Big data in viral hepatitis



Cori Campbell

Green Templeton College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2023

Acknowledgements

Throughout my doctoral studies I have been fortunate to work in various interdisciplinary teams comprised of exceptional colleagues who have inspired and supported me in countless ways. I wish to thank my supervisors Professor Philippa Matthews, Professor Eleanor Barnes and Dr. Iain Gillespie for their unwavering support and guidance, and their excellent scientific direction. My heartfelt thanks and appreciate go to thank Dr. Tingyan Wang, who has been an invaluable mentor and colleague throughout my studies. I extend my sincere gratitude to all my colleagues and peers in the Matthews and Barnes research groups, the Clinical Informatics group, and the Real-World Analytics team at GSK. It has been an exciting and rewarding privilege to work together with a shared passion to produce powerful evidence with direct impacts to patient pathways using real-world data.

Thank you to all of individuals, patients, health care and other professionals working tirelessly to make the National Health Service an exceptional system. The stalwart dedication of millions to provide excellent and accessible health care provision in the United Kingdom has been a constant motivation throughout my studies. Using data from this invaluable resource to address evidence gaps which impact daily clinical care pathways has been a privilege which I am truly grateful for, and which I hope to continue pursuing in future scientific endeavours.

My sincerest thanks are extended to Kelsey, my parents Teena and Garry, Trishna, Gabriel and Bugei. Your unwavering support, friendship, strength, and optimism have supported me through this journey, and I cherish the wonderful fun and adventures we have enjoyed together these years. Thank you for inspiring me to persevere.

Table of Contents

LIST OF TABLES 8

LIST OF FIGURES 11

ABBREVIATIONS 14

1. INTRODUCTION AND BACKGROUND 16

1.1. HEPATITIS B VIRUS (HBV) 17

1.1.1 Virology 17

1.1.2 Diagnosis 18

1.1.3. Treatment 18

1.1.4. Prevention 19

1.2. GLOBAL AND REGIONAL (UK) DISEASE BURDEN OF HBV 22

1.2.1. Prevalence estimates 22

1.2.2. Morbidity and mortality burdens 25

1.3. ELECTRONIC HEALTH RECORD (EHR) DATABASES 27

1.3.1. Primary care databases 28

1.3.1.1. QResearch 29

1.3.1.2. Clinical Practice Research Datalink 31

1.3.2. Secondary care databases 31

1.3.2.1. NIHR Health Informatics Collaborative 32

1.3.3. EHR-based versus traditional observational cohorts 32

1.4. HBV AND EHR – CURRENT EVIDENCE GAPS 33

1.5. AIMS AND OBJECTIVES 35

2. SYSTEMATIC REVIEW AND META-ANALYSIS TO INVESTIGATE RISK FACTORS FOR HCC IN CHB 36

2.1. AVAILABILITY OF CONTENT 36

2.2. BACKGROUND 36

2.3. RESEARCH QUESTION 37

2.4. METHODS 37

2.4.1. Search strategy and selection criteria 37

2.4.2 Data extraction and statistical analysis 40

2.4.3. Definitions 40

2.4.4. Quality appraisal 41

2.4.5. Note on terminology 41
2.5. RESULTS 41
2.5.1. Study characteristics 41
2.5.2. Quality assessment 59
2.5.3. Association of diabetes mellitus with HCC risk 73
2.5.4. Meta-analysis of cohort studies 79
2.5.5. Subgroup and sensitivity analyses 80
2.5.6. Association of hypertension with hepatocellular carcinoma risk 85
2.5.7. Associations of other comorbidities with hepatocellular carcinoma risk .. 88
2.6. DISCUSSION 93
2.6.1. Summary of findings 93
2.6.2. Critical appraisal of included studies – risk of bias, confounding, heterogeneity & measurement error, and comparison with previous meta-analysis 93
2.6.3. Diabetes mellitus and HCC risk in CHB 95
2.6.4. Impact of meta-analysis findings 98
2.6.5. Systematic search update – 2020 to 2023 98
3. CHRONIC HBV COHORT IN THE QRESEARCH PRIMARY CARE DATABASE: A SUMMARY/APPRAISAL OF UK-WIDE CHB PREVALENCE ESTIMATES, CHARACTERISATION OF THE CHB POPULATION AND INVESTIGATION OF RISK FACTORS FOR HEPATOCELLULAR CARCINOMA 99
3.1. AVAILABILITY OF CONTENT 99
3.2. BACKGROUND 99
3.2.1. HBV in the UK: Epidemiology and disease burden 99
3.2.2. QResearch primary care EHR database 101
3.2.3. Aims and objectives 102
3.3. RESEARCH QUESTIONS 103
3.3.1. Research question 1: What is the prevalence of CHB in the UK? A summary and appraisal of previous national CHB prevalence estimates 103
3.3.2. Research question 2: An investigation of demographics/characteristics and risk factors for HCC in the UK CHB population, undertaken in a primary care EHR database (QResearch). 103
3.4. METHODS 103

3.4.1. Methods – Research question 1 103
3.4.2. Methods – Research question 2 104
3.5. RESULTS 114
3.5.1. Research question 1 114
3.5.2. Research question 2 119
3.6. DISCUSSION 140

4. CHRONIC HBV COHORT IN THE NATIONAL INSTITUTE FOR HEALTH RESEARCH HEALTH INFORMATICS COLLABORATIVE SECONDARY CARE DATABASE: ANTIVIRAL TREATMENT COVERAGE AND ELIGIBILITY, AND THE IMPACT OF THE COVID-19 PANDEMIC ON ROUTINE SURVEILLANCE 150

4.1. AVAILABILITY OF CONTENT 150
4.2. BACKGROUND 150
4.2.1. HBV and secondary/tertiary care in the UK 150
4.2.2. NIHR HIC secondary/tertiary care EHR database 152
4.2.3. Aims and objectives 155
4.3. RESEARCH QUESTIONS 155
4.3.1. Research question 1: Antiviral treatment eligibility and coverage across individuals with CHB in the NIHR HIC secondary EHR database; who is appropriately receiving treatment, who is not receiving treatment, and how would modified criteria impact treatment coverage? 155
4.3.2. Research question 2: What was the impact of the COVID-19 pandemic on routine CHB surveillance in secondary care settings? 155
4.4. METHODS 155
4.4.1. Methods – Research question 1 155
4.4.2. Methods – Research question 2 159
4.5. RESULTS 160
4.5.1. Research question 1 160
4.5.2. Research question 2 167
4.6 DISCUSSION 173

5. UK HBV IMMUNISATION IN THE CLINICAL PRACTICE RESEARCH DATALINK PRIMARY CARE ELECTRONIC HEALTH RECORD DATABASE: CHANGES IN COVERAGE IN RESPONSE TO POLICY CHANGES AND THE COVID-19 PANDEMIC 179

5.1 Availability of content 179
5.2 Background 179
5.2.1. HBV immunisation in the UK 179
5.2.2. Clinical Practice Research Datalink (CPRD) primary care EHR database 181
5.2.3. Aims and objectives 182
5.3. Research questions 183
5.3.1. Research question 1. What are the percentages of UK infants receiving 1, 2, 3 or >3 HBV vaccines within the first 14 months of life from 2015 to mid-2020? 183
5.3.2. Research question 2. How have the introduction of the HBV vaccine as a universal infant immunisation and the COVID-19 pandemic impacted HBV vaccination coverage in the UK, and are there differences in coverage based on geographic region or ethnic group? 183
5.4. Methods 184
5.4.1. Study design and follow-up 184
5.4.2. Sample size / power calculations 186
5.4.3. Statistical analysis 188
5.5. Results 189
5.6. Discussion 199
6. DISCUSSION 202
6.1 Summary of work and findings 202
6.2 Future recommendations and next steps 205
6.3 Reflection as an electronic health record researcher 207
6.4. Research output outside doctoral project 207
6.5. Conclusion 209
7. REFERENCES 209
8. APPENDIX 248

Abstract

Chronic infection with hepatitis B virus is one of the leading global causes of chronic liver disease, specifically primary liver cancer. The population is not well characterised in the United Kingdom, with evidence gaps including lack of characterisation of the infected population, the burden of comorbid disease and risk factors for progression to liver cancer. Antiviral treatment eligibility and coverage across the disease population is unclear, as are impacts of COVID-19 and health policy on viral surveillance and vaccination. This thesis presents results from analyses undertaken across multiple electronic health record databases in the United Kingdom to address these gaps.

Systematic literature review and meta-analysis were undertaken to appraise previous studies describing epidemiology and investigating primary liver cancer risk factors. Subsequently the largest cohort of chronically infected individuals to date was characterised in the QResearch database. The chronic hepatitis B virus population in the United Kingdom has been characterised as a young, ethnically diverse and socioeconomically deprived population of individuals. Risk factors for primary liver cancer included male sex, increasing age and comorbidities. A protective association of statin use with risk of primary liver cancer was observed.

In the NIHR HIC database, antiviral treatment coverage across individuals with chronic hepatitis B virus was quantified, reporting treatment initiation practices in line with clinical recommendations. The impact of COVID-19 on surveillance was quantified, reporting a decrease in surveillance as compared to the pre-COVID-19 era. Surveillance has not recovered to pre-COVID coverage. In the Clinical Practice Research Datalink database, it was demonstrated hepatitis B virus vaccination coverage has both increased since the compounds introduction as a universal component in 2017 and has remained high despite the COVID-19 pandemic.

These results have clinical relevance to both care pathways and research investigations. They may facilitate improvements in clinical management, resource allocation and future investigations.

List of Tables

Table 1.1. Summary of UK primary and secondary care electronic health record databases utilised for completion of doctoral studies. 29
Table 2.1. Search terms used in systematic search of Embase, MEDLINE and Web of Science databases. 39
Table 2.2. Summary characteristics of included observational studies. 43
Table 2.3. Covariate adjustment for included observational studies. 52
Table 2.4. Quality appraisal scores of case-control studies identified in meta-analysis. Quality scores awarded according to Newcastle-Ottawa Scale. 60
Table 2.5. Quality appraisal scores of cohort studies identified in meta-analysis. Quality scores awarded according to Newcastle-Ottawa Scale. 62
Table 2.6. Effect estimates for case-control studies investigating the association of diabetes mellitus with hepatocellular carcinoma risk. 74
Table 2.7. Effect estimates for cohort studies investigating the association of diabetes mellitus with hepatocellular carcinoma risk. 76
Table 2.8. Effect estimates for cohort studies investigating the association of hypertension with hepatocellular carcinoma risk. 86
Table 2.9. Effect estimates for associations of dyslipidaemia with hepatocellular carcinoma risk. 89
Table 2.10. Effect estimates for associations of obesity with hepatocellular carcinoma risk. 90
Table 2.11. Effect estimates for associations of statin use and cardiovascular disease with hepatocellular carcinoma risk. 91
Table 2.12. Effect estimates for associations of other factors with hepatocellular carcinoma risk. 92
Table 3.1. Systemised Nomenclature of Medicine primary care codes indicative of hepatocellular carcinoma (HCC) used to identify cases of HCC in adults with CHB from the QResearch database. 107
Table 3.2. Hepatocellular carcinoma (HCC) cases stratified by source of diagnosis. 107
Table 3.3. Follow-up time, stratified by source of chronic hepatitis B virus infection diagnosis. 108

Table 3.4. Quintiles for natural log transformed alanine transaminase (ALT), natural log transformed aspartate transaminase (AST) and platelet count (Plt). ALT and AST were natural log transformed due to their skewed distributions. 109
Table 3.5. Missingness in CHB data in QResearch cohort of adults according to cohort characteristics. 110
Table 3.6. Details of main and sensitivity analyses performed to investigate risk factors for hepatocellular carcinoma in QResearch cohort of adults with chronic hepatitis B virus (CHB) infection. 112
Table 3.7. Tabulation of non-hepatocellular neoplasms in the CHB cohort, overall and stratified by hepatocellular carcinoma (HCC) status. 113
Table 3.8. Prevalence estimates for chronic hepatitis B virus (HBV) infection in sub-populations in the United Kingdom. 116
Table 3.9. Cohort baseline characteristics for all individuals with chronic HBV infection included in this cohort. 119
Table 3.10. Characteristics of adults with chronic hepatitis B virus infection identified from QResearch primary care database, stratified by socioeconomic status. 124
Table 3.11. Baseline characteristics of complete-case and imputed (by multiple imputation with chained equations) datasets used in Cox proportional hazards models. 128
Table 3.12. Cox proportional hazards model using complete imputed dataset generated by multiple imputation with chained equations (n = 7029, HCC cases = 161). 134
Table 3.13. Cox proportional hazards complete-case model conducted on adults with hepatitis B virus infection for whom data was complete in the original QResearch dataset (n = 3648, HCC cases = 68). 138
Table 4.1. Details of main and sensitivity analysis performed to investigate factors associated with antiviral treatment imitation for individuals with chronic hepatitis B virus infection in the NIHR Health Informatics Collaborative database. 158
Table 4.2. Baseline characteristics of individuals included in this NIHR HIC analysis investigating factors associated with antiviral treatment initiation in chronic HBV infection. 160
Table 4.3. Types of hepatic imaging performed in individuals included in this NIHR HIC analysis investigating factors associated with antiviral treatment initiation in chronic HBV infection. 162

Table 4.4. Cox proportional hazards univariable and main multivariable model investigating factors which are associated with antiviral treatment initiation in adults with CHB. 163
Table 4.5. Results from sensitivity analyses undertaken to test robustness of the main model, with patients experiencing the outcome of interest within the first year of follow up excluded. 166
Table 5.1. 95% precision estimates for vaccine uptake across a range of sample sizes and proportional vaccine uptake. Uptake estimates are based on previous Public Health England reports (15–18). 186
Table 5.2. Table of patient characteristics including sex, regional and ethnicity tabulations. 190
Table 5.3. The geographical representation in the entire study sample of infants (n (%)) from across the English regions, Scotland, Wales and Northern Ireland in the cohort over time. 193
Table 5.4. HBV, tetanus and MenB vaccination uptake across three birth periods. 196

List of Figures

Figure 1.1. Key hepatitis B virus targets to be achieved by 2030 outlined in the World Health Organisation's Global Health Sector Strategy. 27
Figure 2.1. Flow chart of study selection. MEDLINE, Embase and Web of Science databases were systematically searched using relevant terms to identify relevant human studies published in English from 1st January 2000 to 24th June 2020. 58
Figure 2.2. Forest plot of hazard ratios from longitudinal cohort studies investigating the association of diabetes mellitus with risk of progression of chronic hepatitis B infection to hepatocellular carcinoma. The study by Yu et al (72) provided an Odds Ratio and was excluded from the meta-analysis. Dashed vertical lines represent HR based on meta-analysis of all studies by fixed effect and random effects models. The studies for pooling the HR had sufficient quality (quality scores ≥5)..... 80
Figure 2.3. Sensitivity analysis of primary meta-analysis demonstrating the influence of restricting the analysis to studies additionally adjusting for HCV coinfection in addition to age and sex. 81
Figure 2.4. Sensitivity analysis of primary meta-analysis demonstrating the influence of restricting the analysis to studies additionally adjusting for HCV coinfection, HIV coinfection and antiviral treatment in addition to age and sex. 81
Figure 2.5. Sensitivity analysis of primary meta-analysis demonstrating the influence of restricting the analysis to studies additionally adjusting for cirrhosis in addition to age and sex. 82
Figure 2.6. Sensitivity analysis of primary meta-analysis stratifying studies by diabetes mellitus disease subtype. 83
Figure 2.7. Sensitivity analysis restricting the primary meta-analysis to studies adjusting for use of metformin. 84
Figure 3.1. Study flowchart depicting identification of adults from the QResearch (QR) primary care database eligible for inclusion an analysis of chronic hepatitis B (CHB) infection..... 105
Figure 3.2. A directed acyclic diagram to demonstrate risk factors for hepatocellular carcinoma development identified via systematic review and meta-analysis (6). Risk factors and their proposed causal pathways are visualised. ..... **Error! Bookmark not defined.**

Figure 3.3. Townsend deprivation quintile breakdown in 8039 adults with chronic hepatitis B virus infection derived from the QResearch primary care database (England) versus the United Kingdom general population. 123
Figure 3.4. Ethnicity breakdown in 8039 adults in a chronic hepatitis B virus cohort characterised from the QResearch primary care database (England) versus all individuals in the QResearch database (~35 million), vs. the United Kingdom general population. General population estimates obtained from 2019 estimates from the Office for National Statistics. 123
Figure 3.5. Forest plot for Cox proportional hazards model to identify risks of hepatocellular carcinoma (HCC) in an adult population with chronic Hepatitis B virus infection derived from the QResearch primary care database. Analysed using a dataset generated by multiple imputation with chained equations (n = 7029, HCC cases = 161). First (least deprived) to fifth (most deprived) Townsend Deprivation Quintiles are denoted by SES1-5, respectively. 133
Figure 4.1. Locations of the 10 National Health Service (NHS) Trusts participating in the National Institute for Health Research (NIHR) Health Informatics Collaborative (HIC) viral hepatitis theme up to February 2022. 154
Figure 4.2. Timeline visualising different follow-up scenarios A, B and C for the analysis. 156
Figure 4.3. Forest plot for main multivariable Cox proportional hazards model from analysis investigating factors associated with antiviral treatment initiation in chronic Hepatitis B virus infected individuals (n = 4771) undertaken in the HIC database. 165
Figure 4.4. Kaplan-Meier plots demonstrating the cumulative proportion of adults with chronic Hepatitis B virus (HBV) infection undergoing routine laboratory surveillance. 168
Figure 4.5. Mean numbers of ALT and HBV DNA VL measurements per 100 patients per month during pre-COVID (2016–19) and COVID-19 (2020-2021) periods. 172
Figure 5.1 Infant vaccination timeline from 0 to 52 weeks of life depicting the vaccine schedule for the hexavalent combination vaccine and standalone Meningitis B vaccine. 181
Figure 5.2. Study timeline depicting calendar years (and their midpoints) with dotted lines, and relevant study periods in shaded areas. Infants are divided into 3 study periods according to month and date of birth. 183

Figure 5.3. Power calculation of difference of two different sample sizes with different proportions equating to a total sample size for investigation of pentavalent DTaP/IPV/Hib/HepB vaccination coverage across calendar years. 188
Figure 5.4. Percentages of infants receiving full vaccine schedules within the first 14 months of life for hepatitis B virus, tetanus and meningitis immunisations for infants born between January 2015 to June 2020. 194
Figure 5.5. Percentages of infants receiving full HBV vaccine schedules within the first 14 months of life for years 2015, 2017 and 2019 throughout the United Kingdom. 195
Figure 5.6. Percentages of infants receiving full vaccinations schedules across 3 separate birth periods: 197

Abbreviations

AASLD	American Association of the Study of Liver Disease
AFP	Alpha-fetoprotein
aHR	Adjusted hazards ratio
APASL	Asian Pacific Association for the Study of the Liver
ARLD	Alcohol-related liver disease
BBV	Blood borne virus
BMI	Body mass index
BRC	Biomedical Research Centre
CHB	Chronic hepatitis B virus infection
CHD	Coronary heart disease
CI	Confidence Interval
COVER	Cover of Vaccination Evaluated Rapidly
COVID-19	Coronavirus Disease 2019
CT	Computed tomography
CVD	Cardiovascular disease
DALY	Disability adjusted life year
DM	Diabetes mellitus
DNA	Deoxyribonucleic acid
EASL	European Association for the Study of the Liver
ECDC	European Centre for Disease Prevention and Control
GBD	Global Burden of Disease
HBeAg	Hepatitis B virus E antigen
HBsAg	Hepatitis B surface antigen
HBsAg	Hepatitis B virus surface antigen
HBIG	Hepatitis B immune globulin
HBV	Hepatitis B virus
HBV VL	Hepatitis B virus DNA viral load
HCC	Hepatocellular carcinoma
HCW	Healthcare worker
HCV	Hepatitis C virus

HES	Hospital episode statistics
HIC	Health Informatics Collaborative
HIV	Human immunodeficiency virus
HR	Hazards ratio
HT	Hypertension
IDPS	Infectious Diseases in Pregnancy Screening
IHD	Ischaemic heart disease
ISOSS	Integrated Screening Outcomes Surveillance Service
MAFLD	Metabolic and fatty liver disease
MAP	Mean arterial pressure
MRI	Magnetic resonance imaging
NA	Nucleos(t)ide analogue
NAFLD	Non-alcoholic fatty liver disease
NCRAS	National Cancer Registration and Analysis Service
NHS	National Health Service
NHS	National Health Service Blood and Transplant
NIHR	National Institute for Health Research
NOS	Newcastle-Ottawa Scale
OPCS	Office of Population Censuses and Surveys
OR	Odds ratio
PHE	Public Health England
PWID	People who inject drugs
RCT	Randomised controlled trial
RR	Risk ratio
SGSS	Second Generation Surveillance System
SLD	Steatotic liver disease
SSBBV	Sentinel surveillance of blood borne virus
TRE	Trusted research environment
UK	United Kingdom
UKHSA	United Kingdom Health Security Agency
VL	Viral load
WHO	World Health Organisation

1. Introduction and background

Hepatitis B virus (HBV) is a viral infection sequestering in the liver which is transmitted via contact with blood or other bodily fluids from an infected person (1). The most frequent route of HBV transmission is from mother to child during birth (1). Additional transmission routes include sexual intercourse, exposure to contaminated sharp objects and needle sharing. HBV causes both acute and chronic infection, and risk of chronic HBV infection (CHB) is associated with age, whereby infection at a younger age confers a higher risk of developing CHB (2).

CHB is both a preventable and treatable condition (1). The World Health Organisation (WHO) recommends universal neonatal HBV vaccination, as well as prophylactic antiviral treatment for chronically infected mothers in order to prevent mother-to-child transmission among those at highest risk of transmission. Antiviral treatment is primarily based on nucleos(t)ide analogue (NA) agents, many of which were originally developed for HIV treatment (3). Because NA treatment does not cure HBV infection but rather supresses viral replication and pathogenesis, people living with HBV infection require lifelong NA treatment to minimise the risks of development of end-stage disease, in combination with vigilant surveillance and monitoring for disease progression.

HBV is a major contributor to liver cancer morbidity and mortality burdens worldwide. Chronic HBV infection increases an individual's risk of developing liver disease including cirrhosis and primary liver cancer (hepatocellular carcinoma, HCC) (4). Despite the existence of efficacious vaccines and treatments, the CHB disease burden is substantial. There are many HBV evidence gaps surrounding crucial questions including who is infected, who is receiving treatment and who is accessing regular healthcare/surveillance services. These questions must be answered to support meaningful advances towards viral elimination as a public health threat.

1.1. Hepatitis B virus (HBV)

The hepatitis B virus (HBV) is a hepatotropic virus which sequesters in liver cells to archive a copy of viral DNA. From here a diverse range of outcomes may arise, ranging from viral DNA integration into host genome, viral interaction with host restriction factors and hepatocytic cell cycle disruption of cell cycle. However, HBV may alternatively exist within a host without causing any pathophysiological processes. There are multiple mechanisms by which HBV causes chronic liver damage which may ultimately result in end-stage outcomes. CHB is diagnosed via serological investigation whereby the sustained serum presence of viral markers indicates chronic infection. Medicines are available to both prevent and treat CHB.

1.1.1 Virology

HBV belongs to the Hepadnaviridae viral family (5,6). It is a partially double-stranded virus comprised of circular deoxyribonucleic acid (DNA). The virus integrates itself into the host genome in hepatocytes, after which it causes genomic instability within host cells as well as epigenetic disruption (7). This is accomplished via many pathways which produce genetic mutations, promote malignant cellular expansion and/or disturb genetic methylation processes, amongst other pathological cascades (7).

HBV virions transport relaxed circular (rc)-DNA to hepatocyte nuclei, where viral genomic material is converted to covalently closed circular DNA (cccDNA) (5,6). Once HBV is residing in hepatocyte nuclei, cccDNA is maintained and transcribed in order to produce pathogenic viral proteins. The ultimate pathogenic process of HBV relies on viral replication to maintain cccDNA stores, with replication carried out via reverse transcription by the viral reverse transcriptase (RT) of viral ribonucleic acid (RNA) into HBV DNA (5). The primary CHB treatment agent is a nucleoside or nucleotide analogue (NA), which is a competitive inhibitor of viral RT, acting to inhibit viral replication and thereby prevent the downstream production of new virions and damaging HBV proteins. However, NA agents have no impact on cccDNA, so treatment is suppressive rather than curative, and has to be taken long term.

1.1.2 Diagnosis

Diagnosis of CHB is made via detection of the serological presence of viral markers, sustained over an extended period of time (8). Robust diagnosis based on National Institute for Health and Care Excellence (NICE) guidelines (8) requires serum detection of HBV surface antigen (HBsAg) or viral DNA (herein referred to as viral load, VL) on two measurements taken ≥6 months apart, in order to identify persistently detectable HBsAg/VL in the blood. Both the presence and magnitude of serum HBsAg and VL concentrations provide valuable insights regarding clinical characteristics and CHB disease progression (9). However, in settings where longitudinal measurement and/or quantification of viral markers is not possible, single antigen tests indicating the binary presence of serum HBsAg are regarded as sufficient to diagnose and monitor HBV (10–12).

1.1.3. Treatment

NA agents are the mainstay pharmacological agents deployed in the treatment of CHB, with use of specific NAs recommended in differing clinical scenarios . Treatment is recommended for a minority of people living with HBV who meet one of multiple sets of eligibility criteria set out by different national and international guidelines including the NICE (8) the European Association for the Study of the Liver (EASL) (13), the American Association of the Study of Liver Disease (AASLD) (14) and Asian Pacific Association for the Study of the Liver (APASL) (15). Treatment eligibility is determined by various factors including age, VL, hepatitis e antigen (HBeAg) status and serum alanine aminotransferase (ALT) levels, alongside imaging with elastography (to determine a fibrosis score) and ultrasound (to identify fibrosis or cirrhosis, and to screen for HCC). These factors are monitored via routine assessment in outpatient clinics (typically secondary care services led by infection, hepatology, or sexual health clinicians). As such, determination of treatment eligibility is possible only in individuals who have received a CHB diagnosis and referral to, and ongoing follow up in specialist clinics.. Therefore, it is likely that some individuals with CHB require treatment but are currently not receiving a NA regimen due to lack of CHB diagnosis or unknown treatment eligibility attributable barriers to care access at outpatient clinics (16).

1.1.4. Prevention

There are multiple strategies for primary HBV prevention (i.e., the minimisation and/or elimination of new incident infections) underway, including vaccination, universal antenatal screening, surveillance actions around HBV cases/clusters, and healthcare initiatives (infection control policies for healthcare settings and healthcare worker clearance) and harm reduction strategies for people who inject drugs (PWID), amongst others (17–23). Vaccinations can be employed both on a population-wide scale to reduce incidence or, where resources are constrained to deploy such a strategy, selective vaccination can be used in certain scenarios in which risk of exposure is high (for example in healthcare workers, high risk travellers, sexual partners of people with chronic infection, and to prevent perinatal mother-to-child transmission for which a timely birth-dose of the vaccine is required) (24). Infant vaccination is the most efficacious means of primary prevention (24). Uptake of childhood HBV vaccination has increased in recent years, whereas infant birth-dose vaccination coverage has been demonstrated a slower increase (24). In 2017, the UK introduced universal infant HBV vaccination (as one component of a hexavalent vaccine including diphtheria/tetanus/acellular pertussis/ inactivated polio vaccine/Haemophilus influenzae type b/hepatitis B (DTaP/IPV/Hib/HepB)) (25). HBV vaccine coverage has been explored by UKHSA (26–30), however further description of neonatal vaccine coverage utilising electronic health records, particularly for vulnerable subgroups and in light of the COVID-19 pandemic and recent changes in vaccine policy, is warranted to both appraise information available in electronic sources and investigate the influence of policy influences on vaccine coverage.

The NHS in England operates the Infectious Diseases in Pregnancy Screening (IDPS) programme to systematically identify and treat pregnant women for human immunodeficiency virus (HIV), syphilis and HBV in order to prevent vertical (mother-to-child) transmission (17–23). The IDPS includes a systematic surveillance programme, the Integrated Screening Outcomes Surveillance Service (ISOSS) (31), which monitors and appraises IDPS outcomes. The IDPS targets each pregnancy for all pregnant women in England. Pregnant women who screen HBV positive are

followed up by a screening team and accordingly referred to specialist care. All screen positive women also contribute blood samples from both initial test date and delivery to the national Virus Reference Department. This serves to determine maternal infectivity risk (via viral load and HBeAg testing (32)) and inform future interventions to prevent vertical transmission, detect mutant HBV strains and monitor factors associated with vertical transmission. If pregnant women are not already receiving antiviral treatment at time of positive HBV screen, treatment initiation is recommended in those for whom viral load is ≥ $2x10^5$ iu/ML (33). Infants born to HBV positive mothers should receive one birth dose administered within 24 hours of delivery, and 3 subsequent doses at 4 weeks, 8 weeks and one year of age), in addition to routine universal infant HBV vaccination programme dosages which commence at age 8 weeks (25). Those born to mothers with higher infectivity markers (HBsAg positivity, viral load is ≥ $1x10^6$ and negative anti-HBe) also receive HBIG administration (25). All infants born to HBV-positive mothers should be tested for HBV infection at one year of age, and infants born to mothers classified as high infectivity are tested at delivery (32,34) Comprehensive guidelines detailing antenatal HBV screening and selective neonatal immunisation practices are available to inform health care professionals (32).

The UKHSA also produces comprehensive guidance on the management and clearance of BBVs, including HBV, in healthcare workers (HCWs) (23). These guidelines provide details on how risk of BBV transmission to HCW from patients and/or colleagues may be minimised. They also include evidence-based recommendations on how to ensure HCWs living with BBVs may be retained in the workforce and how stigma may be reduced.

HBV surveillance in England is comprised of data fed from multiple systems including vaccine monitoring systems, tests undertaken in primary and secondary care, test undertaken to screen blood and tissue donors and recipients, and specialised surveillance systems for high-risk subgroups. Surveillance data on HBV testing and diagnosis is important in both sub-group and population-wide contexts in order to monitor prevention and diagnosis of individuals at risk of HBV infection. HBV surveillance utilises monitoring data from both Cover of vaccination evaluated rapidly (COVER) data collection for routine childhood vaccination coverage, and monitoring data from high-risk infants undergoing selective vaccination (ImmForm) (35).

