1                             | 540400.1                                |
| Person year non exposed                                 | 585520.3                             | 585520.3                                |
| HR for Subjects with normal Fib4 (n= 77,888)            | 1                                    | 1                                       |
| HR (CI) for Subjects with abnormal Fib4 (n= 79,869)     | 5.6 (5.0-6.3)                        | 5.5 (4.9-6.2)                           |

## 8.4 Discussion

In this chapter I have described sensitivity analyses aimed at elucidating the effect of methodological decisions upon the outcomes of my studies. The results are in general robust within the sensitivity analyses, but the examination of variations in cut off values provides important additional insights to the performance of the markers.

### 8.4.1 Changing the index date for the exposed group.

Originally, the start of the follow-up was set at the day of the first test result, irrespective of its normality. However, to

explore the potential biases introduced by different follow-up start times, I also analysed outcomes starting from the first recorded abnormal result. Interestingly, despite these methodological variations potentially affecting the length of follow-up and exposure classification, the results from both approaches were largely consistent.

The similarity of the results between the two methodologies—despite their distinct biases—indicates that the overall conclusions of the study were not substantially skewed by any potential bias introduced by either. Since the updated approach may incorrectly classify unexposed subjects as exposed and the original method may overestimate exposure periods by biasing the length of follow-up, the similarity in results between the two approaches points to an important conclusion: the true effect or outcome most likely lies in the space between the two sets of findings.

The validity of the study's conclusions is supported by the alignment of the two approaches, which shows that the observed effects are a consistent pattern across various analytical strategies and are not just the result of methodological choices.

### **8.4.2 Five-year follow-up sensitivity analysis**

In this analysis of sensitivity within this study, my focus was on the evaluation of the impact of the duration of follow-up on the results, specifically in relation to the variability introduced by participants who joined the cohort in the final year without a complete year of follow-up.

For the AST/ALT ratio, the hazard ratio adjusted modestly from 1.8 to 1.6 for outcomes manifesting after at least five years, reinforcing its sustained utility over time. Similarly, the FIB-4 index showed reliability with a slight change in the hazard ratio from 5.6 to 5.0 over longer durations, confirming its effectiveness as a long-term predictive tool for advanced liver disease. Conversely, the APRI score demonstrated a significant reduction in predictive capability over time, with the hazard ratio decreasing from 30.8 to 18.6. Despite this reduction, the APRI score still showed a stronger association with risk than the AST/ALT ratio and FIB-4 index, indicating its potential for early identification of individuals at heightened risk. However, its diminishing predictability over extended follow-up highlights the complexities of using this biomarker for long-term risk stratification.

Linking these findings with the study by Poynard et al (160), add significant depth to our understanding of the variable performance of liver disease biomarkers over time, particularly in

relation to their prognostic and diagnostic utility. The observed shifts in hazard ratios for AST/ALT, APRI, and FIB-4 across different follow-up durations in my study suggest nuances in how these biomarkers function in the early detection and long-term monitoring of liver disease. Specifically, the diminishing predictability of the APRI score over extended periods, as seen by a notable drop in its hazard ratio, resonates with the findings from Poynard et al., where FibroTest consistently demonstrated a higher prognostic value than APRI for survival without liver-related deaths.

The change in APRI performance over extended follow-up periods may be due to its sensitivity in detecting advanced CLD at earlier stages when symptoms are not yet manifest. This early detection allows for timely intervention, contributing to its strong initial association with liver disease risk. However, as the disease progresses or during longer monitoring periods, the APRI's predictive power diminishes, reducing its relative contribution to risk stratification compared to its initial assessment.

A study by Lin et al. (154), examined the use of APRI in early disease diagnosis, highlighting its moderate utility in identifying clinically significant fibrosis or cirrhosis. This research underscores APRI's value as a non-invasive alternative to liver biopsy, particularly for early detection. While APRI's predictive



accuracy may decrease over time, its effectiveness in early-stage detection remains significant. This highlights the importance of integrating APRI with other biomarkers and clinical indicators, adjusted for follow-up duration and disease progression, to enhance overall liver health assessment. This aspect complements APRI's initial strong association with liver disease risk and its diminishing long-term predictability.

The observed stability in the hazard ratio for FIB-4 across extended follow-up periods underscores its utility as a reliable long-term predictive tool for liver disease progression. This finding is consistent with Poynard et al.'s meta-analysis, which positions FIB-4 as superior in reliability compared to APRI, reinforcing its role not just in initial diagnosis but also in ongoing disease monitoring. The consistent prognostic value of FIB-4, as highlighted in my sensitivity analysis, supports its effectiveness in tracking disease progression over time. This is further corroborated by discussions in section [5.5.4](#), where the application of FIB-4 in disease monitoring was explored in depth.

### **8.4.3 Cut-off variations with the sensitivity, specificity, positive predictive value and negative predictive value.**

This analysis explored the implications of varying cut-offs for the AST/ALT ratio, APRI, and Fib-4 indices in diagnosing advanced CLD. The analysis revealed a clear trend: the hazard ratio of developing adverse liver outcomes increased with the use of higher cut-offs among the exposed group compared to the non-exposed group across all subpopulations.

The selection of cut-offs for the AST/ALT ratio, APRI score, and FIB-4 index in the original study was guided by existing literature and clinical guidelines, ensuring their relevance and robust validation. For the AST/ALT ratio, a cut-off of 1 was chosen based on its widespread acceptance in the diagnosis of liver diseases. Williams and Hoofnagle (139), demonstrated that an AST/ALT ratio of 1 or more is strongly indicative of cirrhosis in patients with NAFLD, making it a pivotal marker in clinical practice. This cut-off helps in distinguishing between various liver conditions, particularly cirrhosis, which aligns with my study's focus on advanced CLD.

Similarly, the cut-off of 1 for the APRI score is recommended by the WHO as a threshold for identifying patients

at higher risk of liver fibrosis and cirrhosis. The FIB-4 index cut-off of 1.3, commonly used in clinical practice, effectively distinguishes patients with minimal or no fibrosis from those with significant fibrosis. Shah et al.'s (158), study confirmed the utility of this threshold in NAFLD patients, highlighting its importance in non-invasive fibrosis assessment.

These cut-offs were selected because they are specifically relevant for identifying cirrhosis and are widely accepted in clinical guidelines. While these cut-offs were selected for their validation and utility, it is crucial to acknowledge that different studies might suggest alternative thresholds based on specific populations or disease severities. These variations can be attributed to differences in study design, population characteristics, and the specific outcomes being measured, whether it be severe fibrosis, cirrhosis, or the likelihood of current liver damage. For example, some studies may propose different cut-offs for viral hepatitis cohorts or general population studies, which could influence the observed diagnostic performance of the biomarkers.

To address these potential variations, my study included a comprehensive sensitivity analysis evaluating different cut-offs for the biomarkers. This analysis demonstrated that the chosen thresholds provided robust diagnostic performance across various scenarios. By assessing alternative cut-offs, we ensured

that my selected thresholds were not only appropriate for the specific populations studied but also adaptable to different clinical contexts. This sensitivity analysis reinforces the validity and reliability of my chosen cut-offs, emphasizing their clinical applicability and the importance of continuous evaluation to maintain diagnostic accuracy.

The sensitivity and specificity of each biomarker varied significantly across different thresholds. As expected, a lower cut-off generally improves sensitivity but at the expense of specificity, leading to more false positives. Conversely, a higher cut-off enhances specificity but reduces sensitivity, potentially missing true cases. For instance, at the lowest cut-offs (AST/ALT 0.3, APRI 0.3, and Fib-4 0.3), sensitivity was maximized (99.7%, 79.7%, and 99.9%, respectively), but specificity was notably low (0.2%, 63.3%, and 0.1%, respectively). This indicates a high number of false positives alongside a low PPV (AST/ALT 1.5%, APRI 2.7%, and Fib-4 1.4%), which could lead to unnecessary follow-up tests and treatments in secondary care settings.

Moreover, considering that selecting different cut-offs for each biomarker can influence which test is preferred, I compared the AST/ALT, APRI, and Fib-4 based on their sensitivity and PPV. This analysis at approximate 50% sensitivity levels highlight

significant differences in their PPV, which is crucial for managing the impact on secondary care.

At a 1.5 cut-off, the AST/ALT ratio has a sensitivity of 45.3% and a PPV of 2.9%, while Fib-4 at a 3.0 cut-off reaches a sensitivity of 49.1% with a PPV of 8.7%. Both tests exhibit moderate sensitivity, but their low PPVs indicate a lower likelihood of confirming disease in those tested positive, leading to significant implications for secondary care.

Conversely, APRI provides a superior combination of sensitivity and PPV. At a 0.7 cut-off, APRI shows a sensitivity of 51.4% and a notably higher PPV of 11.4%. This favourable balance suggests APRI not only detects a significant portion of true cases but also assures a higher likelihood that those identified positively actually have the disease. The higher PPV at this sensitivity level means APRI could effectively reduce unnecessary referrals and treatments in secondary care, optimizing resource utilization.

In general, different studies have proposed different cutoff values for AST/ALT ratio, APRI, and FIB-4 markers depending on the specific stage of fibrosis or cirrhosis being diagnosed. Previous research has mainly focused on identifying and staging liver fibrosis and cirrhosis in patients who already have the

disease rather than predicting the development of CLD (46, 109, 147, 148, 154, 155, 158, 282-286).

The studies by Lin et al. (131) and Shah et al. (158), provide valuable insights into the use of biomarker cutoffs to diagnose stages of liver disease, which directly relate to the findings in my analysis of AST/ALT ratio, APRI, and Fib-4. Lin et al. revealed that varying APRI cutoffs impact the sensitivity and specificity for diagnosing hepatitis C-related fibrosis, noting that a higher APRI cutoff (2.0) increases specificity but reduces sensitivity. This aligns with my findings, where a 0.7 cutoff for APRI offered a balanced sensitivity (51.4%) and specificity (94.8%), demonstrating its efficacy in a clinical setting for optimizing resource use by reducing unnecessary secondary care interventions.

Similarly, Shah et al.'s research on FIB4 and other markers used to diagnose NAFLD confirms that higher cutoffs improve specificity but at the cost of sensitivity. For instance, at a FIB4 threshold of 3.25, sensitivity dropped significantly, whereas specificity increased, underscoring the trade-off between detecting more true cases and avoiding false positives. This complements my observations where a Fib-4 cutoff of 3.0 showed reasonable sensitivity (49.1%) but a higher likelihood of false positives due to its lower PPV (8.7%).

In my study, a fundamental difference exists in the patient population compared to the literature. The majority of studies cited in the literature focused on patients with confirmed diagnoses of advanced CLD. In contrast, this research focusses on those without confirmed diagnoses at the time of biomarker assessment. While this approach may not represent the gold standard for assessing diagnostic accuracy, it offers valuable insights into the potential clinical utility of different biomarker cutoffs for identifying individuals at risk of adverse liver outcomes in a real-world, long-term follow-up scenario. The interpretation of my results must therefore be within this context where the goal is to balance early detection with the risk of overdiagnosis.

#### **8.4.4 The efficacy of biomarkers within a common cohort across three subgroups**

Limiting analyses to subjects with all 3 biomarkers calculable made negligible alterations in the association of the markers with the combined outcome at standard cut-offs suggesting that the variation in the populations was not the reason for differential performance between them.

### **8.5 Conclusion**

In conclusion, the choice of start date for the exposed group did not significantly bias the results. AST/ALT and FIB-4

performed consistently well, even with a five-year follow-up, while APRI's performance declined. Varying cutoffs affected sensitivity, specificity, PPV, and NPV. Optimal cutoffs for future outcome prediction can be determined from this study's results. The variability in populations between the markers did not explain the differences in their performance.



## **Chapter 9      Overall Discussion and Conclusions**

## **9.1 Introduction**

CLD has emerged as a major global health concern due to high mortality rates and the burden placed on healthcare systems. However, CLD often goes unnoticed until it has progressed to an advanced and irreversible stage.

This study has examined NAFLD, as a leading cause of CLD, and estimated its prevalence in the UK and KSA. Furthermore, given certain limitations in exclusively targeting NAFLD, my study adjusted its focus towards exploring the use of the biomarkers in the broader context of CLD. This approach also offers additional advantages in understanding and predicting CLD more comprehensively as discussed in (section [1.7.1](#)). These, therefore, formed the basis for the second part of the study, which used routinely collected data from the CPRD to assess the effectiveness of APRI score, AST/ALT ratio and FIB-4 as prognostic tools in relation to advanced CLD.

## **9.2 Summary of main findings**

My preliminary literature review suggested that, despite the growing concern over NAFLD, comprehensive and consistent data detailing its prevalence in both the UK and KSA were scarce and fragmented. Two systematic reviews conducted to bridge this knowledge gap revealed that definitive population-based studies

are lacking. The pooled prevalence of NAFLD is 23.8% in the UK and 16.8% in KSA among disease-unrestricted populations.

Three additional studies were undertaken to assess the feasibility of using non-invasive biomarkers in predicting the progression to advanced CLD. Among the study population, cirrhosis was the primary outcome observed in the majority of patients with outcomes; the APRI score, although the most frequently calculable biomarker, had the lowest prevalence of abnormal values at standard cut-offs, whereas the AST/ALT ratio, the second most calculable, exhibited the highest prevalence of abnormal values, followed by the FIB-4 index.

My studies have evaluated the effectiveness of these indices in identifying individuals at higher risk of adverse liver outcomes. While all three indices are capable, the APRI showed the strongest correlation with these outcomes at conventional cut-offs. I also explored how performance varies with different cut-offs to help users find the optimal balance between specificity and sensitivity. Additionally, I examined which subpopulations each index performs best in.

### **9.3 Clinical sequences and implications**

NAFLD is increasingly prevalent in both the UK and the KSA, and although the evidence is limited, it suggests a rising

trend. Without intervention, it is reasonable to expect a growing number of patients to present with symptomatic or decompensated disease given this apparent increase in prevalence.

Furthermore, the application of the analysed biomarkers could effectively identify individuals at risk of developing liver disease. My findings demonstrate that the necessary data for this already exists in many GP records. However, implementing a policy to use these biomarkers would potentially necessitate a change in the transaminase levels commonly measured by laboratories.

The incorporation of these biomarkers into routine practice means using them opportunistically during patient visits in primary care settings, rather than as part of a whole population screening program. This approach could lead to proactive management and intervention for patients already undergoing blood tests or with known risk factors, potentially reducing the healthcare burden associated with advanced liver diseases by catching and managing the disease earlier. By integrating these tests into routine care, healthcare providers can identify high-risk individuals during regular visits and initiate timely interventions, ultimately improving patient outcomes and reducing long-term healthcare costs.

This approach suggests promising directions for future interventions, although specific interventions were not tested within the scope of my thesis.

This study was limited to individuals who underwent AST testing. However, the measurement of LFTs, particularly AST, may be generally performed only when patients are at a higher risk for liver diseases and hence some care is needed in extrapolating the results to the general population.

This study demonstrates that APRI is a valuable non-invasive biomarker for predicting advanced liver disease in UK general practice. While APRI showed a strong association with advanced liver outcomes, it is important to contextualize these findings within the framework of the different thresholds used for each biomarker. The thresholds for APRI, AST/ALT, and FIB-4 were not directly comparable, which influenced the observed associations.

My sensitivity analysis revealed that varying the cut-offs for these biomarkers significantly impacts their sensitivity, specificity, PPV, and NPV. For instance, an AST/ALT ratio cut-off of 1.5, an APRI cut-off of 0.7, and a FIB-4 cut-off of 3.0 provided a more balanced comparison. These adjusted cut-offs highlight the critical nature of threshold choice in determining the performance and predictive power of biomarkers.

By considering these more comparable cut-offs, we can see that while APRI remains a strong predictor, its apparent superiority must be viewed with caution. The choice of higher thresholds for APRI in previous analyses led to fewer false positives but also influenced its hazard ratios and associations. Therefore, the initial conclusion that APRI is unequivocally superior should be tempered with the understanding that its performance is closely tied to the threshold used.

Despite these nuances, APRI's effectiveness in early detection still provides significant clinical insights. It enables healthcare providers to identify and prioritize high-risk individuals, facilitating targeted and personalized interventions. However, future work should aim to refine and standardize threshold settings across different biomarkers to enhance comparability and ensure that each biomarker's predictive power is accurately represented.

Additionally, APRI's cost-effectiveness, requiring only AST and platelet count compared to the more complex FIB-4 index, supports its integration into routine clinical practice. Targeting older adults and individuals with high-risk factors could further enhance APRI's effectiveness.

In addition, one of the main features that sets this research apart is the demonstration of a clear correlation between elevated

biomarker levels and severe liver-related outcomes, including liver cirrhosis, increased mortality, HCC, and the need for liver transplants. It underscores the utility of these biomarkers in identifying patients at risk not only of cirrhosis but also of more critical conditions.

This study, therefore, makes a contribution in the following clinical areas:

a) **Improved Risk Stratification:** In this thesis, I examined the utility of non-invasive biomarkers for pragmatic case finding and risk stratification in patients at higher risk of developing advanced CLD. This approach addresses the limitations of current diagnostic pathways in primary care, which often rely on abnormal ALT levels that may not reflect the severity of liver disease.

This study supports recent guidelines that recommend screening for advanced fibrosis in individuals with NAFLD and validates the effectiveness of non-invasive biomarkers in predicting advanced CLD risk in both primary and secondary care settings. This finding confirms that primary care biomarkers can predict liver disease progression, a result further validated in hospital settings. This cross-setting validation underscores the potential for wider adoption of these biomarkers in healthcare.

- b) **Focused Public Health Screening:** Based upon the ability to risk stratify mentioned above, one could realistically hypothesize about the potential of these biomarkers for future screening programs. My results could be useful in the design of studies to investigate this potential. This approach could help in concentrating resources on those most in need, enhancing the efficiency and effectiveness of screenings. While further research is required to justify specific screening strategies, the findings suggest that non-invasive biomarkers may play a crucial role in identifying high-risk individuals for earlier and more timely interventions.
- c) **Resource Reallocation:** The adoption of biomarker-based screening would reallocate resources from general, widespread efforts to more targeted approaches, focusing manpower, advanced diagnostics, and specialist care on those at higher risk of developing advanced CLD based on their biomarker profiles. While this approach may initially increase referrals for previously undetected high-risk patients, the long-term benefits of early detection and intervention could reduce the need for more costly treatments in later stages, thereby impacting healthcare infrastructure positively.
- d) **Cost-Effectiveness:** I have demonstrated that there exist already data from tests previously performed to risk stratify a



subset of the population (who have had an AST test performed). The utilization of this data in general practice could lead to highly effective case finding which may illustrate the benefit of testing AST more widely in clinical practice.

So, this study tentatively recommends the implementation of APRI at a threshold of 0.7 for screening and pragmatic case finding in primary care settings. The quantitative evidence provided by the sensitivity, specificity, PPV, and NPV estimates indicates that APRI offers a balanced performance and could be a cost-effective tool for early detection and intervention for advanced liver disease. Based on the detailed results from Chapter 8, using APRI at this threshold has the potential to enhance early detection and intervention, leading to timely referrals and improved patient outcomes.

While the initial costs of widespread biomarker testing may be higher, the long-term benefits might include a reduced need for more expensive treatments at later stages, ultimately contributing to lower overall healthcare costs. However, it is important to consider that the initial increase in testing could also bring about challenges in terms of healthcare resource allocation and patient management.

It is crucial to acknowledge the potential harms of implementing biomarkers, such as the risk of false positives. The

specificity of APRI at this threshold suggests that while it is effective, there could still be a significant proportion of false positives. For example, with a specificity of around 69.2%, approximately 30.8% of the screened population might be unnecessarily referred for further testing or treatment. This could lead to unnecessary anxiety, additional healthcare costs, and potential strain on healthcare resources.

The specificity of APRI at this threshold suggests that it could help mitigate the risk of false positives, ensuring that a significant proportion of those identified as high risk actually have advanced liver disease. This targeted approach may facilitate better resource allocation by focusing on high-risk individuals and optimizing the efficiency of healthcare interventions. However, the potential impact of false positives, which could result in more unnecessary referrals, must be considered. This could mean that a significant proportion of individuals could be subjected to additional testing and follow-up procedures that are ultimately unnecessary.

It is important to acknowledge that these recommendations are based on the current data and should be viewed with some caution. The findings suggest that APRI could be a valuable tool, but further research and validation studies are needed to confirm its utility and cost-effectiveness in broader clinical practice.

Additionally, the potential detrimental effects, such as the psychological impact of false positives and the logistical challenges of increased testing, should be carefully considered.

#### **9.4 Limitations of the study**

The use of the CPRD in this research confers important strengths, including being representative of the UK population registered in general practice and, by extension, of the UK general population. It is also large, providing reasonable precision in its estimates, and since the data used were recorded electronically through routine practice, the accuracy of the data is generally good.

However, several limitations need to be considered. A key limitation is the generalizability of the study's findings. The focus on a high-risk group, specifically targeting individuals with AST measured, introduces a potential bias regarding the prevalence of CLD within this cohort compared to the general UK population. This selective approach naturally gravitates towards a subgroup with a potentially higher incidence of the disease, limiting the applicability of the findings to the entire UK population and potentially overestimating the risk.

Another limitation concerns the coverage of linked data. This thesis exclusively utilizes data input using Vision software for CPRD GOLD, which provides access to a time when AST testing

was more prevalent. However, the linkage with HES data only began in 1998, while the study period starts from 1990. This discrepancy means that data coverage may not be comprehensive for all patients, particularly for those in the earlier years of the study. Patients with more complete data linkage from 1998 onwards may have different characteristics than those with incomplete linkage, potentially affecting the study's findings.

The selected population for this study consisted of individuals who had blood tests, which might not be representative of the general population. Additionally, the selection criteria and testing practices may have changed over time, further influencing the study results. These changes could introduce bias and affect the comparability of results across different time periods. Moreover, patients entering the study towards its end had a shorter follow-up period, which could impact the detection and assessment of long-term outcomes. This shorter follow-up time may lead to an underestimation of the incidence and progression of advanced liver diseases.

Furthermore, there are general weaknesses to the use of routine practice data in research, which need to be considered in the interpretation of the findings. The data collected as part of routine care, as highlighted in (section [4.7.4](#)), might exhibit some gaps, especially with regard to pertinent information about

previous exposures, and this could impact the thoroughness and depth of the research insights.

Furthermore, the nature of data collection within routine practice is inherently linked to the perspective of the general practice and their focus on the patient's ongoing care. This inherent bias in data collection could result in the omission of certain elements that might be significant from a research standpoint. In addition, when data is collected and analysed, it is important to understand why certain tests or analyses are being conducted; without a clear understanding of the purpose behind these tests, the results might be misinterpreted.

## **9.5 Suggestions for future research**

Given the high heterogeneity in the studies that measured the prevalence of NAFLD in the UK and KSA and the variations in NAFLD prevalence between the UK and KSA, further studies are needed in both countries to increase precision and confidence in estimating disease prevalence.

For future research, it is proposed to conduct population-based studies in both the UK and KSA to refine our understanding of NAFLD prevalence and associated risk factors. Additionally, the need to monitor trends in NAFLD prevalence over time, particularly among high-risk groups, is underscored. This

continuous surveillance is crucial for adapting healthcare strategies and resources effectively.

The use of routinely collected data to measure the prevalence of the three biomarkers (AST/ALT, APRI, and FIB-4) reveals the potential and practicality of a large dataset for future biomarker studies. Moreover, given the improvements in data quality and quantity and the ongoing attempts to integrate the CPRD and other datasets to provide more comprehensive information on patients and their medical diagnoses, there is room for additional research in this area.

Given that not all patients with abnormal markers will eventually manifest adverse liver outcomes, there is a need to validate the utility of these biomarkers in predicting actual disease progression. Consequently, a Randomized Controlled Trial (RCT) is recommended to assess the efficacy of screening using these biomarkers. This RCT aims to fill a crucial gap in current medical research by providing concrete data on the utility and impact of these biomarkers in a real-world clinical context.

The study would encompass a diverse participant group, focusing on adults who are at elevated risk for advanced liver diseases, including but not limited to those with predisposing factors for NAFLD. The trial design involves dividing participants into two groups: one receiving regular screening with the

biomarkers and the other receiving standard care. This setup aims to provide a clear comparison and robust evaluation of the biomarkers' effectiveness in early detection and management of liver diseases. The findings from this RCT are expected to inform clinical guidelines and patient care strategies, potentially enhancing the management of patients at risk of advanced liver diseases in primary care settings.

Moreover, the relevance of findings from the CPRD to KSA context underscores the need for localized pilot studies and research initiatives. The extrapolation of CPRD findings to KSA hinges on understanding the similarities and differences in disease prevalence, healthcare infrastructure, and patient demographics between the two regions. By conducting localized research, including pilot implementations of the biomarker screening within KSA healthcare setting, we can evaluate the adaptability and viability of this approach in the Saudi Arabian environment. This step is crucial for ensuring that the strategies developed are grounded in the local context, enhancing the potential for successful integration into the KSA's healthcare system.

## **9.6 Conclusion**

This research used primary care data linked to the HES, cancer, and death registries, consisting of complete longitudinal records of patients' information, investigations, and diagnoses within the general UK population. The studies reported in this thesis have described, for the first time, the usefulness of using non-invasive biomarkers in predicting CLD for the general population and have clearly shown that abnormal biomarkers are associated with a higher risk of developing CLD. In addition, the findings from the two systematic reviews indicate that results obtained from the CPRD research conducted in the UK could potentially be extrapolated and applied to individuals in the KSA. The future of the use of routine data for this purpose is thus promising.



## References

1. Vernon G, Baranova A, Younossi Z. Systematic review: the epidemiology and natural history of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis in adults. *Alimentary pharmacology & therapeutics*. 2011;34(3):274-85.
2. Sharma A, Nagalli S. Chronic liver disease. StatPearls [Internet]: StatPearls Publishing; 2021.
3. Paik JM, Kabbara K, Eberly KE, Younossi Y, Henry L, Younossi ZM. Global burden of NAFLD and chronic liver disease among adolescents and young adults. *Hepatology*. 2022;75(5):1204-17.
4. Wree A, Broderick L, Canbay A, Hoffman HM, Feldstein AE. From NAFLD to NASH to cirrhosis—new insights into disease mechanisms. *Nature reviews Gastroenterology & hepatology*. 2013;10(11):627-36.
5. Younossi Z, Tacke F, Arrese M, Chander Sharma B, Mostafa I, Bugianesi E, et al. Global perspectives on nonalcoholic fatty liver disease and nonalcoholic steatohepatitis. *Hepatology*. 2019;69(6):2672-82.
6. Allen AM, Kim WR, editors. Epidemiology and healthcare burden of acute-on-chronic liver failure. *Seminars in liver disease*; 2016: Thieme Medical Publishers.
7. Macpherson I, Abeysekera KW, Harris R, Mansour D, McPherson S, Rowe I, et al. Identification of liver disease: why and how. *Frontline Gastroenterology*. 2022;13(5):367-73.
8. 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. *The Lancet*. 2014;384(9958):1953-97.
9. Cotter TG, Rinella M. Nonalcoholic fatty liver disease 2020: the state of the disease. *Gastroenterology*. 2020;158(7):1851-64.
10. Asrani SK, Kouznetsova M, Ogola G, Taylor T, Masica A, Pope B, et al. Increasing health care burden of chronic liver disease compared with other chronic diseases, 2004–2013. *Gastroenterology*. 2018;155(3):719-29. e4.
11. Boursier J, Vergniol J, Guillet A, Hiriart J-B, Lannes A, Le Bail B, et al. Diagnostic accuracy and prognostic significance of blood fibrosis tests and liver stiffness measurement by FibroScan in non-alcoholic fatty liver disease. *Journal of hepatology*. 2016;65(3):570-8.
12. Panel CPG, Berzigotti A, Tsochatzis E, Boursier J, Castera L, Cazzagon N, et al. EASL Clinical Practice Guidelines on non-invasive tests for evaluation of liver disease severity and prognosis–2021 update. *Journal of hepatology*. 2021;75(3):659-89.
13. Fang Y-L, Chen H, Wang C-L, Liang L. Pathogenesis of non-alcoholic fatty liver disease in children and adolescence: From “two hit theory” to “multiple hit model”. *World journal of gastroenterology*. 2018;24(27):2974.
14. Demirjian A, Peng P, Geschwind J-FH, Cosgrove D, Schutz J, Kamel IR, et al. Infiltrating hepatocellular carcinoma: seeing the tree through the forest. *Journal of Gastrointestinal Surgery*. 2011;15:2089-97.

15. Heinrich L, Booiijink R, Khurana A, Weiskirchen R, Bansal R. Lipoxygenases in chronic liver diseases: current insights and future perspectives. *Trends in pharmacological sciences*. 2021.
16. Bataller R, Brenner DA. Liver fibrosis. *The Journal of clinical investigation*. 2005;115(2):209-18.
17. Dijk Fv, Olinga P, Poelstra K, Beljaars L. Targeted therapies in liver fibrosis: combining the best parts of platelet derived growth factor BB and interferon gamma. *Frontiers in medicine*. 2015;2:72.
18. Kang JS, Lee MH. Noninvasive diagnostic and prognostic assessment tools for liver fibrosis and cirrhosis in patients with chronic liver disease. *Liver Cirrhosis-Update and Current Challenges: IntechOpen*; 2017.
19. Aydın MM, Akçalı KC. Liver fibrosis. *The Turkish Journal of Gastroenterology*. 2018;29(1):14.
20. D'Amico G, Garcia-Tsao G, Pagliaro L. Natural history and prognostic indicators of survival in cirrhosis: a systematic review of 118 studies. *Journal of hepatology*. 2006;44(1):217-31.
21. Schuppan D, Afdhal NH. Liver cirrhosis. *The Lancet*. 2008;371(9615):838-51.
22. Moon AM, Singal AG, Tapper EB. Contemporary epidemiology of chronic liver disease and cirrhosis. *Clinical Gastroenterology and Hepatology*. 2020;18(12):2650-66.
23. Balakrishnan M. Public Health Perspectives on Combating Liver Disease. *Clinical Liver Disease*. 2020;16(5):186.
24. Le Couteur DG, McLean AJ. The aging liver: drug clearance and an oxygen diffusion barrier hypothesis. *Clinical pharmacokinetics*. 1998;34(5):359-73.
25. Asrani SK, Devarbhavi H, Eaton J, Kamath PS. Burden of liver diseases in the world. *Journal of hepatology*. 2019;70(1):151-71.
26. Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease—meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology*. 2016;64(1):73-84.
27. James SL, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*. 2018;392(10159):1789-858.
28. Mitra S, De A, Chowdhury A. Epidemiology of non-alcoholic and alcoholic fatty liver diseases. *Translational Gastroenterology and Hepatology*. 2020;5.
29. 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. *Journal of hepatology*. 2006;45(4):529-38.
30. Seeff L, Hoofnagle J. Epidemiology of hepatocellular carcinoma in areas of low hepatitis B and hepatitis C endemicity. *Oncogene*. 2006;25(27):3771-7.

31. Gao E, Hercun J, Heller T, Vilarinho S. Undiagnosed liver diseases. *Translational Gastroenterology and Hepatology*. 2021;6.
32. Health Nif, Excellence C. Non-alcoholic fatty liver disease (NAFLD): assessment and management (NICE Guideline 49). 2016.
33. GOV.UK. "Liver disease profiles: short statistical commentary, November 2018,". 2018.
34. UK CR. Cancer Mortality Statistics. 2021.
35. Estes C, Razavi H, Loomba R, Younossi Z, Sanyal AJ. Modeling the epidemic of nonalcoholic fatty liver disease demonstrates an exponential increase in burden of disease. *Hepatology*. 2018;67(1):123-33.
36. Sayiner M, Koenig A, Henry L, Younossi ZM. Epidemiology of nonalcoholic fatty liver disease and nonalcoholic steatohepatitis in the United States and the rest of the world. *Clinics in liver disease*. 2016;20(2):205-14.
37. Badran M, Laher I. Type II diabetes mellitus in Arabic-speaking countries. *International journal of endocrinology*. 2012;2012.
38. Abdo AA, Sanai FM, Al-Faleh FZ. Epidemiology of viral hepatitis in Saudi Arabia: are we off the hook? *Saudi journal of gastroenterology: official journal of the Saudi Gastroenterology Association*. 2012;18(6):349.
39. Alshaikh MK, Filippidis FT, Baldove JP, Majeed A, Rawaf S. Women in Saudi Arabia and the prevalence of cardiovascular risk factors: a systematic review. *Journal of environmental and public health*. 2016;2016.
40. Alqahtani SA, Broering DC, Alghamdi SA, Bzeizi KI, Alhusseini N, Alabbad SI, et al. Changing trends in liver transplantation indications in Saudi Arabia: from hepatitis C virus infection to nonalcoholic fatty liver disease. *BMC gastroenterology*. 2021;21(1):1-10.
41. Fallatah HI, Al-Dabbagh A, Hiejazi MT, Hanbazazah SA, Hussein AO, Al-Sahafi MA, et al. Prevalence and clinical characteristics of NAFLD in chronic liver disease patients from King Abdulaziz University Hospital, Jeddah. *Saudi Journal of Medicine & Medical Sciences*. 2020;8(2):118.
42. Metrics IfH, Evaluation. Findings from the global burden of disease study 2017. Institute for Health Metrics and Evaluation. 2018.
43. Alswat K, Alijumah AA, Sanai FM, Abaalkhail F, Alghamdi M, Al Hamoudi WK, et al. Nonalcoholic fatty liver disease burden - Saudi Arabia and United Arab Emirates, 2017-2030. *Saudi Journal of Gastroenterology*. 2018;24(4):211-9.
44. Cohen JC, Horton JD, Hobbs HH. Human fatty liver disease: old questions and new insights. *Science*. 2011;332(6037):1519-23.
45. Younossi ZM, Blissett D, Blissett R, Henry L, Stepanova M, Younossi Y, et al. The economic and clinical burden of nonalcoholic fatty liver disease in the United States and Europe. *Hepatology*. 2016;64(5):1577-86.
46. Kaur N, Goyal G, Garg R, Tapasvi C, Chawla S, Kaur R. Potential role of noninvasive biomarkers during liver fibrosis. *World Journal of Hepatology*. 2021;13(12):1919.
47. Denzer UW, Lüth S. Non-invasive diagnosis and monitoring of liver fibrosis and cirrhosis. *Best practice & research Clinical gastroenterology*. 2009;23(3):453-60.

48. Younossi ZM, Stepanova M, Afendy M, Fang Y, Younossi Y, Mir H, et al. Changes in the prevalence of the most common causes of chronic liver diseases in the United States from 1988 to 2008. *Clinical Gastroenterology and Hepatology*. 2011;9(6):524-30. e1.
49. Younossi Z, Anstee QM, Marietti M, Hardy T, Henry L, Eslam M, et al. Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. *Nature reviews Gastroenterology & hepatology*. 2018;15(1):11.
50. Calzadilla Bertot L, Adams LA. The natural course of non-alcoholic fatty liver disease. *International journal of molecular sciences*. 2016;17(5):774.
51. Henry L, Paik J, Younossi ZM. the epidemiologic burden of non-alcoholic fatty liver disease across the world. *Alimentary pharmacology & therapeutics*. 2022;56(6):942-56.
52. Younossi ZM, Golabi P, Paik JM, Henry A, Van Dongen C, Henry L. The most recent and in-depth meta-analytic assessment of the global epidemiology of nonalcoholic fatty liver disease (NAFLD). *South Asia*. 2022;4(5,475):2,016.
53. Sanai FM, Abaalkhail F, Hasan F, Farooqi MH, Al Nahdi N, Younossi ZM. Management of nonalcoholic fatty liver disease in the Middle East. *World Journal of Gastroenterology*. 2020;26(25):3528.
54. Al-Hazzaa HM, Al-Nakeeb Y, Duncan MJ, Al-Sobayel HI, Abahussain NA, Musaiger AO, et al. A cross-cultural comparison of health behaviors between Saudi and British adolescents living in urban areas: gender by country analyses. *International journal of environmental research and public health*. 2013;10(12):6701-20.
55. Cortez-Pinto H, Camilo M, Baptista A, De Oliveira A, De Moura M. Non-alcoholic fatty liver: another feature of the metabolic syndrome? *Clinical Nutrition*. 1999;18(6):353-8.
56. Perumpail BJ, Khan MA, Yoo ER, Cholankeril G, Kim D, Ahmed A. Clinical epidemiology and disease burden of nonalcoholic fatty liver disease. *World journal of gastroenterology*. 2017;23(47):8263.
57. Mofrad P, Contos MJ, Haque M, Sargeant C, Fisher RA, Luketic VA, et al. Clinical and histologic spectrum of nonalcoholic fatty liver disease associated with normal ALT values. *Hepatology*. 2003;37(6):1286-92.
58. Hasan F, Daher HB. The Burden and Clinical Care Pathways of nonalcoholic steatohepatitis in the Middle east. *Clinical Liver Disease*. 2019;14(6):207.
59. Hernaez R, Lazo M, Bonekamp S, Kamel I, Brancati FL, Guallar E, et al. Diagnostic accuracy and reliability of ultrasonography for the detection of fatty liver: a meta-analysis. *Hepatology*. 2011;54(3):1082-90.
60. Bugianesi E, Marzocchi R, Villanova N, Marchesini G. Non-alcoholic fatty liver disease/non-alcoholic steatohepatitis (NAFLD/NASH): treatment. *Best Practice & Research Clinical Gastroenterology*. 2004;18(6):1105-16.
61. Basaranoglu M, Neuschwander-Tetri BA. Nonalcoholic fatty liver disease: clinical features and pathogenesis. *Gastroenterology & hepatology*. 2006;2(4):282.

62. Makri ES, Makri E, Polyzos SA. Combination therapies for nonalcoholic fatty liver disease. *Journal of Personalized Medicine*. 2022;12(7):1166.
63. Munteanu MA, Mircea PA. From NAFLD to cardiovascular disease. Is it (still) the metabolic syndrome? *Clujul Medical*. 2014;87(2):80.
64. Ratziu V, Goodman Z, Sanyal A. Current efforts and trends in the treatment of NASH. *Journal of hepatology*. 2015;62(1):S65-S75.
65. Björck I, Elmståhl HL. The glycaemic index: importance of dietary fibre and other food properties. *Proceedings of the Nutrition Society*. 2003;62(1):201-6.
66. Keane MG, Hensher C, Pereira SP. Improving the identification and monitoring of cirrhosis. *The Practitioner*. 2016;260(1798):25-9.
67. Lee I-M, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT, et al. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *The lancet*. 2012;380(9838):219-29.
68. Hazlehurst JM, Woods C, Marjot T, Cobbald JF, Tomlinson JW. Non-alcoholic fatty liver disease and diabetes. *Metabolism*. 2016;65(8):1096-108.
69. Adams L. The Natural Course of Non-Alcoholic Fatty Liver Disease. *International Journal of Molecular Sciences*. 2016;17(5).
70. Chalasani N, Younossi Z, Lavine JE, Charlton M, Cusi K, Rinella M, et al. The diagnosis and management of nonalcoholic fatty liver disease: practice guidance from the American Association for the Study of Liver Diseases. *Hepatology*. 2018;67(1):328-57.
71. Piscaglia F, Baroni GS, Barchetti A, Pecorelli A, Marinelli S, Tiribelli C, et al. Clinical patterns of hepatocellular carcinoma (HCC) in non alcoholic fatty liver disease (NAFLD): a multicenter case-control study. *Digestive and Liver Disease*. 2015;47:e36-e7.
72. Dowman JK, Tomlinson J, Newsome P. Systematic review: the diagnosis and staging of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis. *Alimentary pharmacology & therapeutics*. 2011;33(5):525-40.
73. Zhang J-Z, Cai J-J, Yu Y, She Z-G, Li H. Nonalcoholic fatty liver disease: an update on the diagnosis. *Gene Expression*. 2019;19(3):187.
74. Liver EAftSoT, Diabetes EAftSo. EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease. *Obesity facts*. 2016;9(2):65-90.
75. Huang T, Behary J, Zekry A. Non-alcoholic fatty liver disease: a review of epidemiology, risk factors, diagnosis and management. *Internal medicine journal*. 2020;50(9):1038-47.
76. Regev A, Berho M, Jeffers LJ, Milikowski C, Molina EG, Pyrsopoulos NT, et al. Sampling error and intraobserver variation in liver biopsy in patients with chronic HCV infection. *The American journal of gastroenterology*. 2002;97(10):2614-8.
77. Åström H, Wester A, Hagström H. Administrative coding for non-alcoholic fatty liver disease is accurate in Swedish patients. *Scandinavian Journal of Gastroenterology*. 2023:1-6.

78. Targher G, Byrne CD, Tilg H. MASLD: a systemic metabolic disorder with cardiovascular and malignant complications. *Gut*. 2024.
79. Sepanlou SG, Safiri S, Bisignano C, Ikuta KS, Merat S, Saberifiroozi M, et al. The global, regional, and national burden of cirrhosis by cause in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet gastroenterology & hepatology*. 2020;5(3):245-66.
80. Fleming KM, Aithal G, Card T, West J. The rate of decompensation and clinical progression of disease in people with cirrhosis: a cohort study. *Alimentary pharmacology & therapeutics*. 2010;32(11-12):1343-50.
81. Moreau R, Jalan R, Gines P, Pavesi M, Angeli P, Cordoba J, et al. Acute-on-chronic liver failure is a distinct syndrome that develops in patients with acute decompensation of cirrhosis. *Gastroenterology*. 2013;144(7):1426-37. e9.
82. Ge PS, Runyon BA. Treatment of patients with cirrhosis. *New England Journal of Medicine*. 2016;375(8):767-77.
83. Xu J-H, Yu Y-Y, Xu X-Y. Management of chronic liver diseases and cirrhosis: current status and future directions. *Chinese Medical Journal*. 2020;133(22):2647-9.
84. Williams R, Ashton K, Aspinall R, Bellis MA, Bosanquet J, Cramp ME, et al. Implementation of the lancet standing commission on liver disease in the UK. *The Lancet*. 2015;386(10008):2098-111.
85. Abeysekera KW, Macpherson I, Glyn-Owen K, McPherson S, Parker R, Harris R, et al. Community pathways for the early detection and risk stratification of chronic liver disease: a narrative systematic review. *The Lancet Gastroenterology & Hepatology*. 2022.
86. Iragorri N, Spackman E. Assessing the value of screening tools: reviewing the challenges and opportunities of cost-effectiveness analysis. *Public health reviews*. 2018;39(1):1-27.
87. Bulliard J-L, Ducros C, Jemelin C, Arzel B, Fioretta G, Levi F. Effectiveness of organised versus opportunistic mammography screening. *Annals of oncology*. 2009;20(7):1199-202.
88. Chalmers J, Wilkes E, Harris R, Kent L, Kinra S, Aithal G, et al. Development and implementation of a commissioned pathway for the identification and stratification of liver disease in the community. *Frontline gastroenterology*. 2020;11(2):86-92.
89. Gilbert R, Logan S, Moyer VA, Elliott EJ. Evidence-Based Case Reviews: Assessing diagnostic and screening tests: Part 2. How to use the research literature on diagnosis. *Western Journal of Medicine*. 2001;175(1):37.
90. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes*: Oxford university press; 2015.
91. Davis TC, Long SW, Jackson RH, Mayeaux E, George RB, Murphy PW, et al. Rapid estimate of adult literacy in medicine: a shortened screening instrument. *Family medicine*. 1993;25(6):391-5.
92. Shi L. The relationship between primary care and life chances. *Journal of health care for the poor and underserved*. 1992;3(2):321-35.

93. Pronovost PJ, Goeschel CA, Marsteller JA, Sexton JB, Pham JC, Berenholtz SM. Framework for patient safety research and improvement. *Circulation*. 2009;119(2):330-7.
94. Trochim W, Kane M. Concept mapping: an introduction to structured conceptualization in health care. *International journal for quality in health care*. 2005;17(3):187-91.
95. Swift A, Heale R, Twycross A. What are sensitivity and specificity? *Evidence-Based Nursing*. 2020;23(1):2-4.
96. DI S. Health measurement scales. A practical guide to their development and use. 1995:15-27.
97. Bravo AA, Sheth SG, Chopra S. Liver biopsy. *New England Journal of Medicine*. 2001;344(7):495-500.
98. Ishak K. Histological grading and staging of chronic hepatitis. *J hepatol*. 1995;22:696-9.
99. Chowdhury AB, Mehta KJ. Liver biopsy for assessment of chronic liver diseases: a synopsis. *Clinical and experimental medicine*. 2022;1-13.
100. Li C, Li R, Zhang W. Progress in non-invasive detection of liver fibrosis. *Cancer biology & medicine*. 2018;15(2):124.
101. Tatler AL. Recent advances in the non-invasive assessment of fibrosis using biomarkers. *Current Opinion in Pharmacology*. 2019;49:110-5.
102. Liver EAfSo. EASL-ALEH Clinical Practice Guidelines: Non-invasive tests for evaluation of liver disease severity and prognosis. *Journal of hepatology*. 2015;63(1):237-64.
103. Adams LA. Biomarkers of liver fibrosis. *Journal of gastroenterology and hepatology*. 2011;26(5):802-9.
104. Rosenberg WM, Voelker M, Thiel R, Becka M, Burt A, Schuppan D, et al. Serum markers detect the presence of liver fibrosis: a cohort study. *Gastroenterology*. 2004;127(6):1704-13.
105. Parkes J, Roderick P, Harris S, Day C, Mutimer D, Collier J, et al. Enhanced liver fibrosis test can predict clinical outcomes in patients with chronic liver disease. *Gut*. 2010;59(9):1245-51.
106. Sharma S, Khalili K, Nguyen GC. Non-invasive diagnosis of advanced fibrosis and cirrhosis. *World journal of gastroenterology: WJG*. 2014;20(45):16820.
107. Manning DS, Afdhal NH. Diagnosis and quantitation of fibrosis. *Gastroenterology*. 2008;134(6):1670-81.
108. Perri RE, Chiorean MV, Fidler JL, Fletcher JG, Talwalkar JA, Stadheim L, et al. A prospective evaluation of computerized tomographic (CT) scanning as a screening modality for esophageal varices. *Hepatology*. 2008;47(5):1587-94.
109. Papastergiou V, Tsochatzis E, Burroughs AK. Non-invasive assessment of liver fibrosis. *Annals of Gastroenterology*. 2012;25(3):218.
110. Gurwin A, Kowalczyk K, Knecht-Gurwin K, Stelmach P, Nowak Ł, Krajewski W, et al. Alternatives for MRI in prostate cancer diagnostics—review of current ultrasound-based techniques. *Cancers*. 2022;14(8):1859.
111. D’Onofrio M, Crosara S, De Robertis R, Canestrini S, Demozzi E, Gallotti A, et al. Acoustic radiation force impulse of the liver. *World journal of gastroenterology: WJG*. 2013;19(30):4841.