Vaccination surveillance data informing COVER and ImmForm come from Child Health Information Systems (CHISs) and/ or from GPs across local authorities (36). HBV testing and diagnosis data comprise various surveillance systems including Second Generation Surveillance System SGSS) laboratory reporting of incident HBV (37), sentinel surveillance of BBV testing (SSBBV) (30), joint NHS Blood and Transplant (NHSBT) screening of blood and tissue donors and transfusion recipients (20,38), BBV testing regimes in secure and detailed settings and unlinked anonymous monitoring of PWID.

Incident HBV diagnoses ascertained from laboratory data are captured by SGSS in England, and while the numbers of individuals HBsAg tested and positive for HBV in the general population, primary care and emergency departments originate from SSBBV. The numbers of individuals tested and positive for HBV in drug services, prisons and sexual health services are also ascertained from SSBBV, as is the proportion of tested individuals positive for additional HBV markers including VL. The joint NHSBT and UKHSA Epidemiology Unit also monitors blood, tissue and organ donors for HBV infection, as well as transfusion recipients.

Infections with blood borne viruses (BBVs) such as HBV pose a substantial health harm risk in the population subgroup of PWID. HBV maybe transmitted by sharing of injecting equipment. PWID are vulnerable individuals who often face socioeconomic structural barriers to access healthcare services, and therefore prevention of HBV transmission is of particular importance. This is accomplished via promotion of HBV vaccination (particular in PWID of younger age) in addition to access to sterile injection equipment, opioid agonist therapy and wound care services (39), A comprehensive strategy that combines both person-centered and whole-system approaches to HBV and other BBV prevention in PWID is utilised in England to work towards minimising the HBV burden in PWID.

1.2. Global and Regional (UK) disease burden of HBV

The attributable disease burden of HBV, both globally and within the UK, is substantial. Estimating both global and regional prevalence of chronic infection is difficult, given the pathophysiology of chronic infection and the patterns by which infection is most concentrated in certain population subgroups who may not be well represented by clinical and public health data (16). Therefore, variable population prevalence estimates have been reported to date for the UK. Furthermore, slow progress has been made on reducing CHB morbidity and mortality burdens in the country, and on meeting targets set in the WHO's Global Health Sector Strategy (GHSS) on viral hepatitis produced in 2016.

1.2.1. Prevalence estimates

The 2022 Global Burden of Disease analysis of HBV epidemiology estimates that 316 (95% confidence intervals (CIs) 284 to 351) million people in total live with CHB worldwide, translating to an all-age prevalence of 4.1% (95% CI 3.7% to 4.5%) (40). Highest 2019 prevalences of infection were reported in the Western Pacific and African regions, with all-age estimates of 7.1% (95% CI 6.3% to 7.9%) and 6.5% (5.8% to 7.3%). The lowest regional prevalence reported was 1.1% (95% CI 1.0% to 1.2%) in the European region. From 1999 to 2019 there was a 31.3% (95% 29.0 to 33.9) decline globally in all-age prevalence. Out of 24 European countries identified as "high income" in the analysis, the UK was one out of five countries (along with Denmark, Germany, the Netherlands, and Switzerland) in which the all-age prevalence did not change between 1999 and 2019, whilst prevalence declined in the remaining 19 countries over the same time period.

CHB prevalence estimates have not been robustly produced in many settings, including the UK (41), for multiple reasons. One reason is a lack of systematically collected HBV serology data from study samples that are generalisable to national-level populations. No disease register currently exists for CHB in the UK. HBV infection is a notifiable disease in the UK, and various data sources are utilised to monitor cases including laboratory testing services, some hospital data sources (including hospital episode statistics (HES)) and registries (national cancer and death registries) (30).

However, a mathematical model taking account of these various data sources to produce an informed CHB prevalence estimate has yet to be undertaken. Additionally, due to the sub-clinical nature of CHB illness during early stages of chronic infection, most infected individuals present in the late stages of symptomatic disease and therefore the majority of cases remain undetected (1). Furthermore, in many settings, CHB populations have not been well-characterised and therefore tailored surveillance of at-risk demographics and subgroups, which might inform prevalence estimation, have not been undertaken. Incorporating migration trends into prevalence estimations has proved challenging in previous investigation (41,42). HBV prevalence measurements in migrants across various European countries have been found to be unrepresentative of country-wide migrant subgroups in these nations and therefore insufficient for inclusion in prevalence estimation. In order to incorporate migration trends in estimations, multipliers have previously been utilised to estimate population-wide prevalence which account for both i) the frequencies of migrant counts by country of origin, and ii) the prevalence of HBV in each given country of origin (42).

The existence of disparate prevalence estimates highlights the absence of robust investigations available to inform HBV population characterisation, health resource allocation, primary and secondary HBV prevention and treatment in the UK. Amongst the limited number of previous investigations reporting population-wide prevalence estimates, most fail to describe the population subgroups in which the CHB disease burden is concentrated. Estimates stratified by age, sex, and/or ethnicity are available (30), however further stratifications by relevant characteristics are omitted. Such relevant characteristics include co-infection with other blood borne viruses (particularly hepatitis C virus (HCV) and HIV), comorbid disease (including diabetes mellitus, metabolic and fatty liver disease (MAFLD), steatotic liver disease (SLD), alcohol-related liver disease (ARLD)), and demographics (including ethnicity, country of birth and socioeconomic deprivation). Additionally, the percentages of infected individuals who are eligible for and/or receiving antiviral treatment are rarely reported. Regional reports produced by the UK Health Security Agency (UKHSA, previously Public Health England) broadly describe CHB epidemiology overall and by population subgroups include sex, ethnicity and by age group. However, previous reports have included neither overall estimates of total individuals with CHB eligible for and/or on treatment, nor proportions eligible/treated stratified by relevant characteristics (43).

The most recent prevalence estimates produced by the UKHSA are proxy estimates for the general population which have been obtained from samples of pregnant women attending antenatal services. ISOSS reporting has found uptake of HBV screening in antenatal services to be >99% (31). All pregnant women attending antenatal care are screened for HBV, as recommended by the National Screening Committee (NSC) (32). As >95% of women attend antenatal services in the UK, prevalence estimates obtained from such services are often generalised to the wider population. Other CHB prevalence estimates have been reported for the UK. The Global Burden of Disease analysis reported an all-age 2019 population prevalence of 0.7% (95% CI 0.6% to 0.8%) for the UK (40). Similarly, a population prevalence of 0.7% (95% CI 0.5% to 0.9%) was reported by the Polaris Institute, which undertook a modelling study that included expert interviews and data from a literature review to inform their estimate (44). The most recent HBV in England from UKHSA reported a 0.45% (95% CI 0.35% to 0.60%) 2023 (30). The most recent HBV report produced by the British Liver Trust, published in 2017, stated there were 180,000 people with CHB in the UK, translating to a population prevalence of 0.27% (45); this estimate referenced a 2002 Department of Health publication which is not available for download. Cohort studies undertaken in primary care electronic health record (EHR) databases in the UK have reported prevalence estimates of <0.10%, likely owing to the poor representation of CHB data in primary care data (46,47).

Estimating the prevalence of CHB is difficult for multiple reasons. Diagnosis requires serum detection of HBsAg or VL on two measurements taken six months apart. Such detection assays are not performed on a population-wide basis in most countries, and therefore it is difficult to comprehensively identify (and thereby treat) CHB on a large-scale basis. VL assays are usually expensive and rapid detection methods not yet widely (42,48), but are becoming increasingly available for estimating the prevalence of HBV viral markers in various settings (10–12). Whilst useful to provide binary indicators of HBsAg positivity, such rapid tests do not capture quantification of HBsAg which may be an important biomarker for risk stratification. Additionally, patterns of HBV infection sequalae vary widely across different environments and settings, and therefore screening approaches must often be tailored to specific populations. Incorporating of migration trend data into models estimating population-

wide HBV prevalence has previously been challenging . This is because measurements of HBV prevalence measurements in migrant sub-populations across various European countries have not been representative of true population trends. Therefore, these trend data could not be included in estimation models, and instead multipliers have been used to account for frequencies of migrant counts by country of origin (and prevalence of HBV in each given country).

Although the UK is regarded as a low-prevalence country, CHB carries substantial disease burden as infection is thought to be concentrated in vulnerable population subgroups including individuals who belong to black and other non-white ethnicities, people who inject drugs, and are socioeconomically deprived (16,39,47). In order to quantify the attributable disease burden in these demographics, existing data sources may be harmonised and leveraged more efficiently in order to utilise their strengths. More specifically, usage of EHR databases is one means by which evidence gaps concerning CHB epidemiology in the UK may be filled. This is because EHR databases contain real-world data from individuals who utilise health services in the UK, therefore allowing for studies to be undertaken in a sampling reference frame which is generalisable to the wider population.

1.2.2. Morbidity and mortality burdens

HBV is responsible for a large morbidity burden worldwide. HBV is the sixth largest contributor to disability-adjusted life years (DALYs) amongst all infectious diseases globally (49), accounting for 18.2 million (95% CI 15.9 to 20.6) DALYs in total (or 219.9 (95% CI 192.9 to 247.8) per 100,000 persons) in 2019 (50). The disease burden associated with CHB is mostly attributed to the end-stages of disease, namely cirrhosis, HCC and hepatic decompensation. Therefore, estimates account for these disease endpoints when quantifying attributable mortality.

CHB is currently the single largest cause of HCC death worldwide (51), and the third largest cause of cirrhosis death (40). It is estimated that HBV-related illness was responsible for 555,000 deaths globally (95% CI 487,000 to 630,000) in 2019 (40). In recent years, crude numbers of attributable deaths have increased whilst age-

standardised death rates for HCC and cirrhosis have decreased (40,51). This is due to both increases in population size, and in the number of CHB adults aging and developing clinical end-stage disease. Prevention and treatment efforts must therefore be scaled up and/or resources accordingly modified in order to address the increasing numbers of total illnesses and deaths.

In the UK, the attributable disease burden of CHB is substantial. From 2017 to 2019, the rate of premature deaths attributable to CHB was 0.13 (95% CI 0.11 to 0.15) per 100,000 population aged <75 years (52). The Global Burden of Disease analysis reported increases in HBV-related death rates both historically and in recent years. Specifically, percentage increases of 110.1% (99.5% to 126.9%) and 1.5% (-0.7% to 3.6%) were observed for all-age death rates from 1990 to 2019 and from 2015 to 2019, respectively (40). This contrasts with percentage decreases of -25.8% (-32.8% to -17.6%) and -0.9% (-3.6 to 2.1%) observed over the same two time periods observed across Western Europe. It is estimated that up to 90% of liver disease cases in England are attributable to four preventable factors, one of which is HBV (with the others being HCV, alcohol consumption and obesity) (52). Furthermore, given that the disease burden of CHB is thought to sequester in vulnerable population subgroups in whom relevant disease endpoints may often go undetected, these estimates of the attributable disease burden are likely underestimations of the true effect.

The WHO's GHSS on viral hepatitis was adopted by the World Health Assembly in 2016, with an updated version published in 2022 (24,53). The report's key aim is the elimination of HBV as a public health problem by 2030, defined specifically as a 95% reduction in HBV incidence and 65% reduction in mortality. Various recommendations for achieving desired targets and reducing the CHB disease burden were made in the GHSS, including strategies to prevent secondary prevention of CHB progression to severe disease endpoints (**Figure 1.1**). These include improving access to and quality of diagnostic, laboratory, prevention and treatment services. Reaching elimination targets requires an understanding of HBV epidemiology and attributable liver disease to focus resources to vulnerable demographics, and track progress. In order to begin making progress towards achieving the GHSS targets, the CHB population in UK must first be robustly characterised, as must the burden of end-stage liver disease in individuals with CHB within the country.

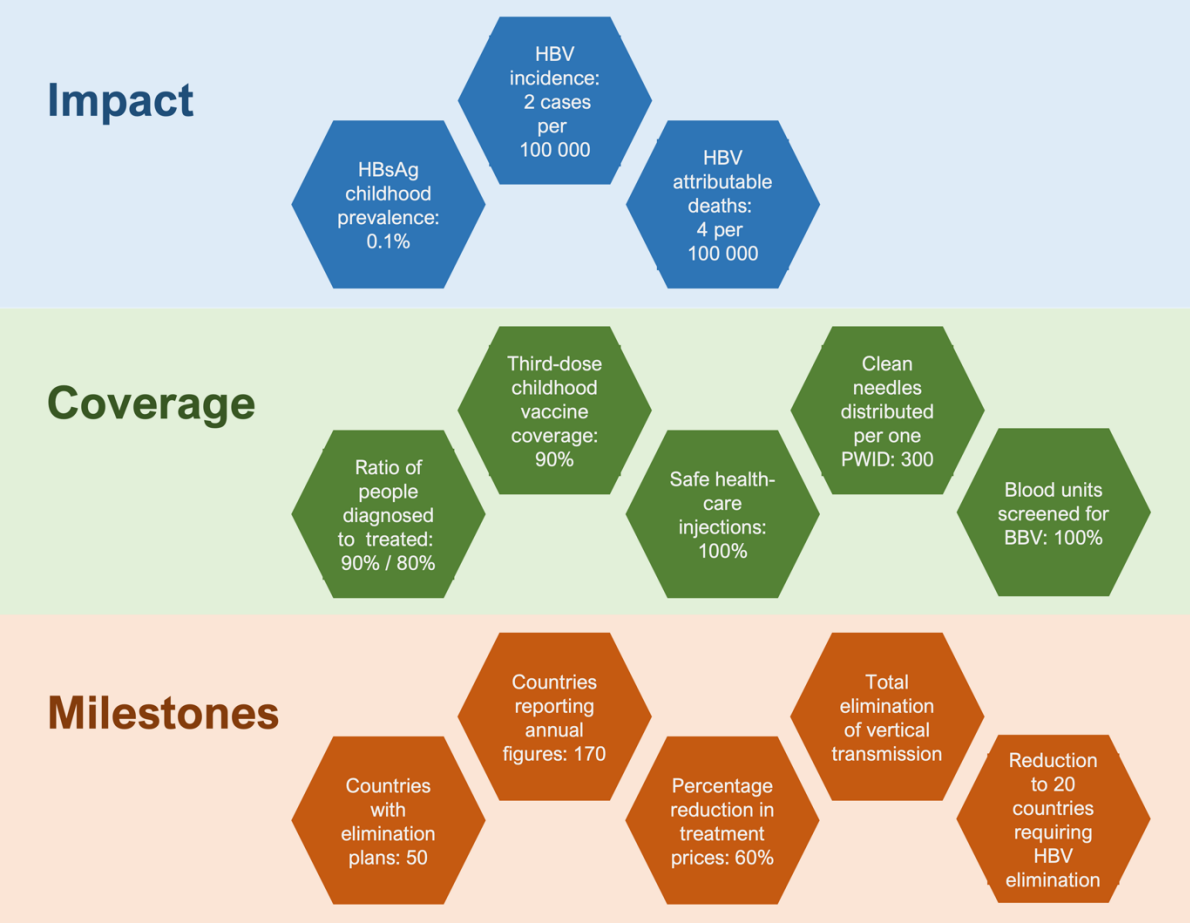


Figure 1.1. Key hepatitis B virus targets to be achieved by 2030 outlined in the World Health Organisation's Global Health Sector Strategy.

1.3. Electronic health record (EHR) databases

Electronic health record (EHR) databases have become a vital data source for various medical research fields, particularly clinical epidemiology, over the last decade. They contain a wealth of individual-level patient data pertaining to various parameters including clinical characteristics, demographics and healthcare utilisation. Information is stored in electronic information systems using clinical coding systems, with unique codes existing for a plethora of conditions and characteristics ranging across disease status, ethnicity, age and geographical location. These data characteristics, along with others which are discussed thereafter, make EHR databases unique and information-rich sources in which epidemiological research can be undertaken. However, primary and secondary care EHRs differ in the coding nomenclatures used, and in the computer software used to log and store information. EHR coding systems have also

not been purpose-built for research purposes; they were originally created to facilitate National Health Service (NHS) finance and billing practices. For the purposes of this doctoral project, “big data” was defined as the parameters collected and stored in primary and secondary care EHR databases as part of routine everyday healthcare practices, as well as external data sources which can be linked to these databases. Databases used in the presented analyses were those “big data” which were deemed suitable to investigate the research questions of interest for this project. These data are not collected for research purposes. One of the overarching aims of this doctoral project was to utilise these data for research purposes and comment on the capacity for them to be used as such.

1.3.1. Primary care databases

Primary care is defined as the first point of contact service-users have with the NHS when initiating access to a care pathway. In the UK, primary healthcare services consist mainly of those provided by general practices (GP), along with additional services such as sexual health outpatient clinics, community pharmacists and some emergency services (54). Utilisation of GP services is high across with country, with >90% of the population registered with a GP (55). General practitioners are the “gatekeepers” of access to specialist services and treatment available in secondary and tertiary care centres, whereby a referral from general practice is the main means of accessing such services.

Data access is under control of GPs at a given practice. Practices opt-in to contribute de-identified patient-level data to EHR databases for research purposes (56). However, patients may opt-out of their information being shared for such purposes (57). Most (>90%) general practices employ primary care EHR IT systems which are provided by The Phoenix Partnership (TPP), Egton Medical Information Systems (EMIS), SystmOne or Vision (58,59). Patient information is stored in primary care EHRs using Read Codes, which enable coded identification of clinical conditions (60). Data from primary care EHR databases can often be linked to additional data sources including hospital admission and outpatient data (hospital statistics data (HES)), HES A&E codes and attendance, vital statistics/Office for National Statistics (ONS) data,

national cancer registry data and local area deprivation measures, amongst others (61–63). Linkage is accomplished using a combination of attributes to construct a unique identifier comprised variables including date of birth, sex and NHS ID.

1.3.1.1. QResearch

The QResearch primary care database is currently the largest primary care database in the UK. Established in 2003, it contains individual-level anonymised patient records from >35 million patients registered across more than 1500 general practices, out of approximately 9,800 which exist throughout the country (64) (**Table 1.1**). QResearch has comprehensive geographical coverage, with inclusion of practices from every health authority in England (62,64)**.** QResearch data span the last 25 years, as some data were retrospectively obtained which predate establishment of the database. Practices employing EMIS or SystmOne software are eligible to contribute to QResearch. The database was established by Professor Julia-Hippisley-Cox and EMIS Health.

Table 1.1. Summary of UK primary and secondary care electronic health record databases utilised for completion of doctoral studies.

Database name (year established)	**Healthcare type**	**Size (individual patients)**	**Parameters available**	**Available linkages**
QResearch (2003)	Primary care	>35 million	Clinical measurements Consultation information Demographics Diagnoses Hospital referrals Prescriptions Registration dates Selected laboratory tests	ONS mortality data HES records National cancer registries COVID-19 testing and vaccination

CPRD (1987)	Primary care	18 million*	Demographics Diagnoses Drug exposures Laboratory tests Hospital referrals Registration dates Specialist care information Symptoms Vaccination history	ONS mortality data HES records National cancer registries Local area demographic data
NIHR HIC (2014)	Secondary care	5991 †	Biopsy reports Demographics Deprivation scores Diagnostic imaging Healthcare utilisation ICD-coded diseases/outcomes Laboratory test information Prescription/treatment information	Not applicable

* Currently registered in the database.

† Patients with chronic hepatitis B virus infection.

ONS, Office for National Statistics; HEs, Hospital Episode Statistics; COVID, Coronavirus disease; CPRD, Clinical Practice Research Datalink; NIHR, National Institute for Health Research; HIC, Health Informatics Collaborative; ICD, International Statistical Classification of Diseases and Related Health Problems.

Only anonymised coded data are uploaded for research use. Ethics approval for the QResearch database has been granted by the East Midlands Derby Research Ethics Committee (reference 18/EM/0400). Available parameters include year of birth, sex, registration dates, date of death, deprivation scores, ethnicity, referrals, consultations, prescriptions, clinical values (e.g. blood pressure), test results (e.g. full blood count) and clinical problems/events such as diagnoses (62,64). Analysis is undertaken for affiliated projects (approved by the scientific steering committee) via a secure remote desktop. Since 2011 data have been linked to ONS mortality data, HES records from secondary care and national cancer registries; more recently, linkage to Coronavirus Disease 2019 (COVID-19) testing and vaccination data has become possible. Researchers access the database via a secure server and remote virtual desktop. Costs of accessing the QResearch database vary based on the size and scale of individual projects, with the minimum cost for one researcher, one project, for one year with GP data only and up to one million records being £30,000 (65).

1.3.1.2. Clinical Practice Research Datalink

The Clinical Practice Research Datalink (CPRD) is a primary care EHR database established in 1987 in the UK containing data from general practices utilising EMIS or Vision IT software (66,67). CPRD contains data from 60 million patients of which 18 million are currently registered with the database. Data are collected from across 2000 general practices, and 25% of patients have ≥ 25 years of follow-up (67).

A minimum set of parameters are available for all patients in CPRD. These include demographic characteristics, diagnoses and symptoms, drug exposures, vaccination history, laboratory tests, referrals to hospital and specialist care information (67). Linkage to additional data sources is available for some patients and is dependent on the IT system employed by their general practice. Examples of linked data sources include HES (Accident and Emergency; Inpatient; Outpatient; Imaging), mortality information from ONS via linkage to the national death registry, national cancer registry and treatment registers, mental health services, socio-economic measures, mother-baby link and pregnancy register algorithms (61).

1.3.2. Secondary care databases

In the UK, secondary care consists of specialist services and treatments which are typically delivered in hospitals via inpatient and outpatient facilities. Services are access via urgent presentation to Accident and Emergency, or via a referral from a general practitioner. Secondary care service delivery and governance is the responsibility of a regional NHS “Trust”, of which there are 25 across the UK (68). Patient data are entered into secondary care EHR IT systems using clinical codes belonging to the Office of Population Censuses and Surveys (OPCS) Classification of Surgical Operations and Procedures code family (69). Data collection and management is controlled at the Trust level. Therefore, there is variation across Trusts in regard to which clinical parameters collected, how characteristics are coded, and how clinician free-text notes are processed.

1.3.2.1. NIHR Health Informatics Collaborative

The National Institute for Health Research (NIHR) Health Informatics Collaborative (HIC), established in 2014, is a "programme of infrastructure development" across the 32 NHS Trusts and health boards (68,70). The programme aims to improve the quality of and access to routinely collected secondary care EHR data for equitable research purposes. Different research themes are hosted across the HIC in respective NIHR Biomedical Research Centres, with the Oxford University Hospitals NHS Foundation Trust hosting the viral hepatitis theme for the HIC database led by Professor Ellie Barnes. At present, 10 NHS Trusts participate in the viral hepatitis theme (71).

Participating NHS Trusts contribute data using a standardised data model. Various parameters are available at individual-level, such as demographics (including year of birth, gender, ethnicity), laboratory test information (including type of test, date of test, numerical results and units), prescription/treatment information, diagnostic imaging (including elastography, ultrasound, computed tomography (CT) and magnetic resonance imaging (MRI)), biopsy reports, healthcare utilisation, deprivation scores and corresponding International Statistical Classification of Diseases and Related Health Problems (ICD) codes for diseases/outcomes. Death information is obtained via linkage to ONS mortality statistics.

1.3.3. EHR-based versus traditional observational cohorts

There are various differences between traditional prospective cohort studies for which participants are actively recruited at baseline, and retrospective EHR-based cohort studies. Data obtained from primary care EHR databases may be more representative of the whole population than traditional prospective cohort samples, whereby individuals who enrol in a cohort may systematically differ to the desired study population and thereby introduce selection bias. Individuals often remain registered with the same general practice for years, and therefore obtaining long durations of follow-up is feasible in EHR-based studies. Additionally, if patients switch general practices, they may switch to another practice which also contributes data to the EHR database. Passive follow-up in EHR-based studies allows for efficient ascertainment

of relevant outcomes. Endpoint identification is also often robust, due to the frequent linkage of EHR databases to vital statistics and cancer registries. Additionally, patients' characteristics (such as blood pressure and serum biomarkers) may be measured more frequently in clinical practice, and therefore EHR data may provide a more accurate characterisation of such measures. This is especially true considering that commonly in prospective cohorts due to resource, infrastructure and feasibility constrains, sample sizes are constrained, or in large studies data collection may be infrequent, with a minority of participants being resurveyed throughout follow-up.

1.4. HBV and EHR – current evidence gaps

International targets calling for the elimination of HBV infection as a public health threat by the year 2030 have been set by the WHO's GHSS on viral hepatitis (53). The aims include substantial reductions in HBV incidence and attributable mortality. One primary aim is the elimination of HBV as a public health threat by 2030. Although substantial progress has been made towards these goals, further investigation is necessary to actualise these targets and benchmark progress.

To reduce attributable disease burdens and achieve HBV elimination, the HBV population in the UK must be well characterised. To date this has not been achieved. Improving quantification and description of the disease population would provide data which may be used towards generating a robust population wide CHB prevalence estimate. It would also provide insight into the disease burden in under-represented population subgroups in whom the HBV burden is concentrated. This is true in Western multi-ethnic populations with immigration from high HBV prevalence countries, as migrants tend to suffer disproportionately from multimorbidity and lack of access to health services. It is also true for sub-Saharan African populations in which the prevalence of CHB is high (40), but the infected population has not been well characterised.

Progress towards the GHSS on viral hepatitis targets also requires improved identification and validation of risk factors for the progression of CHB to disease

endpoints, namely HCC. Existing risk scores (including PAGE-B (72), REACH-B (73,74), GAG-HCC (75,76) and CU-HCC (77–79), use multiple factors to predict HCC risk. However, a minority of the CHB cohorts from which these scores have been derived have been identified diverse populations. Therefore, characterisation of ethnically diverse study samples is required not only to validate or inform existing risk scores, but also to identify novel risk factors that may be unique to under researched populations. This will ultimately improve secondary prevention of CHB progression to end-stage disease outcomes.

Improvement of vaccination and treatment coverage and accessibility are among the primary goals of the GHSS on viral hepatitis. However, in order to achieve these targets, some basic parameters concerning the UK population must first be robustly quantified. Overall uptake of population-wide neonatal HBV vaccination has been reported in the most recent annual report on HBV in the UK (30), but coverage stratified by relevant demographics requires further characterisation. Additionally, the proportions of CHB patients who are both eligible for treatment and who are currently on treatment must be robustly quantified. Utilising EHR databases in the UK for HBV research could serve to identify gaps in vaccination and treatment coverage and provide evidence for targeted strategies.

The COVID-19 pandemic affected clinical management of various diseases, and likely impacted corresponding disease elimination targets. Routinely collected data in EHR databases may be used to quantify the influence of the pandemic on CHB prevention. This includes both primary prevention (how population-wide vaccinations were impacted) and secondary prevention (how surveillance and treatment delivery in this with CHB at risk of progressing to end-stage disease was impacted. As vaccination and treatment information are comprehensively captured in primary and secondary care EHR databases, respectively, the impacts of COVID on both elimination strategies may be efficiently investigated using such data sources.

In order to determine to what extent EHR may fill current evidence gaps and make sufficient progress towards the goals outlined the GHSS on viral hepatitis, research must be undertaken in databases of routinely collected data. This is because EHR databases are comprised of real-world data that provide a uniquely accurate depiction

of current clinical practice, whereby it is possible to identify subgroups who are receiving adequate clinical care and who are currently neglected. Such databases represent an efficient and equitable means for answering major research questions and potentially addressing evidence gaps with robust findings.

1.5. Aims and objectives

The overarching aim of my PhD project is to leverage electronic health records in the UK to provide evidence that can support interventions to reduce the attributable morbidity and mortality burdens of HBV, and to work towards eliminating HBV as a public health threat. My specific objectives are to:

- Summarise and critically appraise previous literature to date reporting risk factors for progression of CHB to HCC.
- Improve characterisation of the CHB population in the UK by identifying and describing CHB cohorts in EHR databases.
- Identify novel HCC risk factors, and validate previously reported risk factors, in a study sample originating from an ethnically diverse sampling population.
- Quantify treatment eligibility and coverage across the UK CHB population.
- Quantify how the COVID-19 pandemic has impacted HBV vaccination and surveillance strategies.
- Quantify how changes in infant vaccination policy, along with the start of the COVID-19 pandemic, have impacted uptake of the complete HBV infant vaccination schedule.

2. Systematic review and meta-analysis to investigate risk factors for HCC in CHB

2.1. Availability of content

I herein present findings from a systematic literature review and meta-analysis undertaken to investigate risk factors for the development of HCC in CHB. This investigation is published in the *Journal of Viral Hepatitis* in December 2020 (80) and has obtained 28 citations at the time of writing this thesis.

2.2. Background

CHB is responsible for substantial morbidity and mortality worldwide. Most CHB deaths are due to primary liver cancer and cirrhosis; these conditions were responsible for over 40% of all viral hepatitis-attributable deaths. A GBD study on the global HCC burden reported a 42% increase in incident cases of HCC attributable to chronic infection between 1990 and 2015, among which CHB infection was the largest contributor, responsible for more than 30% of incident cases in 2015 (51).

Multiple risk factors for HCC in CHB-infected individuals have been established, including male sex, older age, cirrhosis, and co-infection with HIV or other hepatitis viruses (including hepatitis C and D). Previous studies have investigated associations of comorbidities, such as diabetes mellitus (DM) (81–84) and hypertension (HT) (85,86), with risk of HCC in the general population, and The European Association for the Study of the Liver (EASL) recognises DM as a risk factor for HCC in CHB (13). As the global prevalence of comorbidities such as DM (87), renal disease (88), HT (89), coronary heart disease (CHD) (90), non-alcoholic fatty liver disease (NAFLD, also referred to as metabolic and fatty liver disease (MAFLD)) (91–94) continues to rise, these conditions are increasingly relevant to the development of HCC.