112. Sagir A, Ney D, Oh J, Pandey S, Kircheis G, Mayatepek E, et al. Evaluation of acoustic radiation force impulse imaging (ARFI) for the determination of liver stiffness using transient elastography as a reference in children. *Ultrasound international open*. 2015;1(01):E2-E7.
113. Huwart L, Peeters F, Sinkus R, Annet L, Salameh N, ter Beek LC, et al. Liver fibrosis: non-invasive assessment with MR elastography. *NMR in Biomedicine: An International Journal Devoted to the Development and Application of Magnetic Resonance In vivo*. 2006;19(2):173-9.
114. Bohte AE, Garteiser P, De Niet A, Groot PF, Sinkus R, Stoker J, et al. MR elastography of the liver: defining thresholds for detecting viscoelastic changes. *Radiology*. 2013;269(3):768-76.
115. Guechot J, Laudat A, Loria A, Serfaty L, Poupon R, Giboudeau J. Diagnostic accuracy of hyaluronan and type III procollagen amino-terminal peptide serum assays as markers of liver fibrosis in chronic viral hepatitis C evaluated by ROC curve analysis. *Clinical chemistry*. 1996;42(4):558-63.
116. Mölleken C, Sitek B, Henkel C, Poschmann G, Sipos B, Wiese S, et al. Detection of novel biomarkers of liver cirrhosis by proteomic analysis. *Hepatology*. 2009;49(4):1257-66.
117. Saitou Y, Shiraki K, Yamanaka Y, Yamaguchi Y, Kawakita T, Yamamoto N, et al. Noninvasive estimation of liver fibrosis and response to interferon therapy by a serum fibrogenesis marker, YKL-40, in patients with HCV-associated liver disease. *World journal of gastroenterology: WJG*. 2005;11(4):476.
118. Harrison SA, Oliver D, Arnold HL, Gogia S, Neuschwander-Tetri BA. Development and validation of a simple NAFLD clinical scoring system for identifying patients without advanced disease. *Gut*. 2008;57(10):1441-7.
119. Fornis X, Ampurdanes S, Llovet JM, Aponte J, Quintó L, Martínez-Bauer E, et al. Identification of chronic hepatitis C patients without hepatic fibrosis by a simple predictive model. *Hepatology*. 2002;36(4):986-92.
120. Koda M, Matunaga Y, Kawakami M, Kishimoto Y, Suou T, Murawaki Y. FibroIndex, a practical index for predicting significant fibrosis in patients with chronic hepatitis C. *Hepatology*. 2007;45(2):297-306.
121. Angulo P, Hui JM, Marchesini G, Bugianesi E, George J, Farrell GC, et al. The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD. *Hepatology*. 2007;45(4):846-54.
122. Sarin S, Kumar M, Lau G, Abbas Z, Chan H, Chen C, et al. Asian-Pacific clinical practice guidelines on the management of hepatitis B: a 2015 update. *Hepatology international*. 2016;10:1-98.
123. Shiha G, Ibrahim A, Helmy A, Sarin SK, Omata M, Kumar A, et al. Asian-Pacific Association for the Study of the Liver (APASL) consensus guidelines on invasive and non-invasive assessment of hepatic fibrosis: a 2016 update. *Hepatology international*. 2017;11:1-30.
124. Organization WH. Guidelines for the prevention care and treatment of persons with chronic hepatitis B infection: Mar-15: World Health Organization; 2015.
125. Yap CY, Aw TC. Liver function tests (LFTs). *Proceedings of Singapore Healthcare*. 2010;19(1):80-2.



126. Gowda S, Desai PB, Hull VV, Avinash AK, Vernekar SN, Kulkarni SS. A review on laboratory liver function tests. *The Pan african medical journal*. 2009;3.
127. Hall P, Cash J. What is the real function of the liver 'function' tests? *The Ulster medical journal*. 2012;81(1):30.
128. Kwo PY, Cohen SM, Lim JK. ACG clinical guideline: evaluation of abnormal liver chemistries. *Official journal of the American College of Gastroenterology | ACG*. 2017;112(1):18-35.
129. Newsome PN, Cramb R, Davison SM, Dillon JF, Foulerton M, Godfrey EM, et al. Guidelines on the management of abnormal liver blood tests. *Gut*. 2018;67(1):6-19.
130. Huang X-J, Choi Y-K, Im H-S, Yarimaga O, Yoon E, Kim H-S. Aspartate aminotransferase (AST/GOT) and alanine aminotransferase (ALT/GPT) detection techniques. *Sensors*. 2006;6(7):756-82.
131. Gupta M, Choudhury PS, Singh S, Hazarika D. Liver functional volumetry by Tc-99m mebrofenin hepatobiliary scintigraphy before major liver resection: A game changer. *Indian journal of nuclear medicine: IJNM: the official journal of the Society of Nuclear Medicine, India*. 2018;33(4):277.
132. Vaja R, Rana M. Drugs and the liver. *Anaesthesia & Intensive Care Medicine*. 2020;21(10):517-23.
133. Newsome PN, Cramb R, Davison SM, Dillon JF, Foulerton M, Godfrey EM, et al. Guidelines on the management of abnormal liver blood tests. *Gut*. 2017:gutjnl-2017-314924.
134. Lala V, Zubair M, Minter DA. Liver function tests. *StatPearls [internet]: StatPearls Publishing*; 2022.
135. Xi D, Lin H, Shah AA. Overview of autoimmune liver disease: Prevalence, risk factors, and role of autoantibodies. *Clinical Liver Disease*. 2022;20(4):111.
136. Pratt DS, Kaplan MM. Evaluation of abnormal liver-enzyme results in asymptomatic patients. *New England Journal of Medicine*. 2000;342(17):1266-71.
137. Scutt P, Ban L, Card TC, Crooks C, Guha NI, West J, et al. Liver blood marker testing in UK primary care: a UK wide cohort study, 2004-2016. *BMJ Open*. 2022.
138. Alexander M, Loomis AK, Fairburn-Beech J, van der Lei J, Duarte-Salles T, Prieto-Alhambra D, et al. Real-world data reveal a diagnostic gap in non-alcoholic fatty liver disease. *BMC Medicine*. 2018;16 (1) (no pagination)(130).
139. Williams AL, Hoofnagle JH. Ratio of serum aspartate to alanine aminotransferase in chronic hepatitis relationship to cirrhosis. *Gastroenterology*. 1988;95(3):734-9.
140. Giannini E, Botta F, Testa E, Romagnoli P, Polegato S, Malfatti F, et al. The 1-year and 3-month prognostic utility of the AST/ALT ratio and model for end-stage liver disease score in patients with viral liver cirrhosis. *The American journal of gastroenterology*. 2002;97(11):2855-60.
141. Lai X, Chen H, Dong X, Zhou G, Liang D, Xu F, et al. AST to ALT ratio as a prospective risk predictor for liver cirrhosis in patients with chronic HBV

- infection. *European Journal of Gastroenterology & Hepatology*. 2024;36(3):338-44.
142. Giannini E, Risso D, Botta F, Chiarbonello B, Fasoli A, Malfatti F, et al. Validity and clinical utility of the aspartate aminotransferase–alanine aminotransferase ratio in assessing disease severity and prognosis in patients with hepatitis C virus–related chronic liver disease. *Archives of internal medicine*. 2003;163(2):218-24.
  143. Wang F, Gao S, Wu M, Zhao D, Sun H, Yav S, et al. The prognostic role of the AST/ALT ratio in hepatocellular carcinoma patients receiving thermal ablation combined with simultaneous TACE. *BMC gastroenterology*. 2023;23(1):80.
  144. Amernia B, Moosavy SH, Banookh F, Zoghi G. FIB-4, APRI, and AST/ALT ratio compared to FibroScan for the assessment of hepatic fibrosis in patients with non-alcoholic fatty liver disease in Bandar Abbas, Iran. *BMC gastroenterology*. 2021;21(1):1-7.
  145. Angulo P, Keach JC, Batts KP, Lindor KD. Independent predictors of liver fibrosis in patients with nonalcoholic steatohepatitis. *Hepatology*. 1999;30(6):1356-62.
  146. McPherson S, Stewart SF, Henderson E, Burt AD, Day CP. Simple non-invasive fibrosis scoring systems can reliably exclude advanced fibrosis in patients with non-alcoholic fatty liver disease. *Gut*. 2010;59(9):1265-9.
  147. Park GJH, Lin BP, Ngu MC, Jones DB, Katelaris PH. Aspartate aminotransferase: alanine aminotransferase ratio in chronic hepatitis C infection: is it a useful predictor of cirrhosis? *Journal of gastroenterology and hepatology*. 2000;15(4):386-90.
  148. Wai C-T, Greenson JK, Fontana RJ, Kalbfleisch JD, Marrero JA, Conjeevaram HS, et al. A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C. *Hepatology*. 2003;38(2):518-26.
  149. Liu X, Li H, Wei L, Tang Q, Hu P. Optimized cutoffs of gamma-glutamyl transpeptidase-to-platelet ratio, aspartate aminotransferase-to-platelet ratio index, and fibrosis-4 scoring systems for exclusion of cirrhosis in patients with chronic hepatitis B. *Hepatology Communications*. 2022;6(7):1664-72.
  150. Li Q, Ren X, Lu C, Li W, Huang Y, Chen L. Evaluation of APRI and FIB-4 for noninvasive assessment of significant fibrosis and cirrhosis in HBeAg-negative CHB patients with ALT ≤ 2 ULN: A retrospective cohort study. *Medicine*. 2017;96(12).
  151. Zhu X, Wang L-C, Chen E-Q, Chen X-B, Chen L-Y, Liu L, et al. Prospective evaluation of FibroScan for the diagnosis of hepatic fibrosis compared with liver biopsy/AST platelet ratio index and FIB-4 in patients with chronic HBV infection. *Digestive diseases and sciences*. 2011;56:2742-9.
  152. Chou R, Wasson N. Blood tests to diagnose fibrosis or cirrhosis in patients with chronic hepatitis C virus infection: a systematic review. *Annals of internal medicine*. 2013;158(11):807-20.
  153. Festi D, Schiumerini R, Marzi L, Di Biase A, Mandolesi D, Montrone L, et al. the diagnosis of non-alcoholic fatty liver disease—availability and

- accuracy of non-invasive methods. *Alimentary pharmacology & therapeutics*. 2013;37(4):392-400.
154. Lin ZH, Xin YN, Dong QJ, Wang Q, Jiang XJ, Zhan SH, et al. Performance of the aspartate aminotransferase-to-platelet ratio index for the staging of hepatitis C-related fibrosis: an updated meta-analysis. *Hepatology*. 2011;53(3):726-36.
  155. de Oliveira AC, El-Bacha I, Vianna MV, Parise ER. Utility and limitations of APRI and FIB4 to predict staging in a cohort of nonselected outpatients with hepatitis C. *Annals of hepatology*. 2016;15(3):326-32.
  156. Sterling RK, Lissen E, Clumeck N, Sola R, Correa MC, Montaner J, et al. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *Hepatology*. 2006;43(6):1317-25.
  157. Shah AG, Lydecker A, Murray K, Tetri BN, Contos MJ, Sanyal AJ, et al. Comparison of noninvasive markers of fibrosis in patients with nonalcoholic fatty liver disease. *Clinical Gastroenterology and Hepatology*. 2009;7(10):1104-12.
  158. Shah AG, Lydecker A, Murray K, Tetri BN, Contos MJ, Sanyal AJ, et al. Use of the FIB4 index for non-invasive evaluation of fibrosis in nonalcoholic fatty liver disease. *Clinical gastroenterology and hepatology: the official clinical practice journal of the American Gastroenterological Association*. 2009;7(10):1104.
  159. Vallet-Pichard A, Mallet V, Nalpas B, Verkarre V, Nalpas A, Dhalluin-Venier V, et al. FIB-4: an inexpensive and accurate marker of fibrosis in HCV infection. comparison with liver biopsy and fibrotest. *Hepatology*. 2007;46(1):32-6.
  160. Poynard T, Ngo Y, Perazzo H, Munteanu M, Lebray P, Moussalli J, et al. Prognostic value of liver fibrosis biomarkers: a meta-analysis. *Gastroenterology & hepatology*. 2011;7(7):445.
  161. Crossan C, Tsochatzis EA, Longworth L, Gurusamy K, Davidson B, Rodríguez-Perálvarez M, et al. Cost-effectiveness of non-invasive methods for assessment and monitoring of liver fibrosis and cirrhosis in patients with chronic liver disease: systematic review and economic evaluation. 2015.
  162. 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. *The lancet Gastroenterology & hepatology*. 2017;2(4):288-97.
  163. Srivastava A, Gailer R, Tanwar S, Trembling P, Parkes J, Rodger A, et al. Prospective evaluation of a primary care referral pathway for patients with non-alcoholic fatty liver disease. *Journal of hepatology*. 2019;71(2):371-8.
  164. Armstrong MJ, Marchesini G. Referral pathways for NAFLD fibrosis in primary care—No longer a ‘needle in a haystack’. *Journal of Hepatology*. 2019;71(2):246-8.
  165. Pryke R, Guha IN. Time to focus on chronic liver diseases in the community: A review of primary care hepatology tools, pathways of care and reimbursement mechanisms. *Journal of Hepatology*. 2023;78(3):663-71.

166. Ahmed A, Perumpail RB, Harrison SA. High prevalence of hepatic fibrosis in the setting of coexisting diabetes and hepatic steatosis: A case for selective screening in the general population? *Hepatology*. 2016;63(1):20-2.
167. Chalasani N, Younossi Z, Lavine JE, Diehl AM, Brunt EM, Cusi K, et al. The diagnosis and management of non-alcoholic fatty liver disease: practice guideline by the American Gastroenterological Association, American Association for the Study of Liver Diseases, and American College of Gastroenterology. *Gastroenterology*. 2012;142(7):1592-609.
168. Neukam K, Bhagani S, Rodger A, Oben J, Nirmal D, Jain A, et al. High prevalence of non-alcoholic fatty liver disease (NAFLD) among Gujarati Indians in North London: a population-based study. *Clinical Lipidology*. 2017;12(1):33-9.
169. Khalatbari-Soltani S, Imamura F, Brage S, Rolfe ED, Griffin SJ, Wareham NJ, et al. The association between adherence to the Mediterranean diet and hepatic steatosis: cross-sectional analysis of two independent studies, the UK Fenland Study and the Swiss CoLaus Study. *Bmc Medicine*. 2019;17.
170. Wilman HR, Kelly M, Garratt S, Matthews PM, Milanese M, Herlihy A, et al. Characterisation of liver fat in the UK Biobank cohort. *PLoS ONE*. 2017;12 (2) (no pagination)(e0172921).
171. Abeysekera K, Fernandes GS, Hammerton G, Gordon FH, Portal AJ, Heron J, et al. The Prevalence of Non-Alcoholic Fatty Liver Disease in Young Adults in the UK. Available at SSRN 3448013. 2019.
172. el-Hassan AY, Ibrahim EM, al-Mulhim FA, Nabhan AA, Chammas MY. Fatty infiltration of the liver: analysis of prevalence, radiological and clinical features and influence on patient management. *British Journal of Radiology*. 1992;65(777):774-8.
173. Al-Hamoudi W, El-Sabbah M, Ali S, Altuwaijri M, Bedewi M, Adam M, et al. Epidemiological, clinical, and biochemical characteristics of Saudi patients with nonalcoholic fatty liver disease: A hospital-based study. *Annals of Saudi Medicine*. 2012;32(3):288-92.
174. Alshumrani GA, Shawky KA, Assiri YI, Algathradi MA, Mahfouz AA, Mostafa OA. Fatty Liver Disease among Adults in Southwestern Saudi Arabia. *Med J Cairo Univ*. 2013;81:205-9.
175. Alazzeah AY, Epuru S, Azzeah FS, Smadi MM, Banu S, Bano R, et al. Identification of non-alcoholic fatty liver disease in Saudi females and validation of non-invasive indices. *Medical Science*. 2019;23(100):1001-10.
176. Papagianni M, Sofogianni A, Tziomalos K. Non-invasive methods for the diagnosis of nonalcoholic fatty liver disease. *World journal of hepatology*. 2015;7(4):638.
177. Lee SW, Park SH, Kim KW, Choi EK, Shin YM, Kim PN, et al. Unenhanced CT for assessment of macrovesicular hepatic steatosis in living liver donors: comparison of visual grading with liver attenuation index. *Radiology*. 2007;244(2):479-85.
178. Bellentani S, Marino M. Epidemiology and natural history of non-alcoholic liver disease (NAFLD). *Annals of hepatology*. 2009;8(S1):4-8.
179. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. *Systematic Reviews*, 5, 210. 2016.

180. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research synthesis methods*. 2010;1(2):97-111.
181. Armstrong M, Houlihan D, Bentham L, Shaw J, Olliff S, Neuberger J, et al. Prevalence and severity of nonalcoholic fatty liver disease in a large prospective primary care cohort with abnormal liver function tests. *Gut*. 2010;2):A1.
182. El-Gohary M, Moore M, Roderick P, Watkins E, Dash J, Reinson T, et al. Local care and treatment of liver disease (LOCATE) - A cluster-randomized feasibility study to discover, assess and manage early liver disease in primary care. *PLoS ONE*. 2018;13 (12) (no pagination)(e0208798).
183. Lombardi R, Lever R, Smith C, Marshall N, Rodger A, Bhagani S, et al. Liver test abnormalities in patients with HIV mono-infection: Assessment with simple noninvasive fibrosis markers. *Annals of Gastroenterology*. 2017;30(3):349-56.
184. Marjot T, Sbardella E, Moolla A, Hazlehurst JM, Tan GD, Ainsworth M, et al. Prevalence and severity of non-alcoholic fatty liver disease are underestimated in clinical practice: impact of a dedicated screening approach at a large university teaching hospital. *Diabetic Medicine*. 2018;35(1):89-98.
185. Williamson RM, Price JF, Glancy S, Nee LD, Perry E, Hayes PC, et al. Prevalence of hepatic fibrosis and portal hypertension markers in type 2 diabetes: The Edinburgh Type 2 Diabetes Study (ET2DS). *Diabetic Medicine*. 2010;1):7.
186. Amarapurkar D, Kamani P, Patel N, Gupte P, Kumar P, Agal S, et al. Prevalence of non-alcoholic fatty liver disease: population based study. *Annals of hepatology*. 2007;6(3):161-3.
187. Eslam M, George J. Genetic contributions to NAFLD: leveraging shared genetics to uncover systems biology. *Nature reviews Gastroenterology & hepatology*. 2020;17(1):40-52.
188. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *Bmj*. 2003;327(7414):557-60.
189. Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology*. 2016;64(1):73-84.
190. Nguyen M, Li J, Zou B, Fujii H, Yeo YH, Ji F, et al. Prevalence of non-alcoholic fatty liver disease (NAFLD) in Asia: A systematic review and meta-analysis of 195 Studies and 1,753,168 subjects from 13 countries. *Hepatology International*. 2018;12 (2):S232-S3.
191. Im HJ, Ahn YC, Wang J-H, Lee MM, Son CG. Systematic review on the prevalence of nonalcoholic fatty liver disease in South Korea. *Clinics and Research in Hepatology and Gastroenterology*. 2020.
192. Loomba R, Sanyal AJ. The global NAFLD epidemic. *Nature reviews Gastroenterology & hepatology*. 2013;10(11):686-90.
193. Golabi P, Paik J, Reddy R, Bugianesi E, Trimble G, Younossi ZM. Prevalence and long-term outcomes of non-alcoholic fatty liver disease

- among elderly individuals from the United States. *BMC gastroenterology*. 2019;19(1):1-8.
194. Cholongitas E, Pavlopoulou I, Papatheodoridi M, Markakis GE, Bouras E, Haidich A-B, et al. Epidemiology of nonalcoholic fatty liver disease in Europe: a systematic review and meta-analysis. *Annals of Gastroenterology*. 2021;34(3):404.
  195. Rotman Y, Koh C, Zmuda JM, Kleiner DE, Liang TJ, CRN N. The association of genetic variability in patatin-like phospholipase domain-containing protein 3 (PNPLA3) with histological severity of nonalcoholic fatty liver disease. *Hepatology*. 2010;52(3):894-903.
  196. Finelli C, Tarantino G. Is there any consensus as to what diet or lifestyle approach is the right one for NAFLD patients. *J Gastrointest Liver Dis*. 2012;21(3):293-302.
  197. Portillo-Sanchez P, Bril F, Maximos M, Lomonaco R, Biernacki D, Orsak B, et al. High prevalence of nonalcoholic fatty liver disease in patients with type 2 diabetes mellitus and normal plasma aminotransferase levels. *The journal of clinical endocrinology & metabolism*. 2015;100(6):2231-8.
  198. Medina J, Fernández-Salazar LI, García-Buey L, Moreno-Otero R. Approach to the pathogenesis and treatment of nonalcoholic steatohepatitis. *Diabetes care*. 2004;27(8):2057-66.
  199. Mantovani A, Byrne CD, Bonora E, Targher G. Nonalcoholic Fatty Liver Disease and Risk of Incident Type 2 Diabetes: A Meta-analysis. *Diabetes Care*. 2018;41(2):372-82.
  200. Younossi ZM, Golabi P, de Avila L, Paik JM, Srishord M, Fukui N, et al. The global epidemiology of NAFLD and NASH in patients with type 2 diabetes: a systematic review and meta-analysis. *Journal of hepatology*. 2019;71(4):793-801.
  201. Sanyal D, Mukherjee P, Raychaudhuri M, Ghosh S, Mukherjee S, Chowdhury S. Profile of liver enzymes in non-alcoholic fatty liver disease in patients with impaired glucose tolerance and newly detected untreated type 2 diabetes. *Indian journal of endocrinology and metabolism*. 2015;19(5):597.
  202. Pillai SS, Lakhani HV, Zehra M, Wang J, Dilip A, Puri N, et al. Predicting Nonalcoholic Fatty Liver Disease through a Panel of Plasma Biomarkers and MicroRNAs in Female West Virginia Population. *International Journal of Molecular Sciences*. 2020;21(18):6698.
  203. Verma S, Jensen D, Hart J, Mohanty SR. Predictive value of ALT levels for non-alcoholic steatohepatitis (NASH) and advanced fibrosis in non-alcoholic fatty liver disease (NAFLD). *Liver International*. 2013;33(9):1398-405.
  204. Alswat K, Aljumah AA, Sanai FM, Abaalkhail F, Alghamdi M, Al Hamoudi WK, et al. Nonalcoholic fatty liver disease burden -- Saudi Arabia and United Arab Emirates, 2017-2030. *Saudi Journal of Gastroenterology*. 2018;24(4):211-9.
  205. Akbar DH, Kawther AH. Nonalcoholic fatty liver disease in Saudi type 2 diabetic subjects attending a medical outpatient clinic: prevalence and general characteristics. *Diabetes Care*. 2003;26(12):3351-2.

206. Fallatah HI, Akbar HO. Lack of correlation between serum alanine amino transferrase and abdominal ultrasound in the diagnosis non alcoholic fatty liver disease in type 2 diabetes. *Int J Diabetes Metab.* 2010;18:120-3.
207. Elmakki E, Aqeely H, Bani I, Omer H, Solan Y, Taher A, et al. Nonalcoholic fatty liver disease (NAFLD) in Saudi patients with T2DM in Jazan region: prevalence and associated factors. *Journal of Advances in Medicine and Medical Research.* 2015:872-9.
208. Alsabaani AA, Mahfouz AA, Awadalla NJ, Musa MJ, Al Humayed SM. Non-Alcoholic Fatty Liver Disease among Type-2 Diabetes Mellitus Patients in Abha City, South Western Saudi Arabia. *Int J Environ Res Public Health.* 2018;15(11):11.
209. Zelber-Sagi S, Salomone F, Yeshua H, Lotan R, Webb M, Halpern Z, et al. Non-high-density lipoprotein cholesterol independently predicts new onset of non-alcoholic fatty liver disease. *Liver International.* 2014;34(6):e128-e35.
210. Eshraghian A, Dabbaghmanesh MH, Eshraghian H, Fattahi MR, Omrani GR. Nonalcoholic fatty liver disease in a cluster of Iranian population: thyroid status and metabolic risk factors. *Archives of Iranian medicine.* 2013;16(10):0-.
211. Kasapoglu B, Turkay C, Yalcin K, Carlioglu A, Sozen M, Koktener A. Low vitamin D levels are associated with increased risk for fatty liver disease among non-obese adults. *Clinical medicine.* 2013;13(6):576.
212. Kalra S, Vithalani M, Gulati G, Kulkarni C, Kadam Y, Pallivathukkal J, et al. Study of prevalence of nonalcoholic fatty liver disease (NAFLD) in type 2 diabetes patients in India (SPRINT). *J Assoc Physicians India.* 2013;61(7):448-53.
213. Dai W, Ye L, Liu A, Wen SW, Deng J, Wu X, et al. Prevalence of nonalcoholic fatty liver disease in patients with type 2 diabetes mellitus. *Medicine.* 2017;96(39):e8179-e.
214. Almobarak AO, Barakat S, Suliman EA, Elmadhoun WM, Mohamed NA, Abobaker IO, et al. Prevalence of and predictive factors for nonalcoholic fatty liver disease in Sudanese individuals with type 2 diabetes: Is metabolic syndrome the culprit? *Arab Journal of Gastroenterology.* 2015;16(2):54-8.
215. Robert AA, Al Dawish MA. The worrying trend of diabetes mellitus in Saudi Arabia: an urgent call to action. *Current diabetes reviews.* 2020;16(3):204-10.
216. Al-Quwaidhi A, Pearce M, Critchley J, Sobngwi E, O'flaherty M. Trends and future projections of the prevalence of adult obesity in Saudi Arabia, 1992-2022. *Eastern Mediterranean Health Journal.* 2014;20(10).
217. Bellentani S. The epidemiology of non-alcoholic fatty liver disease. *Liver international.* 2017;37:81-4.
218. Loomba R, Adams LA. The 20% rule of NASH progression: the natural history of advanced fibrosis and cirrhosis caused by NASH. *Hepatology (Baltimore, Md).* 2019;70(6):1885.
219. England P. Health matters: obesity and the food environment. London, UK: Public Health England. 2017.

220. Ng SW, Zaghoul S, Ali HI, Harrison G, Popkin BM. The prevalence and trends of overweight, obesity and nutrition-related non-communicable diseases in the Arabian Gulf States. *Obesity Reviews*. 2011;12(1):1-13.
221. Mabry R, Reeves MM, Eakin EG, Owen N. Evidence of physical activity participation among men and women in the countries of the Gulf Cooperation Council: a review. *Obesity reviews*. 2010;11(6):457-64.
222. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, Van Staa T, et al. Data resource profile: clinical practice research datalink (CPRD). *International journal of epidemiology*. 2015;44(3):827-36.
223. Kousoulis AA, Rafi I, de Lusignan S. The CPRD and the RCGP: building on research success by enhancing benefits for patients and practices. *British Journal of General Practice*; 2015. p. 54-5.
224. Evans SJ, Day SJ. Medicines and Healthcare Products Regulatory Agency (MHRA)(Formerly MCA). *Encyclopedia of Biostatistics*. 2005;5.
225. Williams T, van Staa T, Puri S, Eaton S. Recent advances in the utility and use of the General Practice Research Database as an example of a UK Primary Care Data resource. *Therapeutic advances in drug safety*. 2012;3(2):89-99.
226. Jick H, Jick SS, Derby LE. Validation of information recorded on general practitioner based computerised data resource in the United Kingdom. *British Medical Journal*. 1991;302(6779):766-8.
227. Busby J, Schroeder K, Woltersdorf W, Sterne JA, Ben-Shlomo Y, Hay A, et al. Temporal growth and geographic variation in the use of laboratory tests by NHS general practices: using routine data to identify research priorities. *British Journal of General Practice*. 2013;63(609):e256-e66.
228. Wolf A, Dedman D, Campbell J, Booth H, Lunn D, Chapman J, et al. Data resource profile: clinical practice research Datalink (CPRD) aurum. *International journal of epidemiology*. 2019;48(6):1740-g.
229. Boyd A, Cornish R, Johnson L, Simmonds S, Syddall H, Westbury L. *Understanding Hospital episode statistics (HES)*. London, UK: CLOSER. 2017.
230. Statistics OfN. *Mortality statistics: metadata*. ONS Cardiff, UK; 2014.
231. Henson KE, Elliss-Brookes L, Coupland VH, Payne E, Vernon S, Rous B, et al. Data resource profile: national cancer registration dataset in England. *International journal of epidemiology*. 2020;49(1):16-h.
232. NICE. *Quality and Outcomes Framework Indicators* [Website]. [Available from: <https://www.nice.org.uk/standards-and-indicators/qofindicators?categories=&page=3>.
233. Waller P, Cassell JA, Saunders MH, Stevens R. Governance and oversight of researcher access to electronic health data: the role of the independent scientific Advisory Committee for MHRA database research, 2006–2015. *Journal of the Royal College of Physicians of Edinburgh*. 2017;47(1):24-9.
234. Thampanitchawong P, Piratvisuth T. Liver biopsy: complications and risk factors. *World journal of gastroenterology*. 1999;5(4):301.
235. Chalasani N, Younossi Z, Lavine JE, Diehl AM, Brunt EM, Cusi K, et al. *The diagnosis and management of non-alcoholic fatty*. 2012.



236. Bradji L, Boufaïda M, editors. Knowledge based data cleaning for data warehouse quality. International Conference on Digital Information Processing and Communications; 2011: Springer.
237. Washington IM, Van Hoosier G. Clinical biochemistry and hematology. The laboratory rabbit, guinea pig, hamster, and other rodents: Elsevier; 2012. p. 57-116.
238. Daniel S, Ben-Menachem T, Vasudevan G, Ma CK, Blumenkehl M. Prospective evaluation of unexplained chronic liver transaminase abnormalities in asymptomatic and symptomatic patients. The American journal of gastroenterology. 1999;94(10):3010-4.
239. Lilford R, Bentham L, Girling A, Litchfield I, Lancashire R, Armstrong D, et al. Birmingham and Lambeth Liver Evaluation Testing Strategies (BALLETS): a prospective cohort study. Health technology assessment (Winchester, England). 2013;17(28):i-xiv, 1.
240. Wang Y, Hunt K, Nazareth I, Freemantle N, Petersen I. Do men consult less than women? An analysis of routinely collected UK general practice data. BMJ open. 2013;3(8):e003320.
241. Hobbs FR, Bankhead C, Mukhtar T, Stevens S, Perera-Salazar R, Holt T, et al. Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007–14. The Lancet. 2016;387(10035):2323-30.
242. Stroffolini T, Esvan R, Biliotti E, Sagnelli E, Gaeta GB, Almasio PL. Gender differences in chronic HBsAg carriers in Italy: Evidence for the independent role of male sex in severity of liver disease. Journal of medical virology. 2015;87(11):1899-903.
243. Lau-Corona D, Bae WK, Hennighausen L, Waxman DJ. Sex-biased genetic programs in liver metabolism and liver fibrosis are controlled by EZH1 and EZH2. PLoS genetics. 2020;16(5):e1008796.
244. Prati D, Taioli E, Zanella A, Torre ED, Butelli S, Del Vecchio E, et al. Updated definitions of healthy ranges for serum alanine aminotransferase levels. Annals of internal medicine. 2002;137(1):1-10.
245. Giannini EG, Testa R, Savarino V. Liver enzyme alteration: a guide for clinicians. Cmaj. 2005;172(3):367-79.
246. Agency UHS. Liver disease: a preventable killer of young adults 2014 [
247. UK A. Briefing: Health and care of older people in England 2019. February; 2019.
248. Mejlidal A, Andersen K, Behrendt S, Bilberg R, Christensen AI, Lau CJ, et al. History of healthcare use and disease burden in older adults with different levels of alcohol use. A register-based cohort study. Alcoholism: Clinical and Experimental Research. 2021;45(6):1237-48.
249. Tajiri K, Shimizu Y. Liver physiology and liver diseases in the elderly. World journal of gastroenterology: WJG. 2013;19(46):8459.
250. Dong MH, Bettencourt R, Barrett-Connor E, Loomba R. Alanine aminotransferase decreases with age: the Rancho Bernardo Study. PloS one. 2010;5(12):e14254.
251. Cooper KM, Delk M, Devuni D, Sarkar M. Sex differences in chronic liver disease and benign liver lesions. JHEP Reports. 2023:100870.

252. Ratziu HT, Roden M, Gastaldelli A, Yki-Jarvinen H, Schick F. EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease.
253. Mahady SE, Macaskill P, Craig JC, Wong GL, Chu WC, Chan HL, et al. Diagnostic accuracy of noninvasive fibrosis scores in a population of individuals with a low prevalence of fibrosis. *Clinical Gastroenterology and Hepatology*. 2017;15(9):1453-60. e1.
254. Aller R, Sigüenza R, Pina M, Laserna C, Antolín B, Burgueño B, et al. Insulin resistance is related with liver fibrosis in type 2 diabetic patients with non-alcoholic fatty liver disease proven biopsy and Mediterranean diet pattern as a protective factor. *Endocrine*. 2020.
255. Tanajewski L, Harris R, Harman DJ, Aithal GP, Card TR, Gkountouras G, et al. Economic evaluation of a community-based diagnostic pathway to stratify adults for non-alcoholic fatty liver disease: a Markov model informed by a feasibility study. *BMJ open*. 2017;7(6):e015659.
256. Jarvis H, Sanders T, Hanratty B. Liver disease management as routine work in primary care: a qualitative interview study to guide implementation. *British Journal of General Practice*. 2022;72(725):e916-e23.
257. Grattagliano I, Ubaldi E, Bonfrate L, Portincasa P. Management of liver cirrhosis between primary care and specialists. *World journal of gastroenterology: WJG*. 2011;17(18):2273.
258. Muller S, Hider SL, Raza K, Stack RJ, Hayward RA, Mallen CD. An algorithm to identify rheumatoid arthritis in primary care: a Clinical Practice Research Datalink study. *BMJ open*. 2015;5(12):e009309.
259. 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.
260. Park N. Population estimates by output areas, electoral, health and other geographies, England and Wales: mid-2018. Office Natl Stat. 2019.
261. 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. *Journal of hepatology*. 2014;60(2):282-9.
262. Saine ME, Carbonari DM, Newcomb CW, Gallagher AM, Blak BT, Roy JA, et al. Concordance of hospitalizations between Clinical Practice Research Datalink and linked Hospital Episode Statistics among patients treated with oral antidiabetic therapies. *Pharmacoepidemiology and drug safety*. 2019;28(10):1328-35.
263. McDonald L, Sammon CJ, Samnaliev M, Ramagopalan S. Under-recording of hospital bleeding events in UK primary care: a linked Clinical Practice Research Datalink and Hospital Episode Statistics study. *Clinical epidemiology*. 2018:1155-68.
264. Devarbhavi H, Asrani SK, Arab JP, Nartey YA, Pose E, Kamath PS. Global burden of liver disease: 2023 update. *Journal of Hepatology*. 2023.
265. Vaz J, Eriksson B, Strömberg U, Buchebner D, Midlöv P. Incidence, aetiology and related comorbidities of cirrhosis: a Swedish population-based cohort study. *BMC gastroenterology*. 2020;20:1-8.
266. Bolondi L, Gramantieri L. From liver cirrhosis to HCC. *Internal and emergency medicine*. 2011;6:93-8.

267. Lee H, Shin DW, Lee TH, Yang H-K, Ahn E, Yoon J-M, et al. Association between change in serum aminotransferase and mortality: a nationwide cohort study in Korea. *Medicine*. 2016;95(12).
268. Karlsen TH, Sheron N, Zelber-Sagi S, Carrieri P, Dusheiko G, Bugianesi E, et al. The EASL–Lancet Liver Commission: protecting the next generation of Europeans against liver disease complications and premature mortality. *The Lancet*. 2022;399(10319):61-116.
269. Excellence NIfC. Non-alcoholic fatty liver disease (NAFLD): assessment and management (NG49). National Institute for Health and Clinical Excellence (NICE), London. 2016.
270. Smith A, Baumgartner K, Bositis C. Cirrhosis: diagnosis and management. *American family physician*. 2019;100(12):759-70.
271. Venkatesh SK, Torbenson MS. Liver fibrosis quantification. *Abdominal Radiology*. 2022;47(3):1032-52.
272. Oeda S, Tanaka K, Oshima A, Matsumoto Y, Sueoka E, Takahashi H. Diagnostic accuracy of FibroScan and factors affecting measurements. *Diagnostics*. 2020;10(11):940.
273. Yang JD, Abdelmalek MF, Pang H, Guy CD, Smith AD, Diehl AM, et al. Gender and menopause impact severity of fibrosis among patients with nonalcoholic steatohepatitis. *Hepatology*. 2014;59(4):1406-14.
274. Iqbal U, Perumpail BJ, Akhtar D, Kim D, Ahmed A. The epidemiology, risk profiling and diagnostic challenges of nonalcoholic fatty liver disease. *Medicines*. 2019;6(1):41.
275. Qiu J, Kuang M, He S, Yu C, Wang C, Huang X, et al. Gender perspective on the association between liver enzyme markers and non-alcoholic fatty liver disease: insights from the general population. *Frontiers in Endocrinology*. 2023;14:1302322.
276. Kim WR, Brown Jr RS, Terrault NA, El-Serag H. Burden of liver disease in the United States: summary of a workshop. *Hepatology*. 2002;36(1):227-42.
277. Bertakis KD, Azari R. Patient-centered care: the influence of patient and resident physician gender and gender concordance in primary care. *Journal of women's health*. 2012;21(3):326-33.
278. Wong RJ, Kachru N, Martinez DJ, Moynihan M, Ozbay AB, Gordon SC. Real-world comorbidity burden, health care utilization, and costs of nonalcoholic steatohepatitis patients with advanced liver diseases. *Journal of clinical gastroenterology*. 2021;55(10):891-902.
279. Austad SN. Why women live longer than men: sex differences in longevity. *Gender medicine*. 2006;3(2):79-92.
280. McMaughan DJ, Oloruntoba O, Smith ML. Socioeconomic status and access to healthcare: interrelated drivers for healthy aging. *Frontiers in public health*. 2020;8:231.
281. Alvarez PM, McKeon JF, Spitzer AI, Krueger CA, Pigott M, Li M, et al. Socioeconomic factors affecting outcomes in total knee and hip arthroplasty: a systematic review on healthcare disparities. *Arthroplasty*. 2022;4(1):1-11.

282. Nallagangula KS, Nagaraj SK, Venkataswamy L, Chandrappa M. Liver fibrosis: a compilation on the biomarkers status and their significance during disease progression. *Future Science OA*. 2017;4(1):FSO250.
283. Sheth SG, Flamm SL, Gordon FD, Chopra S. AST/ALT ratio predicts cirrhosis in patients with chronic hepatitis C virus infection. *The American journal of gastroenterology*. 1998;93(1):44-8.
284. Matteoni CA, Younossi ZM, Gramlich T, Boparai N, Liu YC, McCullough AJ. Nonalcoholic fatty liver disease: a spectrum of clinical and pathological severity. *Gastroenterology*. 1999;116(6):1413-9.
285. Gallacher J, McPherson S. Practical diagnosis and staging of nonalcoholic fatty liver disease: A narrative review. *EMJ*. 2018;3(2):108-18.
286. Kim BK, Kim DY, Park JY, Ahn SH, Chon CY, Kim JK, et al. Validation of FIB-4 and comparison with other simple noninvasive indices for predicting liver fibrosis and cirrhosis in hepatitis B virus-infected patients. *Liver International*. 2010;30(4):546-53.
287. Shearer JE, Gonzalez JJ, Min T, Parker R, Jones R, Su GL, et al. Systematic review: development of a consensus code set to identify cirrhosis in electronic health records. *Alimentary Pharmacology & Therapeutics*. 2022;55(6):645-57.

# Appendices

## Appendix A

JBI critical appraisal checklist for prevalence studies.

### Quality assessment

Reviewer

Date

Author

Year

Record Number

|                                                                       | Yes | No | Unclear | Not applicable |
|-----------------------------------------------------------------------|-----|----|---------|----------------|
| 1. Was the sample frame appropriate to address the target population? |     |    |         |                |
| 2. Were study participants sampled in an appropriate way?             |     |    |         |                |
| 3. Was the sample size adequate?                                      |     |    |         |                |
| 4. Were the study subjects and the setting described in detail?       |     |    |         |                |
| 5. Was the data analysis conducted with                               |     |    |         |                |

|                                                                                                 |  |  |  |  |
|-------------------------------------------------------------------------------------------------|--|--|--|--|
| sufficient coverage of the identified sample?                                                   |  |  |  |  |
| 6. Were valid methods used for the identification of the condition?                             |  |  |  |  |
| 7. Was the condition measured in a standard, reliable way for all participants?                 |  |  |  |  |
| 8. Was there an appropriate statistical analysis?                                               |  |  |  |  |
| 9. Was the response rate adequate, and if not, was the low response rate managed appropriately? |  |  |  |  |

Overall appraisal:            Include       Exclude       Seek further info

Comments (Including reason for exclusion)

## **Appendix B**

# CPRD GOLD Data Specification

Version 2.3

Date: 1 February 2021





### Documentation Control Sheet

During the course of the project it may be necessary to issue amendments or clarifications to parts of this document. This form must be updated whenever changes are made and should be filed inside the front cover of the new or amended document.

| Version | Summary of Change | Prepared By                   | Date       | Reviewed By                      | Date       |
|---------|-------------------|-------------------------------|------------|----------------------------------|------------|
| 1.0     | Initial Draft     |                               |            |                                  |            |
| 1.1     | Modified          | Shivani Padmanabhan           | 01/06/2009 | Nick Wilson                      | 22/07/2009 |
| 1.2     | Modified          | Shivani Padmanabhan           | 28/07/2009 | Arlene Gallagher                 | 30/07/2009 |
| 1.3     | Modified          | Shivani Padmanabhan           | 06/01/2011 | Nick Wilson                      | 07/01/2011 |
| 1.4     | Modified          | Shivani Padmanabhan           | 11/01/2013 | Nick Wilson                      | 11/01/2013 |
| 1.5     | Modified          | Shivani Padmanabhan           | 31/07/2013 | Nick Wilson                      | 03/08/2013 |
| 1.6     | Formatted         | Arlene Gallagher              | 11/12/2013 | Shivani Padmanabhan              | 30/12/2013 |
| 1.7     | Modified          | Jennifer Chapman              | 19/11/2014 | Darren Lunn                      | 20/11/2014 |
| 1.8     | Modified          | Mark Hobbs                    | 06/05/2015 | Jennifer Chapman                 | 12/06/2015 |
| 1.9     | Formatted         | Grant Lee                     | 02/07/2015 | Jennifer Chapman                 | 03/07/2015 |
| 2.0     | Modified          | Shivani Padmanabhan           | 01/09/2017 | Darren Lunn                      | 01/09/2017 |
| 2.1     | Modified          | Hilary Shepherd               | 09/09/2020 | Rachael Williams                 | 09/09/2020 |
| 2.2     | Modified          | Hilary Shepherd/ Kirsty Syder | 17/11/2020 | Arlene Gallagher/Eleanor Yelland | 27/01/2021 |
| 2.3     | Modified          | Eleanor Yelland               | 01/02/2021 | Hilary Shepherd                  | 01/02/2021 |

### Summary of Changes

#### Version 1.1

- Refined wordings

#### Version 1.2

- Acceptable field in Patient file equals 1 if patient is acceptable, else 0 (Lookup reference incorrectly labelled as Y\_N in previous versions)
- UTS field in Practice file has been derived using a CPRD algorithm that looks at death recording at the practice, and gaps in the data (prior to August 2009, this field was populated with the practice UTS date as was generated in the old FF-CPRD system)
- The ndd field in the Therapy file has been populated for the most common occurring dosage strings in the data (field was set to '0' prior to August 2009)
- Descriptions of all fields have been revised for clarity

#### Version 1.3

- Field name 'attendtype' in Referral table modified to 'attendance'

#### Version 1.4

- Reference to Multilex product code system has been changed to Gemscript

#### Version 1.5

- Added batch number as a field in immunisation
- Description and Mapping for 'data8' in Test file has been amended

#### Version 1.6

- New CPRD branding and formatting

#### Version 1.7

- Mapping of 'data8' field in Test file has been amended to remove reference to GEN\_SDC
- Description of date formatting has been updated to reflect new format
- Description of entity types has been updated to include date values

#### Version 1.8

- Merged information from Spec 1.5 and Spec 1.7
- All description tables now list Column Name, Field Name, Description, Mapping, Type, Format

#### Version 1.9

- Minor formatting of document

#### Version 2.0

- Removed textid fields from the clinical, immunisation, test and referral tables
- Removed textid and ndd fields, and added dosageid as a field in the therapy table
- Removed ses field from the patient table

#### Version 2.1

- Added medical and product dictionary tables and information on data dictionaries
- Birth year description updated to reflect change in mapping
- Changed "CHAR" data type to "TEXT" to align with the CPRD Aurum specification document

#### Version 2.2

- Updated data type for patient and practice IDs
- Aligned CPRD GOLD and CPRD Aurum specifications
- Updated for 5-digit patid
- Updated branding and formatting

#### Version 2.3

- Harmonised updates to documentation

### **Dataset Format**

1. The **Patient** file (Patient\_ANN.txt) contains basic patient demographics and patient registration details for the patients.
2. The **Practice** file (Practice\_ANN.txt) contains details of each practice, including region and collection information.
3. The **Staff** file (Staff\_ANN.txt) contains practice staff details, with one record per member of staff.
4. The **Consultation** file (Consultation\_ANN.txt) contains information relating to the type of consultation as entered by the GP from a pre-determined list. Consultations can be linked to the events that occur as part of the consultation via the consultation identifier (consid).
5. The **Clinical** file (Clinical\_ANN.txt) contains medical history events. This file contains all the medical history data entered on the GP system, including symptoms, signs and diagnoses. This can be used to identify any clinical diagnoses, and deaths. Patients may have more than one row of data. The data is coded using Read codes, which allow linkage of codes to the medical terms provided.
6. The **Additional Clinical Details** file (Additional\_ANN.txt) contains information entered in the structured data areas in the GP's software. Patients may have more than one row of data. Data in this file is linked to events in the clinical file through the additional details identifier (adid).
7. The **Referral** file (Referral\_ANN.txt) contains referral details recorded on the GP system. These files contain information involving patient referrals to external care centres (normally to secondary care locations such as hospitals for inpatient or outpatient care), and include speciality and referral type.
8. The **Immunisation** file (Immunisation\_ANN.txt) contains details of immunisation records on the GP system.
9. The **Test** file (Test\_ANN.txt) contains records of test data on the GP system. The data is coded using a Read code, chosen by the GP, which will generally identify the type of test used. The test name is identified via the *entity type*, a numerical code, which is determined by the test result item chosen by the GP at source. There are three types of test records, involving 4, 7 or 8 data fields (data1 - data8). The data must be managed according to which sort of test record it is. Data can denote either qualitative entries (for example 'Normal' or Abnormal') or quantitative entries involving a numeric value.
10. The **Therapy** file (Therapy\_ANN.txt) contains details of all prescriptions on the GP system. This file contains data relating to all prescriptions (for drugs and appliances) issued by the GP. Patients may have more than one row of data. Drug products and appliances are recorded by the GP using the Gemscript product code system.