Various risk scores have been developed to predict HCC risk: the PAGE-B risk score was developed to predict HCC risk in Caucasian patients on antiviral treatment (90), while the REACH-B (73,74), GAG-HCC (75,76) and CU-HCC (77–79) risk scores

apply to untreated Asian populations. Existing risk scores use age, sex and VL to predict risk. CU-HCC and GAG-HCC include parameters for cirrhosis, however no score accounts for comorbidities such as DM or HT. It is possible that risk prediction could be improved by accounting for these comorbid conditions.

Despite rises in the global prevalence of relevant comorbidities, evidence concerning associations of comorbidities with HCC risk in CHB is poor and heterogeneous. Neither EASL, American Association of the Study of Liver Disease (AASLD) (14) or Asian Pacific Association for the Study of the Liver (APASL) (15) guidelines for HBV management include recommendations for enhanced screening or DM management in CHB. However, there has been recent clinical interest in the potential utility of metformin in preventing and treating various cancers (95,96). Furthermore, there are few studies investigating associations of other potentially relevant comorbidities (such as HT, CHD and renal disease) and their metabolic risk factors (such as obesity and dyslipidaemia) with HCC risk.

2.3. Research question

I undertook a systematic review and meta-analysis to summarise and critically appraise studies investigating associations of relevant comorbidities and metabolic factors with risk of HCC in CHB-infected individuals.

2.4. Methods

2.4.1. Search strategy and selection criteria

In June 2020 I systematically searched three literature databases (Web of Science, Embase and MEDLINE) in accordance with PRISMA guidelines (97) (**Table S1**); search terms are listed in **Table 2.1**. Databases were searched from 1st January 2000 until 24th June 2020, without application of any restrictions for study design applied to search terms or results but including only full-text human studies published in English.

Search results were combined and deduplicated from the three databases prior to screening for eligibility. Articles not investigating associations of comorbidities with risk of HCC and/or not restricted to CHB-infected participants were excluded. Reference lists of relevant systematic reviews/meta-analyses and eligible studies were searched to identify additional studies for inclusion.

Table 2.1. Search terms used in systematic search of Embase, MEDLINE and Web of Science databases.

Criteria	Database	Search terms
Population	MEDLINE and Embase (searched via Ovid)	("hepatitis b" or "hbv" or "hepatitis").ab.s
	Web of Science	AB=("Hepatitis B" or "hbv" or ("hepatitis" and "b"))
Exposure	MEDLINE and Embase (searched via Ovid)	("diabetes mellitus" or "diabetes" or "hypertensi*" or "hyperlipidaemia" or "hyperlipidemia" or "dyslipidaemia" or "dyslipidemia" or "obes*" or "overweight" or "osteoporosis" or "osteopenia" or "osteopaenia" or "renal disease" or "kidney disease" or "vitamin d" or "comorb*" or "comorbidit*").ab.
	Web of Science	AB=("diabetes mellitus" or "diabetes" or "hypertensi*" or "hypertensi*" or "hyperlipidaemia" or "hyperlipidemia" or "dyslipidaemia" or "dyslipidemia" or "obes*" or "overweight" or "osteoporosis" or "osteopenia" or "osteopaenia" or "renal disease" or "kidney disease" or "vitamin d" or "comorb*" or "comorbidit*")
Outcome	MEDLINE and Embase (searched via Ovid)	("hepatocellular carcinoma" or "hcc" or "hepatoma" or ("hepatocellular" and "carcinoma")).ab.
	Web of Science	AB=("hepatocellular carcinoma" or "hcc" or "hepatoma" or ("hepatocellular" and "carcinoma"))
Final search	MEDLINE and Embase (searched via Ovid)	(("hepatitis b" or "hbv" or "hepatitis").ab.) AND (("diabetes mellitus" or "diabetes" or "hypertensi*" or "hyperlipidaemia" or "hyperlipidemia" or "dyslipidaemia" or "dyslipidemia" or "obes*" or "overweight" or "osteoporosis" or "osteopenia" or "osteopaenia" or "renal disease" or "kidney disease" or "vitamin d" or "comorb*" or "comorbidit*").ab.) AND (("hepatocellular carcinoma" or "hcc" or "hepatoma" or ("hepatocellular" and "carcinoma")).ab.)
	Web of Science	(AB=("Hepatitis B" or "hbv" or ("hepatitis" and "b"))) AND (AB=("diabetes mellitus" or "diabetes" or "hypertensi*" or "hypertensi*" or "hyperlipidaemia" or "hyperlipidemia" or "dyslipidaemia" or "dyslipidemia" or "obes*" or "overweight" or "osteoporosis" or "osteopenia" or "osteopaenia" or "renal disease"

	or "kidney disease" or "vitamin d" or "comorb*" or "comorbidit*")) AND (AB=("hepatocellular carcinoma" or "hcc" or "hepatoma" or ("hepatocellular" and "carcinoma")))

MH, medical subject heading term.
Search results were restricted to human studies published in English from 2000 to 2020.

2.4.2 Data extraction and statistical analysis

I extracted the following summary characteristics from included studies: country, publication year, study design, follow-up period, comorbidities investigated, number of participants, number of HCC cases, sex, age at baseline, risk ratio and covariates adjusted for.

I performed meta-analysis in R (version 3.5.1) using the "meta" package (version 4.12-0) (98), including only hazard ratios (HRs) minimally adjusting for age and sex. I calculated pooled summary effect estimates using the inverse-variance weighting of HRs on the natural logarithmic scale and quantified between-study heterogeneity using the I^2 statistic; significance of heterogeneity was investigated using Cochran's Q test (p threshold=0.05). Where I^2 was >0 and heterogeneity were significant, both fixed- and random-effects summary estimates are presented. I performed multiple sensitivity analyses whereby analyses were restricted to studies adjusting for various additional confounders and for DM treatment, and stratified by DM type, in order to investigate robustness of observed associations.

2.4.3. Definitions

For DM, diagnoses of type 1 and type 2 diabetes were considered as well as unspecified DM, for pooling the effect, followed by further stratification by subtypes of diabetes if enough studies were eligible. HT was defined by either a diagnosis recorded as part of the medical history or current health assessment, or a

measurement with mean arterial pressure (MAP) above a specified threshold which varied by study. Obesity was based on BMI values, by referring the cut-off in the included studies, where 25, 27, 30 kg/m^2 were the common threshold values used. Cardiovascular disease (CVD) was defined broadly as an umbrella term including any of the following disease subtypes: ischaemic heart disease (IHD)/coronary heart disease (CHD), cerebrovascular disease. Dyslipidaemia was defined according to serum lipid concentrations above a certain threshold (thresholds defined by the healthcare setting in which they were recorded).

2.4.4. Quality appraisal

I used the Newcastle-Ottawa Scale (NOS) to assess the quality of non-randomised studies, including cohort and case-control studies (99), judging studies based on points awarded for selection of study groups, comparability of groups, and exposure/outcome ascertainment. Studies with scores of <5, 5-7 and >7 points were considered to be of low, sufficient and high quality, respectively.

2.4.5. Note on terminology

I note that studies included in this systematic review and meta-analysis utilise the term non-alcoholic fatty liver disease (NAFLD). Nomenclature has recently evolved, with some studies published in the last few years differentiating NAFLD into MAFLD and SLD. Nomenclature used in this chapter reflects that which was used in studies identified for inclusion via systematic literature database search.

2.5. Results

2.5.1. Study characteristics

My total systematic literature search identified 1,814 articles (899 from MEDLINE, 407 from Embase and 508 from Web of Science) (**Figure 2.1**). After deduplication, I screened 1,136 individual articles by title/abstract, from which I identified 140 articles for full-text assessment. After exclusion of ineligible articles and reference list

searching of relevant articles, I identified 40 articles for inclusion in this review. Summary characteristics of included studies are reported in **Tables 2.2 and 2.3**.

Table 2.2. Summary characteristics of included observational studies.

Country, Author (Year)	Study design	Population	Median follow-up period, years	Risk factors investigated	Participants, n	HCC Cases, n	Age at baseline, years	Sex (% male)
UK, Ferreira (2020) (46)	Retrospective cohort study	CHB	8.7§ (DM) 6.2§ (non-DM)	T2DM	3927	16-19	39.2 (mean, non-DM) 54.9 (mean, DM)	50
Korea, Goh (2020) (100)	Retrospective cohort study	CHB patients undergoing liver biopsy from one hospital	7.2	DM (type unspecified), HT, Cholesterol >200mg/dL,	7713	702	50 (statin users) 47 (statin nonusers)	66
Korea, Kim (2020) (101)	Prospective cohort study	CHB patients treatment-naïve at baseline	4.9	DM (type unspecified)	3277	292	48.7 (mean)	63
Singapore, Lim (2020) (102)	Retrospective cohort study	CHB patients undergoing liver biopsy from one hospital	9.3	T2DM	289	27	43.2 (mean, Nonhepatic steatosis) 46.4 (mean, Hepatic steatosis)	72
United States and Asia-Pacific,	Prospective cohort study	Treated CHB patients	5.6§	DM (type unspecified)	5365	378	48.4 (mean)	69

Country, Author (Year)	Study design	Population	Median follow-up period, years	Risk factors investigated	Participants, n	HCC Cases, n	Age at baseline, years	Sex (% male)
Yang (2020) (103)								
Hong Kong, Yip (2020) (104)	Retrospective cohort study	CHB patients with DM	7.1	Thiazolidinedione use, Metformin use, Sulphonylurea use, Insulin use, Statin use, Aspirin/clopidogrel use	28,999	2,655	57.7 (mean)	61
France, Brichler (2019) (105)	Prospective cohort study	Cirrhotic CHB patients	5.4	Obesity (BMI ≥ 30 kg/m^2), HT	317	27	53 (median)	82
Korea Cho (2019) (106)	Retrospective cohort study	Treated CHB patients	3.6	DM (type unspecified), NAFLD, HT, Dyslipidaemia, CVD	826	86	52 (mean, NAFLD) 54 (mean, non-NAFLD)	61
Taiwan, Shyu (2019) (107)	Nested case-control	Treated and untreated CHB patients	11.4	DM (type unspecified)	5,932	731	40-90 (range)	56.7

Country, Author (Year)	Study design	Population	Median follow-up period, years	Risk factors investigated	Participants, n	HCC Cases, n	Age at baseline, years	Sex (% male)
China, Tan (2019) (108)	Prospective cohort study	CHB patients treatment-naïve at baseline	6.3§	T2DM, HT Obesity (> 30kg/m^2), Hypertriglyceridaemia, Hypercholesteraemia	4454	89	45.4 (mean)	68
China, Wang (2019) (109)	Prospective cohort study	CHB patients treatment-naïve at baseline	4.1	DM (type unspecified)	1325	105	50 (median)	73
Taiwan, Hong Kong, Hsu (2018) (110)	Retrospective cohort study	CHB patients receiving ETV or TDF	2.2 (Taiwan) 2.8 (Hong Kong)	DM (I and II)	23,851 (Taiwan) 19,321 (Hong Kong)	596 (Taiwan) 383 (Hong Kong)	47.5 (median, Taiwan) 52.1 (median, Hong Kong)	74 (Taiwan) 66.05 (Hong Kong)
Taiwan, Hsu (2018) (111)	Retrospective cohort study	CHB patients receiving ETV or TDF for ≥ 3 months	2.1	DM (I and II), HT	27,820	802	48.1 (median)	74
US, Kennedy	Case-control (3:1 ratio)	Cases and controls selected from two	-	DM (I and II)	1,101	278	64	78

Country, Author (Year)	Study design	Population	Median follow-up period, years	Risk factors investigated	Participants, n	HCC Cases, n	Age at baseline, years	Sex (% male)
(2018) (112)		independent medical centres						
Korea, Kim (2018) (113)	Retrospective cohort study	Treated and untreated male CHB patients	8.0 years‡	Fasting serum glucose, T2DM, HT	214,167	11,241	N/A	100
China, Li (2018) (114)	Case-control	Cases and controls selected from one hospital	-	DM (type unspecified)	322	112	52 (median, HCC cases), 51(median, non-HCC controls)	18.8 (HCC cases) 26.2 (non-HCC controls)
Hong Kong, Yip (2018) (104)	Retrospective cohort study	CHB patients who have achieved HBsAg seroclearance	3.4	DM (type unspecified), Statin use	4568	54	56.7 (mean)	63
Hong Kong, Chan (2017) (115)	Retrospective cohort study	CHB patients undergoing liver biopsy from one hospital	6.7	DM (type unspecified), HT	270	11	43.6 (mean)	75.2
US, Chayanupatkul (2017) (116)	Retrospective Cohort study	CHB patients attending Veterans Affairs hospitals	4.7 years§	DM (type unspecified), Obesity (BMI > 30 kg/m^2) ¶, HT¶, Hyperlipidaemia¶,	8539	317	N/A	N/A

Country, Author (Year)	Study design	Population	Median follow-up period, years	Risk factors investigated	Participants, n	HCC Cases, n	Age at baseline, years	Sex (% male)
				IHD¶				
China, Han (2017) (117)	Case-control	Cirrhotic patients with one positive HBsAg	-	T2DM	182	73	56.18 (mean)	79.1%
Korea, Kim (2017) (118)	Retrospective cohort study	CHB patients not on treatment	5.0	Insulin resistance (estimated using serum glucose), HT	1696	24	50 (median)	56.8
France, Mallet (2017) (119)	Retrospective cohort study	CHB patients	5.0‡	DM (type unspecified), Renal disease (off dialysis, no transplant), Renal disease (on dialysis, no transplant), Cardiovascular disease, Respiratory disease	48,189	3145	44 (median)	59
Korea, Shim (2017) (120)	Retrospective cohort study	CHB patients treatment-naïve at baseline	3.6	DM (type unspecified)	356	45	52	60

Country, Author (Year)	Study design	Population	Median follow-up period, years	Risk factors investigated	Participants, n	HCC Cases, n	Age at baseline, years	Sex (% male)
Taiwan, Yu (2017) (121)	Retrospective cohort study	Male civil servants with CHB	19.0	Metabolic risk factors (obesity, DM (type unspecified), hypertriglyceridemia and HT), with exposure groups split into groups of 0, 1, 2 and ≥3 risk factors	1690	158	48.4 (median)	100
Korea, Lee (2016) (122)	Retrospective cohort study	CHB patients treatment-naïve at baseline	3.8	Obesity (BMI ≥ 25 kg/m^2)	102	7	46.4 (mean)	66
Netherlands, Brouwer (2015) (123)	Retrospective cohort study	Untreated CHB patients	10.1	Obesity (BMI ≥ 27.5 kg/m^2)	531	13	N/A for cohort	67
Taiwan, Fu (2015) (124)	Matched prospective cohort	Treated and untreated CHB patients	5.4§ (DM) 5.5§ (non-DM)	DM (type unspecified)	4179	111 42 (non-DM)	49.19 (mean, DM group) 49.05 (mean, non-DM)	58.55 (DM), 58.37 (non-DM)
New Zealand, Hsiang	Retrospective cohort study	Treated and untreated CHB patients	5.0§	T2DM	223	36	51 (mean)	66.8

Country, Author (Year)	Study design	Population	Median follow-up period, years	Risk factors investigated	Participants, n	HCC Cases, n	Age at baseline, years	Sex (% male)
(2015) (125)								
Taiwan, Hsu (2014) (126)	Retrospective cohort study	Treated CHB patients with cirrhosis	2.1	DM (I and II), HT, Dyslipidaemia	210	35	52.8 (median)	73.3
South Korea, Kim (2014) (127)	Retrospective cohort study	Cirrhotic CHB patients treated with ETV for ≥ 12 months	3.1	DM (type unspecified)	306	45	49.4 (mean)	68.3
Taiwan, Wu (2014) (128)	Retrospective cohort study	Treated and untreated CHB patients	3.5§ (treated) 5.2§ (untreated)	DM (type unspecified), COPD, ACS, Cerebrovascular disease, Renal failure, Hypercholesterolaemia	43190	5446	43.5 (mean, same for both treated and untreated)	76.9 (treated), 75.5 (untreated)
Taiwan, Chen (2013) (129)	Retrospective cohort study	CHB patients recruited from national cancer screening programme	1-3‡	DM (type unspecified), Obesity (BMI ≥ 27 kg/m^2),	5606	57	>40	N/A

Country, Author (Year)	Study design	Population	Median follow-up period, years	Risk factors investigated	Participants, n	HCC Cases, n	Age at baseline, years	Sex (% male)
				Metabolic syndrome				
China, Gao (2013) (130)	Cross sectional case-control (stratified by cirrhosis)	Cirrhotic CHB patients treated at a single hospital for cirrhosis and/or HCC	-	DM (type unspecified), MAP	370	122	54.7 (mean)	86.9
China, Li (2012) (131)	Case-control	Cases and controls selected from one hospital (untreated for CHB)	-	T2DM	6275	1105	53.8 (mean, cases) 44.9 (mean, controls)	84.7 (cases) 73.8 (controls)
Taiwan, Chao (2011) (132)	Case-cohort	Male civil servants with CHB	17.0‡	Insulin resistance (estimated using serum glucose)	1142	124	30-65	100
Australia, Walter (2011) (133)	Retrospective cohort	Participants with ≥1 record of positive HBsAg or VL in the national notifiable diseases database	3.0	DM (type unspecified)	43,892	242	34.9 (median)	N/A
Taiwan, Wang (2009) (134)	Prospective cohort study	Treated and untreated CHB patients	8.0‡	T2DM	696	24	49.3 (mean)	51.3

Country, Author (Year)	Study design	Population	Median follow-up period, years	Risk factors investigated	Participants, n	HCC Cases, n	Age at baseline, years	Sex (% male)
Taiwan, Chen (2008) (135)	Prospective cohort study	CHB patients recruited from national cancer screening	12.3§	History of DM (type unspecified), HT	3931	187	45.9	59.3
Taiwan, Yu (2008) (136)	Prospective cohort study	Male civil servants with CHB (treated and untreated)	14.7§	DM (type unspecified)	2903	134	N/A	100
Taiwan, Lai (2006) (137)	Prospective cohort study	CHB patients recruited from national cancer screening	2.8§	T2DM	6545	46	N/A	N/A

FU, follow-up; RF, risk factors; DM, diabetes mellitus; ETV, entecavir; TDF, tenofovir disoproxil fumarate; HT, hypertension; HCV, hepatitis C virus; N/A, not available; HIV, human immunodeficiency virus; ALT, alanine aminotransferase; BMI, body mass index; NAFLD, non-alcoholic fatty liver disease; NSAID, non-steroidal anti-inflammatory drug; HCC, hepatocellular carcinoma; MELD, model for end-stage liver disease; T1DM, type 1 diabetes mellitus; HBeAg, hepatitis B E antigen; HDV, hepatitis delta virus; PIIINP, Procollagen 3 N-terminal peptide; COPD, chronic obstructive pulmonary disease; ACS, acute coronary syndrome; AFP, alpha-feto protein; GGT, gamma-glutamyl transpeptidase; INR, international normalised ratio; MAP, mean arterial pressure; BCP, basal core promoter; CHF, congestive heart failure; HBsAg, hepatitis B surface antigen.

† Comorbidity index calculated according to presence of any of the following comorbidities: myocardial infarction, congestive heart failure, peripheral vascular disease, cerebrovascular disease, dementia, chronic pulmonary disease, rheumatic disease, peptic ulcer disease, hemiplegia or paraplegia, renal disease, any malignancy, metastatic solid tumour and HIV/AIDS.

‡ Length from index date to end of observation, as mean/median follow-up not reported in study.

§ Mean length of follow-up.

¶ RRs for these factors not adjusted to listed covariates.

Table 2.3. Covariate adjustment for included observational studies.

Country, Author (Year)	Risk factors investigated	Covariate adjustment
UK, Ferreira (2020) (46)	T2DM	Age, BMI, cigarette smoking, calendar year, dyslipidaemia, ethnicity, HBV vaccination, HCV coinfection, other liver disease (excluded), sex
Korea, Goh (2020) (100)	DM (unspecified) HT, Cholesterol >200mg/dL	Age, ALT, antiplatelet therapy, antiviral therapy, any other malignancy (excluded), cirrhosis, DM, VL, HCV coinfection (excluded), HIV coinfection (excluded), sex, statin use
Korea, Kim (2020) (101)	DM (unspecified)	Age, cirrhosis, compromised liver function (excluded), HCV coinfection, other significant medical illness (excluded), sex
Singapore, Lim (2020) (102)	T2DM	Age, autoimmune disease, HCV coinfection (excluded), HIV coinfection (excluded), metabolic factors, primary biliary cirrhosis (excluded), primary sclerosing cholangitis (excluded), steatosis-inducing medication use (excluded), substantial alcohol intake (excluded)
United States and Asia-Pacific, Yang (2020) (103)	DM (unspecified)	AFP, age, albumin, alcohol use, ALT, ascites or encephalopathy, bilirubin, cirrhosis, HBeAg, HCV coinfection (excluded), HDV coinfection (excluded), HIV coinfection (excluded), Immunosuppressant use (excluded), platelet count, sex
Hong Kong, Yip (2020) (104)	Thiazolidinedione use, metformin use, sulphonyl urea use, Insulin use, statin use, aspirin/clopidogrel use	Age, albumin, ALT, Antiviral treatment, autoimmune disease (excluded), bilirubin, cirrhosis, DM, HbA1c , HCV coinfection (excluded), HDV coinfection (excluded), HIV coinfection (excluded), HT, other metabolic liver disease (excluded), platelets, renal disease (excluded), sex
France, Brichler (2019) (105)	Obesity (BMI ≥ 30 kg/m^2), HT	Age, HCV coinfection (excluded), platelets
Korea, Cho (2019) (106)	DM (type unspecified), NAFLD, HT, Dyslipidaemia, CVD	Age, autoimmune liver disease (excluded), cirrhosis, compromised liver function (excluded), HBeAg, VL, other liver disease (excluded), other viral infections (excluded), steatosis-inducing medication use (excluded), substantial alcohol intake (excluded)

Country, Author (Year)	Risk factors investigated	Covariate adjustment
Taiwan, Shyu (2019) (107)	DM (type unspecified)	Age (matched), alcohol-related liver disease (matched), antiviral treatment (excluded), baseline liver cirrhosis (matched), Deyo-Charlson comorbidity index(42)†, HCV (excluded), obese/overweight, sex (matched)
China, Tan (2019) (108)	T2DM, HT, obesity (> 30kg/m^2), hypertriglyceridaemia, hypercholesteraemia	Age, alcohol consumption, antiviral therapy (excluded), any other malignancy, AST (≥40 U/L), cigarette smoking, cirrhosis, HCV coinfection (excluded), high HBV viral load (excluded), HIV coinfection (excluded), sex
China, Wang (2019) (109)	DM (type unspecified)	Age, fibrosis, HCV coinfection (excluded), HIV coinfection (excluded), immunosuppressant therapy (excluded), low HBV viral load (excluded), other liver disease (excluded)
Taiwan, Hong Kong, Hsu (2018) (110)	DM (I and II)	Age, antiviral treatment (excluded), cirrhosis, DM, HCV coinfection (excluded), sex
Taiwan, Hsu (2018) (111)	DM (I and II), HT	Age, antiviral treatment (excluded), cirrhosis, DM, HCV, hyperlipidaemia, HT, interferon use, metformin use, sex, statin use
US, Kennedy (2018) (112)	DM (I and II)	Age (matched), alcohol misuse, cirrhosis, DM, ethnicity, HCV, HIV, sex (matched)
Korea, Kim (2018) (113)	Fasting serum glucose, T2DM, HT	Age, alcohol consumption, alcohol liver disease, ALT, antiviral therapy, any other malignancies, aspirin use, BMI, cerebrovascular disease, chronic pulmonary disease, cigarette smoking, cirrhosis, congestive heart failure, dementia, HCV, hemiplegia or paraplegia, HIV/AIDS, insurance premium, metastatic solid tumour, metformin use, myocardial infarction, NAFLD, NSAID use, peptic ulcer disease, peripheral vascular disease, physical activity, renal disease, residential area, rheumatic disease, sex (restricted to males), total cholesterol level
China, Li (2018) (114)	DM (type unspecified)	Age (matched), any other malignancies (excluded), autoimmune hepatitis (excluded), cigarette smoking, cirrhosis, family history of HCC, VL, HCV (excluded), HIV (excluded), NAFLD (excluded), sex (matched)

Country, Author (Year)	Risk factors investigated	Covariate adjustment
Hong Kong, Yip (2018) (104)	DM (type unspecified), statin use	Age, antiviral treatment use prior to seroclearance, cirrhosis, HCV coinfection(excluded), HDV coinfection (excluded), liver transplant (excluded), sex, statin use
Hong Kong, Chan (2017) (115)	DM (type unspecified), HT	No adjustments
US, Chayanupatkul (2017) (116)	DM (type unspecified),Obesity (BMI > 30 kg/m^2) ¶, HT¶, hyperlipidaemia¶, IHD¶	Age, ALT, family history of HCC, HCV coinfection, race
China, Han (2017) (117)	T2DM	Age (matched), any other malignancies (excluded), Child-Pugh score (matched), sex (matched)
Korea, Kim (2017) (118)	Insulin resistance (estimated using serum glucose), HT	Age, cirrhosis, DM, VL levels, HCV coinfection (excluded), HIV coinfection (excluded), obesity, sex
France, Mallet (2017) (119)	DM (type unspecified), renal disease (off dialysis, no transplant), renal disease (on dialysis, no transplant), cardiovascular disease, respiratory disease	Age, alcohol consumption, any other malignancy, HCV coinfection, HDV coinfection, HIV coinfection, immunosuppressant use (excluded), obesity, other causes of cirrhosis, residency area, sex
Korea, Shim (2017) (120)	DM (type unspecified)	No adjustment
Taiwan, Yu (2017) (121)	Metabolic risk factors (obesity, DM (type unspecified), hypertriglyceridemia and HT), with exposure groups split into groups of 0, 1, 2 and ≥3 risk factors	Age, alcohol consumption, first-degree family history of HCC, sex (restricted to men), tobacco use
Korea, Lee (2016) (122)	Obesity (BMI ≥ 25 kg/m^2)	No adjustment
Netherlands, Brouwer (2015) (123)	Obesity (BMI ≥ 27.5 kg/m^2)	Antiviral treatment (excluded), autoimmune liver disease (excluded), HCV coinfection (excluded), HDV coinfection (excluded), HEV coinfection (excluded), HIV coinfection (excluded), other primary liver disease (excluded), primary biliary cirrhosis (excluded), Wilson's disease (excluded)
Taiwan, Fu (2015) (124)	DM (type unspecified)	Age (matched), antiviral treatment, cirrhosis, Deyo comorbidity index, DM onset date (matched), hyperlipidaemia, obesity, sex (matched), statin use

Country, Author (Year)	Risk factors investigated	Covariate adjustment
New Zealand, Hsiang (2015) (125)	T2DM	Age, alcohol use, antiviral therapy initiated (yes/no), baseline VL, oesophageal varices at baseline, HCV coinfection (excluded), HDV infection (excluded), HIV coinfection (excluded), MELD score (< or >- 11), positive HBeAg at baseline, sex, sustained viral suppression during follow-up (yes/no), T1DM (excluded)
Taiwan, Hsu (2014) (126)	DM (I and II), HT, Dyslipidaemia	Age, antiviral therapy, any other malignancies (excluded), cirrhosis (restricted to cirrhotic patients), VL (restricted to those with >2000 IU/mL viral load), HCV coinfection (excluded), HIV coinfection (excluded), MELD score, previous organ transplantation, sex
South Korea, Kim (2014) (127)	DM (type unspecified)	Age, antiviral treatment (excluded untreated), autoimmune hepatitis (excluded), cirrhosis (restricted to cirrhotic patients), decompensated cirrhosis, VL, HCV coinfection (excluded), HDV coinfection (excluded), HIV coinfection (excluded), no virological response at 12 months after initiation of treatment, serum PIIINP, sex
Taiwan, Wu (2014) (128)	DM (type unspecified), COPD, ACS, Cerebrovascular disease, renal failure, hypercholesterolaemia	Age, antiviral treatment, any other malignancies, cirrhosis, DM, HCV coinfection (excluded), HIV coinfection (excluded), metformin use, NSAIDs/aspirin use, sex, statin use
Taiwan, Chen (2013) (129)	DM (type unspecified), obesity (BMI ≥ 27 kg/m^2), metabolic syndrome	No adjustment
China, Gao (2013) (130)	DM (type unspecified), MAP	AFP, albumin, ALT, any other malignancies, autoimmune disorders, bilirubin, cigarette smoking, drug- or poison-induced liver damage, ethnicity, GGT, HCV coinfection, HDV coinfection, heavy alcohol consumption, haemoglobin, HIV coinfection, INR, platelet count, portal vein diameter, presence of ascites, serum sodium, sex
China Li (2012) (131)	T2DM	Age, antiviral treatment (patients on treatment excluded), any other malignancies (excluded), Child-Pugh grade of cirrhosis severity, cirrhosis,

Country, Author (Year)	Risk factors investigated	Covariate adjustment
		family history of HCC, HBeAg, HCV coinfection (excluded), residential area, sex
Taiwan, Chao (2011) (132)	Insulin resistance (estimated using serum glucose)	Age, antiviral therapy (excluded), alcohol consumption, BCP double mutations, cigarette smoking, first-degree family history of HCC, HBeAg, VL, HBV genotype, HCV coinfection (excluded), number of visits
Australia, Walter (2011) (133)	DM (type unspecified)	Age, any other malignancies, cerebrovascular disease, chronic pulmonary disease, CHF, dementia, Deyo-Charlson comorbidity index(42)†, HCV coinfection, peptic ulcer disease, peripheral vascular disease, renal disease, rheumatic disease, sex
Taiwan, Wang (2009) (134)	T2DM	Age, alcohol consumption, BMI, cigarette smoking, sex
Taiwan, Chen (2008) (135)	History of DM (type unspecified), HT	Age, alcohol consumption, BMI, cigarette smoking, DM, educational level, obesity, sex, total cholesterol, triglycerides
Taiwan, Yu (2008) (136)	DM (type unspecified)	Age, alcohol consumption, BMI, number of visits, sex (restricted to males), tobacco use
Taiwan, Lai (2006) (137)	T2DM	Age, alcohol consumption, any other malignancies (excluded), cigarette smoking, HCV coinfection, sex

FU, follow-up; RF, risk factors; DM, diabetes mellitus; VL, viral load; ETV, entecavir; TDF, tenofovir disoproxil fumarate; HT, hypertension; HCV, hepatitis C virus; N/A, not available; HIV, human immunodeficiency virus; ALT, alanine aminotransferase; BMI, body mass index; NAFLD, non-alcoholic fatty liver disease; NSAID, non-steroidal anti-inflammatory drug; HCC, hepatocellular carcinoma; MELD, model for end-stage liver disease; T1DM, type 1 diabetes mellitus; HBeAg, hepatitis B E antigen; HDV, hepatitis delta virus; PIIINP, Procollagen 3 N-terminal peptide; COPD, chronic obstructive pulmonary disease; ACS, acute coronary syndrome; AFP, alpha-feto protein; GGT, gamma-glutamyl transpeptidase; INR, international normalised ratio; MAP, mean arterial pressure; BCP, basal core promoter; CHF, congestive heart failure; HBsAg, hepatitis B surface antigen.