### ***Data dictionaries***

CPRD GOLD dictionaries are provided as text files that can be imported into standard statistical software to enable code searching. The dictionaries are also available through the CPRD Code Browser. The CPRD Code Browser and a user guide can be requested by contacting [enquiries@cprd.com](mailto:enquiries@cprd.com). If you are already using the code browser to search the CPRD Aurum dictionaries you will still need to contact us to download the latest browser containing the CPRD GOLD dictionaries.

- I. The **medical** dictionary contains information on all medical history observations recorded by the GP using Read version (v) 2 codes and the descriptions of all medical codes referenced in the data files as 'medcode'. The medical dictionary file is provided in tab-delimited text format.
- II. The **product** dictionary contains information on drug and appliance prescriptions recorded by the GP using the Genscript product code system (brand and generic name) and descriptions of all product codes referenced in the data files as 'prodcode'. The product dictionary file is provided in tab-delimited text format.

### Field descriptions

Full descriptions of fields in each file are provided in the tables below. All files can be linked using the encrypted patient identifier (patid). The last five digits of the patient identifier (patid) and staff identifier (staffid) denote the identifier of the practice (pracid) that the patient or staff member belongs to. All pracids in CPRD GOLD start with 1 as the first digit. The mapping column lists lookup files with further information on decoding numerical values. A mapping of 'None' indicates the existence of raw data in the field.

#### 1. Patient

| Column name               | Field name | Description                                                                                                                                                                                                                                                                | Mapping    | Type    | Format     |
|---------------------------|------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|---------|------------|
| Patient Identifier        | patid      | Encrypted unique identifier given to a patient in CPRD GOLD                                                                                                                                                                                                                | None       | TEXT    | 20         |
| VAMP Identifier           | vmid       | Old VM id for the patient when the practice was using the VAMP system                                                                                                                                                                                                      | None       | INTEGER | 20         |
| Patient Gender            | gender     | Patient's gender                                                                                                                                                                                                                                                           | Lookup SEX | INTEGER | 1          |
| Birth Year                | yob        | Patient's year of birth. This is actual year of birth e.g. 1984                                                                                                                                                                                                            | None       | INTEGER | 4          |
| Birth Month               | mob        | Patient's month of birth (for those aged under 16). 0 indicates no month set                                                                                                                                                                                               | None       | INTEGER | 2          |
| Marital Status            | marital    | Patient's current marital status                                                                                                                                                                                                                                           | Lookup MAR | INTEGER | 3          |
| Family Number             | famnum     | Family ID number                                                                                                                                                                                                                                                           | None       | INTEGER | 20         |
| CHS Registered            | chsreg     | Value to indicate whether the patient is registered with Child Health Surveillance                                                                                                                                                                                         | Lookup Y_N | INTEGER | 1          |
| CHS Registration Date     | chsdate    | Date of registration with Child Health Surveillance                                                                                                                                                                                                                        | DD/MM/YYYY | DATE    | DD/MM/YYYY |
| Prescription Exemption    | prescr     | Type of prescribing exemption the patient has currently (e.g. medical / maternity)                                                                                                                                                                                         | Lookup PEX | INTEGER | 3          |
| Capitation Supplement     | capsup     | Level of capitation supplement the patient has currently (e.g. low, medium, high)                                                                                                                                                                                          | Lookup CAP | INTEGER | 3          |
| First Registration Date   | frd        | Date the patient first registered with the practice. If patient only has 'temporary' records, the date is the first encounter with the practice; if patient has 'permanent' records it is the date of the first 'permanent' record (excluding preceding temporary records) | DD/MM/YYYY | DATE    | DD/MM/YYYY |
| Current Registration Date | crd        | Date the patient's current period of registration with the practice began (date of the first 'permanent' record after the latest transferred out period). If there are no 'transferred out periods', the date is equal to 'frd'                                            | DD/MM/YYYY | DATE    | DD/MM/YYYY |

|                         |           |                                                                                                                |                       |         |            |
|-------------------------|-----------|----------------------------------------------------------------------------------------------------------------|-----------------------|---------|------------|
| Registration Status     | regstat   | Status of registration detailing gaps and temporary patients                                                   | PAT_STAT <sup>1</sup> | INTEGER | 2          |
| Registration Gaps       | reggap    | Number of days missing in the patient's registration details                                                   | PAT_GAP <sup>2</sup>  | INTEGER | 5          |
| Internal Transfer       | internal  | Number of internal transfer out periods, in the patient's registration details                                 | None                  | INTEGER | 2          |
| Transfer Out Date       | tod       | Date the patient transferred out of the practice, if relevant. Empty for patients who have not transferred out | DD/MM/YYYY            | DATE    | DD/MM/YYYY |
| Transfer Out Reason     | toreason  | Reason the patient transferred out of the practice. Includes 'Death' as an option                              | Lookup TRA            | INTEGER | 3          |
| Death Date              | deathdate | Date of death of patient – derived using a CPRD algorithm                                                      | DD/MM/YYYY            | DATE    | DD/MM/YYYY |
| Acceptable Patient Flag | accept    | Flag to indicate whether the patient has met certain quality standards: 1 = acceptable, 0 = unacceptable       | Boolean               | INTEGER | 1          |

<sup>1</sup> **PAT\_STAT:** Transferred out period is the time between a patient transferring out and re-registering at the same practice. If the patient has transferred out for a period of more than 1 day, and the transfer is not internal, this value is incremented. 0 means continuous registration, 1 means one 'transferred out period', 2 means two periods, etc. If the patient only has 'temporary' records, then this value is set to 99.

<sup>2</sup> **PAT\_GAP:** Number of days between patient's transferred out date and re-registration date for the patient's 'transferred out periods', regardless of whether the transfer was internal or not.

#### 6. Additional Clinical Details

| Column name                   | Field name | Description                                                                                                                | Mapping                    | Type         | Format          |
|-------------------------------|------------|----------------------------------------------------------------------------------------------------------------------------|----------------------------|--------------|-----------------|
| Patient Identifier            | patid      | Encrypted unique identifier given to a patient in CPRD GOLD                                                                | None                       | TEXT         | 20              |
| Entity Type                   | enttype    | Identifier that represents the structured data area in Vision where the data was entered                                   | Lookup Entity              | INTEGER      | 5               |
| Additional Details Identifier | adid       | Identifier that allows information about the original clinical event to be retrieved, when used in combination with pracid | Link <b>Clinical table</b> | INTEGER      | 20              |
| Data 1                        | data1      | Depends on entity type *                                                                                                   | Lookup Entity              | NUMERIC DATE | 15.3 DD/MM/YYYY |
| Data 2                        | data2      | Depends on entity type *                                                                                                   | Lookup Entity              | NUMERIC DATE | 15.3 DD/MM/YYYY |
| Data 3                        | data3      | Depends on entity type *                                                                                                   | Lookup Entity              | NUMERIC DATE | 15.3 DD/MM/YYYY |
| Data 4                        | data4      | Depends on entity type *                                                                                                   | Lookup Entity              | INTEGER DATE | 12 DD/MM/YYYY   |
| Data 5                        | data5      | Depends on entity type *                                                                                                   | Lookup Entity              | INTEGER DATE | 12 DD/MM/YYYY   |
| Data 6                        | data6      | Depends on entity type *                                                                                                   | Lookup Entity              | INTEGER DATE | 12 DD/MM/YYYY   |
| Data 7                        | data7      | Depends on entity type *                                                                                                   | Lookup Entity              | INTEGER DATE | 4 DD/MM/YYYY    |

\* Each entity type may be associated with up to seven data fields. Content of each data field is dependent on the entity type – the fields may contain raw data values, dates in the form dd/mm/yyyy, or may be encoded values that represent read codes, text etc. The file Entity.xls contains information on all entity types, and provides the number of data fields associated with the entity, description of the data in each field, and details of the lookups needed to decode the data.

## 7. Referral

| Column name             | Field name | Description                                                                                                     | Mapping                   | Type    | Format     |
|-------------------------|------------|-----------------------------------------------------------------------------------------------------------------|---------------------------|---------|------------|
| Patient Identifier      | patid      | Encrypted unique identifier given to a patient in CPRD GOLD                                                     | None                      | TEXT    | 20         |
| Event Date              | eventdate  | Date associated with the event, as entered by the GP                                                            | DD/MM/YYYY                | DATE    | DD/MM/YYYY |
| System Date             | sysdate    | Date the event was entered into Vision                                                                          | DD/MM/YYYY                | DATE    | DD/MM/YYYY |
| Consultation Type       | constype   | Code for the category of event recorded within the GP system (e.g. management or administration)                | Lookup SED                | INTEGER | 3          |
| Consultation Identifier | consid     | Identifier that allows information about the consultation to be retrieved, when used in combination with pracid | Link Consultation table   | INTEGER | 20         |
| Medical Code            | medcode    | CPRD unique code for the medical term selected by the GP                                                        | Lookup Medical Dictionary | INTEGER | 20         |
| Staff Identifier        | staffid    | Identifier of the practice staff member entering the data. A value of 0 indicates that the staffid is unknown   | Link Staff table          | INTEGER | 20         |
| Source                  | source     | Classification of the source of the referral e.g. GP, Self                                                      | Lookup SOU                | INTEGER | 2          |
| NHS Speciality          | nhsspec    | Referral speciality according to the National Health Service (NHS) classification                               | Lookup DEP                | INTEGER | 3          |
| FHSA Speciality         | fhsaspec   | Referral speciality according to the Family Health Services Authority (FHSA) classification                     | Lookup SPE                | INTEGER | 3          |
| In Patient              | inpatient  | Classification of the type of referral, e.g. Day case, In patient                                               | Lookup RFT                | INTEGER | 2          |
| Attendance Type         | attendance | Category describing whether the referral event is the first visit, a follow-up etc.                             | Lookup ATT                | INTEGER | 2          |
| Urgency                 | urgency    | Classification of the urgency of the referral e.g. Routine, Urgent                                              | Lookup URG                | INTEGER | 2          |



## 8. Immunisation

| Column name             | Field name | Description                                                                                                     | Mapping                   | Type    | Format     |
|-------------------------|------------|-----------------------------------------------------------------------------------------------------------------|---------------------------|---------|------------|
| Patient Identifier      | patid      | Encrypted unique identifier given to a patient in CPRD GOLD                                                     | None                      | TEXT    | 20         |
| Event Date              | eventdate  | Date associated with the event, as entered by the GP                                                            | DD/MM/YYYY                | DATE    | DD/MM/YYYY |
| System Date             | sysdate    | Date the event was entered into Vision                                                                          | DD/MM/YYYY                | DATE    | DD/MM/YYYY |
| Consultation Type       | constype   | Code for the category of event recorded within the GP system (e.g. intervention)                                | Lookup SED                | INTEGER | 3          |
| Consultation Identifier | consid     | Identifier that allows information about the consultation to be retrieved, when used in combination with pracid | Link Consultation table   | INTEGER | 20         |
| Medical Code            | medcode    | CPRD unique code for the medical term selected by the GP                                                        | Lookup Medical Dictionary | INTEGER | 20         |
| Staff Identifier        | staffid    | Identifier of the practice staff member entering the data. A value of 0 indicates that the staffid is unknown   | Link Staff table          | INTEGER | 20         |
| Type                    | immstype   | Individual components of an immunisation, e.g. Mumps, Rubella, Measles                                          | Lookup IMT                | INTEGER | 4          |
| Stage                   | stage      | Stage of the immunisation given, e.g. 1, 2, B2                                                                  | Lookup IST                | INTEGER | 2          |
| Status                  | status     | Status of the immunisation e.g. Advised, Given, Refusal                                                         | Lookup IMM                | INTEGER | 3          |
| Compound                | compound   | Immunisation compound administered – may be a single or multi-component preparation, e.g. MMR                   | Lookup IMC                | INTEGER | 4          |
| Source                  | source     | Location where the immunisation was administered, e.g. In this practice                                         | Lookup INP                | INTEGER | 3          |
| Reason                  | reason     | Reason for administering the immunisation, e.g. Routine measure                                                 | Lookup RIN                | INTEGER | 3          |
| Method                  | method     | Route of administration for the immunisation, e.g. Oral, Intramuscular                                          | Lookup IME                | INTEGER | 3          |
| Batch Number            | batch      | Immunisation batch number                                                                                       | Lookup BatchNumber        | INTEGER | 20         |

#### 9. Test

| Column name             | Field name | Description                                                                                                     | Mapping                   | Type    | Format     |
|-------------------------|------------|-----------------------------------------------------------------------------------------------------------------|---------------------------|---------|------------|
| Patient Identifier      | patid      | Encrypted unique identifier given to a patient in CPRD GOLD                                                     | None                      | TEXT    | 20         |
| Event Date              | eventdate  | Date associated with the event, as entered by the GP                                                            | DD/MM/YYYY                | DATE    | DD/MM/YYYY |
| System Date             | sysdate    | Date the event was entered into Vision                                                                          | DD/MM/YYYY                | DATE    | DD/MM/YYYY |
| Consultation Type       | constype   | Code for the category of event recorded within the GP system (e.g. examination)                                 | Lookup SED                | INTEGER | 3          |
| Consultation Identifier | consid     | Identifier that allows information about the consultation to be retrieved, when used in combination with pracid | Link Consultation table   | INTEGER | 20         |
| Medical Code            | medcode    | CPRD unique code for the medical term selected by the GP                                                        | Lookup Medical Dictionary | INTEGER | 20         |
| Staff Identifier        | staffid    | Identifier of the practice staff member entering the data. A value of 0 indicates that the staffid is unknown   | Link Staff table          | INTEGER | 20         |
| Entity Type             | enttype    | Identifier that represents the structured data area in Vision where the data was entered                        | Lookup Entity             | INTEGER | 5          |

Depending on the Test entity type, tests have 4, 7, or 8 data fields:

4 fields:

|        |       |                    |            |         |      |
|--------|-------|--------------------|------------|---------|------|
| Data 1 | data1 | Qualifier          | Lookup TQU | INTEGER | 3    |
| Data 2 | data2 | Normal range from  | None       | NUMERIC | 16.3 |
| Data 3 | data3 | Normal range to    | None       | NUMERIC | 16.3 |
| Data 4 | data4 | Normal range basis | None       | NUMERIC | 16.3 |

7 fields:

|        |       |                                                                 |                        |         |      |
|--------|-------|-----------------------------------------------------------------|------------------------|---------|------|
| Data 1 | data1 | Operator                                                        | Lookup OPR             | INTEGER | 3    |
| Data 2 | data2 | Value                                                           | None                   | NUMERIC | 16.3 |
| Data 3 | data3 | Unit of measure                                                 | Lookup SUM             | INTEGER | 4    |
| Data 4 | data4 | Qualifier                                                       | Lookup TQU             | INTEGER | 3    |
| Data 5 | data5 | Normal range from                                               | None                   | NUMERIC | 16.3 |
| Data 6 | data6 | Normal range to                                                 | None                   | NUMERIC | 16.3 |
| Data 7 | data7 | Normal range basis<br>(or peak flow device for entity type 311) | Lookup POP<br>(or PFD) | INTEGER | 2    |

8 fields:

|        |       |                                                                       |                    |                 |                  |
|--------|-------|-----------------------------------------------------------------------|--------------------|-----------------|------------------|
| Data 1 | data1 | Operator                                                              | Lookup OPR         | INTEGER         | 3                |
| Data 2 | data2 | Value                                                                 | None               | NUMERIC         | 16.3             |
| Data 3 | data3 | Unit of measure                                                       | Lookup SUM         | INTEGER         | 4                |
| Data 4 | data4 | Qualifier                                                             | Lookup TQU         | INTEGER         | 3                |
| Data 5 | data5 | Normal range from                                                     | None               | NUMERIC         | 16.3             |
| Data 6 | data6 | Normal range to                                                       | None               | NUMERIC         | 16.3             |
| Data 7 | data7 | Normal range basis                                                    | Lookup POP         | INTEGER         | 2                |
| Data 8 | data8 | Expected delivery date (entity type 284) /<br>Weeks (entity type 154) | DD/MM/YYYY<br>None | DATE<br>INTEGER | DD/MM/YYYY<br>10 |

#### 10. Therapy

| Column name             | Field name | Description                                                                                                                                                                                                                      | Mapping                        | Type    | Format     |
|-------------------------|------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------|---------|------------|
| Patient Identifier      | patid      | Encrypted unique identifier given to a patient in CPRD GOLD                                                                                                                                                                      | None                           | TEXT    | 20         |
| Event Date              | eventdate  | Date associated with the event, as entered by the GP                                                                                                                                                                             | DD/MM/YYYY                     | DATE    | DD/MM/YYYY |
| System Date             | sysdate    | Date the event was entered into Vision                                                                                                                                                                                           | DD/MM/YYYY                     | DATE    | DD/MM/YYYY |
| Consultation Identifier | consid     | Identifier that allows information about the consultation to be retrieved, when used in combination with pracid                                                                                                                  | Link <b>Consultation</b> table | INTEGER | 20         |
| Product Code            | prodcode   | CPRD unique code for the treatment selected by the GP                                                                                                                                                                            | Lookup Product Dictionary      | INTEGER | 20         |
| Staff Identifier        | staffid    | Identifier of the practice staff member entering the data. A value of 0 indicates that the staffid is unknown                                                                                                                    | Link <b>Staff</b> table        | INTEGER | 20         |
| Dosage Identifier       | dosageid   | Identifier that allows dosage information on the event to be retrieved. Use the Common Dosages Lookup to obtain the anonymised dosage text and extracted numerical information such as daily dose.                               | Lookup Common Dosages          | TEXT    | 64         |
| BNF Code                | bnfcode    | Code representing the chapter & section from the British National Formulary for the product selected by GP                                                                                                                       | Lookup BNF Codes               | INTEGER | 5          |
| Total Quantity          | qty        | Total quantity entered by the GP for the prescribed product                                                                                                                                                                      | None                           | INTEGER | 20         |
| Number of Days          | numdays    | Number of treatment days prescribed for a specific therapy event                                                                                                                                                                 | None                           | INTEGER | 20         |
| Number of Packs         | numpacks   | Number of individual product packs prescribed for a specific therapy event                                                                                                                                                       | None                           | INTEGER | 8          |
| Pack Type               | packtype   | Pack size or type of the prescribed product                                                                                                                                                                                      | Lookup PackType                | INTEGER | 10         |
| Issue Sequence Number   | issueseq   | Number to indicate whether the event is associated with a repeat schedule. Value of 0 implies the event is not part of a repeat prescription. A value ≥ 1 denotes the issue number for the prescription within a repeat schedule | None                           | INTEGER | 20         |

I. Medical dictionary

| Column name | Description                                              | Mapping | Type    | Format |
|-------------|----------------------------------------------------------|---------|---------|--------|
| medcode     | CPRD unique code for the medical term selected by the GP | None    | INTEGER | 20     |
| readcode    | Read Code                                                | None    | TEXT    | 7      |
| desc        | Description of the medical term                          | None    | TEXT    | 100    |

II. Product dictionary

| Column name   | Description                                                                                                                                     | Mapping | Type    | Format |
|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------|---------|---------|--------|
| prodcode      | CPRD unique code for the treatment selected by the GP                                                                                           | None    | INTEGER | 20     |
| dmdcode       | Unique product identifier from the NHS Dictionary of Medicines and Devices (dm+d) – the NHS standard dictionary for products licensed in the UK | None    | TEXT    | 100    |
| gemsriptcode  | Gemsript product code for the corresponding product name - should be treated as a string field as it contains leading '0's                      | None    | TEXT    | 8      |
| productname   | Product name as entered at the practice                                                                                                         | None    | TEXT    | 500    |
| drugsubstance | Drug substance                                                                                                                                  | None    | TEXT    | 1500   |
| strength      | Strength of the product                                                                                                                         | None    | TEXT    | 500    |
| formulation   | Form of the product e.g. tablets, capsules etc                                                                                                  | None    | TEXT    | 100    |
| route         | Route of administration of the product                                                                                                          | None    | TEXT    | 100    |
| bnfcode       | British National Formulary (BNF) code                                                                                                           | None    | TEXT    | 100    |
| bnfchapter    | British National Formulary (BNF) chapter                                                                                                        | None    | TEXT    | 500    |

## Appendix C

# ISAC APPLICATION FORM

## PROTOCOLS FOR RESEARCH USING THE CLINICAL PRACTICE RESEARCH DATALINK (CPRD)

|                              |       |                                                                                                                                                                                                                                                                                                                                                                                 |
|------------------------------|-------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| For ISAC use only            |       |                                                                                                                                                                                                                                                                                                                                                                                 |
| Protocol No.                 | ..... | <p style="text-align: center; margin: 0;"><b>IMPORTANT</b></p> <p style="margin: 0;"><i>Please refer to the guidance for 'Completing the ISAC application form' found on the CPRD website (<a href="http://www.cprd.com/isac">www.cprd.com/isac</a>). If you have any queries, please contact the ISAC Secretariat at <a href="mailto:isac@cpdr.com">isac@cpdr.com</a>.</i></p> |
| Submission date (DD/MM/YYYY) | ..... |                                                                                                                                                                                                                                                                                                                                                                                 |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                                                            |                                             |                                           |                                            |                                                          |                                                    |                                                                      |                                                  |                                                                            |                                 |                   |                       |         |                          |                         |  |                               |           |  |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------|---------------------------------------------|-------------------------------------------|--------------------------------------------|----------------------------------------------------------|----------------------------------------------------|----------------------------------------------------------------------|--------------------------------------------------|----------------------------------------------------------------------------|---------------------------------|-------------------|-----------------------|---------|--------------------------|-------------------------|--|-------------------------------|-----------|--|
| <b>SECTION A: GENERAL INFORMATION ABOUT THE PROPOSED RESEARCH STUDY</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |                                                            |                                             |                                           |                                            |                                                          |                                                    |                                                                      |                                                  |                                                                            |                                 |                   |                       |         |                          |                         |  |                               |           |  |
| <p><b>1. Study Title<sup>§</sup></b> (Please state the study title below)</p> <p><b>Determining the frequency of abnormal hepatic fibrosis measures and liver cirrhosis/decompensation in UK primary care</b></p> <p><small><sup>§</sup>Please note: This information will be published on the CPRD's website as part of its transparency policy.</small></p> <p><b>2. Has any part of this research proposal or a related proposal been previously submitted to ISAC?</b></p> <p>Yes* <input type="checkbox"/> No <input checked="" type="checkbox"/></p> <p><small>*If yes, please provide the previous protocol number/s below. Please also state in your current submission how this/these are related or relevant to this study.</small></p> <p><b>3. Has this protocol been peer reviewed by another Committee? (e.g. grant award or ethics committee)</b></p> <p>Yes* <input checked="" type="checkbox"/> No <input type="checkbox"/></p> <p><small>*If Yes, please state the name of the reviewing Committee(s) below and provide an outline of the review process and outcome as an Appendix to this protocol :</small></p> <p>MRC Clinician Scientist Fellowship application (successful, Nov 2016), see Appendix A for details</p> <p><b>4. Type of Study</b> (please tick all the relevant boxes which apply)</p> <table style="width: 100%; border: none;"> <tr> <td style="width: 50%;">Adverse Drug Reaction/Drug Safety <input type="checkbox"/></td> <td style="width: 50%;">Drug Effectiveness <input type="checkbox"/></td> </tr> <tr> <td>Drug Utilisation <input type="checkbox"/></td> <td>Pharmacoeconomics <input type="checkbox"/></td> </tr> <tr> <td>Disease Epidemiology <input checked="" type="checkbox"/></td> <td>Post-authorisation Safety <input type="checkbox"/></td> </tr> <tr> <td>Health care resource utilisation <input checked="" type="checkbox"/></td> <td>Methodological Research <input type="checkbox"/></td> </tr> <tr> <td>Health/Public Health Services Research <input checked="" type="checkbox"/></td> <td>Other* <input type="checkbox"/></td> </tr> </table> <p><small>*If Other, please specify the type of study in the lay summary</small></p> <p><b>5. Health Outcomes to be Measured<sup>§</sup></b></p> <p><small><sup>§</sup>Please note: This information will be published on CPRD's website as part of its transparency policy.</small></p> <p><u>Please summarise below the primary/secondary health outcomes to be measured in this research protocol:</u></p> <table style="width: 100%; border: none;"> <tr> <td style="width: 33%;">• Liver cirrhosis</td> <td style="width: 33%;">• Oesophageal varices</td> <td style="width: 33%;">• Death</td> </tr> <tr> <td>• Advanced liver disease</td> <td>• Hepatocellular cancer</td> <td></td> </tr> <tr> <td>• Decompensated liver disease</td> <td>• Ascites</td> <td></td> </tr> </table> <p><small>[Please add more bullet points as necessary]</small></p> | Adverse Drug Reaction/Drug Safety <input type="checkbox"/> | Drug Effectiveness <input type="checkbox"/> | Drug Utilisation <input type="checkbox"/> | Pharmacoeconomics <input type="checkbox"/> | Disease Epidemiology <input checked="" type="checkbox"/> | Post-authorisation Safety <input type="checkbox"/> | Health care resource utilisation <input checked="" type="checkbox"/> | Methodological Research <input type="checkbox"/> | Health/Public Health Services Research <input checked="" type="checkbox"/> | Other* <input type="checkbox"/> | • Liver cirrhosis | • Oesophageal varices | • Death | • Advanced liver disease | • Hepatocellular cancer |  | • Decompensated liver disease | • Ascites |  |
| Adverse Drug Reaction/Drug Safety <input type="checkbox"/>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | Drug Effectiveness <input type="checkbox"/>                |                                             |                                           |                                            |                                                          |                                                    |                                                                      |                                                  |                                                                            |                                 |                   |                       |         |                          |                         |  |                               |           |  |
| Drug Utilisation <input type="checkbox"/>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | Pharmacoeconomics <input type="checkbox"/>                 |                                             |                                           |                                            |                                                          |                                                    |                                                                      |                                                  |                                                                            |                                 |                   |                       |         |                          |                         |  |                               |           |  |
| Disease Epidemiology <input checked="" type="checkbox"/>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | Post-authorisation Safety <input type="checkbox"/>         |                                             |                                           |                                            |                                                          |                                                    |                                                                      |                                                  |                                                                            |                                 |                   |                       |         |                          |                         |  |                               |           |  |
| Health care resource utilisation <input checked="" type="checkbox"/>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | Methodological Research <input type="checkbox"/>           |                                             |                                           |                                            |                                                          |                                                    |                                                                      |                                                  |                                                                            |                                 |                   |                       |         |                          |                         |  |                               |           |  |
| Health/Public Health Services Research <input checked="" type="checkbox"/>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | Other* <input type="checkbox"/>                            |                                             |                                           |                                            |                                                          |                                                    |                                                                      |                                                  |                                                                            |                                 |                   |                       |         |                          |                         |  |                               |           |  |
| • Liver cirrhosis                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | • Oesophageal varices                                      | • Death                                     |                                           |                                            |                                                          |                                                    |                                                                      |                                                  |                                                                            |                                 |                   |                       |         |                          |                         |  |                               |           |  |
| • Advanced liver disease                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | • Hepatocellular cancer                                    |                                             |                                           |                                            |                                                          |                                                    |                                                                      |                                                  |                                                                            |                                 |                   |                       |         |                          |                         |  |                               |           |  |
| • Decompensated liver disease                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | • Ascites                                                  |                                             |                                           |                                            |                                                          |                                                    |                                                                      |                                                  |                                                                            |                                 |                   |                       |         |                          |                         |  |                               |           |  |

08 August 2016 Version 1.0

|                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                                                                                                                           |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------|
| <b>6. Publication: This study is intended for</b> (please tick all the relevant boxes which apply):                                                                                                                                                                                                                                                                                                                                                      |                                                                                                                           |
| Publication in peer-reviewed journals <input checked="" type="checkbox"/><br>Presentation at company/institutional meetings <input checked="" type="checkbox"/><br>Other <input type="checkbox"/>                                                                                                                                                                                                                                                        | Presentation at scientific conference <input checked="" type="checkbox"/><br>Regulatory purposes <input type="checkbox"/> |
| <i>*If Other, please provide further information:</i>                                                                                                                                                                                                                                                                                                                                                                                                    |                                                                                                                           |
| <b>SECTION B: INFORMATION ON INVESTIGATORS AND COLLABORATORS</b>                                                                                                                                                                                                                                                                                                                                                                                         |                                                                                                                           |
| <b>7. Chief Investigator</b><br>Please state the full name, job title, organisation name & e-mail address for correspondence - see guidance notes for eligibility. Please note that there can only be one Chief Investigator per protocol.<br><br>Dr Joanne Morling, MRC Clinician Scientist (Honorary Consultant in Public Health Medicine)<br>Division of Epidemiology and Public Health, University of Nottingham.<br>Joanne.morling@nottingham.ac.uk |                                                                                                                           |
| CV has been previously submitted to ISAC <input type="checkbox"/> <b>CV number:</b><br>A new CV is being submitted with this protocol <input checked="" type="checkbox"/><br>An updated CV is being submitted with this protocol <input type="checkbox"/>                                                                                                                                                                                                |                                                                                                                           |
| <b>8. Affiliation of Chief Investigator</b> (full address)<br><br>Division of Epidemiology and Public Health<br>University of Nottingham<br>C120, Clinical Sciences Building Ph2<br>City Hospital Campus, Hucknall Road<br>Nottingham NG5 1PB                                                                                                                                                                                                            |                                                                                                                           |
| <b>9. Corresponding Applicant</b><br>Please state the full name, affiliation(s) and e-mail address below:<br><br>Dr Joanne Morling, Division of Epidemiology and Public Health, University of Nottingham.<br>Joanne.morling@nottingham.ac.uk                                                                                                                                                                                                             |                                                                                                                           |
| Same as chief investigator <input checked="" type="checkbox"/> <b>CV number:</b><br>CV has been previously submitted to ISAC <input type="checkbox"/><br>A new CV is being submitted with this protocol <input type="checkbox"/><br>An updated CV is being submitted with this protocol <input type="checkbox"/>                                                                                                                                         |                                                                                                                           |
| <b>10. List of all investigators/collaborators</b><br>Please list the full name, affiliation(s) and e-mail address* of all collaborators, other than the Chief Investigator below:<br><br>Other investigator:<br><br>Dr Tim Card, Division of Epidemiology and Public Health, University of Nottingham.<br>Tim.card@nottingham.ac.uk                                                                                                                     |                                                                                                                           |
| CV has been previously submitted to ISAC <input checked="" type="checkbox"/> <b>CV number:</b> 041_15CESL<br>A new CV is being submitted with this protocol <input type="checkbox"/><br>An updated CV is being submitted with this protocol <input type="checkbox"/>                                                                                                                                                                                     |                                                                                                                           |
| <b>11. Conflict of interest statement</b><br>Please provide a draft of the conflict (or competing) of interest (COI) statement that you intend to include in any publication which might result from this work<br><br>Dr Joanne Morling is funded by a four year Medical Research Council Clinician Scientist Fellowship.                                                                                                                                |                                                                                                                           |

08 August 2016 Version 1.0



|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                                                                                                                                                                                 |                                                                                                                                                                                                                                                  |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| There is no other support (financial or otherwise) from any other organisation for this work.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                                                                                                                                                                                 |                                                                                                                                                                                                                                                  |
| <b>12. Experience/expertise available</b><br>Please complete the following questions to indicate the experience/ expertise available within the team of investigators/collaborators actively involved in the proposed research, including the analysis of data and interpretation of results.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                                                                                                                                                                                 |                                                                                                                                                                                                                                                  |
| <b>Previous GPRD/CPRD Studies</b><br>None <input type="checkbox"/><br>1-3 <input type="checkbox"/><br>> 3 <input checked="" type="checkbox"/>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | <b>Publications using GPRD/CPRD data</b><br><input type="checkbox"/><br><input type="checkbox"/><br><input checked="" type="checkbox"/>                                         |                                                                                                                                                                                                                                                  |
| <b>Experience/Expertise available</b><br><b>Is statistical expertise available within the research team?</b><br><i>If yes, please indicate the name(s) of the relevant investigator(s)</i><br>Drs Morling and Card have both completed PhDs using extensive statistical skills and published widely. We do not anticipate using any statistical technique beyond our ability/knowledge.                                                                                                                                                                                                                                                                                                                                                                                                                                                | Yes                                                                                                                                                                             | No                                                                                                                                                                                                                                               |
| <b>Is experience of handling large data sets (&gt;1 million records) available within the research team?</b><br><i>If yes, please indicate the name(s) of the relevant investigator(s)</i><br>Dr Card has published a number of studies based on GPRD/CPRD data and HES data both in linked and unlinked forms. Dr Morling has published studies based on Scottish routinely collected data (SMR) both in linked and unlinked forms.                                                                                                                                                                                                                                                                                                                                                                                                   | <input checked="" type="checkbox"/>                                                                                                                                             | <input type="checkbox"/>                                                                                                                                                                                                                         |
| <b>Is experience of practising in UK primary care available to or within the research team?</b><br><i>If yes, please indicate the name(s) of the relevant investigator(s)</i><br>Drs Morling and Card have both have both completed attachments in general practice during their medical training. Additional support is available through wider project initiatives.                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | <input checked="" type="checkbox"/>                                                                                                                                             | <input type="checkbox"/>                                                                                                                                                                                                                         |
| <b>13. References relating to your study</b><br>Please list up to 3 references (most relevant) relating to your proposed study:<br><br>Williams, R. 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. The Lancet 2014;384:1953–1997.<br><br>Ratib S, West J, Crooks CJ and Fleming KM. Diagnosis of Liver Cirrhosis in England, a Cohort Study, 1998-2009: A Comparison With Cancer. American Journal of Gastroenterology 2014;109:190-198.<br><br>Fleming K, Aithal GP, Card TR and West J. The rate of decompensation and clinical progression of disease in people with cirrhosis: a cohort study. Alimentary pharmacology & therapeutics 2010;32:11-1343-1350. |                                                                                                                                                                                 |                                                                                                                                                                                                                                                  |
| SECTION C: ACCESS TO THE DATA                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                                                                                                                                                                                 |                                                                                                                                                                                                                                                  |
| <b>14. Financial Sponsor of study<sup>§</sup></b><br><i>§Please note: The name of the source of funding will be published on CPRD's website as part of its transparency policy</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                                                                                                                                                                                 |                                                                                                                                                                                                                                                  |
| Pharmaceutical Industry <input type="checkbox"/><br>Academia <input type="checkbox"/><br>Government / NHS <input checked="" type="checkbox"/><br>Charity <input type="checkbox"/><br>Other <input type="checkbox"/><br>None <input type="checkbox"/>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | <input type="checkbox"/><br><input type="checkbox"/><br><input checked="" type="checkbox"/><br><input type="checkbox"/><br><input type="checkbox"/><br><input type="checkbox"/> | <i>Please specify name and country:</i><br><i>Please specify name and country:</i><br><i>Please specify name and country:</i> Medical Research Council, UK<br><i>Please specify name and country:</i><br><i>Please specify name and country:</i> |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                                     |                                                               |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------|---------------------------------------------------------------|
| <b>15. Type of Institution conducting the research</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                     |                                                               |
| Pharmaceutical Industry                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | <input type="checkbox"/>            | Please specify name and country:                              |
| Academia                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | <input checked="" type="checkbox"/> | Please specify name and country: University of Nottingham, UK |
| Government Department                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | <input type="checkbox"/>            | Please specify name and country:                              |
| Research Service Provider                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | <input type="checkbox"/>            | Please specify name and country:                              |
| NHS                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | <input type="checkbox"/>            | Please specify name and country:                              |
| Other                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | <input type="checkbox"/>            | Please specify name and country:                              |
| <b>16. Data access arrangements</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                                     |                                                               |
| The financial sponsor/ collaborator* has a licence for CPRD GOLD and will extract the data                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                                     | <input type="checkbox"/>                                      |
| The institution carrying out the analysis has a licence for CPRD GOLD and will extract the data**                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                                     | <input checked="" type="checkbox"/>                           |
| A data set will be provided by the CPRD*€                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                                     | <input type="checkbox"/>                                      |
| CPRD has been commissioned to extract the data and perform the analyses€                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                                     | <input type="checkbox"/>                                      |
| Other:                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                     | <input type="checkbox"/>                                      |
| If Other, please specify:                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                                     |                                                               |
| <p>*Collaborators supplying data for this study must be named on the protocol as co-applicants.</p> <p>**If data sources other than CPRD GOLD are required, these will be supplied by CPRD</p> <p>€Please note that datasets provided by CPRD are limited in size; applicants should contact CPRD (<a href="mailto:kc@cprd.com">kc@cprd.com</a>) if a dataset of &gt;300,000 patients is required.</p> <p>€Investigators must discuss their request with a member of the CPRD Research team before submitting an ISAC application. Please contact the CPRD Research Team on +44 (20) 3080 6383 or email (<a href="mailto:kc@cprd.com">kc@cprd.com</a>) to discuss your requirements. Please also state the name of CPRD Research team with whom you have discussed this request (provide the date of discussion and any relevant reference information):</p> |                                     |                                                               |
| Name of CPRD Researcher                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | Reference number (where available)  | Date of contact                                               |
| <b>17. Primary care data</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                                     |                                                               |
| Please specify which primary care data set(s) are required)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                                     |                                                               |
| Vision only (Default for CPRD studies)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | <input checked="" type="checkbox"/> | Both Vision and EMIS®* <input type="checkbox"/>               |
| EMIS® only*                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | <input type="checkbox"/>            |                                                               |
| <p>Note: Vision and EMIS are different practice management systems. CPRD has traditionally collected data from Vision practice. Data collected from EMIS is currently under evaluation prior to wider release.</p> <p>*Investigators requiring the use of EMIS data <b>must</b> discuss the study with a member of the CPRD Research team before submitting an ISAC application</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                                     |                                                               |
| Please state the name of the CPRD Researcher with whom you have discussed your request for EMIS data:                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                                     |                                                               |
| Name of CPRD Researcher                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | Reference number (where available)  | Date of contact                                               |
| <b>SECTION D: INFORMATION ON DATA LINKAGES</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                     |                                                               |
| <b>18. Does this protocol seek access to linked data</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                                     |                                                               |
| Yes* <input checked="" type="checkbox"/>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | No <input type="checkbox"/>         | If No, please move to section E.                              |
| <p>*Research groups which have not previously accessed CPRD linked data resources <b>must</b> discuss access to these resources with a member of the CPRD Research team, before submitting an ISAC application. Investigators requiring access to HES Accident and Emergency data, HES Diagnostic Imaging Dataset and PROMS data <b>must</b> also discuss this with a member of the CPRD Research team before submitting an ISAC application. Please contact the CPRD Research Team on +44 (20) 3080 6383 or email <a href="mailto:kc@cprd.com">kc@cprd.com</a> to discuss your requirements <b>before</b> submitting your application.</p>                                                                                                                                                                                                                  |                                     |                                                               |
| Name of CPRD Researcher: Tarita Murray-Thomas                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                     |                                                               |
| Reference number CPRD00012420 Date of contact 31.07.2017                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                                     |                                                               |
| Please note that as part of the ISAC review of linkages, your protocol may be shared - in confidence - with a representative of the requested linked data set(s) and summary details may be shared - in confidence - with the Confidentiality Advisory Group of the Health Research Authority.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                     |                                                               |

**19. Please select the source(s) of linked data being requested<sup>§</sup>**

<sup>§</sup>Please note: This information will be published on the CPRD's website as part of its transparency policy.

- |                                                                 |                                                                              |
|-----------------------------------------------------------------|------------------------------------------------------------------------------|
| <input checked="" type="checkbox"/> ONS Death Registration Data | <input type="checkbox"/> MINAP (Myocardial Ischaemia National Audit Project) |
| <input checked="" type="checkbox"/> HES Admitted Patient Care   | <input checked="" type="checkbox"/> Cancer Registration Data*                |
| <input type="checkbox"/> HES Outpatient                         | <input type="checkbox"/> PROMS (Patient Reported Outcomes Measure)**         |
| <input checked="" type="checkbox"/> HES Accident and Emergency  | <input type="checkbox"/> CPRD Mother Baby Link                               |
| <input type="checkbox"/> HES Diagnostic Imaging Dataset         |                                                                              |
- ☒ Practice Level Index of Multiple Deprivation (Standard)  
☐ Practice Level Index of Multiple Deprivation (Bespoke)  
☒ Patient Level Index of Multiple Deprivation\*\*\*  
☐ Patient Level Townsend Score \*\*\*  
☐ Other\*\*\*\* Please specify:

\*Applicants seeking access to cancer registration data must complete a Cancer Dataset Agreement form (available from CPRD). This should be submitted to the ISAC as an appendix to your protocol. Please also note that applicants seeking access to cancer registry data must provide consent for publication of their study title and study institution on the UK Cancer Registry website.

\*\*Assessment of the quality of care delivered to NHS patients in England undergoing four procedures: hip replacement, knee replacement, groin hernia and varicose veins. Please note that patient level PROMS data are only accessible by academics

\*\*\* Patient level IMD and Townsend scores will not be supplied for the same study

\*\*\*\*If "Other" is specified, please provide the name of the individual in the CPRD Research team with whom this linkage has been discussed.

Name of CPRD Researcher

Reference number (where available)

Date of contact

**20. Total number of linked datasets requested including CPRD GOLD**

Number of linked datasets requested (practice/ 'patient' level Index of Multiple Deprivation, Townsend Score or the CPRD Mother Baby Link should not be included in this count) 5

Please note: Where ≥5 linked datasets are requested, approval may be required from the Confidentiality Advisory Group (CAG) to access these data

**21. Is linkage to a local<sup>‡</sup> dataset with <1 million patients being requested?**

Yes\* ☐ No ☒

\*If yes, please provide further details:

<sup>‡</sup> Data from defined geographical areas i.e. non-national datasets.

**22. If you have requested one or more linked data sets, please indicate whether the Chief Investigator or any of the collaborators listed in question 5 above, have access to these data in a patient identifiable form (e.g. full date of birth, NHS number, patient post code), or associated with an identifiable patient index.**

Yes\* ☐ No ☒

\* If yes, please provide further details:

**23. Does this study involve linking to patient *identifiable* data (e.g. hold date of birth, NHS number, patient post code) from other sources?**

Yes ☐ No ☒

**SECTION E: VALIDATION/VERIFICATION****24. Does this protocol describe a purely observational study using CPRD data?**

Yes\* ☒ No\*\* ☐

\* Yes: If you will be using data obtained from the CPRD Group, this study does not require separate ethics approval from an NHS

08 August 2016 Version 1.0

Research Ethics Committee.

**\*\* No:** You may need to seek separate ethics approval from an NHS Research Ethics Committee for this study. The ISAC will provide advice on whether this may be needed.

**25. Does this protocol involve requesting any additional information from GPs?**

Yes\* ☐ No ☒

*\* If yes, please indicate what will be required:*

|                                                             |                              |                             |
|-------------------------------------------------------------|------------------------------|-----------------------------|
| Completion of questionnaires by the GP <sup>v</sup>         | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| Is the questionnaire a validated instrument?                | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| If yes, has permission been obtained to use the instrument? | Yes <input type="checkbox"/> | No <input type="checkbox"/> |
| Please provide further information:                         |                              |                             |

Other (please describe)

<sup>v</sup> Any questionnaire for completion by GPs or other health care professional must be approved by ISAC before circulation for completion.

**26. Does this study require contact with patients in order for them to complete a questionnaire?**

Yes\* ☐ No ☒

*\* Please note that any questionnaire for completion by patients must be approved by ISAC before circulation for completion.*

**27. Does this study require contact with patients in order to collect a sample?**

Yes\* ☐ No ☒

*\* Please state what will be collected:*

**SECTION F: DECLARATION**

**28. Signature from the Chief Investigator**

- I have read the guidance on '**Completion of the ISAC application form**' and '**Contents of CPRD ISAC Research Protocols**' and have understood these;
- I have read the submitted version of this research protocol, including all supporting documents, and confirm that these are accurate.
- I am suitably qualified and experienced to perform and/or supervise the research study proposed.
- I agree to conduct or supervise the study described in accordance with the relevant, current protocol
- I agree to abide by all ethical, legal and scientific guidelines that relate to access and use of CPRD data for research
- I understand that the details provided in sections marked with (S) in the application form and protocol will be published on the CPRD website in line with CPRD's transparency policy.
- I agree to inform the CPRD of the final outcome of the research study: publication, prolonged delay, completion or termination of the study.