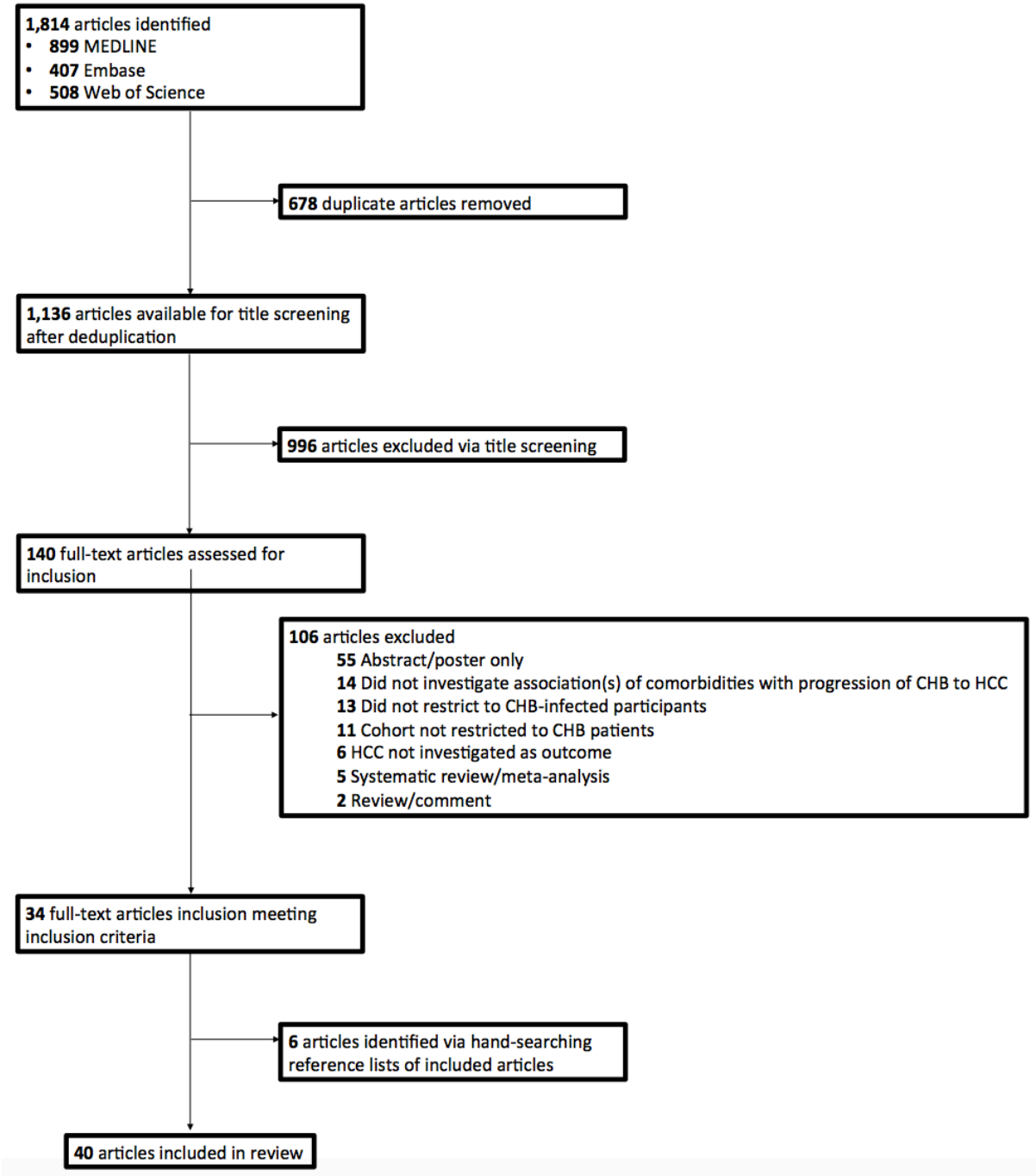


Figure 2.1. Flow chart of study selection. MEDLINE, Embase and Web of Science databases were systematically searched using relevant terms to identify relevant human studies published in English from 1st January 2000 to 24th June 2020.

All studies were observational in design, with 33 cohort and 7 case-control studies included (**Tables 2.2, 2.3**). Thirty-two studies were conducted in Asian countries. Four studies were restricted to male cohorts and 36 were undertaken in mixed-sex cohorts. All studies recruited participants from health centres, healthcare or prescription databases, or pre-existing cohorts or cancer screening programmes. All studies were undertaken in adults, with mean/median ages of cohorts ranging between 40 and 65 years in 33 studies. Thirty-three studies investigated DM/insulin resistance/fasting serum glucose, 11 studies investigated blood hypertension/blood pressure, 7 investigated dyslipidaemia, 5 investigated obesity and cardiovascular disease. Fewer than 5 studies investigated other factors including renal disease, statin use and use of antidiabetic drugs.

In the 40 studies including 536,456 adults, >30,500 HCC events occurred (I am unable to report an exact number, because one study did not report a precise number of HCC cases (46)). Sample sizes of cohort studies varied widely, ranging from 102 to 214,167 (median 3,090), with corresponding numbers of HCC cases ranged from 7 (arising among 102 participants) to 11,241 (arising among 214,167 participants). Case-control sample sizes ranged from 182 to 6,275 (median 1,122) with corresponding numbers of HCC cases ranged from 73 (out of 182 participants) and 1105 (out of 6,275 participants).

2.5.2. Quality assessment

Among 40 studies, 39 had quality scores ≥5 (**Tables 2.4 and 2.5**). All cohort studies were of sufficient quality with 13 of these being scored as high-quality. Six case-control studies were of sufficient quality and one of poor quality. Inclusion criteria varied widely and therefore study populations were heterogeneous. In most studies, exposures and outcomes were ascertained using health assessment, imaging or record linkage. Twenty-three cohort studies and 7 case-control studies accounted for age and sex. HCC typically arise after long durations of infection, and therefore prolonged follow up allows for detection of more HCC events; among 23 cohort studies identified, only 5 cohort studies had lengths of follow-up ≥ 10 years).

Table 2.4. Quality appraisal scores of case-control studies identified in meta-analysis. Quality scores awarded according to Newcastle-Ottawa Scale.

	Selection			Comparability		Exposure			
Case-control study	**Is case definition adequate?**	**Representativeness of cases**	**Selection of controls**	**Definition of controls**	**Comparability of cases and controls**	**Method of exposure ascertainment**	**Same method for cases and controls?**	**Non-response rate**	**Score, study quality**
Shyu (2019) (107)	Yes, reference to ICD codes	Consecutive or obviously representative series of cases	Cohort controls	No history of disease	Study controls for age and sex	Secure record	Yes	Not mentioned	7, sufficient
Kennedy (2018) (112)	Yes, reference to ICD codes	Consecutive or obviously representative series of cases	Hospital controls	No history of disease	Study controls for age and sex	Secure record	Yes	Not mentioned	6, sufficient
Li (2018) (114)	Yes, reference to primary record source	Consecutive or obviously representative series of cases	Hospital controls	No mention of outcome history	Study controls for age and sex	Self-report or medical record	Yes	Not mentioned	5, sufficient
Han (2017) (117) †	No	Consecutive or obviously representative series of cases	Hospital controls	No mention of outcome history	Study controls for age and sex	Medical record	Yes	Not mentioned	4, low
Gao (2013) (130)	Yes, reference to primary	Consecutive or obviously representative series of cases	Hospital controls	No history of disease	Study controls for sex but not age	Secure record	Yes	Same for both groups†	7, sufficient

Case-control study	Selection			Comparability		Exposure			Score, study quality
	Is case definition adequate?	Representativeness of cases	Selection of controls	Definition of controls	Comparability of cases and controls	Method of exposure ascertainment	Same method for cases and controls?	Non-response rate	
	record source								
Li (2012) (131)	Yes, reference to primary record source	Consecutive or obviously representative series of cases	Hospital controls	No mention of outcome history	Study controls for age and sex	Secure record	Yes	Same for both groups	7, sufficient
Chao (2011) (132)	Yes, reference to primary record source	Consecutive or obviously representative series of cases	Cohort controls	No mention of outcome history	Study controls for age and sex	No description	No description	Not mentioned	5, sufficient

† Participants with missing values excluded

Table 2.5. Quality appraisal scores of cohort studies identified in meta-analysis. Quality scores awarded according to Newcastle-Ottawa Scale.

	Selection				Comparability	Outcome			
Cohort study	**Representativeness of exposed cohort**	**Selection of non-exposed cohort**	**Ascertainment of exposure**	**Participants outcome-free at baseline**	**Comparability of cohorts**	**Outcome assessment**	**Sufficient length of follow-up?‡**	**Adequacy of follow-up**	**Score, study quality**
Ferreira (2020) (46)	Representative of average CHB in community	From same community as exposed	Secure record	Yes	Study controls for age and sex	Record linkage	No	No/minimal loss to follow up	8, high
Goh (2020) (100)	Selected group (CHB patients undergoing liver biopsy)	From same community as exposed	Health assessment/ health record	Yes	Study controls for age and sex	Imaging assessment and pathology	No	No/minimal loss to follow up	7, sufficient
Kim (2020) (101)	Somewhat representative of average CHB in community (restricted to patients who were treatment-naïve at baseline)	From same community as exposed	Health assessment	Yes	Study controls for age and sex	Imaging assessment and pathology	No	No statement	7, sufficient
Lim (2020) (102)	Selected group (CHB patients undergoing liver biopsy)	From same community as exposed	Health assessment	Yes	Study controls for age	Imaging assessment	No	No/minimal loss to follow up	6, sufficient

	Selection				Comparability		Outcome		
Cohort study	**Representativeness of exposed cohort**	**Selection of non-exposed cohort**	**Ascertainment of exposure**	**Participants outcome-free at baseline**	**Comparability of cohorts**	**Outcome assessment**	**Sufficient length of follow-up?‡**	**Adequacy of follow-up**	**Score, study quality**
Yang (2020) (103)	Somewhat representative of average CHB in community (restricted to participants receiving antiviral treatment)	From same community as exposed	Health assessment/ health record	Yes	Study controls for age and sex	Cytology or pathology	No	No statement	7, sufficient
Yip (2020) (104)	Somewhat representative of average CHB in community (restricted to CHB with DM)	From same community as exposed	Secure record	Yes	Study controls for age and sex	Record linkage	No	No/minimal loss to follow up	8, high
Brichler (2019) (105)	Somewhat representative of average CHB in community (restricted to CHB patients with cirrhosis)	From same community as exposed	Health assessment	Yes	Study controls for age	Imaging assessment, pathology and non-invasive assessment	No	Description of patients lost to follow up	7, sufficient

	Selection			Comparability		Outcome			
Cohort study	**Representativeness of exposed cohort**	**Selection of non-exposed cohort**	**Ascertainment of exposure**	**Participants outcome-free at baseline**	**Comparability of cohorts**	**Outcome assessment**	**Sufficient length of follow-up?‡**	**Adequacy of follow-up**	**Score, study quality**
Cho (2019) (106)	Somewhat representative of average CHB in community (restricted to participants receiving antiviral treatment)	From same community as exposed	Health assessment	Yes	Study controls for age	Pathology and clinical assessment	No	Description of patients lost to follow up	7, sufficient
Tan (2019) (108)	Somewhat representative of average CHB in community (restricted to civil servants who were treatment-naïve at baseline)	From same community as exposed	Health assessment	Yes	Study controls for age and sex	Imaging assessment and pathology	No	No/minimal loss to follow up	8, high
Wang (2019) (109)	Somewhat representative of average CHB in community (restricted to	From same community as exposed	Health assessment/ health record	Yes	Study controls for age	Imaging assessment and pathology	No	No statement	6, sufficient

Cohort study	Selection			Comparability		Outcome			Score, study quality
	Representativeness of exposed cohort	Selection of non-exposed cohort	Ascertainment of exposure	Participants outcome-free at baseline	Comparability of cohorts	Outcome assessment	Sufficient length of follow-up?‡	Adequacy of follow-up	
	patients who were treatment-naïve at baseline)								
Yip (2018) (104)	Somewhat representative of average CHB in community (restricted to patients who have achieved HBsAg seroclearance)	From same community as exposed	Secure record	Yes	Study controls for age and sex	Record linkage	No	No/minimal loss to follow up	8, high
Hsu (2018) (110)	Somewhat representative of average CHB in community (restricted to participants receiving antiviral treatment)	From same community as exposed	Secure record	Yes	Study controls for age and sex	Independent assessment/record linkage	Yes	No/minimal loss to follow up	9, high

	Selection				Comparability		Outcome		
Cohort study	**Representativeness of exposed cohort**	**Selection of non-exposed cohort**	**Ascertainment of exposure**	**Participants outcome-free at baseline**	**Comparability of cohorts**	**Outcome assessment**	**Sufficient length of follow-up?‡**	**Adequacy of follow-up**	**Score, study quality**
Hsu (2018) (111)	Somewhat representative of average CHB in community (restricted to participants receiving antiviral treatment)	From same community as exposed	Secure record	Yes	Study controls for age and sex	Record linkage	No	No/minimal loss to follow up	8, high
Kim (2018) (113)	Somewhat representative of average CHB in community (restricted to males)	From same community as exposed	Secure record	Yes	Study controls for age and sex	Record linkage	No	No/minimal loss to follow up	8, high
Chan (2017) (115)	Selected group (CHB patients undergoing liver biopsy)	From same community as exposed	Comprehensive health assessment	Yes	None	Assessment using ≥ 2 imaging techniques	No	Description of patients lost to follow up	5, sufficient
Chayanupatkul	Somewhat representative of average CHB in	From same community as exposed	Secure record	Not specified	Study controls for age	Record linkage,	No	No statement	5, sufficient

	Selection			Comparability		Outcome			
Cohort study	**Representativeness of exposed cohort**	**Selection of non-exposed cohort**	**Ascertainment of exposure**	**Participants outcome-free at baseline**	**Comparability of cohorts**	**Outcome assessment**	**Sufficient length of follow-up?‡**	**Adequacy of follow-up**	**Score, study quality**
(2017) (116)	community (patients attending Veterans Affairs hospitals)					imaging and pathology			
Kim (2017) (118)	Representative of average CHB in community	From same community as exposed	Health assessment/ health record	Yes	Study controls for age and sex	Assessment using imaging	No	No statement	6, sufficient
Mallet (2017) (119)	Representative of average CHB in community	From same community as exposed	Secure record	Not specified	Study controls for age and sex	Record linkage	No	No/minimal loss to follow up	7, sufficient
Shim (2017) (120)	Somewhat representative of average CHB in community (restricted to patients who were treatment-naïve at baseline)	From same community as exposed	Secure record	Yes	None	Imaging assessment and pathology	No	No/minimal loss to follow up	6, sufficient

	Selection				Comparability		Outcome		
Cohort study	**Representativeness of exposed cohort**	**Selection of non-exposed cohort**	**Ascertainment of exposure**	**Participants outcome-free at baseline**	**Comparability of cohorts**	**Outcome assessment**	**Sufficient length of follow-up?‡**	**Adequacy of follow-up**	**Score, study quality**
Yu (2017) (121)	Somewhat representative of average CHB in community (restricted to untreated males)	From same community as exposed	Health assessment/ interview	Not specified	Study controls for age and sex	Record linkage	Yes	No/minimal loss to follow up	8, high
Lee (2016) (122)	Somewhat representative of average CHB in community (restricted to patients who were treatment-naïve at baseline)	From same community as exposed	Health assessment	Yes	None	Imaging assessment and/or biopsy	No	No/minimal loss to follow up	6, sufficient
Brouwer (2015) (123)	Somewhat representative of average CHB in community (restricted to untreated CHB patients)	From same community as exposed	Secure record	Not specified	None	Imaging assessment and pathology	Yes	Description of patients lost to follow up	6, sufficient

	Selection				Comparability	Outcome			
Cohort study	**Representativeness of exposed cohort**	**Selection of non-exposed cohort**	**Ascertainment of exposure**	**Participants outcome-free at baseline**	**Comparability of cohorts**	**Outcome assessment**	**Sufficient length of follow-up?‡**	**Adequacy of follow-up**	**Score, study quality**
Fu (2015) (124)	Representative of average CHB in community	From same community as exposed	Secure record	Yes	Study controls for age and sex	Record linkage	No	No/minimal loss to follow up	8, high
Hsiang (2015) (125)	Somewhat representative of average CHB in community (patients with transient elastography or radiological features of cirrhosis who sought care at a single hospital)	From same community as exposed	Health assessment/ health record	Not specified	Study controls for age and sex	Assessment using imaging	No	Description of patients lost to follow up	7, sufficient
Hsu (2014) (126)	Somewhat representative of average CHB in community (restricted to participants	From same community as exposed	Secure record	Yes	Study controls for age and sex	Assessment using imaging and AFP testing	No	No statement	7, sufficient

Cohort study	Selection			Comparability		Outcome			Score, study quality
	Representativeness of exposed cohort	Selection of non-exposed cohort	Ascertainment of exposure	Participants outcome-free at baseline	Comparability of cohorts	Outcome assessment	Sufficient length of follow-up?‡	Adequacy of follow-up	
	receiving antiviral treatment)								
Kim (2014) (127)	Somewhat representative of average CHB in community (restricted to cirrhotic to participants receiving antiviral treatment)	From same community as exposed	No description	Yes	Study controls for age and sex	Assessment using imaging and AFP testing	No	No/minimal loss to follow up	7, sufficient
Wu (2014) (128)	Representative of average CHB in community	From same community as exposed	Secure record	Yes	Study controls for age and sex	Record linkage	No	No/minimal loss to follow up	8, high
Chen (2013) (129)	Representative of average CHB in community	From same community as exposed	Health assessment	Yes	None	Record linkage	No	No/minimal loss to follow up	6, sufficient
Walter (2011) (133)	Representative of average CHB in community	From same community as exposed	Hospital record	Yes	Study controls for age and sex	Record linkage	No	No/minimal loss to follow up	8, high

	Selection				Comparability		Outcome		
Cohort study	**Representativeness of exposed cohort**	**Selection of non-exposed cohort**	**Ascertainment of exposure**	**Participants outcome-free at baseline**	**Comparability of cohorts**	**Outcome assessment**	**Sufficient length of follow-up?‡**	**Adequacy of follow-up**	**Score, study quality**
Wang (2009) (134)	Representative of average CHB in community	From same community as exposed	Health assessment/ health record	Yes	Study controls for age and sex	Imaging and record linkage	No	Response rate of 68.4%	7, sufficient
Chen (2008) (135)	Representative of average CHB in community	From same community as exposed	Interview	Yes	Study controls for age and sex	Record linkage	Yes	No statement	8, high
Yu (2008) (136)	Somewhat representative of average CHB in community (restricted to males)	From same community as exposed	Health assessment/ interview	Not specified	Study controls for age and sex	Imaging	Yes	No statement	7, sufficient
Lai (2006) (137)	Representative of average CHB in community	From same community as exposed	Health assessment	Yes	Study controls for age and sex	Imaging assessment and pathology	No	No/minimal loss to follow up	8, high

CHB, chronic hepatitis B; HBsAg, hepatitis B surface antigen; AFP, alpha-fetoprotein.

2.5.3. Association of diabetes mellitus with HCC risk

Thirty-six studies investigated the association of DM with risk of CHB progression to HCC, comprising 7 case-control studies (**Table 2.6**) and 29 cohort studies (**Table 2.7**). Four studies were restricted to males and the others included both sexes (**Table 2.2**). Mean ages at baseline in all studies were ≥40 years, respectively. Study populations were heterogeneous with variable inclusion criteria, and definitions of DM were not consistent between studies. Four case-control and four cohort studies investigated type 2 DM/insulin resistance, three case-control and seven cohort studies investigated unspecified DM, and one case-control and three cohort studies investigated both type 1 and 2 DM as a composite potential risk factor.

Table 2.6. Effect estimates for case-control studies investigating the association of diabetes mellitus with hepatocellular carcinoma risk.

Country, Author (Year)	Participants, n	HCC Cases, n	Age at baseline, years	Sex (% male)	Source of controls	Relative risk measure†	Risk Ratio (95% Confidence intervals)
Taiwan, Shyu (2019) (107)	5,932	731	40-90 (range)	56.7	Same insurance database as cases	aHR	1.35 (1.16-1.57)
US, Kennedy (2018) (112)	1,101	278	64	78	Same clinical data repository as cases	aOR	0.6 (0.4-0.9)
China, Li (2018) (114)	322	112	52 (median, HCC cases), 51(median, non-HCC controls)	18.8 (HCC cases) 26.2 (non-HCC controls)	Hospital controls	aOR	2.042 (1.150-5.018)
China, Han (2017) (117)	182	73	56.18 (mean)	79.1	Hospital controls	N/A	N/A
China, Gao (2013) (130)	370	122	54.7 (mean)	86.9	Hospital controls	aOR‡	0.376 (0.175-0.807) (compensated cirrhosis) 0.192 (0.054-0.679) (decompensated cirrhosis)
China Li (2012) (131)	6275	1105	53.8 (mean, cases) 44.9 (mean, controls)	84.7 (cases) 73.8 (controls)	Hospital controls	aOR	0.8 (0.6-1.1)

Taiwan, Chao (2011) (132)	1142	124	30-65	100	Control group is random sample of non-HCC participants from total cohort	aHR	Those with homeostasis model assessment-insulin resistance (HOMA-IR, fasting insulin (μU/ml) X fasting glucose (mmol/l)/22.5) index scores of <0.46, 0.78-1.22 and >1.22 had HRs of 1.48 (0.90-2.44), 0.92 (0.54-1.55) and 1.96 (1.23-3.10) as compared to those with a HOMA-IR index of 0.46-0.77

aHR, adjusted Hazards Ratio; aOR, adjusted Odds Ratio; N/A, not available.

† Adjusted risk ratios are minimally adjusted for age and sex.

‡ Adjusted for age but not sex.

Table 2.7. Effect estimates for cohort studies investigating the association of diabetes mellitus with hepatocellular carcinoma risk.

Country, Author (Year)	Participants, n	HCC Cases, n	Age at baseline, years	Sex (% male)	Relative risk measure†	Risk Ratio (95% Confidence intervals)
UK, Ferreira (2020) (46)	3927	16-19	39.2 (mean, non-DM) 54.9 (mean, DM)	50	aHR	1.40 (0.47 to 4.24)
Korea, Goh (2020) (100)	7713	702	50 (statin users) 47 (statin nonusers)	66	aHR	1.31 (1.09 to 1.58)
Korea, Kim (2020) (101)	3277	292	48.7 (mean)	63	aHR	1.35 (0.99 to 1.84)
Singapore, Lim (2020) (102)	289	27	43.2 (mean, Nonhepatic steatosis) 46.4 (mean, Hepatic steatosis)	72	aHR‡	2.69 (1.07 to 6.76)
United States and Asia-Pacific, Yang (2020) (103)	5365	378	48.4 (mean)	69	aHR	1.69 (1.28 to 2.22)
Korea Cho (2019) (106)	826	86	52 (mean, NAFLD) 54 (mean, non-NAFLD)	61	uHR	1.51 (0.90 to 2.55)
China, Tan (2019) (108)	4454	89	45.4 (mean)	68	aHR	2.28 (1.36 to 3.80)
China, Wang (2019) (109)	1325	105	50 (median)	73	aHR‡	1.90 (1.19 to 3.05)

Hong Kong, Yip (2018) (104)	4568	54	56.7 (mean)	63	aHR	1.85 (1.04 to 3.28)
Taiwan & Hong Kong, Hsu (2018) (110)	23,851 (Taiwan) 19,321 (Hong Kong)	596 (Taiwan) 383 (Hong Kong)	47.5 (median, Taiwan) 52.1 (median, Hong Kong)	74 (Taiwan) 66.05 (Hong Kong)	aHR	1.30 (1.10 to 1.60)
Taiwan, Hsu (2018) (111)	27,820	802	48.1 (median)	74	aHR	1.25 (1.06 to 1.47)
Korea, Kim (2018) (113)	214,167	11,241	N/A	100	aHR	1.23 (1.15 to 1.34)
Hong Kong, Chan (2017) (115)	270	11	43.6 (mean)	75.2	uHR	0.53 (0.07 to 4.33)
US, Chayanupatkul (2017) (116)	8539	317	N/A	N/A	uHR	1.44 (0.63 to 3.32)
Korea, Kim (2017) (118)	1696	24	50 (median)	56.8	aHR	1.75 (0.51 to 6.01)
France, Mallet (2017) (119)	48,189	3145	44 (median)	59	aHR	1.26 (1.15 to 1.37)
Korea, Shim (2017) (120)	356	45	52	60	uHR	1.66 (0.86 to 3.21)
Netherlands, Brouwer (2015) (123)	531	13	N/A for cohort	67	uHR	6.80 (2.10 to 22.10)
Taiwan, Fu (2015) (124)	4179	111 42 (non-DM) 69 (DM)	49.19 (mean, DM group) 49.05 (mean, non-DM)	58.55 (DM), 58.37 (non-DM)	aHR	1.80 (1.19 to 2.71)

New Zealand, Hsiang (2015) (125)	223	36	51 (mean)	66.8	aHR	2.36 (1.14 to 4.85)
Taiwan, Hsu (2014) (126)	210	35	52.8 (median)	73.3	aHR	3.49 (1.54 to 7.91)
South Korea, Kim (2014) (127)	306	45	49.4 (mean)	68.3	aHR	1.23 (0.59 to 2.57)
Taiwan, Wu (2014) (128)	43190	5446	43.5 (mean, same for both treated and untreated)	76.9 (treated), 75.5 (untreated)	aHR	1.05 (0.93 to 1.17)
Taiwan, Chen (2013) (129)	5606	57	>40	N/A	uHR	1.61 (0.73 to 3.58)
Australia, Walter (2011) (133)	43,892	242	34.9 (median)	N/A	N/A	Effect size N/A
Taiwan, Wang (2009) (134)	696	24	49.3 (mean)	51.3	aHR	1.3 (0.3 to 5.6)
Taiwan, Chen (2008) (135)	3931	187	45.9	59.3	aHR	2.41 (1.17 to 4.95)
Taiwan, Yu (2008) (136)	2903	134	N/A	100	aOR	1.16 (0.55 to 2.42)
Taiwan, Lai (2006) (137)	6545	46	N/A	N/A	aHR	1.04 (0.36 to 3.02)

aHR, adjusted Hazards Ratio; uHR, unadjusted Hazards Ratio; aOR, adjusted Odds Ratio.

† Adjusted risk ratios are minimally adjusted for age and sex.

‡ Adjusted for age but not sex.

Of the 7 case-control studies that reported effect estimates, there was directional inconsistency between effect estimates reported in case-control studies, with 4 studies reporting an increased risk of HCC in those with DM as compared to those without, 3 studies reporting a decreased risk of HCC in those with DM, and one study failing to provide an effect estimate. Risk ratios (RRs) >1 ranged from 1.35 to 2.04, and all were statistically significant. RRs <1 ranged from 0.19 to 0.80, of which two were statistically significant. Among 28 cohort studies providing effect estimates (27 HRs and 1 odds ratio (OR)), there was directional consistency with 27 of the reported RRs >1. Effect sizes >1 ranged from 1.05 to 6.80, with 15 RRs being statistically significant. The single RR that was <1 was nonsignificant.

Minimal adjustment for confounders differed between case-control and cohort studies. Most case-control studies adjusted for age, sex, HCV coinfection, HIV coinfection and cirrhosis. Twenty cohort studies minimally adjusted for age and sex. Of these, 15 adjusted for HCV coinfection, 13 for cirrhosis, 12 for antiviral treatment, 10 for HIV coinfection, 9 for alcohol consumption, 7 each for HBV viral DNA load and cigarette smoking and 6 for other liver disease (including alcoholic liver disease). Eight studies excluded participants who developed HCC within the first 3 to 12 months of follow-up in their main analyses. One study did so in sensitivity analysis and found this did not modify associations observed.

2.5.4. Meta-analysis of cohort studies

DM was associated with an increased risk of progression to HCC by meta-analysis restricted to HRs minimally adjusted for age and sex (**Figure 2.2**). As there was significant heterogeneity (I^2=49%, p<0.01), results from both fixed- and random-effects analyses are presented. In random-effects analysis risk of HCC was 36% (summary RR 1.36; 95% CI 1.23-1.49) significantly higher in DM compared to non-DM.

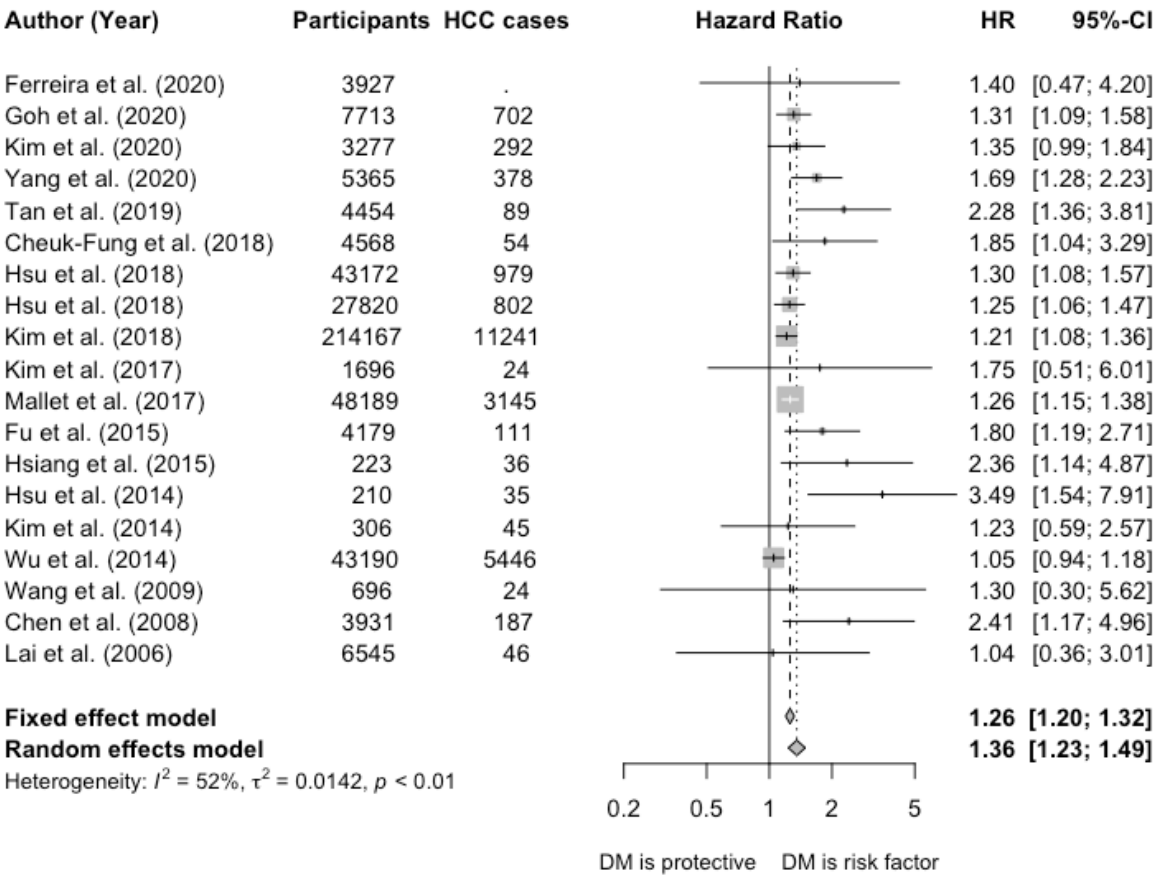


Figure 2.2. Forest plot of hazard ratios from longitudinal cohort studies investigating the association of diabetes mellitus with risk of progression of chronic hepatitis B infection to hepatocellular carcinoma. The study by Yu et al provided an Odds Ratio and was excluded from the meta-analysis. Dashed vertical lines represent HR based on meta-analysis of all studies by fixed effect and random effects models. The studies for pooling the HR had sufficient quality (quality scores ≥5).
HR, hazard ratio; CI, confidence interval; DM, diabetes mellitus.

2.5.5. Subgroup and sensitivity analyses

I undertook sensitivity analyses to investigate the robustness of pooled estimates to additional adjustment for HCV or HIV coinfection, cirrhosis, and DM treatment. After restricting meta-analysis to 16 studies adjusting for HCV coinfection in addition to age and sex (**Figure 2.3**), pooled HRs did not change materially. Considering 8 studies adjusting for HIV and antiviral treatment (**Figure 2.4**), pooled HR from the fixed-effects analysis was attenuated towards the null slightly but still remained significant. To investigate the robustness of the association of DM with HCC to adjustment for cirrhosis, a potential mediator, I restricted meta-analysis to studies adjusting for cirrhosis (**Figure 2.5**). This did not change pooled HRs materially.

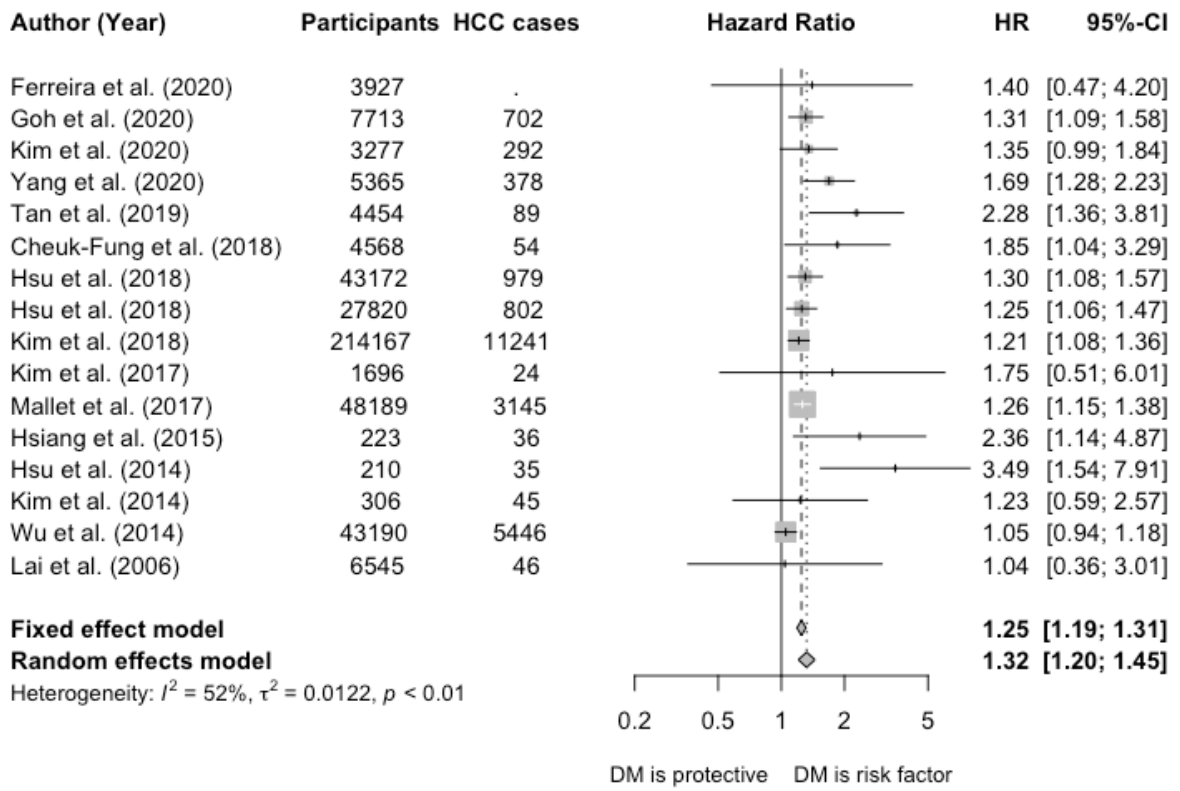


Figure 2.3. Sensitivity analysis of primary meta-analysis demonstrating the influence of restricting the analysis to studies additionally adjusting for HCV coinfection in addition to age and sex.
HR, hazard ratio; CI, confidence interval; DM, diabetes mellitus.

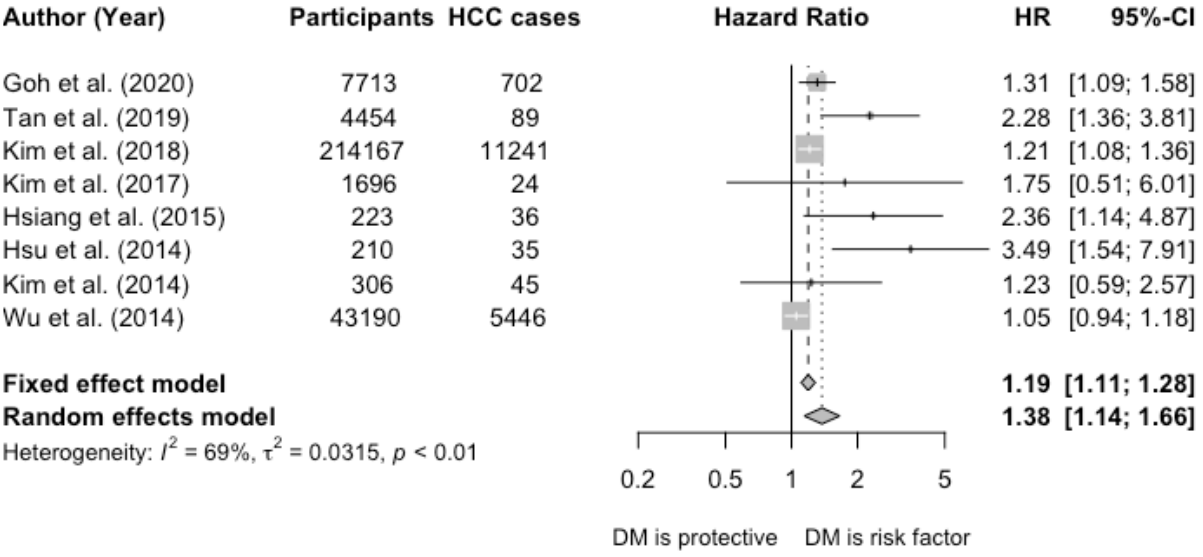


Figure 2.4. Sensitivity analysis of primary meta-analysis demonstrating the influence of restricting the analysis to studies additionally adjusting for HCV coinfection, HIV coinfection and antiviral treatment in addition to age and sex.
HR, hazard ratio; CI, confidence interval; DM, diabetes mellitus.

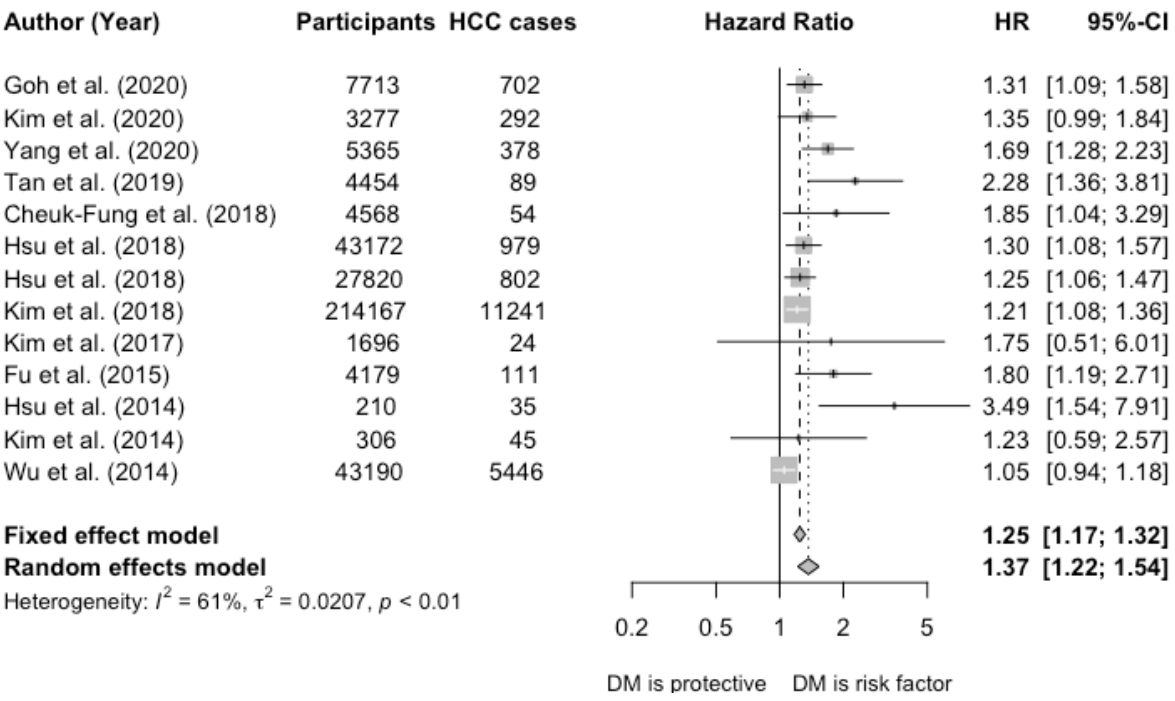


Figure 2.5. Sensitivity analysis of primary meta-analysis demonstrating the influence of restricting the analysis to studies additionally adjusting for cirrhosis in addition to age and sex.
HR, hazard ratio; CI, confidence interval; DM, diabetes mellitus.

To investigate heterogeneity between type 2 DM and unspecified DM, I performed sensitivity analysis whereby studies were stratified by DM type. Amongst studies investigating type 2 DM, heterogeneity was 33% (p=0.18) (**Figure 2.6**). However the HR did not differ materially to that observed in the primary meta-analysis.

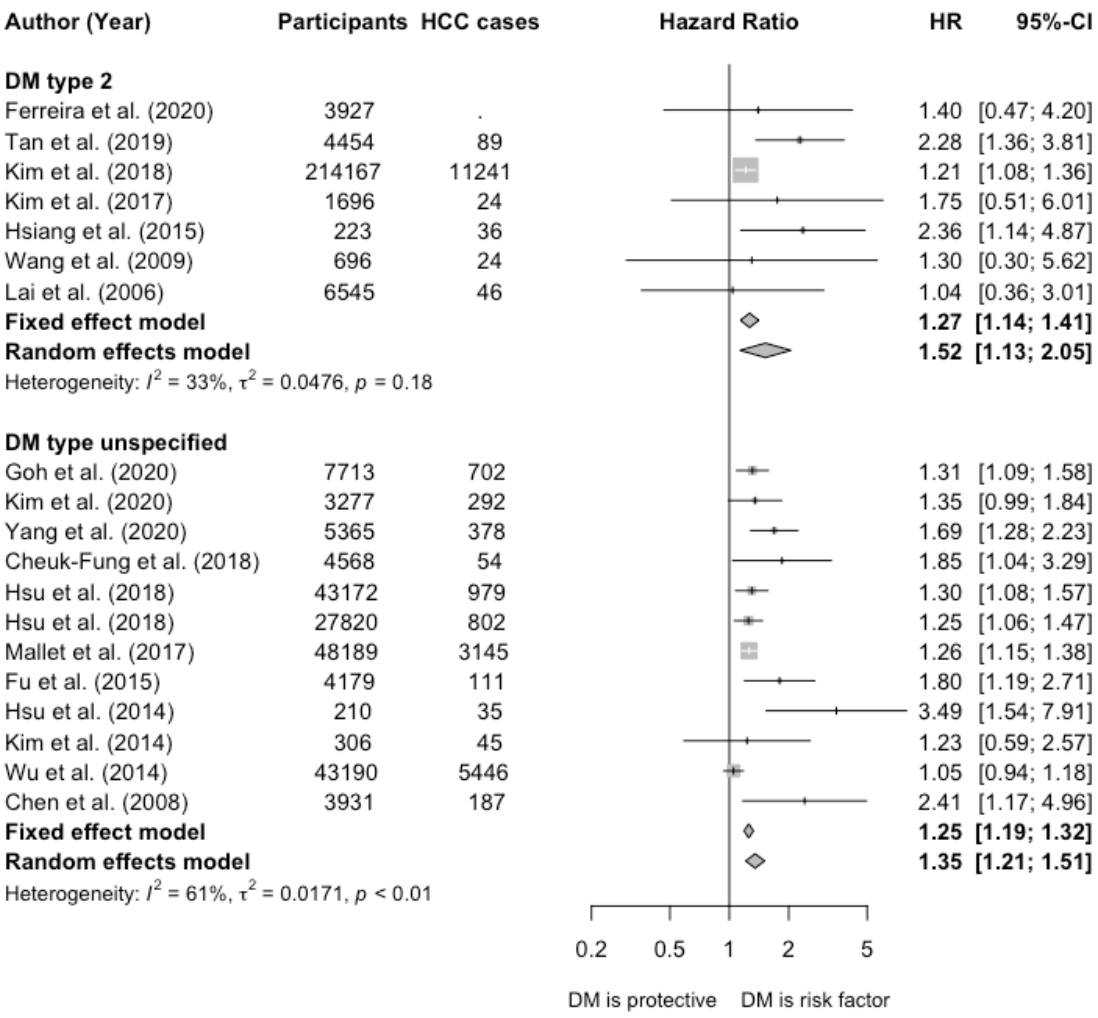


Figure 2.6. Sensitivity analysis of primary meta-analysis stratifying studies by diabetes mellitus disease subtype.
HR, hazard ratio; CI, confidence interval; DM, diabetes mellitus.

The association of DM with HCC risk was attenuated towards the null in studies that adjusted for metformin use, with risk of HCC 16% higher in DM participants as compared to non-DM (random effects HR 1.16, 95% CI 1.04-1.29) in analysis restricted to studies adjusting for metformin use (**Figure 2.7**). After restricting to studies adjusting to DM treatment, pooled HRs remained statistically significant.

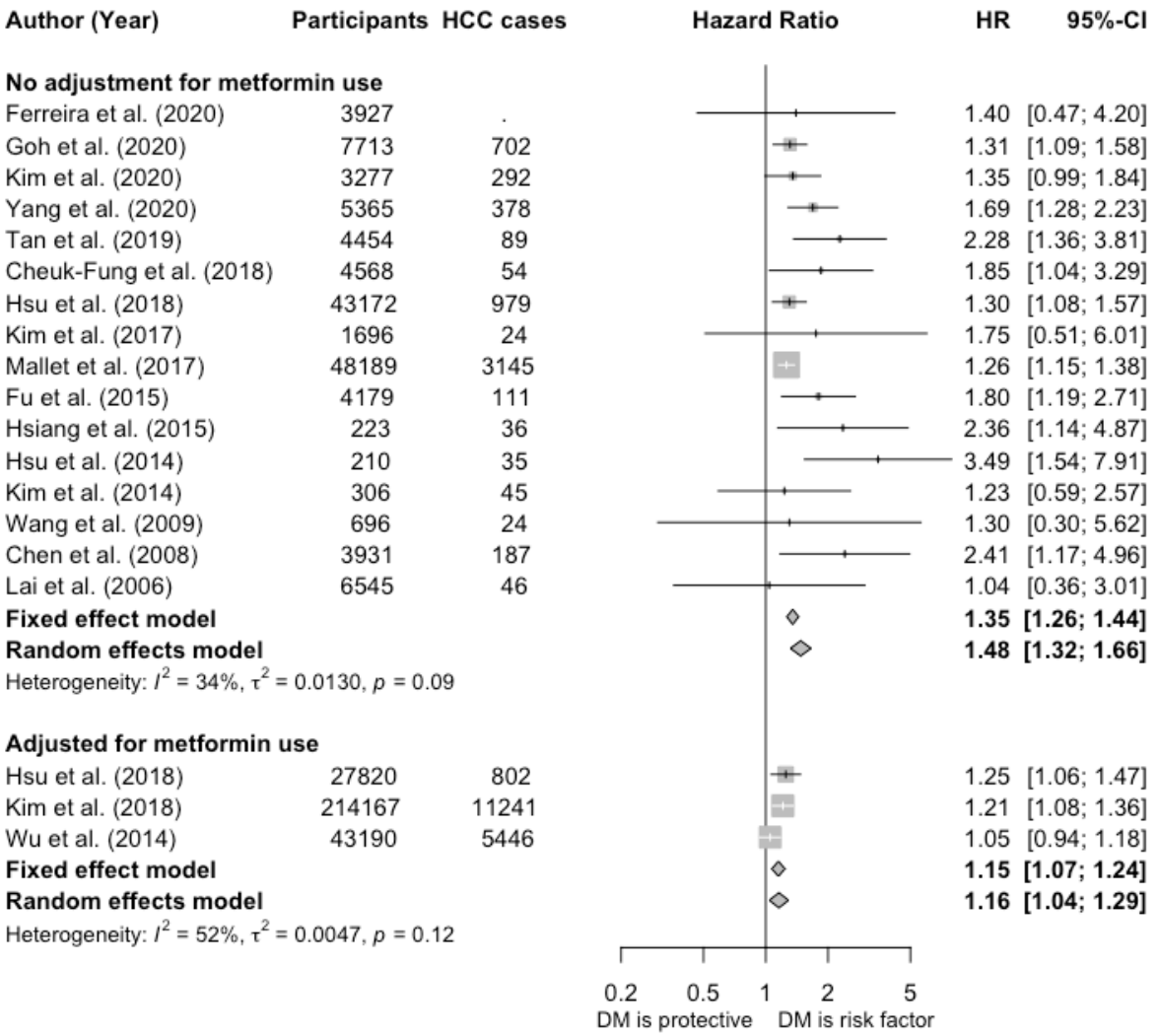


Figure 2.7. Sensitivity analysis restricting the primary meta-analysis to studies adjusting for use of metformin.
HR, hazard ratio; CI, confidence interval; DM, diabetes mellitus

2.5.6. Association of hypertension with hepatocellular carcinoma risk

Eleven studies investigated the association of HT with risk of CHB progression to HCC, one case-control study and 10 cohort studies (**Table 2.8**). All studies were mixed-sex samples in which mean/median age at baseline was ≥40 years (**Table 2.8**). Definitions of HT were heterogeneous; most studies ascertained HT via record linkage, but others used health assessment or interview. Few studies defined clinical thresholds for hypertension classification. “Higher” MAP was the primary exposure of interest in the case-control study, for which a threshold was not defined.

Table 2.8. Effect estimates for cohort studies investigating the association of hypertension with hepatocellular carcinoma risk.

Country, Author (Year)	Participants, n	HCC Cases, n	Age at baseline, years	Sex (% male)	Relative risk measure†	Risk Ratio (95% Confidence intervals)
UK, Ferreira (2020) (46)	3927	16-19	39.2 (mean, non-DM) 54.9 (mean, DM)	50	aHR	1.19 (0.95, 1.48)
France, Brichler (2019) (105)	317	27	53 (median)	82	uHR	2.91 (1.33-6.35)
Korea Cho (2019) (106)	826	86	52 (mean, NAFLD) 54 (mean, non-NAFLD)	61	aHR‡	1.38 (0.85-2.24)
China, Tan (2019) (108)	4454	89	45.4 (mean)	68	aHR	0.96 (0.60–1.53)
Taiwan, Hsu (2018) (111)	27,820	802	48.1 (median)	74	uHR	2.13 (1.85-2.45)
Hong Kong, Chan (2017) (115)	270	11	43.6 (mean)	75.2	uHR	2.33 (0.68-7.98)
US, Chayanupatkul (2017) (116)	8539	317	N/A	N/A	aHR‡	3.15 (1.02–9.75)
Korea, Kim (2017) (118)	1696	24	50 (median)	56.8	aHR	1.70 (0.71–4.05)
China, Gao (2013) (130) §	370	122	54.7 (mean)	86.9	aOR‡	N/A (p>0.1)

Taiwan, Hsu (2014) (126)	210	35	52.8 (median)	73.3	uHR	1.63 (0.74-3.60)
Taiwan, Chen (2008) (135)	3931	187	45.9	59.3	aHR	0.45 (0.16–1.21)

aHR, adjusted Hazards Ratio; uHR, unadjusted Hazards Ratio.

† Adjusted risk ratios are minimally adjusted for age and sex.

‡ Adjusted for age but not sex.

§ Case-control study using hospital controls.

Among 10 studies reporting hazards of HCC associated with HT, only three identified significantly increased risks, with two unadjusted and one adjusted for age. Another five studies reported an effect in the same direction, but effect sizes were not statistically significant. Adjusted HRs >1 ranged from 1.19 to 1.70 and <1 from 0.04 to 0.96. Adjustment for confounders was poor, with only four HRs minimally adjusted for age and sex.

2.5.7. Associations of other comorbidities with hepatocellular carcinoma risk

Seven studies investigated the association of dyslipidaemia with HCC risk in CHB patients (**Table 2.9**). All studies reported reduced risks of HCC in participants with dyslipidaemia as compared to those without, however only one HR was statistically significant. Clinical definitions of dyslipidaemia were often not reported, and only four studies minimally adjusted for age and sex. Six studies investigated the association of obesity with HCC risk (**Table 2.10**). Clinical definitions of obesity varied greatly, and out of four studies reporting increased risks of HCC with obesity, only one HR was statistically significant.

Table 2.9. Effect estimates for associations of dyslipidaemia with hepatocellular carcinoma risk.

Comorbidity	Relative risk measure†	Risk Ratio (95% CI)
Dyslipidaemia		
UK, Ferreira (2020) (46) ‡	aHR	0.77 (0.63, 0.95)
Korea, Cho (2019) (106)	uHR	0.60 (0.26-1.39)
China, Tan (2019) (108) ‡ §	aHR	0.67 (0.39–1.16)
China, Tan (2019) (108) ‡ ¶	aHR	0.49 (0.12–2.00)
US, Chayanupatkul (2017) (116) ‡	uHR	0.86 (0.38–2.00)
Taiwan, Hsu (2014) (126)	uHR	0.34 (0.05-2.50)
Taiwan, Wu (2014) (128) ‡ ¶	aHR	0.87 (0.61-1.23)

CI, confidence intervals; aHR, adjusted Hazards Ratio; uHR, unadjusted Hazards Ratio; BMI, body mass index; CVD, cardiovascular disease; IHD, ischaemic heart disease; ACS, acute coronary syndrome; NAFLD, non-alcoholic fatty liver disease; COPD, chronic obstructive pulmonary disease.

† Adjusted risk ratios are minimally adjusted for age and sex.

‡ Defined specifically as hyperlipidaemia.

§ Defined specifically as hypertriglyceridaemia.

¶ Defined specifically as hypercholesteraemia.

Table 2.10. Effect estimates for associations of obesity with hepatocellular carcinoma risk.

Comorbidity	Relative risk measure†	Risk Ratio (95% CI)
Obesity (BMI ≥ 30 kg/m^2)		
France, Brichler (2019) (105)	aHR	2.67 (1.04-6.84)
US, Chayanupatkul (2017) (116)	aHR ‡	2.21 (0.75–6.56)
Obesity (BMI > 30 kg/m^2)		
China, Tan (2019) (108)	aHR	1.59 (0.73–3.44)
Obesity (BMI ≥ 27.5 kg/m^2)		
Netherlands, Brouwer (2015) (123)	uHR	1.90 (0.60-6.20)
Obesity (BMI ≥ 27 kg/m^2)		
Taiwan, Chen (2013) (129)	uHR	0.84 (0.44–1.60)
Obesity (BMI ≥ 25 kg/m^2)		
Korea, Lee (2016) (122)	uHR	0.90 (0.17-4.62)

CI, confidence intervals; aHR, adjusted Hazards Ratio; uHR, unadjusted Hazards Ratio; BMI, body mass index; CVD, cardiovascular disease; IHD, ischaemic heart disease; ACS, acute coronary syndrome; NAFLD, non-alcoholic fatty liver disease; COPD, chronic obstructive pulmonary disease.

† Adjusted risk ratios are minimally adjusted for age and sex.

‡ Adjusted for age but not sex.

Three studies investigated the association of statin use with HCC risk in CHB (**Table 2.11**). All studies reported HRs <1, and two of these HRs were statistically significant. HRs reported in 5 studies for HCC risk associated with CVD varied, likely due to the variable definitions of CVD used across studies. Associations for other variables, including respiratory disease and renal disease, were reported by <2 studies each (**Table 2.12**).

Table 2.11. Effect estimates for associations of statin use and cardiovascular disease with hepatocellular carcinoma risk.

Comorbidity	Relative risk measure†	Risk Ratio (95% CI)
Statin use		
UK, Ferreira (2020) (46)	aHR	0.36 (0.19, 0.68)
Hong Kong, Yip (2020) (104)	aHR	0.81 (0.73-0.90)
Hong Kong, Yip (2018) (104)	aHR	0.52 (0.26–1.01)
CVD		
Korea, Cho (2019) (106) – CVD	uHR	1.46 (0.64-3.34)
US, Chayanupatkul (2017) (116) – IHD	uHR	1.19 (0.39–3.65)
France, Mallet (2017) (119) - CVD	aHR	0.57 (0.53–0.62)
Taiwan, Wu (2014) (128) - ACS	aHR	0.66 (0.55-0.79)
Taiwan, Wu (2014) (128) – Cerebrovascular disease	aHR	0.50 (0.40-0.64)

CI, confidence intervals; aHR, adjusted Hazards Ratio; uHR, unadjusted Hazards Ratio; BMI, body mass index; CVD, cardiovascular disease; IHD, ischaemic heart disease; ACS, acute coronary syndrome; NAFLD, non-alcoholic fatty liver disease; COPD, chronic obstructive pulmonary disease.

† Adjusted risk ratios are minimally adjusted for age and sex.

Table 2.12. Effect estimates for associations of other factors with hepatocellular carcinoma risk.

Comorbidity	Relative risk measure†	Risk Ratio (95% CI)
Other		
Hong Kong, Yip (2020) (104) - Thiazolidinedione use)	aHR	0.57 (0.28-1.14)
Hong Kong, Yip (2020) (104) - Metformin use	aHR	0.93 (0.85-1.02)
Hong Kong, Yip (2020) (104) - Sulphonylurea use	aHR	1.19 (1.09-1.30)
Hong Kong, Yip (2020) (104) - Insulin use	aHR	1.32 (1.19-1.46)
Hong Kong, Yip (2020) (104) - Aspirin/clopidogrel use	aHR	0.76 (0.68-0.85)
Korea, Cho (2019) (106) - NAFLD	aHR	1.67 (1.05-2.63)
France, Mallet (2017) (119) - Renal disease (off dialysis, no transplant	aHR	0.34 (0.28–0.41)
France, Mallet (2017) (119) - Renal disease (on dialysis, no transplant)	aHR	0.36 (0.22–0.60)
France, Mallet (2017) (119) - Respiratory disease	aHR	0.37 (0.33–0.42)
Taiwan, Yu (2017) (121) - Metabolic risk factors ‡	aHR	HRs for obese or diabetic, obese and diabetic, and ≥3 metabolic risk factors were 1.29 (95% CI 0.92-1.81), 1.18 (0.16-8.54) and 2.61 (1.34-5.08), respectively, compared to the non-obese and non-diabetic reference group.
Taiwan, Wu (2014) (128) COPD	aHR	0.71 (0.59-0.87)
Taiwan, Wu (2014) (128) - Renal failure	aHR	0.84 (0.68-1.04)
Taiwan, Chen (2013) (129) - Metabolic syndrome	uHR	1.19 (0.63–2.26)

CI, confidence intervals; aHR, adjusted Hazards Ratio; uHR, unadjusted Hazards Ratio; BMI, body mass index; CVD, cardiovascular disease; IHD, ischaemic heart disease; ACS, acute coronary syndrome; NAFLD, non-alcoholic fatty liver disease; COPD, chronic obstructive pulmonary disease.