Name: Joanne Morling

Date: 10.10.2017

e-Signature: Joanne Morling



## PROTOCOL INFORMATION REQUIRED

The following sections below **must** be included in the CPRD ISAC research protocol. Please refer to the guidance on 'Contents of CPRD ISAC Research Protocols' ([www.cprd.com/isac](http://www.cprd.com/isac)) for more information on how to complete the sections below. Pages should be numbered. All abbreviations must be defined on first use.

| Applicants must complete all sections listed below<br>Sections which do not apply should be completed as 'Not Applicable' |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|---------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>A. Study Title</b>                                                                                                     | <b>Determining the frequency of abnormal hepatic fibrosis diagnostic measures and liver cirrhosis/decompensation in UK primary care</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| <b>B. Lay Summary (Max. 200 words)<sup>§</sup></b>                                                                        | <p><sup>§</sup>Please note: This information will be published on CPRD's website as part of its transparency policy</p> <p>Liver problems in people in the UK and internationally are becoming increasingly common, but are difficult to identify early in the course of the disease. As the liver becomes more diseased scarring forms within it (fibrosis) and this is termed advanced liver disease. We intend to look at the potential of using blood test indicators of advanced liver disease as screening in primary care in the UK.</p> <p>Firstly, we plan to use general practice data (CPRD) to determine how using different blood test indicators of advanced liver disease (fibrosis) as a screening tool would affect NHS health services. We will work out how many people would need to be referred to see a hospital consultant as a result of the screening test.</p> <p>Secondly, we plan to look and see how many people who we know end up in hospital with advanced liver disease and its complications (such as liver cancer) would have been identified earlier if we used the possible blood test indicators of advanced liver disease (fibrosis) as a screening test.</p> <p>The results will help us decide if and how to screen for advanced liver disease in general practice and will guide the future health care of patients.</p> <p style="text-align: right;">[199]</p> |
| <b>C. Technical Summary (Max. 200 words)</b>                                                                              | <p><b>Overall aim:</b></p> <p>To investigate the potential burden on UK health services of screening for chronic liver disease in UK primary care.</p> <p><b>Methodology:</b></p> <p>We will conduct a population based cohort study to examine firstly the prevalence of probable hepatic fibrosis as determined by non-invasive markers, and secondly rates of liver cirrhosis and hepatic decompensation.</p> <p>We will extract all patients with the necessary measures available to determine the prevalence of hepatic fibrosis phenotyping them using non-invasive markers (AST/ALT ratio, APRI, FIB4, NAFLD Fibrosis Score). The components of each marker will be extracted as laboratory measures or READ</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |

| Applicants must complete all sections listed below<br>Sections which do not apply should be completed as 'Not Applicable'                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |  |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
| <p>codes from the linked clinical file and the composite score calculated.</p> <p>Hepatic decompensation will be defined as a composite outcome including liver cirrhosis, oesophageal varices, hepatocellular cancer and ascites.</p> <p>Ability of each of the markers to predict hepatic decompensation will be determined separately using Cox proportional hazards regression and net benefit analyses for diagnostic test prediction models (as described by Vickers et al BMJ 2016;352:i6).</p> <p><b>Scientific and medical opportunities:</b></p> <p>This research will improve the applicability and reliability of clinical guidelines for the diagnosis of CLD in primary care.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |  |
| [185]                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |  |
| <b>D. Objectives, Specific Aims and Rationale</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |  |
| <p><b>Rationale:</b></p> <p>Both in the UK and internationally, chronic liver disease (CLD) is one of the few diseases with an increasing incidence. However, effective methods for early identification are lacking, leading to potentially preventable complications and mortality from CLD. The use of non-invasive measures of liver fibrosis/cirrhosis to screen people with liver disease who encounter primary care services has been proposed to facilitate earlier identification. However there are significant gaps in current knowledge regarding how to implement such a process.</p> <p><b>Aim:</b></p> <p>To investigate the potential burden on UK health services of screening for chronic liver disease in UK primary care.</p> <p><b>Objectives:</b></p> <ol style="list-style-type: none"><li>1. To determine the frequency of abnormal hepatic fibrosis indicators in primary care in the UK and the potential impact on the NHS of onward referral to secondary care using different diagnostic strategies.</li><li>2. To determine the proportion of liver cirrhosis and hepatic decompensation that would be identified earlier using different CLD case finding strategies and whether earlier treatment may have influenced outcome.</li></ol> |  |
| <b>E. Study Background</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |  |
| <p><b>The burden of chronic liver disease (CLD)</b></p> <p>In the UK liver disease is a significant and growing burden on the National Health Service (NHS). In recent decades huge investments have been made in primary and secondary prevention of cardiovascular disease, contributing to marked decreases in morbidity and mortality. Opposite patterns are seen for liver disease with limited investment despite 400% increases in liver disease mortality in the population as a whole since 1970 and nearly 500% increases in mortality observed in working age populations over in this period<sup>1</sup>. The Secretary of State for Health recently named CLD third (behind ischaemic heart disease and self-harm) in the "Big Five" causes of premature</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |  |

**Applicants must complete all sections listed below  
Sections which do not apply should be completed as 'Not Applicable'**

mortality<sup>2</sup> with 62,000 years of working life lost every year<sup>3</sup>. Given both the high UK alcohol consumption rates and levels of obesity<sup>4</sup>, and their known synergism<sup>5</sup>, the burden of CLD can be expected to rise further.

This is an issue of international importance. Worldwide the incidence of CLD is rising. The World Health Organisation (WHO) Global Burden of Diseases programme found a 33% rise in liver cirrhosis related mortality (779,000 to 1031,000) and a 62% rise in hepatocellular carcinomas (HCC) related deaths (463,000 to 752,000) between 1990 and 2010<sup>6</sup>.

#### **Chronic liver disease**

The three main forms of CLD in the UK are: alcohol related liver disease (ALD) obesity related non-alcoholic fatty liver disease (NAFLD) and chronic viral hepatitis (CVH). These reflect the wide range of causes of CLD. Irrespective of the aetiology, the natural history of disease and changes in the liver are broadly similar for all forms of CLD: inflammation leads to hepatic fibrosis and ultimately cirrhosis and liver failure. The complications of CLD are well defined and are largely related to the consequences of portal hypertension (e.g. ascites, variceal bleeding and encephalopathy), sepsis and hepatocellular carcinoma (HCC). New NICE guidance<sup>1</sup> (currently under consultation) recommends that following a diagnosis of cirrhosis the patient be offered 6-monthly US surveillance for HCC (to allow early detection and potentially curative treatment) alongside upper gastrointestinal endoscopy to detect oesophageal varices and 3-yearly screening for varices in patients with none on the initial screen.

In 2014 Public Health England announced its seven public health priority areas for improving the health of the UK over the coming years – one of which was liver disease. Work has been established on quantifying the burden of chronic liver disease with the launch of regional liver profiles in October 2014, and future aspirations committed to in the awaited UK Liver Strategy document. There are, however, significant gaps in our understanding of the practicalities and costs of implementing strategies to address CLD.

#### **Identification of CLD**

Three independent reports have highlighted the need for the early detection of liver disease including the Chief Medical Officer report (2012)<sup>7</sup>, the All-Party Parliamentary Hepatology Group Inquiry<sup>8</sup> and the Lancet commission in 2014<sup>9</sup>, in order to allow intervention and change the course of the disease. Despite this, the existing diagnostic pathways for detection and onward referral of suspected CLD are based on traditional liver enzyme tests which lack accuracy and result in delays to diagnosis. With no nationally agreed assessment guidance, the approach used by general practitioners (GPs) to identify patients with CLD varies widely.

Calls for screening<sup>10,11</sup> for CLD acknowledge that there are still significant gaps in our knowledge of the diagnostic pathway. A randomised controlled trial of the effect of CLD screening is not currently feasible and decisions regarding earlier detection will have to be made using observational data and modelling, in order to meet the UK National Screening Committee criteria<sup>12</sup>. Major gaps in knowledge exist in developing the patient pathway, specifically in relation to diagnostic criteria and the cost-effectiveness of screening and early diagnosis of CLD. Also of concern are the potential harms (e.g.



**Applicants must complete all sections listed below**  
**Sections which do not apply should be completed as 'Not Applicable'**

patient concerns relating to false positive results, insufficient resource to meet demand) of implementing poorly thought through piecemeal screening programmes, which differ across the UK. Whether screening or opportunistic strategies are to be used, properties of diagnostic tests need to be better understood.

The gold standard for the diagnosis of CLD is liver biopsy, but this has a number of limitations. Liver biopsy is expensive<sup>13</sup> and has significant risks to the patient, with major complications reported in 0.6%, and directly attributable mortality in 0.01% of procedures<sup>14</sup>. Due to the small size of biopsy needles, measuring only 1/50,000<sup>th</sup> of the liver, findings may not be representative of the whole liver<sup>15,16</sup>. Histological examination of specimens is subject to intra- and inter-observer variation<sup>17</sup>. Therefore, increasingly, non-invasive approaches are being proposed.

Current non-invasive approaches can be divided into indirect 'simple' markers which are routinely collected by clinicians (e.g. transaminases, gamma-glutamyl transferase, platelet count), direct 'complex' markers and more specialist measures (e.g. procollagen peptides I/III, Type IV collagen), cytokines (e.g. interleukin-10, transforming growth factor alpha) and imaging (e.g. transient elastography), which can be used singly or in combination. Serum markers are more easily implemented than imaging, however they can nonetheless be costly and implementing algorithms within health services is complex. Imaging is the most specific for liver pathology and can be used in isolation, but requires significant operator training and capital costs.

To date these approaches have largely been validated for accuracy only in select secondary care populations, with limited use and validation in community settings. The focus to date has been confined to diagnostic test accuracy, without considering the effects of novel tests on morbidity, mortality or opportunity-costs.

Reflecting this uncertainty, referral to secondary care shows substantial variation, and there are no established recommendations for which groups of patients are likely to benefit from referral in terms of mortality, morbidity and quality of life. Though analysis of existing health records cannot provide all the information needed to make these assessments, it can help to start filling some of the gaps in our knowledge.

Data exists in the routine GP records of many patients to calculate a number of 'simple' liver fibrosis markers, giving an opportunity to examine the performance of these tests in an opportunistic manner (objective 2), as well as to determine the burden which would be placed on secondary care by the volume of onward referral which the results suggest (objective 1). We will also be able to determine from the details held at what stage the liver disease was diagnosed (e.g. early vs late) and whether it would be amenable to intervention if diagnosed earlier.

The 'simple' markers which have been validated and can be generated include:

- Aspartate aminotransferase to alanine aminotransferase ratio (AST/ALT) calculated as  $\text{AST(IU/L)} / \text{ALT(IU/L)}$ .
- AST to Platelet Ratio Index (APRI)<sup>18</sup> calculated as  $[[\text{AST(IU/L)} / \text{upper limit normal}] / \text{platelets}(\times 10^9/\text{L})] \times 100$ .
- Fibrosis 4 Index [FIB4]<sup>19</sup> calculated as  $[\text{age(years)} \times \text{AST(IU/L)}] / [\text{platelets}(\times 10^9/\text{L}) \times$



| Applicants must complete all sections listed below<br>Sections which do not apply should be completed as 'Not Applicable'                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                       |                                         |                                          |                          |                                                       |                                         |                                          |   |           |         |         |     |                 |                |                |     |                |              |                |      |               |              |              |     |              |            |            |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------|-----------------------------------------|------------------------------------------|--------------------------|-------------------------------------------------------|-----------------------------------------|------------------------------------------|---|-----------|---------|---------|-----|-----------------|----------------|----------------|-----|----------------|--------------|----------------|------|---------------|--------------|--------------|-----|--------------|------------|------------|
| $\sqrt{\text{AST(IU/L)}}$ <ul style="list-style-type: none"> <li>Non-alcohol fatty liver disease fibrosis score [NFS]<sup>20</sup> calculated as <math>-1.675 + 0.037 \times \text{age(years)} + 0.094 \times \text{BMI(kg/m}^2\text{)} + 1.13 \times \text{IFG/diabetes(yes=1,no=0)} + 0.99 \times \text{AST/ALT ratio} - 0.013 \times \text{platelet}(\times 10^9/\text{L}) - 0.66 \times \text{albumin(g/dl)}</math></li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                                                       |                                         |                                          |                          |                                                       |                                         |                                          |   |           |         |         |     |                 |                |                |     |                |              |                |      |               |              |              |     |              |            |            |
| <b>F. Study Type</b><br><br>Descriptive<br>Exploratory                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                                                       |                                         |                                          |                          |                                                       |                                         |                                          |   |           |         |         |     |                 |                |                |     |                |              |                |      |               |              |              |     |              |            |            |
| <b>G. Study Design</b><br><br>Cohort study.<br>Participants will enter the study on the occasion of their first abnormal liver marker.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                                                       |                                         |                                          |                          |                                                       |                                         |                                          |   |           |         |         |     |                 |                |                |     |                |              |                |      |               |              |              |     |              |            |            |
| <b>H. Feasibility counts</b><br><br>A feasibility assessment of CPRD to ensure that there will be sufficient data available has been undertaken. Using data for a single year (2012, most recently available data) the number of patients with data available to allow the calculation of a number of markers of hepatic fibrosis as described in Section E is shown in the table below. Full analysis will be undertaken on all available years (2000 to 2016/17). <table border="1" data-bbox="443 1160 1219 1368"> <thead> <tr> <th>Liver fibrosis indicator</th> <th>Age 18-65 with at least 1 episode on CPRD in 2012 [a]</th> <th>[a] AND a record of alcohol excess ever</th> <th>[a] AND a record of type 2 diabetes ever</th> </tr> </thead> <tbody> <tr> <td>n</td> <td>3,086,007</td> <td>162,353</td> <td>115,602</td> </tr> <tr> <td>ALT</td> <td>616,471 (20.0%)</td> <td>38,862 (23.9%)</td> <td>73,592 (63.7%)</td> </tr> <tr> <td>AST</td> <td>108,556 (3.5%)</td> <td>6,866 (4.2%)</td> <td>12,477 (10.8%)</td> </tr> <tr> <td>APRI</td> <td>39,420 (1.3%)</td> <td>3,253 (2.0%)</td> <td>4,209 (3.6%)</td> </tr> <tr> <td>NFS</td> <td>2,623 (0.1%)</td> <td>407 (0.3%)</td> <td>767 (0.7%)</td> </tr> </tbody> </table> <p><b>Note:</b> Both APRI and NFS are not routinely recorded in CPRD however the component elements are available allowing calculation.</p> |                                                       |                                         |                                          | Liver fibrosis indicator | Age 18-65 with at least 1 episode on CPRD in 2012 [a] | [a] AND a record of alcohol excess ever | [a] AND a record of type 2 diabetes ever | n | 3,086,007 | 162,353 | 115,602 | ALT | 616,471 (20.0%) | 38,862 (23.9%) | 73,592 (63.7%) | AST | 108,556 (3.5%) | 6,866 (4.2%) | 12,477 (10.8%) | APRI | 39,420 (1.3%) | 3,253 (2.0%) | 4,209 (3.6%) | NFS | 2,623 (0.1%) | 407 (0.3%) | 767 (0.7%) |
| Liver fibrosis indicator                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | Age 18-65 with at least 1 episode on CPRD in 2012 [a] | [a] AND a record of alcohol excess ever | [a] AND a record of type 2 diabetes ever |                          |                                                       |                                         |                                          |   |           |         |         |     |                 |                |                |     |                |              |                |      |               |              |              |     |              |            |            |
| n                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | 3,086,007                                             | 162,353                                 | 115,602                                  |                          |                                                       |                                         |                                          |   |           |         |         |     |                 |                |                |     |                |              |                |      |               |              |              |     |              |            |            |
| ALT                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | 616,471 (20.0%)                                       | 38,862 (23.9%)                          | 73,592 (63.7%)                           |                          |                                                       |                                         |                                          |   |           |         |         |     |                 |                |                |     |                |              |                |      |               |              |              |     |              |            |            |
| AST                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | 108,556 (3.5%)                                        | 6,866 (4.2%)                            | 12,477 (10.8%)                           |                          |                                                       |                                         |                                          |   |           |         |         |     |                 |                |                |     |                |              |                |      |               |              |              |     |              |            |            |
| APRI                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | 39,420 (1.3%)                                         | 3,253 (2.0%)                            | 4,209 (3.6%)                             |                          |                                                       |                                         |                                          |   |           |         |         |     |                 |                |                |     |                |              |                |      |               |              |              |     |              |            |            |
| NFS                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | 2,623 (0.1%)                                          | 407 (0.3%)                              | 767 (0.7%)                               |                          |                                                       |                                         |                                          |   |           |         |         |     |                 |                |                |     |                |              |                |      |               |              |              |     |              |            |            |
| <b>I. Sample size considerations</b><br><br>Due to the nature of this study (descriptive) it is not possible to perform a formal sample size calculation. The feasibility study undertaken (see section H) found large numbers of people with plasma measures available, and given we intend this study to explore 10 years of data (only one year was used in the feasibility study) we are confident there will be adequate numbers available. With respect to outcome measures. Our research group identified ~5000 cases of cirrhosis for the 10 years to 2009 (Fleming et al, AP&T 2010). Given we are planning to add an additional 7 years of data and rates of advanced liver disease are increasing we hypothesise in the region of 10,000 cases                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                       |                                         |                                          |                          |                                                       |                                         |                                          |   |           |         |         |     |                 |                |                |     |                |              |                |      |               |              |              |     |              |            |            |

| <p align="center"><b>Applicants must complete all sections listed below</b><br/> <b>Sections which do not apply should be completed as 'Not Applicable'</b></p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>of cirrhosis will be available.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| <p><b>J. Data Linkage Required (if applicable):</b></p> <p>We will use HES, ONS and Cancer Registry data to maximise ascertainment of outcomes i.e. to calculate the number of episodes of cirrhosis/hepatic decompensation recorded in HES that are not captured in the primary care data by additionally seeking them in linked data available using ICD10 codes – see Appendix B</p> <p>Specifically the linked data sets required and their uses are:</p> <ol style="list-style-type: none"> <li>1) ONS Death Registration Data – death is an outcome</li> <li>2) HES Admitted Patient Care – admissions related to hepatic decompensation are an outcome</li> <li>3) HES Accident and Emergency – attendances related to hepatic decompensation are an outcome</li> <li>4) Cancer Registration Data – hepatocellular carcinoma (HCC) is an outcome</li> <li>5) Practice/Patient Level IMD – deprivation is a known confounder to be controlled for</li> </ol>                                                                                                                                                                                                                                                                                                                                             |
| <p><b>K. Study population</b></p> <p>Patients aged 18 years or over, registered in the CPRD from the 1990 to 2016, and having at least one year of registration on CPRD after the latest date of either the date of current registration or the date the practice was labelled “up to standard” (UTS) will be potentially eligible for inclusion in the study.</p> <p>From within this population we will identify all those with the necessary data present to calculate a non-invasive marker of liver fibrosis. (See M below for list of markers)</p> <p>Follow-up time will begin on the first date (index date) that an abnormal non-invasive hepatic fibrosis measure was recorded.</p> <p>Patients will be censored at the first outcome event recorded in their record or at date of death, date of leaving the practice, date of the latest download of data, or the study end date, whichever is earliest.</p> <p>The study will be undertaken among all CPRD patients.</p> <p>Exclusion criteria include:</p> <ul style="list-style-type: none"> <li>- Age &lt;18years at the index date</li> <li>- Previously diagnosed chronic liver disease prior to the index date. Chronic liver disease as defined by codes in appendix C. NB acute liver disease (e.g. DILI) is not an exclusion.</li> </ul> |
| <p><b>L. Selection of comparison group(s) or controls</b></p> <p>From within all those identified as having the necessary data for non-invasive score calculation: exposed are those with a score indicating high risk, and the comparator are those with low risk scores.</p> <p>In effect we will undertake a series of four cohort studies here (one for each marker), with controls those negative for the individual marker (i.e. one control group per study).</p> <p>In order to ensure that controls are suitable participants at risk of liver disease their entry to the</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |

| Applicants must complete all sections listed below<br>Sections which do not apply should be completed as 'Not Applicable'                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| study will be at a random time between their first normal marker and censorship.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| <p><b>M. Exposures, Health Outcomes<sup>§</sup> and Covariates</b></p> <p><sup>§</sup>Please note: Summary information on health outcomes (as included on the ISAC application form above) will be published on CPRD's website as part of its transparency policy</p> <p><b>Exposure</b> – Screening positive for likely hepatic fibrosis will be defined as:</p> <ul style="list-style-type: none"> <li>a) AST/ALT &gt;1.0</li> <li>b) APRI score &gt;1.0</li> <li>c) FIB4 index &gt;1.30</li> <li>d) NAFLD Fibrosis score &gt;0.676</li> </ul> <p>Variation around these cut-offs will be examined in sensitivity analyses.</p> <p>READ codes for identification are listed in Appendix C</p> <p><b>Outcome</b> – development of liver cirrhosis and hepatic decompensation</p> <p>The outcomes to be studied include: liver cirrhosis (Child Pugh stage where available), oesophageal varices (bleeding/non-bleeding), hepatocellular cancer (tumour stage, size and treatment), ascites and a composite of all four.</p> <p>We will identify the first occurrence of each outcome during follow-up which will be defined as a READ code in the clinical file or coding in HES data.</p> <p>READ and ICD10 codes for identification are listed in Appendix B</p> <p>Specifically for HCC:</p> <p>The outcome of an HCC diagnosis is dependent upon the treatment modality. It is essential for the study to determine whether the HCCs are amenable to curative treatment or not (in which case earlier diagnosis could influence outcome). Tumours will be categorised by treatment indication i) curative - surgical resection or transplant (for solitary tumours &lt;5cm in diameter or &lt;= 3 tumours all &lt;= 3cm diameter); ii) palliative-medical – ablation therapy; iii) palliative-medical-chemoembolization and others. Treatment is determined by the size, number and site of the tumour. [This requires Cancer registry fields related to tumour size, number, and multifocality].</p> <p>Additionally knowledge of whether the tumour was detected in an existing screening programme will indicate whether there is any potential for earlier diagnosis.</p> <p>In order to determine this the Cancer registry fields related to existing surveillance will be used.</p> <p>Given that the majority of liver tumours are metastases and some are cholangiocarcinomas a sensitivity analyses will be undertaken on <i>proven</i> HCC.</p> |

| <p align="center"><b>Applicants must complete all sections listed below</b><br/> <b>Sections which do not apply should be completed as 'Not Applicable'</b></p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>HCC inclusion criteria:</p> <ul style="list-style-type: none"> <li>- Histologically proven HCC or,</li> <li>- Radiology consistent with HCC in the presence of an elevated alpha fetoprotein level</li> </ul> <p>[This requires Cancer registry fields related to histology and basis of diagnosis].</p> <p>As such this information is required to classify the outcome appropriately) In order to utilise the outcome effectively the following information is required: i) fact of diagnosis, ii) date of diagnosis (for time to event analysis), iii) staging of disease (to ensure like is compared with like) and iv) fact of treatment (the research involves screening for potential disease so identification at a treatable time is important)</p> <p><b>Covariates</b></p> <p>The presence of potential associations will be assessed by reviewing the medical records for any evidence of the following covariates at the baseline (index date); use of alcohol, body mass index (BMI), socioeconomic status (using individual IMD quintiles), presence of type 2 diabetes, other causes of thrombocytopenia, concurrent diagnosis of any underlying disease that may act as possible confounders for each outcome. Co-morbidities will be identified by the recording of a READ code in the patient's clinical file.</p> <p>READ codes for identification are listed in Appendix D</p> |
| <p><b>N. Data/ Statistical Analysis</b></p> <p>Analyses will be undertaken for each potential screening marker separately.</p> <p>Initial data investigation will be undertaken to determine the patterns of marker measurement. This will allow the optimal period to be chosen for the analyses when markers recording is steady to ensure any results found are not an artefact of changes in marker measurement recording within the dataset.</p> <ol style="list-style-type: none"> <li>(1) Descriptive statistics will be used to summarise the prevalence of hepatic fibrosis risk by each non-invasive marker and their associated characteristics. The prevalence of abnormal markers will be calculated as the number of people with an abnormal marker / the number of people with either a normal or abnormal marker. For the initial analyses the abnormality of markers will be determined using the literature based cut-offs as described in Section M.</li> <li>(2) In line with national waiting time directives patients with an abnormal marker and a diagnosis of advanced liver disease occurring more than 18weeks after the index marker will be defined as having their diagnosis brought forward by the screening marker. The proportion of patients</li> </ol>                                                                                                              |