† Adjusted risk ratios are minimally adjusted for age and sex.

‡ Metabolic risk factors (obesity, diabetes, hypertriglyceridemia and HT), with exposure groups split into groups of 0, 1, 2 and ≥3 risk factors.

2.6. Discussion

2.6.1. Summary of findings

My meta-analysis suggests that DM is a risk factor for HCC in CHB infected individuals, with hazards of HCC >20% higher in the presence of DM; however, I report significant between-study heterogeneity. This association did not materially change after restriction to studies adjusting for relevant confounders but did suggest a favourable impact of DM treatment with metformin. Pooled effect estimates remained significant in sensitivity analyses. Few studies investigated other comorbidities, and some comorbidity search terms included in the systematic literature search returned few or no results. This highlights the need for future investigation of these comorbidities, as antiviral treatment cannot eliminate the risk of HCC entirely and therefore novel risk factors must be identified in order to inform interventions. Although EASL (13) and APASL (15) guidelines recognise the association with DM, it is not currently consistently described in other recommendations (e.g. AASLD guidelines (14) do not list DM as a risk factor for HCC).

2.6.2. Critical appraisal of included studies – risk of bias, confounding, heterogeneity & measurement error, and comparison with previous meta-analysis

Some studies investigating comorbidities and their metabolic risk factors reported significantly reduced hazards of participants with these conditions as compared to those without. This association may be confounded by the requirement for treatment in secondary care, whereby CHB-infected individuals may be more likely to receive screening and antiviral treatment.

Findings from case-control and cohort studies were not consistent; whilst the majority of cohort studies reported increased risks of HCC associated with DM, case-control findings were inconsistent, and indeed three studies reported a significant reduction of HCC risks in association with DM. Explanations for such findings including

confounding, selection bias associated with the study of hospital control groups that enrich for DM (130,131), and chance, especially in small studies (112,130–132).

My findings are consistent with a previous meta-analysis (138); I provide a comprehensive review of all cohort studies and include a larger number of studies. I restricted to studies reporting HRs minimally adjusted for age and sex. However, adjustment for covariates and inclusion criteria varied considerably between studies, and this may explain some of the between-study heterogeneity. Substantial heterogeneity remained in sensitivity analyses restricted to studies adjusting for additional key confounders, as adjustment for confounders was variable within these studies and populations may not have been comparable. Although baseline age and sex characteristics were comparable across studies, there was variability regarding exclusion of those with additional comorbidities and those on antiviral treatment.

It Is possible that varied lengths of follow-up also contributed to between-study heterogeneity, although HRs did not significantly vary with length of follow-up in sensitivity analysis. This is because cancer is a chronic disease with a slow development, and preclinical disease can be present for many years before clinical manifestation; follow-up times <10 years may be insufficient to detect HCC outcomes. I was unable to provide effect estimates across most potential patient subgroups because the subgroups contained small numbers of studies, putting subgroup analyses at greater risk of chance findings as well as being subject to the influence of multiple testing.

Additional study characteristics which may have potentially contributed to between-study heterogeneity are worthy of discussion. There was substantial diversity across cohorts regarding study location, with the included studies undertaken across 11 countries and 4 continents. There are multiple factors that vary with geography, both within and between countries, which may influence the association of DM with HCC. One such factor is mode of HBV transmission (occurring via perinatal exposure or contact with infectious blood/bodily fluids healthcare and/or sexual settings varies across countries (1)) resulting in regional variations in length of infection for individuals who develop CHB. Longer durations of infection may influence the hepatic pathophysiological mechanisms of comorbidities, such as DM, which influence HCC

risk. Additional factors which differ greatly across countries include socioeconomic deprivation, access to healthcare and clinical guidelines/practices. These factors impact how and if both comorbid diseases are diagnosed and treated, which may in turn impact how they associate with HCC risk in individuals with CHB.

The sampling references from which cohorts were selected also varied widely, potentially contributing to observed between-study heterogeneity. Inclusion criteria ranged from a broad inclusion of any individual with CHB, to individuals with CHB identified in hospitals/medical centres, to those attending healthcare facilities to undergo liver biopsy. Therefore studies included individuals at variable points along the CHB disease pathway. Furthermore, some studies restricted to treated, untreated or treatment-naïve samples. Treatment status impacts CHB by inhibiting progression of viral pathophysiology in the liver, possibly influencing the mechanisms by which comorbid disease affect HCC risk. These study attributes may have potentially contributing to observed heterogeneity.

The association I report in this meta-analysis is weaker than those observed in patients with chronic HCV infection. In previous studies of individuals with chronic HCV infection, risk of HCC was elevated ~2-fold in the presence of DM (139,140). Previous studies also report increased risks of DM in HCV-infected individuals as compared to non-infected individuals (141–144). However, this is likely due to the various extra-hepatic manifestations of HCV which are not present in HBV infection.

2.6.3. Diabetes mellitus and HCC risk in CHB

I noted variable definitions of DM, with some studies restricting investigation to type 2 DM whereas others included participants with unspecified DM. Risk factors for types 1 and 2 diabetes mellitus vary, and heterogeneity in DM definitions could therefore contribute to variable study populations and outcomes. Global prevalence and incidence estimates for specific DM types do not exist, as distinguishing between types often requires laboratory resources that are not available in many settings. However most cases of type 1 diabetes are found in Europe and North America, and the large

majority of studies included in this systematic review and meta-analysis were conducted in Asian countries.

In my sensitivity analysis restricted to studies adjusting for cirrhosis, the observed association of DM with HCC was attenuated towards the null. This may be explained by a confounding of the association by cirrhosis, accounted for by an independent association of cirrhosis with both DM and HCC, and the absence of cirrhosis from the causal pathway that associates DM with HCC. However, if cirrhosis is located along this causal pathway, then it can be characterised as a mediator rather than a confounder. If cirrhosis is a mediator then adjusting for it would be incorrect.

Past studies support a positive association of DM with HCC risk in non-CHB patients (83,84,145), and aetiological investigation has suggested that DM can lead to cirrhosis and thereby HCC via fatty liver disease (146), as a result of accumulation of fatty acids causing oxidative stress driving inflammation and tissue necrosis (147,148), and longer-term fibrosis and cirrhosis, thereby increasing HCC risk. However alternative pathways causally associating DM with HCC have been suggested, including increased hepatocyte proliferation induced by hyperinsulinaemia (149,150) and production of pro-inflammatory cytokines that increase cell survival via apoptosis inhibition (151–153). It is possible that multiple disease pathways associating DM with HCC operate simultaneously. Elucidation of the aetiological mechanisms underpinning this association will inform future epidemiological studies and disease management. Characterising the impact of glycaemic control on HCC risk is also an important question for future research.

Three studies adjusted for metformin use (111,113,128) , and in sensitivity analysis restricted to these studies, the association between DM and HCC remained significant but was attenuated towards the null. It is not known the extent to which this is a result of glycaemic regulation by metformin, accomplished by inhibition of hepatic gluconeogenesis and improvement of insulin sensitivity in tissues leading to reduced oxidative stress in the liver (154), and/or a direct impact of metformin in reducing cancer risk via regulation of cellular signalling. Evidence from observational studies (155–157) and randomised controlled trials (RCTs) (158) supports a protective effect of metformin against the development and progression of cancer in diabetic

individuals. There is also some RCT evidence for protective effects of metformin against progression of certain cancer types in non-diabetic individuals (159) although this is not consistent. Multiple large-scale phase III RCTs are currently underway (160–163) and will provide further information regarding the roles of DM and metformin in cancer development.

I included all studies investigating the association of comorbidities with risk of CHB progression to HCC that minimally adjusted for age and sex in order to provide a comprehensive review of available evidence. However, few studies investigated non-DM comorbidities, preventing meta-analysis for these comorbidities. Additionally I was unable to restrict the meta-analysis of DM and HCC to studies adjusting for confounders other than age and sex, as few studies minimally adjusted for all relevant factors. Publication bias may influence the outcome, as I restricted the search to peer-reviewed literature, and studies that do not report an association of DM with HCC may be less likely to be published. These results may not be generalisable to the global CHB population, as there were a limited number of studies from non-Asian countries. The lack of studies from any African countries is of concern, given that the region carries both the highest HBV prevalence (164) and largest mortality burdens for cirrhosis and HCC (165,166), and that risk factors and disease phenotypes may differ in African cohorts.

My finding that DM is a risk factor for HCC in CHB-infected individuals suggests that enhanced cancer surveillance may be justified in patients with CHB and DM to enable early detection and treatment. Improvements in guidelines could help to inform more consistent approaches to risk reduction. After adjustment for metformin use, this association remained significant, but was attenuated suggesting a potential benefit of metformin that warrants further study. Ongoing investigation is required in order to identify and characterise risk factors for HCC, to extend these analyses to diverse global populations, and to elucidate disease mechanisms in order to inform prevention, screening and therapeutic intervention.

2.6.4. Impact of meta-analysis findings

My findings have provided a robust summary of the previous literature investigating risk factors for progression of CHB to HCC. This demonstrates the value and impact of the results to CHB research, and the importance of the investigation in enabling further investigation regarding risk of end-stage disease progression in chronic infection.

2.6.5. Systematic search update – 2020 to 2023

Since I have published this paper, a further seven relevant cohort studies have been produced (based on repeat search using terms in **Table 2.1** between 24th June 2020 until 15th April 2023) (167–173). Studies investigated risk factors for HCC in both ethnically homogenous (168,170–173) and diverse (167,169) cohorts. No novel risk factors for progression of CHB to HCC were identified. Findings across studies were concordant with those reported in the meta-analysis. Specifically, diabetes mellitus (168,170–173) HT (168,170,172), obesity (168,170,172) and dyslipidaemia (168,170,172) were reported as HCC risk factors.

These more recent studies support findings from the systematic review and meta-analysis, providing further evidence that the presence of comorbid (particularly metabolic) disease in CHB adverse associates with HCC risk. Further investigation is warranted in order to elucidate mechanisms underpinning the observed associations, and to identify potential clinical intervention points along the pathophysiological pathway. Ultimately, these findings can inform clinical management, resource allocation and end-stage disease prevention in CHB.

3. Chronic HBV cohort in the QResearch primary care database: a summary/appraisal of UK-wide CHB prevalence estimates, characterisation of the CHB population and investigation of risk factors for hepatocellular carcinoma

3.1. Availability of content

I herein present findings from a cohort of CHB individuals I have characterised in the QResearch primary care database. This investigation is published in *Public Health* (47). I also present findings from a literature review I undertook to summarise available UK CHB prevalence estimates. This investigation has been peer-reviewed and published on *Wellcome Open Research* (41).

3.2. Background

3.2.1. HBV in the UK: Epidemiology and disease burden

Despite the substantial disease burden of CHB in the UK, many evidence gaps exist surrounding its epidemiology and impact as a risk factor for HCC. CHB is one of four main risk factors for HCC in the UK, along with HCV, alcohol consumption and obesity/metabolic syndrome which together account for >90% of HCC cases (52). The morbidity and mortality burdens of HCC in the UK are disproportionately high compared to the incidence of HCC; while HCC has the 18th highest incidence amongst all cancers in the UK, it is the 8th commonest cause of cancer death (174). Despite the modifiable nature of CHB (whereby it is both preventable and treatable, albeit not curable), its attributable disease burden persists, largely due to the lack HBV diagnosis/detection and subsequent antiviral treatment in the UK. Specifically, evidence gaps exist regarding the demographics, access to and uptake of health services, and burden of other comorbid disease amongst the CHB population. Furthermore, a robust estimate for HBV prevalence in the UK has not been produced to date.

The UK Health Security Agency (UKHSA, formerly Public Health England (PHE)) has produced individual reports (updated in 2020) detailing HBV epidemiology across the following UK regions: Yorkshire and the Humber, the North West, the South West, the West Midlands, the South East and London (43), using data from 2016/17. UKHSA reports utilise SGSS and SSBBV surveillance findings from laboratory data (previously detailed in Chapter 1) to summarise new HBV diagnoses and HBV test positivity (primarily HBsAg). Laboratory data contributing from these surveillance systems come from various sources including emergency departments, blood and transplant screening, sexual health services and prison services, in addition to primary and secondary care testing, Overall HBsAg positivity frequencies (reported in government sources, aiming for data that reflect general population samples on which to base a prevalence estimate) is reported, as well as frequencies stratified by selected characteristics, e.g. sex, ethnicity, upper-tier local authority, PWID, pregnancy. However details descriptions of regional CHB sub-populations, for example description of socioeconomic demographics and burdens of comorbid diseases, are not reported. Additionally no information concerning antiviral treatment eligibility and coverage is provided.

There has been increasing interest in identifying risk factors for progression of CHB to cirrhosis and HCC. Age, sex, HBV viral load and viral genotype are recognised determinants of HCC risk (74–79,175), and recent studies have reported associations between HCC and various comorbidities, including type 2 diabetes mellitus (T2DM) and hypertension (80). Existing risk scores, including PAGE-B (72), REACH-B (73,74), GAG-HCC (75,76) and CU-HCC (77–79), use such risk factors to predict HCC However, most studies focus on sub-groups of the CHB population, including in ethnically homogenous sampling references and hospital settings. Few cohorts have been characterised in European countries and/or in ethnically diverse populations. Therefore, findings may not be generalisable, especially not to ethnically and socioeconomically diverse settings. Additionally, the development of risk scores in secondary care/hospital settings poses difficulties to their broad-scale validation and implementation on a population-wide scale.

In addition to the lack of knowledge surrounding characteristics of the UK CHB population and the magnitude of the disease burden, an evidence gap concerning

HCC risk factors also exists in the UK. The most recent prevalence estimate produced by UKHSA is based on methodology reliant on HBsAg positivity rate amongst individuals with an available measurement (30). Therefore, this estimate does not account for positive undiagnosed individuals (who contribute to the CHB burden but are not accessing any healthcare or surveillance services). These marginalised individuals may be substantial contributors to the overall disease burden and are also at high risk of developing end-stage disease complications (16). Estimates produced by NICE (176) and Department for Health and Social Care (177,178) reported ≥10 years ago rely on similar methodology and have also not been updated to reflect the current population at risk of CHB in the UK. Similarly, disease population subgroups at the highest risk of end-stage disease have not been well characterised.

3.2.2. QResearch primary care EHR database

Primary care electronic health record (EHR) databases are increasingly being used to characterise the burden and epidemiology of various diseases in the UK. Studies based on EHRs enable retrospective characterisation of large populations. This not only enhances statistical power, but also allows for identification of a study sample which is more representative of an entire disease population compared to many clinical trials. EHR databases in the UK are strengthened by long durations of follow-up, with exposures and outcomes often re-measured longitudinally. Such databases can also usually be linked to additional data sources (such as national cancer registries and vital statistics), allowing for ascertainment of a sufficient number of relevant endpoints to power investigations.

The English QResearch primary care EHR database was established in 2003 by Professor Julia Hippisley-Cox, with ethics approval granted by the East Midlands Derby Research Ethics Committee (reference 18/EM/0400). Originally founded at the University of Nottingham, it is now hosted within the Nuffield Department of Primary Care Health Sciences as a not-for-profit collaboration between the University of Oxford and EMIS, one of the main suppliers of primary care IT systems. The database contains the records of >35 million individuals from >1800 general practices across England, for whom records date back as far as 1989. Decisions regarding primary

care data are taken by the QResearch management board. The board convenes multiple times in a year to direct strategy/mission, governance/security, resource and data management for the database.

The QResearch database contains only anonymised individual-level patient data. Data subsets for specific projects are deposited on a Trusted Research Environment (TRE) for analysis by researchers. A range of variables are available for patients including demographic information, clinical and physiological measurements (including blood pressure, height, and weight), GP referrals to secondary or tertiary care services, prescriptions made in the primary care setting, test results (e.g. full blood count) and clinical events which are ICD-10 coded (including diagnoses and endpoints) (179).

Patient-level data in QResearch may be linked to a range of other data sources. From 2011 onwards, this includes mortality data from the ONS, HES records from secondary/tertiary healthcare settings and the National Cancer Registration and Analysis Service (NCRAS). There is a monetary cost associated with accessing QResearch data, with the specific amount dependent on various project parameters including the size of the sample required, the number of primary care parameters required for analysis and linkage of the data to additional sources.

3.2.3. Aims and objectives

Given the substantial evidence gaps concerning HBV epidemiology, disease burden and risk factors for progression to HCC, I set out to summarise and appraise UK-wide CHB prevalence estimates which have been produced to date, identify and describe a CHB cohort by collating data from a large-scale primary care database in England (180), and investigate risk factors for progression of CHB to HCC in this cohort.

3.3. Research questions

3.3.1. Research question 1: What is the prevalence of CHB in the UK? A summary and appraisal of previous national CHB prevalence estimates

3.3.2. Research question 2: An investigation of demographics/characteristics and risk factors for HCC in the UK CHB population, undertaken in a primary care EHR database (QResearch).

3.4. Methods

3.4.1. Methods – Research question 1

I searched for estimates of CHB case numbers in the UK (incorporating incidence and/or prevalence-like data) across a range of available sources in a narrative (rather than systematic) review approach. This approach was undertaken in order to review evidence rapidly, and because a systematic approach was not feasible due to insufficient data quality and quantity across many studies whereby only a handful of large-scale investigations have been carried out to date. I included UK-wide reports from government bodies, publications from independent bodies (including medical charities and non-governmental organisations) and articles in peer-reviewed scientific journals, excluding smaller epidemiological studies when there was overlap of data sources with larger samples from other studies. HBV 'positivity rates' are presented from each respective data source, but should be interpreted with caution given that these estimates are not representative of the true UK-wide population prevalence or incidence. Details of study samples/denominator are provided in the Results section. I used the Office for National Statistics (ONS) UK population estimates as a point of reference for the overall denominator (181–183). I also generated a new prevalence estimate, utilising data from the CHB cohort I characterised in QResearch. We have also drawn on findings from a similar investigation previously undertaken in the Clinical Practice Research Datalink (CPRD) (46), which is another UK primary care database containing EHRs for over 16 million patients. This previous investigation identified people living with CHB from all those registered in the database between 2000 and 2015.

3.4.2. Methods – Research question 2

Data source and study population/design

I used data from the QResearch database (version 45), utilising linkage to NCRAS to maximise the number of HCC cases identified for analysis. I first identified individuals from the QResearch database between 01 January 1999 and 31 December 2019 who were age ≥18 years and had confirmed CHB based on any one of the following conditions: (i) record of a diagnostic SNOMED/Read code indicating CHB; (ii) record of a diagnostic International Classification of Disease (ICD)-9 or -10 code indicating CHB, and/or; (iii) Presence of HBsAg or VL on at least two recordings ≥6 months apart (**Figure 3.1**).

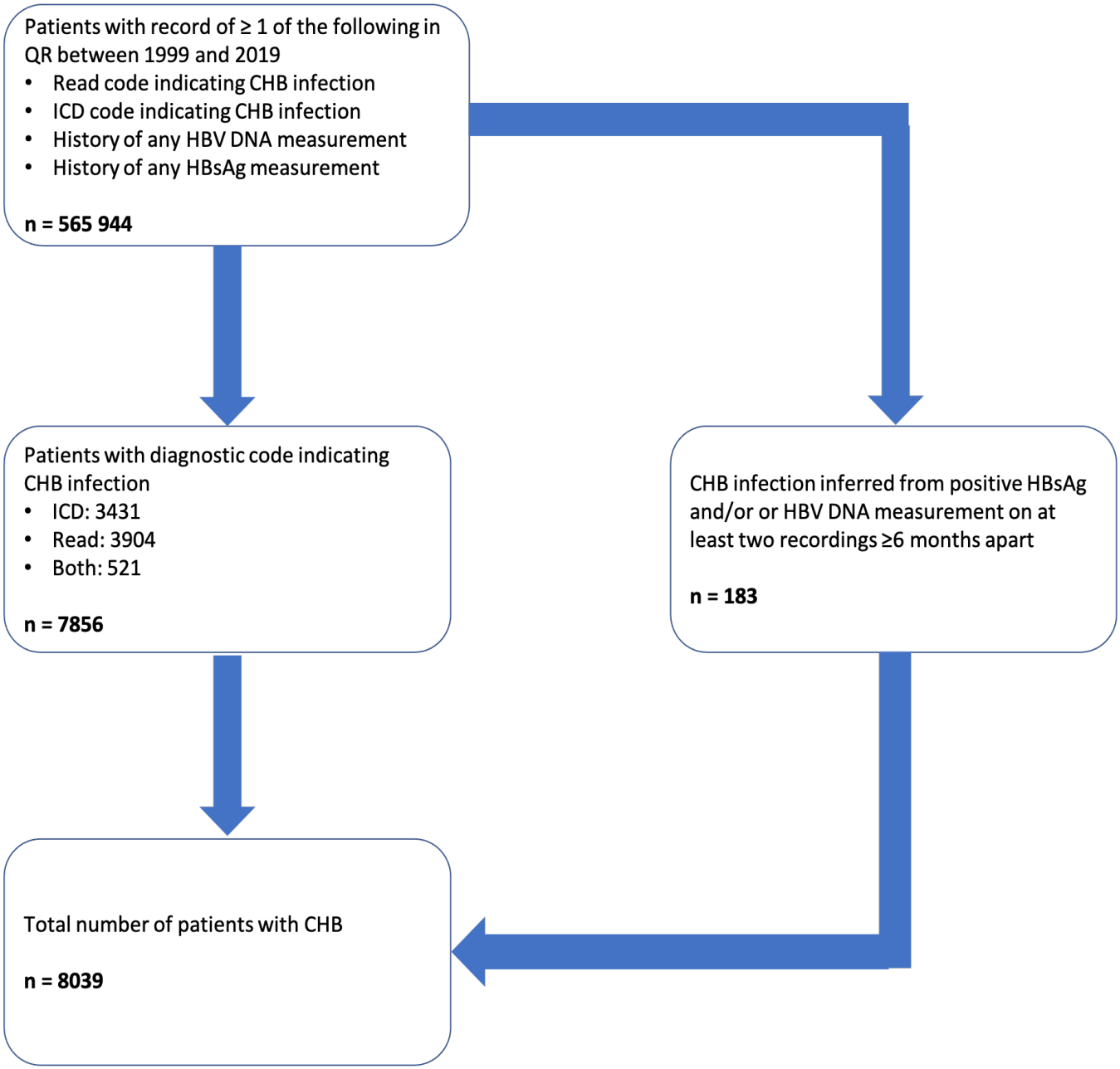


Figure 3.1. Study flowchart depicting identification of adults from the QResearch (QR) primary care database eligible for inclusion an analysis of chronic hepatitis B (CHB) infection.
HBV DNA = hepatitis B virus DNA viral load; HBsAg = hepatitis B surface antigen; ICD = international classification of disease

Covariate selection and ascertainment

I identified relevant covariates for extraction *a priori* based on previous literature (46,47,72,184–189). Clinical relevance and a previous meta-analysis I undertook to

investigate risk factors for HCC in CHB was also used to inform variable selection (79) (**as presented in Chapter 2**). Along with CHB, various other lifestyle factors or diseases have been suggested to influence HCC risk, including obesity, and associated cardiometabolic diseases (including DM, steatotic liver disease (SLD) and metabolic liver disease (MLD), identified in Read/ICD coding as NAFLD, as well as hypertension (HT). Specifically, hyperglycaemia and HT resulting from these conditions may result in direct physiological damage to hepatocytes, ultimately leading to cirrhosis and HCC. Chronic kidney disease (CKD) was also identified via meta-analysis as a comorbidity associating with HCC risk via hypertension and resultant cirrhosis. Alcohol and cigarette consumption have both been identified as risk factors for HCC development by the WHO's international Agency for Research on Cancer (190,191), whereby consumption causes direct pathophysiological damage to hepatocytes. Finally comorbid conditions were found to associate with cirrhosis and thereby HCC risk in my meta-analysis, including IHD and cerebrovascular disease.

Lifestyle factors, demographics and relevant numeric biomarkers were characterised from relevant SNOMED/Read codes. Comorbidities were characterised from relevant SNOMED/Read and ICD-9 and -10 codes. Ethnicity is categorised in QResearch as per 2011 census categories (192). Townsend Deprivation quintile is available as a measure of socioeconomic status in QResearch. Body mass index (BMI, kg/m^2) was categorised (underweight, <18.5 kg/m^2; healthy weight, 18.5-2.49 kg/m^2; overweight, 25.0-29.9 kg/m^2; obese, ≥ 30 kg/m^2) based on World Health Organization (WHO) categories (193). Covariate measurements made within ±3 years of earliest CHB diagnosis and before HCC diagnosis were used as proxy baseline measurements. Where patients had >1 measurement taken within 3 years of the earliest CHB diagnosis, measurements taken closest to diagnosis date were used.

Outcome ascertainment

The primary endpoint of interest was HCC, ascertained via SNOMED/Read or ICD codes corresponding to HCC, and by linkage of the cohort to NCRAS (62,194). To maximise outcome ascertainment, I used multiple codes relevant to HCC (**Table 3.1**). I performed sensitivity analysis (further details below) whereby all patients with non-

HCC neoplasms were excluded, to investigate robustness of main analysis using a broad HCC definition. A tabulation of HCC cases across source of diagnosis is presented in **Table 3.2**.

Table 3.1. Systemised Nomenclature of Medicine primary care codes indicative of hepatocellular carcinoma (HCC) used to identify cases of HCC in adults with CHB from the QResearch database.

SNOMED ConceptID	SNOMED Fully Specified Name
92644006	Carcinoma in situ of liver (disorder)
93870000	Malignant neoplasm of liver (disorder)
95214007	Primary malignant neoplasm of liver (disorder)
109841003	Liver cell carcinoma (disorder)
187767006	Malignant neoplasm of liver and intrahepatic bile ducts (disorder)
187769009	Primary carcinoma of liver (disorder)
271525004	Carcinoma in situ of liver and/or biliary system (disorder)
620441000000107	Malignant neoplasm of liver and intrahepatic bile ducts NOS (disorder)

SNOMED, Systemised Nomenclature of Medicine.

Table 3.2. Hepatocellular carcinoma (HCC) cases stratified by source of diagnosis.

Source of HCC diagnosis	Number of cases
ICD	20
NCRAS	22
SNOMED/Read	24
ICD and NCRAS	41
ICD and SNOMED/Read	15
NCRAS and SNOMED/Read	13
ICD and NCRAS and SNOMED/Read	75

HCC, hepatocellular carcinoma; ICD, international classification of disease; NCR, National Cancer Registry and Analysis Service; SNOMED, Systemised Nomenclature of Medicine.

Follow-up

Earliest date of CHB diagnosis was regarded as cohort entry and initiation of follow-up for each individual. For patients who developed HCC, date of HCC diagnosis was regarded as the end of follow-up. For those who did not develop HCC (i.e., patients

who were censored), follow-up ended at cohort exit date (either due to leaving their general practice and switching to a practice which does not contribute to QResearch, or death) or 31 December 2019, whichever occurred earlier. Patients in whom database exit date preceded or was equal to first recorded CHB diagnosis date (n = 1010) whereby follow-up time ≤0 years were excluded from longitudinal analysis.

In some patients, HCC diagnosis date or cohort exit date preceded or was equal to CHB diagnosis date (**Table 3.3**). Data from these individuals were excluded from analyses of HCC risk factors.

Table 3.3. Follow-up time, stratified by source of chronic hepatitis B virus infection diagnosis.

Source of CHB diagnosis	**Follow-up time (years), mean (SD) or median (IQR)***	**CHB diagnosis date precedes cohort exit date OR HCC diagnosis date, n**	**Cohort exit date precedes CHB diagnosis date, n**	**HCC diagnosis date precedes CHB diagnosis date, n**	**Cohort exit date equates to CHB diagnosis date, n**	**HCC diagnosis date equates to CHB diagnosis date, n**
Total	5.12 (4.9)	7029	960	44	<5	5
ICD	7.0 (4.9)	2430	958	29	0	<5
ICD and laboratory test	7.1 (3.0)†	12	0	0	0	0
ICD and SNOMED/Read	8.8 (7.3)	516	0	<5	0	<5
Laboratory test	5.8 (2.5)	183	0	0	0	0
SNOMED/Read code	5.0 (5.6)	3831	≤5	11	<5	<5
SNOMED/Read code and laboratory test	8.7 (6.2)	57	0	0	0	0

SD, standard deviation; IQR. Interquartile range; CHB, chronic hepatitis b virus; HCC, hepatocellular carcinoma; ICD, International classification of disease; SNOMED, Systemised Nomenclature of Medicine.

*Mean and SD or median and IQR calculated for patients for whom CHB diagnosis date precedes cohort exit date OR HCC diagnosis date

†Median and IQR presented

Statistical analysis

Statistical analyses were carried out in R (version 4.1.0). Baseline characteristics were summarised for all CHB patients (regardless of length of follow-up) using descriptive statistics. Means and standard deviations (SDs) or medians and interquartile ranges (IQRs) were presented for continuous measures and were compared using *t* or Wilcoxon rank-sum tests, respectively. Patient counts and percentages were presented for continuous and binary variables and were compared using chi-squared or Fisher's exact tests.

Univariable and multivariable Cox proportional hazards models were used to investigate risk factors for progression of CHB to HCC. Variables were included in the multivariable model based on significance of univariable associations (where $P \leq 0.1$) and/or based on biological/clinical relevance and previous literature (46,47,72,184–189).

Continuous laboratory parameters which were right-skewed were transformed with a natural logarithm for inclusion in multivariable models. Laboratory parameters were divided into quintiles for inclusion in multivariable models. Means and SDs for log AST, log alanine transaminase (ALT) and platelet count (Plt) quintiles are presented in **Table 3.4**. Hazard ratios and 95% confidence intervals (95% CI) were reported for Cox proportional hazards model outputs. Analysis on the imputed dataset was used for main models.

Table 3.4. Quintiles for natural log transformed alanine transaminase (ALT), natural log transformed aspartate transaminase (AST) and platelet count (Plt). ALT and AST were natural log transformed due to their skewed distributions.

	Quintile				
Characteristic	**First**	**Second**	**Third**	**Fourth**	**Fifth**
Log ALT					
N (%)	1429	1416	1400	1408	1376
Mean log ALT (SD)	2.69 (0.23)	3.12 (0.09)	3.45 (0.10)	3.85 (0.14)	4.79 (0.64)

Univariable HR (95% CI)	0.89 (0.50 to 1.60)	1.24 (0.73 to 2.11)	1.00 (ref)	1.19 (0.70 to 2.04)	2.17 (1.35 to 3.49)**
Log AST					
N (%)	1378	1415	1420	1409	1407
Mean log AST (SD)	2.89 (0.14)	3.18 (0.07)	3.39 (0.06)	3.67 (1.12)	4.55 (0.69)
Univariable HR (95% CI)	1.11 (0.66 to 1.88)	0.95 (0.55 to 1.65)	1.00 (ref)	1.16 (0.69 to 1.96)	1.96 (1.23 to 3.12)**
Plt					
N (%)	1410	1402	1393	1422	1402
Mean PL (SD)	4.80 (0.37)	5.21 (0.05)	5.37 (0.04)	5.52 (0.05)	5.76 (0.15)
Univariable HR (95% CI)	4.18 (2.50 to 7.00)**	1.18 (0.63 to 2.22)	1.00 (ref)	1.18 (0.63 to 2.21)	1.38 (0.75 to 2.54)

ALT, alanine aminotransferase; SD, standard deviation; HR, hazards ratio; CI, confidence intervals; AST, aspartate aminotransferase; Plt, platelets.
* $P < 0.05$
** $P < 0.01$

Handling of missing data

Values were missing for Townsend Deprivation Quintile, ethnicity, alcohol consumption, cigarette consumption, BMI, Plt, ALT measurement, aspartate transaminase (AST) measurement, Hepatitis B surface antigen (HBsAg), and HBV viral load (VL) **(Table 3.5)**.

Table 3.5. Missingness in CHB data in QResearch cohort of adults according to cohort characteristics.

Characteristic	**Missingness, n (%)**
Age	0 (0%)
Sex	0 (0%)
Townsend Deprivation Quintile	22 (0.3%)
Ethnicity	1289 (16.0%)
Alcohol consumption	3527 (43.9%)
Cigarette consumption	2287 (28.5%)
BMI	3111 (38.7%)

Characteristic	Missingness, n (%)
ALT	3497 (43.5%)
AST	6370 (79.2%)
Plt	3217 (40.0%)
HBsAg	7964 (99.1%)
VL	7656 (95.2%)

The missingness for HBV laboratory tests (VL and HBsAg) may be explained by the measurement of these parameters in secondary/tertiary care centres rather than primary care centres. Therefore the missingness presented here does not indicate a lack of measurement in CHB patients, but rather is an artefact of the existence of such laboratory data in secondary/tertiary care electronic systems rather than primary care systems.

BMI, body mass index; ALT, alanine transaminase; AST, aspartate transaminase; Plt, platelets; HBsAg, hepatitis B virus surface antigen; VL, hepatitis B virus viral load.

Multiple imputation by chained equations (MICE) was used to impute missing data. The assumption of missing at random was made for imputed variables. This is in accordance with previous handling of missing data in cohorts utilising QResearch data (195–197), and current recommendations for imputation of missing data (198). Characteristics with >90% missingness were not imputed. Ten imputed datasets were generated, and results from univariable and multivariable Cox proportional hazards models from each dataset were pooled according to Rubin's rules (199,200).

Sensitivity analyses

To test robustness of my main model, I performed three sensitivity analyses (**Table 3.6**), as follows (i) main results model fit to complete-case cohort subset (i.e. the subset of patients with completeness for all variables); (ii) exclusion of patients in the CHB population with diagnosed non-HCC neoplasms (presented in **Table 3.7**) to control for unmeasured outcome misclassification whereby secondary liver cancer has been misclassified as primary HCC, and (iii) addition of ALT, AST and Plt in the main model fit to the imputed dataset, as the percentage of missingness in these exposures was too high for them to be included in main analysis.

Table 3.6. Details of main and sensitivity analyses performed to investigate risk factors for hepatocellular carcinoma in QResearch cohort of adults with chronic hepatitis B virus (CHB) infection.

Analysis	Main or sensitivity	Variables included	Corresponding table(s)/figure
Cox proportional hazards model on imputed dataset	Main	Age; Alcohol consumption; Antidiabetic drug use; Antihypertensive use; Antiviral treatment; ARLD; Ascites; BMI; Cerebrovascular disease; CHF; Cigarette consumption; Cirrhosis; CKD; ESLD; Ethnicity; Hypertension; IHD; NAFLD; Non-HCC neoplasm; NSAID use; Peptic ulcer; Sex; Statin use; T2DM; Townsend deprivation quintile	Figure 5, Table 12
Cox proportional hazards model on complete-case dataset	Sensitivity	Age; Alcohol consumption; Antidiabetic drug use; Antihypertensive use; Antiviral treatment; ARLD; Ascites; BMI; Cerebrovascular disease; CHF; Cigarette consumption; Cirrhosis; CKD; ESLD; Ethnicity; Hypertension; IHD; NAFLD; Non-HCC neoplasm; NSAID use; Peptic ulcer; Sex; Statin use; T2DM; Townsend deprivation quintile	Table 13
Cox proportional hazards model on imputed dataset whereby individuals with history of non-HCC neoplasm were excluded	Sensitivity	Age; Alcohol consumption; Antidiabetic drug use; Antihypertensive use; Antiviral treatment; ARLD; Ascites; BMI; Cerebrovascular disease; CHF; Cigarette consumption; Cirrhosis; CKD; ESLD; Ethnicity; Hypertension; IHD; NAFLD;NSAID use; Peptic ulcer; Sex; Statin use; T2DM; Townsend deprivation quintile	Table 12
Cox proportional hazards model on imputed dataset with AST, ALT and Plt additionally added to model	Sensitivity	Age; Alcohol consumption; ALT; Antidiabetic drug use; Antihypertensive use; Antiviral treatment; ARLD; Ascites; AST; BMI; Cerebrovascular disease; CHF; Cigarette consumption; Cirrhosis; CKD; ESLD; Ethnicity; Hypertension; IHD; NAFLD; Non-HCC neoplasm; NSAID use; Peptic ulcer; Plt; Sex; Statin use; T2DM; Townsend deprivation quintile	Table 12

BMI, body mass index; T2DM, type 2 diabetes mellitus; CHF, congestive heart failure; CKD, chronic kidney disease; IHD, ischaemic heart disease; ARLD, alcohol-related liver disease; ESLD, end-stage liver disease; NAFLD, non-alcoholic fatty liver disease; NSAID, non-steroidal anti-inflammatory drug; AST, aspartate transaminase; ALT, alanine transaminase; Plt, platelets.

Table 3.7. Tabulation of non-hepatocellular neoplasms in the CHB cohort, overall and stratified by hepatocellular carcinoma (HCC) status.

Type of Non-HCC neoplasm, n (%)	Overall	Non-HCC patients	HCC patients
Total	381 (4.7 %)	345 (4.4%)	36 (17.1%)
Bone, connective tissue, skin and breast	124 (1.5%)	115 (1.5%)	9 (4.3%)
Digestive organs and peritoneum	44 (0.5%)	31 (0.4%)	13 (6.2%)
Eye, brain, meninges and other parts of the CNS	<5	<5	<5
Genitourinary organs	91 (1.1%)	81 (1.0%)	10 (4.8%)
Lip, oral cavity and pharynx	16 (0.2%)	16 (0.2%)	<5
Lymphatic and haemopoietic tissue	81 (1.0%)	79 (1.0%)	<5
Respiratory tract and intrathoracic organs	13 (0.2%)	11 (0.1%)	<5
Thyroid gland and other endocrine gland structures	8 (0.1%)	8 (0.1%)	<5

HCC, hepatocellular carcinoma; CNS, central nervous system.

3.5. Results

3.5.1. Research question 1

UK data for CHB epidemiology are summarised in **Table 3.8.** Four of seven estimates report information concerning population demographics, including number of infected individuals across age, sex and ethnicity categories. Among sources setting out to report CHB prevalence, estimates varied from 0.27% (British Liver Trust / Department of Health and Social Care (DHSC) 2002 estimate) to 0.73% (estimate by the Polaris Institute). An alternative proxy for population prevalence is obtained via the UK antenatal screening programme, which achieves over 95% coverage of every pregnant woman annually (approx. 700,000 women in the UK), with a CHB prevalence of <0.5% (213) The most recent study was a country-level modelling investigation undertaken by the European Centre for Disease Prevention and Control (ECDC), which utilised HBsAg positivity data for both the general population and population subgroups of interest to generate an overall HBsAg positivity prevalence estimate of 0.83% (48).

Prevalence estimates have been produced by a variety of organisations/authorities including public health bodies, charities, independent research organisations and institutional university research teams. Although the reported estimates differ in magnitude, they also vary by definition of the measure being reported. Differences between sources highlight varied sources of error, bias and missingness, problems with data linkage, and substantial “blind spots” in consistent testing and registration of HBV diagnoses.

Estimates vary due to the differing estimate types reported, all of which employ specific methodology and data sources for estimation. PHE/UKHSA studies report two distinct estimates. The first estimate utilises laboratory surveillance data collected across various sources (laboratories carrying out tests ordered in primary and secondary care, emergency departments, blood and transplant screening, sexual health services and prison services) to report HBV test positivity amongst those tested. This estimate is not strictly a prevalence estimate, because the sample from which data are drawn

does not represent the general population. The second PHE/UKHSA estimate is the HBV prevalence of women attending HBV screening via antenatal screening. This is a prevalence estimate for this subsection of the population (i.e. pregnant women) and provides an insight to prevalence in women of childbearing age, since update of this HBV screening arm is >99% (31). Estimates derived from CPRD and QResearch represent the number of people diagnosed with CHB in each respective database. These figures are not prevalence estimates, since individuals included in the database and subsequently investigated for HBV in primary care are not necessarily representative of the general population. There are two studies the Polaris Institute and ECDC which employ mathematical modelling to produce i) a population HBV prevalence estimate and ii) a population HBsAg positivity prevalence estimate, respectively.

Table 3.8. Prevalence estimates for chronic hepatitis B virus (HBV) infection in sub-populations in the United Kingdom.

Estimate source	Population and/or data source (study period/year)	Sources and citation(s)	Estimate type	Denominator (population tested)	Number of adults testing positive for HBV infection	Percentage infected with HBV in study/ survey population (%)	Information on relevant population characteristics ‡
PHE / UKHSA	Individuals tested for HBsAg in 19 participating laboratories (excludes testing in antenatal services and reference lab testing; deduplicated using identifiers where available) (2020)	Sentinel surveillance of BBV testing (201)	Positivity amongst those tested (diagnosed prevalence)	302,135	2611	0.86%	Yes: information by ethnicity, sex, age, region
Polaris Observatory	Modelling study based on literature review and expert interviews (2016)	Razavi-Shearer *et al.,* 2018 (44)	UK HBV population prevalence	2016 UK population of 65,648,000 (181)	440,000	0.73%	No
PHE Infectious Diseases in Pregnancy Screening Programme	Pregnant women attending antenatal services in England (April 2019-March 2020)	Antenatal screening standards data report 1 April 2019 – 31 March 2020 (202)	Antenatal population prevalence	Pregnant women attending antenatal services 661,281 (screening	2493	0.38%	No

Estimate source	Population and/or data source (study period/year)	Sources and citation(s)	Estimate type	Denominator (population tested)	Number of adults testing positive for HBV infection	Percentage infected with HBV in study/ survey population (%)	Information on relevant population characteristics ‡
				coverage of 99.8%)			
BLT (44) *	Source unavailable (likely 2002) §	NICE guidelines (203,204)**	Number of people in the UK positive for HBV - prevalence	Not clearly stated – likely 2002 UK population of 59,365,677 (178)	180,000	0.27%	No
CPRD	CPRD primary care database consisting of EHRs from >16 million patients (2000 to 2015)	Ferreira *et al.* (45)	Number of CHB diagnoses in database	CPRD subset of EHRs from 4.4 million patients ¶	3927	0.09%	Yes
QResearch cohort	QResearch primary care database (1999 to 2019)	CCCC-UK cohort (47)	Number of CHB diagnoses in database	QResearch database subset of EHRs from 16.9 million patients (62)	8039	0.05%	Yes

Estimate source	Population and/or data source (study period/year)	Sources and citation(s)	Estimate type	Denominator (population tested)	Number of adults testing positive for HBV infection	Percentage infected with HBV in study/ survey population (%)	Information on relevant population characteristics ‡
ECDC	Country-level (ecological) modelling study based on literature review and expert data collection (2019)	Trickey et al (48)	HBsAg positivity prevalence	2019 UK population of 66,796,807 (182)	554,413	0.83%	Yes

HBV, Hepatitis B virus; UKHSA, United Kingdom Health Security Agency; PHE, Public Health England; PWID, people who inject drugs; BLT, British Liver Trust; NICE, National Institute for Health and Care Excellence; CPRD, Clinical Practice Research Datalink; CHB chronic Hepatitis B virus infection; EHR, electronic health record; CCCC, Characteristics of Chronic Hepatitis B associated with Cirrhosis and Cancer; ECDC, European Centre for Disease Prevention and Control; HBsAg, hepatitis B surface antigen.
* The British Liver Trust published a HBV report in 2017 stating that approximately 180,000 people live with chronic HBV infection in the UK. This report cited a governmental migrant health guide citing this prevalence estimate. The source which this government report cited was a National Institute for Health and Care Excellence (NICE) testing guideline which stated this prevalence estimate and dated it to 2002.

† May include but is not limited to: age, sex, ethnicity and socioeconomic status.

‡ Sources include laboratory data and sentinel surveillance data.

§ Department of Health 2002 report cited is unavailable for download. No clear methodology reported.

¶ Subset of total (16.9 million) patients in CPRD with valid EHR information and temporality who were acceptable for research.

3.5.2. Research question 2

Individuals with CHB are under-represented in primary care records

I identified 8039 CHB patients in the QResearch database between 1999 and 2019 from a database-wide denominator of ~35 million individuals, which translates to a CHB prevalence of 0.023%. This is at least an order of magnitude less than current estimates for CHB prevalence in the UK, of 0.27% to 0.73%. Thus, through QResearch I have only had access to a dataset representing a fraction of all CHB cases. This discrepancy highlights under-diagnosis of HBV infection, and poor linkage of data between specialist care services and primary care.

The majority of patients with CHB were male, of non-white ethnicity, in the fourth and fifth (most deprived) Townsend deprivation quintiles, and not receiving HBV treatment

Most of the adults represented in this study were identified by a SNOMED/Read or ICD CHB diagnostic code (7856/8039, 97.7%), with a remaining 2.3% (252/8039) identified by HBsAg and/or VL measurements (in the absence of a diagnostic code) (**Table 3.9**). Median follow-up duration was 3.87 years (IQR 6.30 years), with differential follow-up between individuals who developed HCC (median follow-up 1.47 years, IQR 5.13 years) and those who did not (median follow-up duration 3.93 years, IQR 6.28 years). Mean age at baseline was 38.3 years (SD 11.6 years), and at baseline >75% were ≤45 years of age. The majority were male (4856/8039, 60.4%).

Table 3.9. Cohort baseline characteristics for all individuals with chronic HBV infection included in this cohort.

Characteristic	Total, n (%)	HCC cases, n (%)	Non-HCC cases, n (%)
N	8039	210	7829
Follow-up time (years), median (IQR)	3.87 (6.30)	1.47 (5.13)	3.93 (6.28)
CHB diagnosis source, n (%) *			
Any ICD	3952 (49.2%)	47 (22.4%)	3789 (48.4%)
Any SNOMED/Read	4425 (55.0%)	68 (32.4%)	4357 (55.7%)
Any laboratory test	240 (3.0%)	<5	238 (3.0%)
Age (years), mean (SD)	38.3 (11.6)	50.3 (12.4)	37.9 (11.4)
Age group, n (%)			
18-25 years	860 (10.7%)	<5	857 (10.9%)
26-35 years	2930 (36.4%)	22 (10.5%)	2908 (37.1%)
36-45 years	2346 (29.2%)	52 (24.8%)	2294 (29.3%)

Characteristic	Total, n (%)	HCC cases, n (%)	Non-HCC cases, n (%)
46-55 years	1182 (14.7%)	63 (30.0%)	1119 (14.3%)
56-65 years	502 (6.2%)	41 (19.5%)	461 (5.9%)
≥66 years	219 (2.7%)	29 (13.8%)	190 (2.4%)
Sex, n (%)			
Female	3183 (39.6%)	24 (11.4%)	3159 (40.3%)
Male	4856 (60.4%)	186 (88.6%)	4670 (59.7%)
Townsend Deprivation Quintile, n (%)			
First (least deprived)	467 (5.8%)	27 (12.9%)	440 (5.6%)
Second	731 (9.1%)	20 (9.5%)	711 (9.1%)
Third	1226 (15.3%)	27 (12.9%)	1199 (15.3%)
Fourth	2040 (25.4%)	48 (22.9%)	1992 (25.4%)
Fifth (most deprived)	3553 (44.3%)	87 (41.4%)	3466 (44.3%)
Missing	22 (0.3%)	<5	21 (0.3%)
Ethnicity, n (%)			
Black African	1714 (25.4%)	35 (22.3%)	1679 (25.5%)
Bangladeshi	325 (4.8%)	11 (7.0%)	314 (4.8%)
Caribbean	104 (1.5%)	7 (4.5%)	97 (1.5%)
Chinese	871 (12.9%)	21 (13.4%)	850 (12.9%)
Indian	203 (3.0%)	7 (4.5%)	196 (3.0%)
Other	832 (12.3%)	15 (9.6%)	817 (12.4%)
Other Asian	384 (5.7%)	5 (3.2%)	379 (5.7%)
Pakistani	401 (5.9%)	7 (4.5%)	394 (6.0%)
White	1916 (28.4%)	49 (31.2%)	1867 (28.3%)
Missing	1289 (16.0%)	53 (25.2%)	1236 (15.8%)
Cigarette consumption, n (%)			
Non-smoker	3810 (66.2%)	78 (54.2%)	3732 (66.5%)
Ex-smoker	768 (13.4%)	25 (17.4%)	743 (13.2%)
Light smoker (1-9 cigarettes/day)	908 (15.8%)	31 (21.5%)	877 (15.6%)

Characteristic	Total, n (%)	HCC cases, n (%)	Non-HCC cases, n (%)
Moderate to heavy smoker (≥10 cigarettes per day)	266 (4.6%)	10 (6.9%)	256 (4.6%)
Missing	2287 (28.5%)	66 (31.4%)	2221 (28.4%)
Alcohol consumption, n (%)			
Non-drinker	3697 (81.9%)	100 (83.3%)	3597 (81.9%)
Trivial drinker (<1 unit per day)	485 (10.7%)	13 (10.8%)	472 (10.7%)
Light drinker (1-2 units per day)	209 (4.6%)	<5	206 (4.7%)
Moderate to heavy drinker (≥3 units per day)	121 (2.7%)	<5	117 (2.7%)
Missing	3527 (43.9%)	90 (42.9%)	3439 (43.9%)
BMI (kg/m^2), mean (SD)	26.05 (5.06)	26.13 (4.63)	26.05 (5.07)
BMI, n (%)			
Underweight (<18.5 kg/m^2)	159 (3.2%)	<5	156 (3.2%)
Normal weight (18.5-24.9 kg/m^2)	2092 (42.5%)	49 (39.8%)	2043 (42.5%)
Overweight (25.0-29.9 kg/m^2)	1734 (35.2%)	52 (42.3%)	1682 (35.0%)
Obese (≥30 kg/m^2)	943 (19.1%)	19 (15.4%)	924 (19.2%)
Missing	3111 (38.7%)	87 (41.4%)	3024 (38.6%)
Antiviral initiation, n (%)			
No initiation	7076 (88.0%)	140 (66.7%)	6936 (88.6%)
Before CHB diagnosis	138 (1.7%)	11 (5.2%)	127 (1.6%)
Within 1 year of CHB diagnosis	221 (2.7%)	24 (11.4%)	197 (2.5%)
Within 2 years of CHB diagnosis	120 (1.5%)	7 (3.3%)	113 (1.4%)
Within 3 years of CHB diagnosis	78 (1.0%)	7 (3.3%)	71 (0.9%)

Characteristic	Total, n (%)	HCC cases, n (%)	Non-HCC cases, n (%)
≥4 years after CHB diagnosis	406 (5.1%)	21 (10.0%)	385 (4.9%)
Type 2 DM, n (%)	715 (8.9%)	21 (10.0%)	664 (8.5%)
Congestive heart failure, n (%)	75 (0.9%)	7 (3.3%)	68 (0.9%)
Hypertension, n (%)	1229 (15.3%)	71 (33.8%)	1158 (14.8%)
Chronic kidney disease, n (%)	287 (3.6%)	29 (13.8%)	258 (3.3%)
Alcohol-related liver disease, n (%)	93 (1.2%)	19 (9.0%)	74 (0.9%)
Ascites, n (%)	114 (1.4%)	32 (15.2%)	82 (1.0%)
Autoimmune hepatitis, n (%)	9 (0.1%)	<5	9 (0.1%)
Cerebrovascular disease, n (%)	310 (3.9%)	21 (10.0%)	289 (3.7%)
Cirrhosis, n (%)	689 (8.6%)	121 (57.6%)	568 (7.3%)
End-stage liver disease, n (%)	85 (1.1%)	16 (7.6%)	69 (0.9%)
Ischaemic heart disease, n (%)	255 (3.2%)	20 (9.5%)	235 (3.0%)
Non-alcoholic fatty liver disease, n (%)	404 (5.0%)	8 (3.8%)	396 (5.1%)
Non-HCC neoplasm, n (%)	381 (4.7 %)	345 (4.4%)	36 (17.1%)
Peptic ulcer, n (%)	248 (3.1%)	32 (15.2%)	216 (2.8%)
Antidiabetic drug use, n (%)	737 (9.2%)	49 (23.3%)	688 (8.8%)
Antihypertensive use, n (%)	408 (5.1%)	22 (10.5%)	386 (4.9%)
NSAID use, n (%)	472 (5.9%)	24 (11.4%)	448 (5.7%)
Statin use, n (%)	457 (5.7%)	19 (9.0%)	438 (5.6%)

CHB, chronic hepatitis B virus infection; ICD, International classification of disease; SNOMED, Systemised Nomenclature of Medicine; SD, standard deviation; BMI, body mass index; DM, diabetes mellitus; HCC, hepatocellular carcinoma; NSAID, non-steroidal anti-inflammatory drug.

* Some patients have diagnostic indicators of CHB from multiple sources and therefore there overlap between categories.

Black Africans represented 25.4% of individuals, with 12.9% of Chinese ethnicity, 5.9% of Pakistani ethnicity, 3.0% of Indian ethnicity and 28.4% of White ethnicity (**Table 3.9**). Proportions of Black and ethnic minorities in this CHB cohort were greater than those in both the database (205) and the general English population (206) (**Figure 3.3, 3.4**).

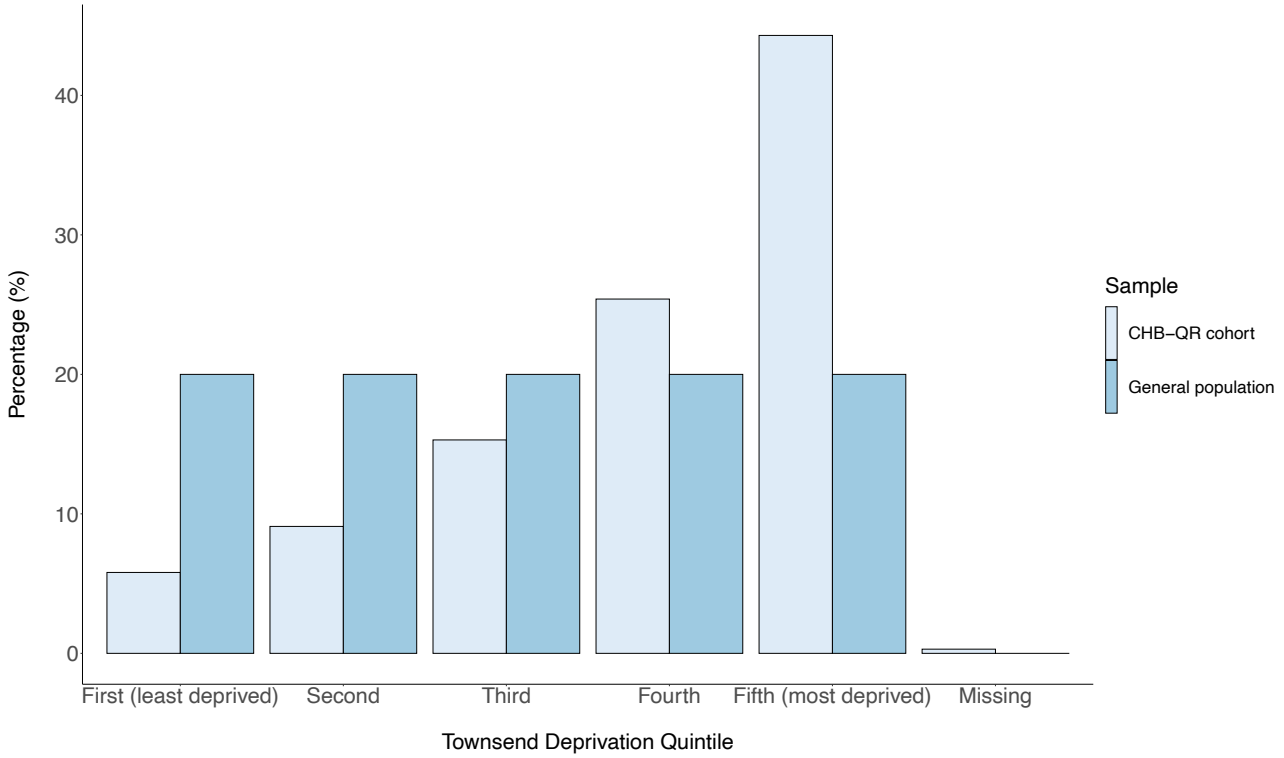


Figure 3.2. Townsend deprivation quintile breakdown in 8039 adults with chronic hepatitis B virus infection derived from the QResearch primary care database (England) versus the United Kingdom general population.

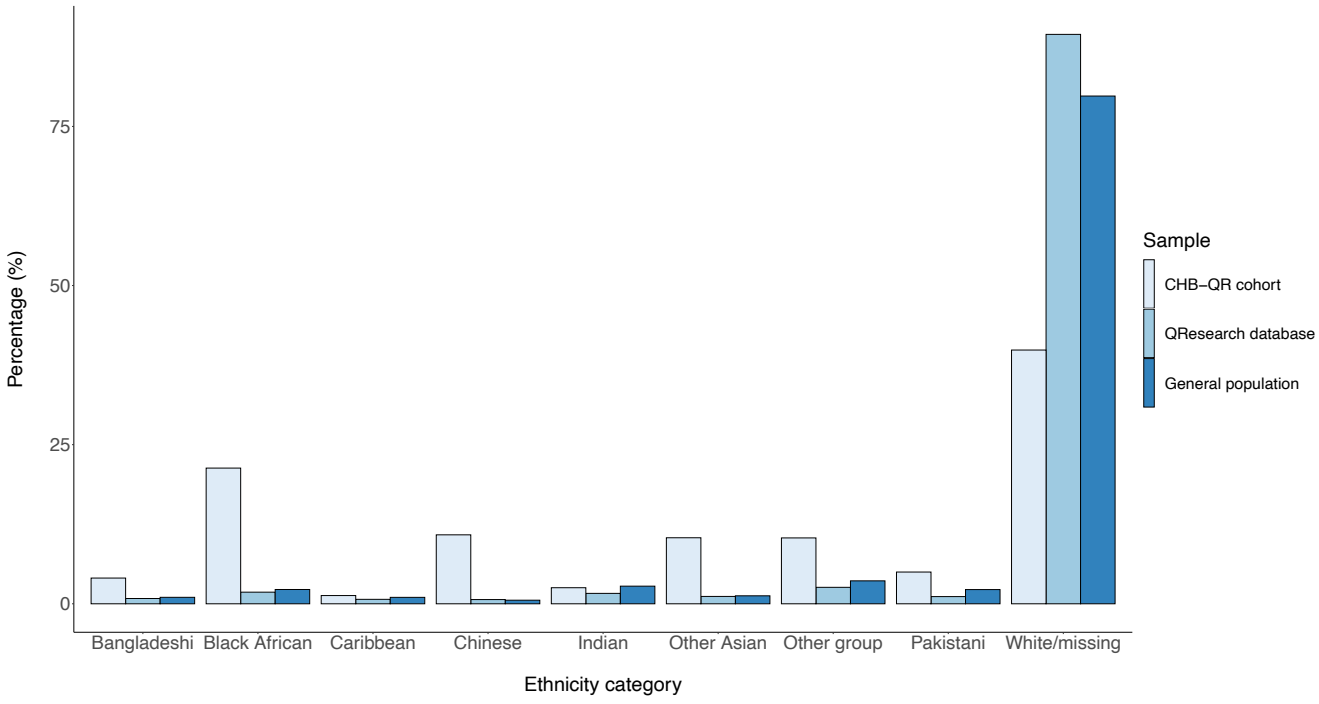


Figure 3.3. Ethnicity breakdown in 8039 adults in a chronic hepatitis B virus cohort characterised from the QResearch primary care database (England) versus all individuals in the QResearch database (~35 million), vs. the United Kingdom general population. General population estimates obtained from 2019 estimates from the Office for National Statistics.