| <b>Applicants must complete all sections listed below</b><br><b>Sections which do not apply should be completed as 'Not Applicable'</b> |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
|-----------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                                                                                                         | <p>whose diagnosis would be expedited using the marker will be calculated as the number of people with an abnormal marker and a diagnosis of CLD beyond 18 weeks / the number of people with a marker available. The proportion whose diagnosis would not have been brought forward will be calculated as the number of people with an abnormal marker and a diagnosis of CLD within 18 weeks / the number of people with a marker available.</p> <p>(3) Cox proportion hazards models will be used to estimate crude and adjusted hazard ratios (HRs) with 95% CIs of the first recorded diagnosis of each outcomes (in separate model) in those with abnormal markers compared to those with normal markers.</p> <p>(4) The most appropriate cut-off for use in primary care will be determined using net benefit analyses for diagnostic test prediction models (as described by Vickers et al BMJ 2016;352:i6).</p> <p>Risk stratification will be undertaken using the groups: presence of type 2 DM, hazardous alcohol use, obesity (BMI&gt;30), any 2 of the 3 risk factors, all 3 risk factors.</p> <p>Sensitivity analyses will be undertaken adjusting the cut-offs of each of the screening markers to produce a range of possible results. Cut-offs will be examined using i) a standard decile approach and ii) the cut-off determined using the novel methodology in (4).</p> |
| <b>O. Plan for addressing confounding</b>                                                                                               | <p>This study aims to understand the patterns of abnormal liver markers in general practice in order to investigate the feasibility of case finding within UK primary care.</p> <p>Assuming satisfactory markers and cut-offs are identified the intention would be to use the modelling described in (3) and (4) to determine the optimal predictive model of CLD. As such, no causal inferences are intended and any associations determined (confounding or otherwise) will be incorporated into the predictive model.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| <b>P. Plans for addressing missing data</b>                                                                                             | <p>We anticipate few issues of data known to be missing, beyond the fact that the components required for the calculation of non-invasive hepatic fibrosis markers will not have been checked in many of the subjects. In order to be eligible for the study participants must have all of the data necessary to calculate the non-invasive fibrosis risk score.</p> <p>Amongst confounding variables we anticipate BMI and alcohol have some missing data.</p> <p>We plan to undertake complete dataset analysis, where necessary making variables categorical with a missing category (as data will not be missing at random).</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| <b>Q. Patient or user group involvement (if applicable)</b>                                                                             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |

| <p align="center"><b>Applicants must complete all sections listed below</b><br/> <b>Sections which do not apply should be completed as 'Not Applicable'</b></p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>This study is supported by a group of six patient representatives drawn from the Nottingham Digestive Diseases Centre Biomedical Research Centre Patient Advisory Group. They have been involved with the design of the study and will continue to be involved in its implementation.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| <p><b>R. Plans for disseminating and communicating study results, including the presence or absence of any restrictions on the extent and timing of publication</b></p> <p><b>Research community</b></p> <p>This work is relevant to researchers working in the areas of Hepatology, General Practice and Public Policy. We have established strong collaborations with researchers working across these fields, who will facilitate access to these audiences by identifying relevant journals, conferences and networks through which to disseminate this work.</p> <p><b>Public Health community</b></p> <p>Through the Chief Investigators Honorary role at Public Health England there will be direct input into the East Midlands Liver Disease Framework and also national discussions. To date there have been three stakeholder events focused purely on CLD pathways and this proposed work has already been raised at them. This route and resultant networks will allow for the communication of research findings and their practical implications to wide range of stakeholders including: local councillors, local authority public health workforce, alcohol and drug misuse services, hepatologists, general practitioners, clinical commissioning groups, NHS England specialised commissioners, long term conditions services, strategic and service development personnel.</p> <p><b>Clinicians</b></p> <p>Whilst the above paragraph focuses on the Public Health community, this will be approached in a multi-disciplinary fashion. For example, the stakeholder events noted above include hepatologists and general practitioners. Through clinical colleagues, there will be two-way communications with local hospital clinicians to ensure any proposed pathway is well understood and any concerns are raised early (i.e. before the end of the project).</p> <p><b>Patients and public</b></p> <p>As noted in section Q this study is supported by a patient representative group. This group and the wider PAG would facilitate dissemination of the findings for example through assisting with lay summaries, press releases and social media content</p> <p>There are currently no restrictions on the extent and timing of publications.</p> |
| <p><b>S. Limitations of the study design, data sources, and analytic methods</b></p> <p>The key limitation of this study relates to the use of opportunistic screening tests for case finding. As</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |

**Applicants must complete all sections listed below  
Sections which do not apply should be completed as 'Not Applicable'**

such not all patients will have the relevant tests available. This will likely bias the sampling towards more unwell patients in whom it was felt such tests were needed.

The effect of missing data will also potentially bias the sampling towards a more unwell cohort.

There are alternative options for selection of controls. The study proposes selecting different control groups of each of the different screening tests. With controls being selected from those negative for the test under study. An alternative option would be to include those negative for all markers (a single control group). This second option would reduce the risk of allocating someone to the control group who may be positive for a different marker and hence actually have advanced liver disease (misallocation), however it will reduce the numbers available for the study notably and further bias the sample towards the most unwell patients as they will have required to have had all the component tests undertaken. In an attempt to understand the impact of such decisions we will re-analyse the data using the second approach to look at the range of possible results.

#### **T. References**

1. Guideline Development Group. NICE Guideline - Methods, evidence and recommendations Cirrhosis Assessment and management of cirrhosis. [Draft for consultation]. (National Institute for Health and Clinical Excellence, 2015).
2. Public Health Policy & Strategy Unit/NHS Commissioning Unit. Living Well for Longer: A call to action to reduce avoidable premature mortality. (Department of Health, 2013).
3. Office for National Statistics. Mortality statistics: deaths registered in England and Wales, 2010. (2011).
4. Craig, R., Fuller, E. & Mindell, J. Health Survey for England 2014: Health, social care and lifestyles. (Health and Social Care Information Centre, 2015).
5. Hart, C. L., Morrison, D. S., Batty, G. D., Mitchell, R. J. & Davey Smith, G. Effect of body mass index and alcohol consumption on liver disease: analysis of data from two prospective cohort studies. *BMJ* **340**, (2010).
6. Lozano, R. et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet* **380**, 2095–2128 (2012).
7. Davies, S. Chief Medical Officer annual report 2011: volume 1, chapter 2c, p 163. Dep. Health Novemb. 2012 Available [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/141773/CMOAnnualReport2011Chapter2c.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/141773/CMOAnnualReport2011Chapter2c.pdf) Accessed 01/07/2015 (2012).
8. Liver Disease: Today's complacency, tomorrow's catastrophe. All-Party Parliamentary Hepatology Group (APPHG) Inquiry into Improving Outcomes in Liver Disease. Available <http://www.asorg.uk/uploads/APPHG20report20March20201420FINAL.pdf> Accessed 01/07/2015 (2014).

**Applicants must complete all sections listed below  
Sections which do not apply should be completed as 'Not Applicable'**

9. Williams, R. 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. *The Lancet* **384**, 1953–1997 (2014).
10. Ahmed, A., Perumpail, R. B. & Harrison, S. A. High prevalence of hepatic fibrosis in the setting of coexisting diabetes and hepatic steatosis: A case for selective screening in the general population? *Hepatology* n/a-n/a (2015). doi:10.1002/hep.28277
11. Chalasani, N. et al. The diagnosis and management of non-alcoholic fatty liver disease: Practice Guideline by the American Association for the Study of Liver Diseases, American College of Gastroenterology, and the American Gastroenterological Association. *Hepatology* **55**, 2005–2023 (2012).
12. UK National Screening Committee. Criteria for appraising the viability, effectiveness and appropriateness of a screening programme. (2015).
13. Pasha, T., Gabriel, S., Therneau, T., Dickson, E. R. & Lindor, K. D. Cost-effectiveness of ultrasound-guided liver biopsy. *Hepatology* **27**, 1220–1226 (1998).
14. West, J. & Card, T. R. Reduced Mortality Rates Following Elective Percutaneous Liver Biopsies. *Gastroenterology* **139**, 1230–1237 (2010).
15. Regev, A. et al. Sampling error and intraobserver variation in liver biopsy in patients with chronic HCV infection. *Am. J. Gastroenterol.* **97**, 2614–2618 (2002).
16. Bedossa, P., Dargere, D. & Paradis, V. Sampling variability of liver fibrosis in chronic hepatitis C. *Hepatology* **38**, 1449–1457 (2003).
17. Theodossi, A. et al. Observer variation in assessment of liver biopsies including analysis by kappa statistics. *Gastroenterology* **79**, 232–241 (1980).
18. Wai, C.-T. et al. A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C. *Hepatology* **38**, 518–526 (2003).
19. Vallet-Pichard, A. et al. FIB-4: An inexpensive and accurate marker of fibrosis in HCV infection. comparison with liver biopsy and fibrotest. *Hepatology* **46**, 32–36 (2007).
20. Angulo, P. et al. The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD. *Hepatology* **45**, 846–854 (2007).
21. O'Sullivan, J. W. et al. Temporal trends in use of tests in UK primary care, 2000-15: retrospective analysis of 250 million tests. *bmj* **363**, k4666 (2018).
22. NICE guidelines [NG49] Non-alcoholic fatty liver disease (NAFLD): assessment and management. (National Institute for Health and Care Excellence (NICE), 2016).



| <p align="center"><b>Applicants must complete all sections listed below</b><br/> <b>Sections which do not apply should be completed as 'Not Applicable'</b></p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>23. NICE guidelines [NG50] Cirrhosis in over 16s: assessment and management. (National Institute for Health and Care Excellence (NICE), 2016).</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| <p><b>List of Appendices</b> <i>(Submit all appendices as separate documents to this application)</i></p> <p><b>Appendix A</b> – MRC Application and explanation of review process<br/> <b>Appendix B</b> – Liver biomarker (exposure) coding<br/> <b>Appendix C</b> – Liver disease (outcome) coding<br/> <b>Appendix D</b> – Covariate coding</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| <p><b>Amendment – 5<sup>th</sup> September 2019</b></p> <p><b>Assessment of temporal trends in liver enzyme testing</b></p> <p>Based on the original application CPRD have provided linked data on ~600,000 people with inclusion criteria:</p> <ul style="list-style-type: none"> <li>- Age &gt;40</li> <li>- Have an AST measure</li> <li>- Do not have chronic liver disease</li> </ul> <p>We would like to expand this to all adults (HES-CPRD data only) to look at overall testing trends and then look at high- risk liver disease sub groups (have liver disease/t2dm/obesity/alcohol excess).</p> <p><b>Objectives, specific aims and rationale:</b></p> <p><i>Objective:</i><br/> To understand the temporal changes in individual liver function testing over the period 1990 to most recent in UK general practice.</p> <p><i>Rationale:</i><br/> During the initial exploration of the data for the approved project 17_135R it was noted that the frequency of AST testing had dropped significantly over recent years. This is in contrast to what would be anticipated given the rise in liver disease and increasing pressure on GPs to identify such disease early. Additionally, Sullivan et als 2018<sup>i</sup> publication using CPRD refers to liver function tests collectively as increasing, which may mast differing trends in the individual components. This is of particular importance as in order to calculate liver fibrosis markers, as advised by NICE guidance<sup>ii,iii</sup>, AST and not ALT is more frequently required. It would therefore be beneficial as a precursor to the fibrosis biomarker project to undertake a preliminary study into the prevalence of liver enzyme testing over the period 1990 to most recently available.</p> <p><b>Background:</b><br/> Unchanged - See original protocol</p> <p><b>Study Type:</b><br/> Unchanged - Descriptive/exploratory</p> <p><b>Study design:</b><br/> Cross-sectional study.</p> <p><b>Feasibility counts:</b></p> |

**Applicants must complete all sections listed below**  
**Sections which do not apply should be completed as 'Not Applicable'**

**Sample size:**

Unchanged – see original protocol

**Data linkage required:**

For this additional element of the study no data linkage is required. We will extract our own data from CPRD Gold via our existing licence.

**Study population:**

Patients aged 18 years or over, registered in the CPRD from the 1990 to most recently available, and having at least one day of registration on CPRD after the latest date of either the date of current registration or the date the practice was labelled "up to standard" (UTS) will be potentially eligible for inclusion in the study. Additionally to maintain the source population for the existing approved study (above) only patients from HES linked practices will be included.

From within this population we will identify all those with the necessary data present to calculate the prevalence of liver enzyme testing.

Exclusion criteria include:

- Age <18years at the index date

**Selection of control group:**

No control group

**Exposures, health outcomes and covariate:**

*Exposure:*

Liver enzymes

- Alanine aminotransferase (ALT)
- Aspartate aminotransferase (AST)
- Gamma glutamyl transferase (GGT)
- Albumin
- Bilirubin
- Platelets

To be identified from the entity file

*Outcome:*

No outcome, exposure prevalence study

*Covariates:*

Unchanged - see original protocol

**Data/Statistical analysis:**

Temporal trends in test use, and crude and age and sex standardised rates of each test use.

Subgroup analyses will be as in the existing application (type 2 diabetes, alcohol excess and obesity) and also amongst those with known liver disease.

These subgroups will be derived from read codes as described in the original protocol.

**Plans for addressing confounding:**

Unchanged - see original protocol

**Plans for addressing missing data:**

Unchanged - see original protocol

**Patient involvement:**

## Appendix D

*Appendix D.1: Medcodes and Read codes for identifying patients with CLD in CPRD for the codes approved in the ISAC form.*

| MedCode       | Read Code | Read Term                                                    |
|---------------|-----------|--------------------------------------------------------------|
| <b>6863</b>   | J61..00   | Cirrhosis and chronic liver disease                          |
| <b>1638</b>   | J615z13   | Cirrhosis of liver NOS                                       |
| <b>40953</b>  | H55..11   | Cirrhosis of lung                                            |
| <b>16725</b>  | J615.00   | Cirrhosis - non alcoholic                                    |
| <b>69204</b>  | J615100   | Multilobular portal cirrhosis                                |
| <b>40567</b>  | J615600   | Capsular portal cirrhosis                                    |
| <b>26319</b>  | G852200   | Oesophageal varices in cirrhosis of the liver                |
| <b>58630</b>  | J616z00   | Biliary cirrhosis NOS                                        |
| <b>68376</b>  | J612.11   | Florid cirrhosis                                             |
| <b>6015</b>   | Jyu7100   | [X]Other and unspecified cirrhosis of liver                  |
| <b>55454</b>  | J615y00   | Portal cirrhosis unspecified                                 |
| <b>47257</b>  | J615.11   | Portal cirrhosis                                             |
| <b>15424</b>  | J616100   | Secondary biliary cirrhosis                                  |
| <b>22841</b>  | J615z11   | Macronodular cirrhosis of liver                              |
| <b>44676</b>  | J615400   | Fatty portal cirrhosis                                       |
| <b>8363</b>   | G852300   | Oesophageal varices in alcoholic cirrhosis of the liver      |
| <b>5638</b>   | J616000   | Primary biliary cirrhosis                                    |
| <b>92909</b>  | J615500   | Hypertrophic portal cirrhosis                                |
| <b>4743</b>   | J612.00   | Alcoholic cirrhosis of liver                                 |
| <b>16455</b>  | J615z00   | Non-alcoholic cirrhosis NOS                                  |
| <b>3450</b>   | J615300   | Diffuse nodular cirrhosis                                    |
| <b>18739</b>  | J615z12   | Cryptogenic cirrhosis of liver                               |
| <b>9494</b>   | J616.00   | Biliary cirrhosis                                            |
| <b>24220</b>  | 7609.00   | Open operations on oesophageal varices                       |
| <b>73139</b>  | G852100   | Oesophageal varices without bleeding in diseases EC          |
| <b>10797</b>  | G858.00   | Oesophageal varices NOS                                      |
| <b>107975</b> | 7609y00   | Other specified open operation on oesophageal varices        |
| <b>20912</b>  | 7609400   | Open injection sclerotherapy to oesophageal varices          |
| <b>20233</b>  | 7609z00   | Open operation on oesophageal varices NOS                    |
| <b>30655</b>  | G851.00   | Oesophageal varices without bleeding                         |
| <b>11960</b>  | 760C300   | Fibreoptic endoscopic injection sclerotherapy oesoph varices |
| <b>46647</b>  | 760F400   | Rigid oesophagoscopy banding of oesophageal varices          |
| <b>1641</b>   | G85..11   | Oesophageal varices                                          |
| <b>11972</b>  | G85..12   | Varices - other                                              |
| <b>89717</b>  | 761D800   | Fibreopt endoscop rubber band ligation of upper GIT varices  |
| <b>28929</b>  | G857.00   | Gastric varices                                              |
| <b>24989</b>  | G850.00   | Oesophageal varices with bleeding                            |
| <b>47214</b>  | 760F300   | Rigid oesophagoscopy injection sclerotherapy oesoph varices  |
| <b>105611</b> | Gyu9400   | [X]Oesophageal varices in diseases classified                |

|              |         |                                                      |
|--------------|---------|------------------------------------------------------|
|              |         | elsewhere                                            |
| <b>43404</b> | 7609300 | Local ligation of oesophageal varices                |
| <b>16759</b> | 760C500 | Fibreoptic endoscopic banding of oesophageal varices |
| <b>96756</b> | G852000 | Oesophageal varices with bleeding in diseases EC     |
| <b>44424</b> | G852.00 | Oesophageal varices in diseases EC                   |
| <b>22411</b> | J622.11 | Encephalopathy - hepatic                             |
| <b>1508</b>  | R095.00 | [D]Ascites                                           |
| <b>58736</b> | 25OZ.00 | O/E - ascites NOS                                    |
| <b>48138</b> | 7H2B113 | Insertion of peritoneal to venous shunt for ascites  |
| <b>3503</b>  | 7H2B200 | Drainage of ascites NEC                              |
| <b>44222</b> | 7H2B100 | Peritoneal to venous drainage for ascites            |
| <b>37930</b> | R095z00 | [D]Ascites NOS                                       |
| <b>16113</b> | 25O..00 | O/E - ascites                                        |
| <b>19438</b> | 7H2B000 | Paracentesis abdominis for ascites                   |
| <b>52838</b> | 25O4.00 | O/E -ascites-shifting dullness                       |
| <b>73305</b> | 25O3.00 | O/E-ascites-fluid thrill shown                       |

*Appendix D.2: ICD10 codes for identifying patients with CLD in HES.*

| <b>Code</b>  | <b>Definition</b>                                                     |
|--------------|-----------------------------------------------------------------------|
| <b>K70.3</b> | Alcoholic cirrhosis of liver                                          |
| <b>K74</b>   | Fibrosis and cirrhosis of liver                                       |
| <b>K76.6</b> | Portal hypertension                                                   |
| <b>K76.7</b> | Hepatorenal syndrome                                                  |
| <b>C22.0</b> | Liver cell carcinoma                                                  |
| <b>I98.2</b> | Oesophageal varices without bleeding in diseases classified elsewhere |
| <b>I98.3</b> | Oesophageal varices with bleeding in diseases classified elsewhere    |
| <b>R18</b>   | Ascites                                                               |

*Appendix D.3: consensus codes List for liver cirrhosis adapted from (287).*

| <b>ICD10 code</b> | <b>Description</b>                                   |
|-------------------|------------------------------------------------------|
| <b>K74.6</b>      | Other and unspecified cirrhosis of the liver         |
| <b>K70.3</b>      | Alcoholic cirrhosis of the liver                     |
| <b>I85</b>        | Oesophageal varices                                  |
| <b>I85.0</b>      | With bleeding                                        |
| <b>I85.9</b>      | Without bleeding                                     |
| <b>I98</b>        | Oesophageal varices in diseases classified elsewhere |
| <b>I98.2</b>      | Without bleeding                                     |
| <b>I98.3</b>      | With bleeding                                        |

|              |                              |
|--------------|------------------------------|
| <b>K76.6</b> | Portal hypertension          |
| <b>K72.9</b> | Hepatic failure, unspecified |
| <b>K76.7</b> | Hepatorenal syndrome         |

*Appendix D.4 :Mapped codes list from consensus codes list for Liver cirrhosis*

| <b>Med code</b> | <b>Description</b>                                           |
|-----------------|--------------------------------------------------------------|
| <b>1638</b>     | Cirrhosis of liver NOS                                       |
| <b>1641</b>     | Oesophageal varices                                          |
| <b>3450</b>     | Diffuse nodular cirrhosis                                    |
| <b>4743</b>     | Alcoholic cirrhosis of liver                                 |
| <b>6015</b>     | [X]Other and unspecified cirrhosis of liver                  |
| <b>6863</b>     | Cirrhosis and chronic liver disease                          |
| <b>8206</b>     | Pigmentary cirrhosis of liver                                |
| <b>8363</b>     | Oesophageal varices in alcoholic cirrhosis of the liver      |
| <b>10797</b>    | Oesophageal varices NOS                                      |
| <b>11960</b>    | Fibreoptic endoscopic injection sclerotherapy oesoph varices |
| <b>16455</b>    | Non-alcoholic cirrhosis NOS                                  |
| <b>16725</b>    | Cirrhosis - non alcoholic                                    |
| <b>16759</b>    | Fibreoptic endoscopic banding of oesophageal varices         |
| <b>18739</b>    | Cryptogenic cirrhosis of liver                               |
| <b>19512</b>    | Glycogenosis with hepatic cirrhosis                          |
| <b>20912</b>    | Open injection sclerotherapy to oesophageal varices          |
| <b>22841</b>    | Macronodular cirrhosis of liver                              |
| <b>24220</b>    | Open operations on oesophageal varices                       |
| <b>30655</b>    | Oesophageal varices with bleeding                            |
| <b>26319</b>    | Oesophageal varices in cirrhosis of the liver                |
| <b>27438</b>    | Cardiac portal cirrhosis                                     |
| <b>28929</b>    | Gastric varices                                              |
| <b>73139</b>    | Oesophageal varices without bleeding                         |
| <b>40567</b>    | Capsular portal cirrhosis                                    |
| <b>40963</b>    | Portal fibrosis without cirrhosis                            |
| <b>44120</b>    | Toxic liver disease with fibrosis and cirrhosis of liver     |
| <b>44424</b>    | Oesophageal varices in diseases EC                           |
| <b>44676</b>    | Fatty portal cirrhosis                                       |
| <b>46647</b>    | Rigid oesophagoscopy banding of oesophageal varices          |

|              |                                                             |
|--------------|-------------------------------------------------------------|
| <b>47214</b> | Rigid oesophagoscopy injection sclerotherapy oesoph varices |
| <b>48928</b> | Infectious cirrhosis NOS                                    |
| <b>55333</b> | Sengstaken intubation of oesophagus                         |
| <b>62582</b> | Oesophageal varices in diseases EC NOS                      |