At baseline, most (88.0%) of patients had no record of antiviral treatment. Within 1, 2, 3 and ≥ 4 years of CHB diagnoses, cumulatively 2,7%, 4.0 %, 5.0% and 10.1% of patients had a record of antiviral treatment initiation, respectively.

Cohort characteristics stratified by Townsend deprivation quintile are presented in **Table 3.10**. Age and sex associated significantly with Townsend deprivation quintile whereby mean ages were younger ($P < 0.001$) and proportions of males were higher ($P = 0.021$) in more deprived quintiles. Ethnicity significantly differed across quintiles ($P < 0.001$) whereby the proportions of Bangladeshi and Black African ethnicity patients increased with increasing deprivation and the proportion of White and Chinese patients decreased. Alcohol and cigarette consumption were significantly associated with Townsend deprivation quintile ($P < 0.001$ and $P = 0.04$, respectively), but no obvious trends across quintiles were apparent. The only comorbidities differing significantly in prevalence across quintiles were cerebrovascular disease, peptic ulcer disease and non-HCC neoplasms. Associations of antidiabetic drug, antihypertensive, NSAID and statin use with Townsend Deprivation Quintile were nonsignificant.

Table 3.10. Characteristics of adults with chronic hepatitis B virus infection identified from QResearch primary care database, stratified by socioeconomic status.

	Townsend Deprivation Quintile						
Characteristic	**First (least deprived)**	**Second**	**Third**	**Fourth**	**Fifth (most deprived)**	**Missing**	***P***
N	467	731	1226	2040	3553	22	
Follow-up time (years), mean (SD)	6.33 (6.59)	6.15 (7.65)	5.40 (5.48)	4.94 (5.25)	5.10 (5.20)	3.79 (3.56)	<0.001
HCC cases, n (%)	27 (5.8%)	20 (2.7%)	27 (2.2%)	48 (2.4%)	87 (2.4%)	<5	0.001
Age (years), mean (SD)	43.64 (13.10)	40.11 (12.26)	38.98 (11.91)	37.35 (11.22)	37.45 (11.08)	38.18 (9.52)	<0.001
Age group, n (%)							<0.001
18-25 years	110 (23.6%)	247 (33.8%)	433 (35.3%)	799 (39.2%)	1330 (37.4%)	11 (50.0%)	
26-35 years	25 (5.4%)	64 (8.8%)	112 (9.1%)	244 (12.0%)	415 (11.7%)	<5	
36-45 years	153 (32.8%)	207 (28.3%)	379 (30.9%)	561 (27.5%)	1041 (29.3%)	5 (22.7%)	
46-55 years	86 (18.4%)	124 (17.0%)	178 (14.5%)	281 (13.8%)	509 (14.3%)	<5	
56-65 years	58 (12.4%)	64 (8.8%)	76 (6.2%)	116 (5.7%)	186 (5.2%)	<5	
≥66 years	35 (7.5%)	25 (3.4%)	48 (3.9%)	39 (1.9%)	72 (2.0%)	<5	
Sex, n (%)							0.021

	Townsend Deprivation Quintile						
Characteristic	**First (least deprived)**	**Second**	**Third**	**Fourth**	**Fifth (most deprived)**	**Missing**	***P***
Female	214 (45.8%)	300 (41.0%)	505 (41.2%)	802 (39.3%)	1355 (38.1%)	7 (31.8%)	
Male	253 (54.2%)	431 (59.0%)	721 (58.8%)	1238 (60.7%)	2198 (61.9%)	15 (68.2%)	
Ethnicity, n (%)							<0.001
Black African	26 (7.2%)	69 (11.9%)	155 (15.8%)	386 (22.5%)	1073 (34.6%)	≤5	
Bangladeshi	<5	10 (1.7%)	26 (2.6%)	56 (3.3%)	229 (7.4%)	<5	
Caribbean	<5	<5	14 (1.4%)	35 (2.0%)	49 (1.6%)	<5	
Chinese	84 (23.3%)	121 (20.9%)	156 (15.9%)	201 (11.7%)	305 (9.8%)	<5	
Indian	10 (2.8%)	17 (2.9%)	55 (5.6%)	80 (4.7%)	40 (1.3%)	<5	
Other	49 (13.6%)	58 (10.0%)	93 (9.5%)	216 (12.6%)	415 (13.4%)	<5	
Other Asian	24 (6.7%)	46 (7.9%)	67 (6.8%)	99 (5.8%)	147 (4.7%)	<5	
Pakistani	19 (5.3%)	41 (7.1%)	98 (10.0%)	153 (8.9%)	90 (2.9%)	<5	
White	143 (39.7%)	213 (36.8%)	318 (32.4%)	486 (28.4%)	755 (24.3%)	<5	
Cigarette consumption, n (%)							0.04
Non-smoker	211 (67.8%)	326 (66.3%)	581 (66.3%)	978 (66.4%)	1704 (65.9%)	10 (66.7%)	
Ex-smoker	51 (16.4%)	84 (17.1%)	122 (13.9%)	167 (11.3%)	343 (13.3%)	<5	
Light smoker (1-9 cigarettes/day)	37 (11.9%)	62 (12.6%)	132 (15.1%)	248 (16.8%)	425 (16.4%)	<5	
Moderate to heavy smoker (≥10 cigarettes per day)	12 (3.9%)	20 (4.1%)	41 (4.7%)	80 (5.4%)	113 (4.4%)	<5	
Alcohol consumption, n (%)							<0.001
Non-drinker	183 (81.0%)	263 (76.9%)	553 (82.2%)	959 (82.9%)	1732 (82.3%)	7 (70.0%)	
Trivial drinker (<1 unit per day)	32 (14.2%)	42 (12.3%)	70 (10.4%)	120 (10.4%)	221 (10.5%)	<5	
Light drinker (1-2 units per day)	<5	21 (6.1%)	31 (4.6%)	39 (3.4%)	113 (5.4%)	<5	
Moderate to heavy drinker (≥3 units per day)	8 (3.5%)	16 (4.7%)	19 (2.8%)	39 (3.4%)	38 (1.8%)	<5	
BMI (kg/m^2), mean (SD)	25.34 (4.76)	25.72 (4.92)	25.81 (4.94)	26.30 (5.31)	26.14 (5.01)	26.56 (4.47)	0.031
BMI, n (%)							0.433

	Townsend Deprivation Quintile						
Characteristic	**First (least deprived)**	**Second**	**Third**	**Fourth**	**Fifth (most deprived)**	**Missing**	***P***
Underweight (<18.5 kg/m^2)	11 (4.3%)	13 (3.2%)	27 (3.7%)	38 (3.0%)	70 (3.1%)	<5	
Normal weight (18.5-24.9 kg/m^2)	125 (48.6%)	176 (43.2%)	322 (43.6%)	533 (42.3%)	933 (41.4%)	<5	
Overweight (25.0-29.9 kg/m^2)	79 (30.7%)	146 (35.9%)	260 (35.2%)	422 (33.5%)	822 (36.5%)	5 (50.0%)	
Obese (≥30 kg/m^2)	42 (16.3%)	72 (17.7%)	129 (17.5%)	268 (21.3%)	430 (19.1%)	<5	
Antiviral initiation, n (%)							0.445
No initiation	393 (84.2%)	647 (88.5%)	1071 (87.4%)	1808 (88.6%)	3138 (88.3%)	19 (86.4%)	
Before CHB diagnosis	13 (2.8%)	17 (2.3%)	24 (2.0%)	32 (1.6%)	52 (1.5%)	<5	
Within 1 year of CHB diagnosis	19 (4.1%)	16 (2.2%)	39 (3.2%)	46 (2.3%)	100 (2.8%)	<5	
Within 2 years of CHB diagnosis	7 (1.5%)	9 (1.2%)	14 (1.1%)	29 (1.4%)	61 (1.7%)	<5	
Within 3 years of CHB diagnosis	7 (1.5%)	6 (0.8%)	15 (1.2%)	21 (1.0%)	28 (0.8%)	<5	
≥4 years after CHB diagnosis	28 (6.0%)	36 (4.9%)	63 (5.1%)	104 (5.1%)	17,4 (4.9%)	<5	
Type 2 DM, n (%)	39 (8.4%)	51 (7.0%)	119 (9.7%)	180 (8.8%)	323 (9.1%)	<5	0.383
Congestive heart failure, n (%)	6 (1.3%)	10 (1.4%)	12 (1.0%)	19 (0.9%)	28 (0.8%)	<5	0.675
Hypertension, n (%)	85 (18.2%)	97 (13.3%)	183 (14.9%)	298 (14.6%)	563 (15.8%)	<5	0.212
Chronic kidney disease, n (%)	20 (4.3%)	21 (2.9%)	46 (3.8%)	75 (3.7%)	123 (3.5%)	<5	0.553
Alcohol-related liver disease, n (%)	10 (2.1%)	7 (1.0%)	13 (1.1%)	27 (1.3%)	36 (1.0%)	<5	0.336
Ascites, n (%)	12 (2.6%)	10 (1.4%)	20 (1.6%)	24 (1.2%)	47 (1.3%)	<5	0.188
Autoimmune hepatitis, n (%)	<5	<5	<5	<5	<5	<5	NA
Cerebrovascular disease, n (%)	30 (6.4%)	29 (4.0%)	54 (4.4%)	74 (3.6%)	123 (3.5%)	<5	0.035
End-stage liver disease, n (%)	8 (1.7%)	6 (0.8%)	12 (1.0%)	22 (1.1%)	36 (1.0%)	<5	0.414

	Townsend Deprivation Quintile						
Characteristic	**First (least deprived)**	**Second**	**Third**	**Fourth**	**Fifth (most deprived)**	**Missing**	***P***
Ischaemic heart disease, n (%)	20 (4.3%)	29 (4.0%)	41 (3.3%)	67 (3.3%)	98 (2.8%)	<5	0.28
Non-alcoholic fatty liver disease, n (%)	22 (4.7%)	33 (4.5%)	65 (5.3%)	96 (4.7%)	188 (5.3%)	<5	0.728
Non-HCC neoplasm, n (%)	132 (28.3%)	161 (22.0%)	272 (22.2%)	359 (17.6%)	699 (19.7%)	<5	<0.001
Peptic ulcer, n (%)	26 (5.6%)	13 (1.8%)	43 (3.5%)	58 (2.8%)	108 (3.0%)	<5	0.008
Antidiabetic drug use, n (%)	41 (8.8%)	54 (7.4%)	125 (10.2%)	183 (9.0%)	330 (9.3%)	<5	0.242
Antihypertensive use, n (%)	25 (5.4%)	36 (4.9%)	66 (5.4%)	107 (5.2%)	174 (4.9%)	<5	0.865
NSAID use, n (%)	26 (5.6%)	29 (4.0%)	86 (7.0%)	112 (5.5%)	217 (6.1%)	<5	0.106
Statin use, n (%)	21 (4.5%)	30 (4.1%)	78 (6.4%)	114 (5.6%)	213 (6.0%)	<5	0.268

CHB, chronic hepatitis B virus infection; ICD, International classification of disease; SNOMED, Systemised Nomenclature of Medicine; SD, standard deviation; BMI, body mass index; DM, diabetes mellitus; HCC, hepatocellular carcinoma; NSAID, non-steroidal anti-inflammatory drug.

Prevalence of diabetes and hypertension in adults with CHB was higher than in the general population

Baseline prevalence of T2DM and hypertension in people living with CHB were 8.9% and 15.3%, respectively. This differs from the prevalence of T2DM and hypertension in the wider QResearch database of <8% and <3%, respectively (207). Prevalence of other comorbidities (including congestive heart failure, chronic kidney disease, alcohol-related liver disease, ascites, autoimmune hepatitis, cerebrovascular disease, end-stage liver disease, ischaemic heart disease, no-alcoholic fatty liver disease and peptic ulcer disease; **Table 3.9**) ranged from 0.1% to 5%. Baseline cirrhosis is difficult to define from primary care EHR due to lack of relevant laboratory and imaging data, but 8.6% of the cohort had a diagnostic code indicating cirrhosis. Medication use was as follows: antidiabetic drugs (9.2%), antihypertensives (5.1%), non-steroidal anti-inflammatory drugs (NSAIDs) (5.9%) and statins (5.7%). Prevalence of non-HCC neoplasm was 4.7% in the overall CHB cohort.

Data for smoking consumption, alcohol consumption and BMI were available for 71.5%, 56.1% and 61.3% of the cohort (**Table 3.9**).

Risk factors for HCC included male sex, older age, increased deprivation, Caribbean ethnicity and peptic ulcer disease

Baseline characteristics of the imputed dataset used in analysis of HCC risk factors are presented in **Table 3.11**. Multivariable Cox proportional hazards models were constructed for 7029 patients in whom 161 HCC cases developed throughout 41 147 person-years of follow-up (**Figure 3.5, Table 3.12**). This translated to an HCC incidence rate of 5.10 cases per 1000 person-years (95% Confidence Interval (95CIs) 4.46 to 5.84).

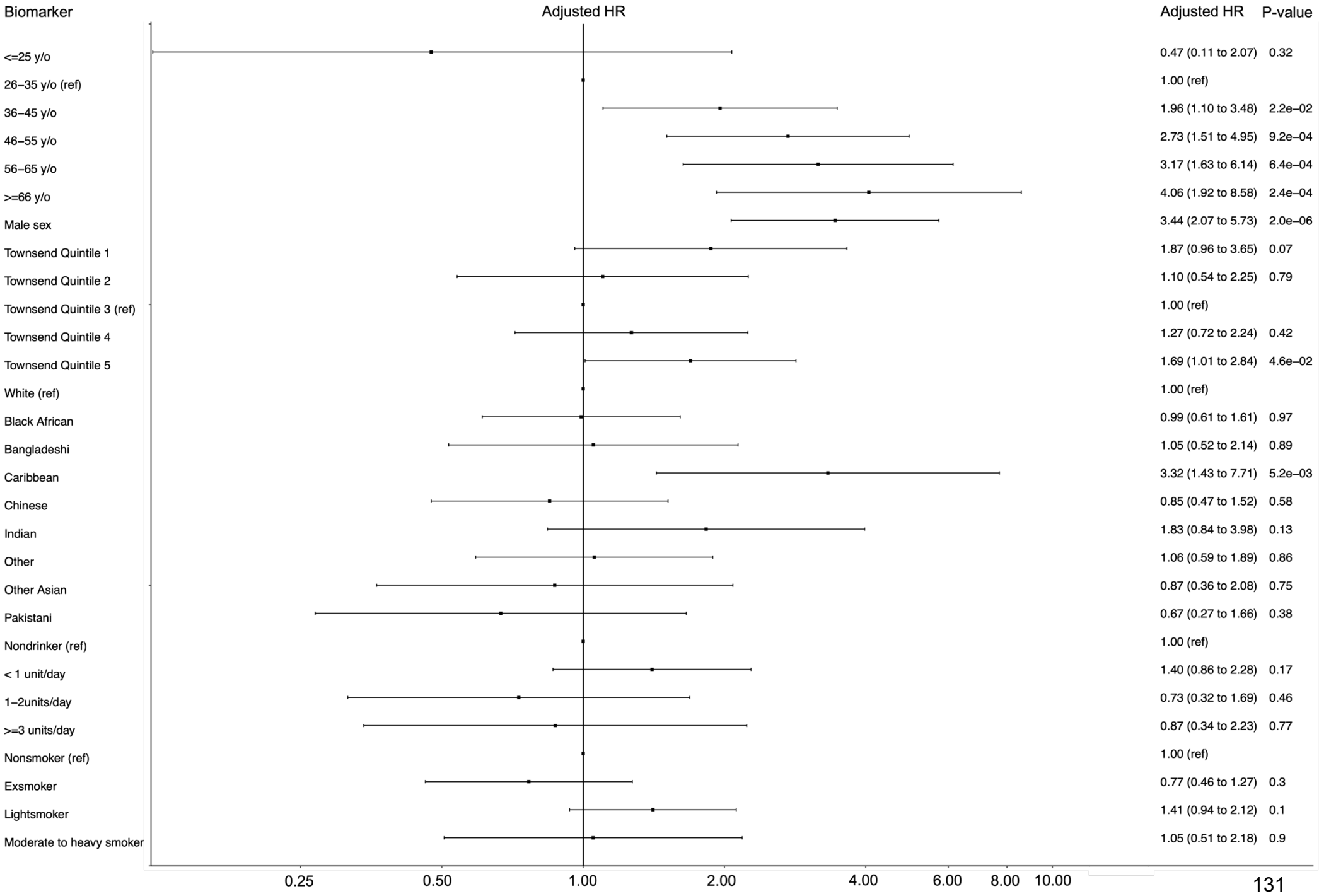
Table 3.11. Baseline characteristics of complete-case and imputed (by multiple imputation with chained equations) datasets used in Cox proportional hazards models.

Characteristic	**Complete-case dataset, n (%)**	**Imputed dataset, n (%)**
N	3648	7029*
HCC cases, n (%)	68 (1.9%)	161 (2.3%)
Age group, n (%)		
18-25 years	292 (8.0)	659 (9.4)
26-35 years	1253 (34.3)	2500 (35.6)
36-45 years	1126 (30.9)	2108 (30.0)
46-55 years	593 (16.3)	1086 (15.5)
56-65 years	270 (7.4)	468 (6.7)
≥66 years	114 (3.1)	208 (3.0)
Sex, n (%)		
Male	2082 (57.1)	4186 (59.6)
Townsend Deprivation Quintile, n (%)		
First (least deprived)	174 (4.8)	419 (6.0)
Second	261 (7.2)	645 (9.2)
Third	523 (14.3)	1094 (15.6)
Fourth	924 (25.3)	1781 (25.3)
Fifth (most deprived)	1766 (48.4)	3090 (44.0)
Ethnicity, n (%)		
Black African	956 (26.2)	1772 (25.2)
Bangladeshi	213 (5.8)	322 (4.6)
Caribbean	58 (1.6)	112 (1.6)
Chinese	428 (11.7)	894 (12.7)

Characteristic	Complete-case dataset, n (%)	Imputed dataset, n (%)
Indian	131 (3.6)	224 (3.2)
Other	441 (12.1)	863 (12.3)
Other Asian	194 (5.3)	397 (5.6)
Pakistani	241 (6.6)	420 (6.0)
White	986 (27.0)	2025 (28.8)
Cigarette consumption, n (%)		
Non-smoker	2484 (68.1)	4661 (66.3)
Ex-smoker	480 (13.2)	924 (13.1)
Light smoker (1-9 cigarettes/day)	531 (14.6)	1110 (15.8)
Moderate to heavy smoker (≥10 cigarettes per day)	153 (4.2)	334 (4.8)
Alcohol consumption, n (%)		
Non-drinker	3013 (82.6)	5784 (82.3)
Trivial drinker (<1 unit per day)	371 (10.2)	737 (10.5)
Light drinker (1-2 units per day)	170 (4.7)	318 (4.5)
Moderate to heavy drinker (≥3 units per day)	94 (2.6)	190 (2.7)
BMI (kg/m^2), mean (SD)	26.14 (5.07)	26.05 (5.02)
BMI, n (%)		
Underweight (<18.5 kg/m^2)	113 (3.1)	222 (3.2)
Normal weight (18.5-24.9 kg/m^2)	1530 (41.9)	2948 (41.9)
Overweight (25.0-29.9 kg/m^2)	1292 (35.4)	2527 (36.0)
Obese (≥30 kg/m^2)	713 (19.5)	1332 (19.0)
Antiviral initiation, n (%)		
No initiation	3168 (86.8)	6092 (86.7)
Before CHB diagnosis	67 (1.8)	121 (1.7)
Within 1 year of CHB diagnosis	116 (3.2)	214 (3.0)

Characteristic	Complete-case dataset, n (%)	Imputed dataset, n (%)
Within 2 years of CHB diagnosis	73 (2.0)	120 (1.7)
Within 3 years of CHB diagnosis	49 (1.3)	78 (1.1)
≥4 years after CHB diagnosis	175 (4.8)	404 (5.7)
Type 2 DM, n (%)	503 (13.8)	678 (9.6)
Congestive heart failure, n (%)	41 (1.1)	71 (1.0)
Hypertension, n (%)	754 (20.7)	1165 (16.6)
Chronic kidney disease, n (%)	175 (4.8)	271 (3.9)
Alcohol-related liver disease, n (%)	39 (1.1)	74 (1.1)
Ascites, n (%)	34 (0.9)	88 (1.3)
Autoimmune hepatitis, n (%)	3 (0.1)	9 (0.1)
Cerebrovascular disease, n (%)	181 (5.0)	290 (4.1)
Cirrhosis, n (%)	301 (8.3)	608 (8.6)
End-stage liver disease, n (%)	28 (0.8)	78 (1.1)
Ischaemic heart disease, n (%)	143 (3.9)	229 (3.3)
Non-alcoholic fatty liver disease, n (%)	248 (6.8)	393 (5.6)
Non-HCC neoplasm, n (%)	167 (4.6)	340 (4.8)
Peptic ulcer, n (%)	113 (3.1)	221 (3.1)
Antidiabetic drug use, n (%)	495 (13.6)	689 (9.8)
Antihypertensive use, n (%)	293 (8.0)	389 (5.5)
NSAID use, n (%)	331 (9.1)	446 (6.3)
Statin use, n (%)	342 (9.4)	443 (6.3)

CHB, chronic hepatitis B virus infection; ICD, International classification of disease; SNOMED, Systemised Nomenclature of Medicine; IQR, interquartile range; SD, standard deviation; BMI, body mass index; DM, diabetes mellitus; HCC, hepatocellular carcinoma; NSAID, non-steroidal anti-inflammatory drug.


Biomarker
Adjusted HR
Adjusted HR
P-value
<=25 y/o
0.47 (0.11 to 2.07) 0.32
26–35 y/o (ref)
1.00 (ref)
36–45 y/o
1.96 (1.10 to 3.48) 2.2e–02
46–55 y/o
2.73 (1.51 to 4.95) 9.2e–04
56–65 y/o
3.17 (1.63 to 6.14) 6.4e–04
>=66 y/o
4.06 (1.92 to 8.58) 2.4e–04
Male sex
3.44 (2.07 to 5.73) 2.0e–06
Townsend Quintile 1
1.87 (0.96 to 3.65) 0.07
Townsend Quintile 2
1.10 (0.54 to 2.25) 0.79
Townsend Quintile 3 (ref)
1.00 (ref)
Townsend Quintile 4
1.27 (0.72 to 2.24) 0.42
Townsend Quintile 5
1.69 (1.01 to 2.84) 4.6e–02
White (ref)
1.00 (ref)
Black African
0.99 (0.61 to 1.61) 0.97
Bangladeshi
1.05 (0.52 to 2.14) 0.89
Caribbean
3.32 (1.43 to 7.71) 5.2e–03
Chinese
0.85 (0.47 to 1.52) 0.58
Indian
1.83 (0.84 to 3.98) 0.13
Other
1.06 (0.59 to 1.89) 0.86
Other Asian
0.87 (0.36 to 2.08) 0.75
Pakistani
0.67 (0.27 to 1.66) 0.38
Nondrinker (ref)
1.00 (ref)
< 1 unit/day
1.40 (0.86 to 2.28) 0.17
1–2units/day
0.73 (0.32 to 1.69) 0.46
>=3 units/day
0.87 (0.34 to 2.23) 0.77
Nonsmoker (ref)
1.00 (ref)
Exsmoker
0.77 (0.46 to 1.27) 0.3
Lightsmoker
1.41 (0.94 to 2.12) 0.1
Moderate to heavy smoker
1.05 (0.51 to 2.18) 0.9
0.25
0.50
1.00
2.00
4.00
6.00
8.00
10.00

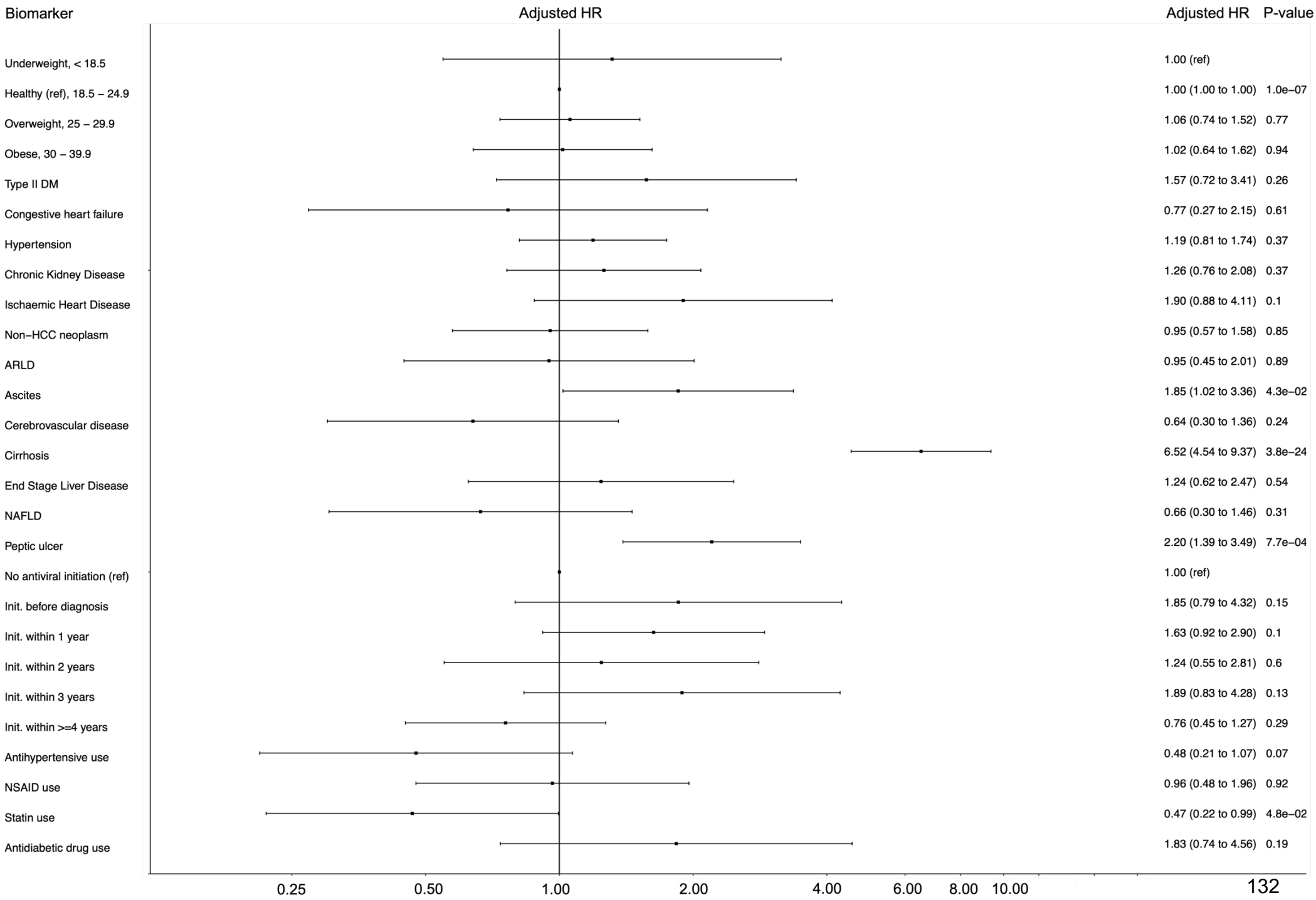

Biomarker
Adjusted HR
Adjusted HR P-value
Underweight, < 18.5
1.00 (ref)
Healthy (ref), 18.5 – 24.9
1.00 (1.00 to 1.00) 1.0e−07
Overweight, 25 – 29.9
1.06 (0.74 to 1.52) 0.77
Obese, 30 – 39.9
1.02 (0.64 to 1.62) 0.94
Type II DM
1.57 (0.72 to 3.41) 0.26
Congestive heart failure
0.77 (0.27 to 2.15) 0.61
Hypertension
1.19 (0.81 to 1.74) 0.37
Chronic Kidney Disease
1.26 (0.76 to 2.08) 0.37
Ischaemic Heart Disease
1.90 (0.88 to 4.11) 0.1
Non−HCC neoplasm
0.95 (0.57 to 1.58) 0.85
ARLD
0.95 (0.45 to 2.01) 0.89
Ascites
1.85 (1.02 to 3.36) 4.3e−02
Cerebrovascular disease
0.64 (0.30 to 1.36) 0.24
Cirrhosis
6.52 (4.54 to 9.37) 3.8e−24
End Stage Liver Disease
1.24 (0.62 to 2.47) 0.54
NAFLD
0.66 (0.30 to 1.46) 0.31
Peptic ulcer
2.20 (1.39 to 3.49) 7.7e−04
No antiviral initiation (ref)
1.00 (ref)
Init. before diagnosis
1.85 (0.79 to 4.32) 0.15
Init. within 1 year
1.63 (0.92 to 2.90) 0.1
Init. within 2 years
1.24 (0.55 to 2.81) 0.6
Init. within 3 years
1.89 (0.83 to 4.28) 0.13
Init. within >=4 years
0.76 (0.45 to 1.27) 0.29
Antihypertensive use
0.48 (0.21 to 1.07) 0.07
NSAID use
0.96 (0.48 to 1.96) 0.92
Statin use
0.47 (0.22 to 0.99) 4.8e−02
Antidiabetic drug use
1.83 (0.74 to 4.56) 0.19
0.25
0.50
1.00
2.00
4.00
6.00
8.00
10.00

Figure 3.4. Forest plot for Cox proportional hazards model to identify risks of hepatocellular carcinoma (HCC) in an adult population with chronic Hepatitis B virus infection derived from the QResearch primary care database. Analysed using a dataset generated by multiple imputation with chained equations (n = 7029, HCC cases = 161). First (least deprived) to fifth (most deprived) Townsend Deprivation Quintiles are denoted by SES1-5, respectively.

Table 3.12. Cox proportional hazards model using complete imputed dataset generated by multiple imputation with chained equations (n = 7029, HCC cases = 161).

Characteristic	Univariable HR (95% CI)	Multivariable HR (95% CI)	Multivariable HR sensitivity analysis I (95% CI) †	Multivariable HR sensitivity analysis II (95% CI) ‡
Age group, n (%)				
18-25 years	0.47 (0.11 to 2.05)	0.47 (0.11 to 2.07)	0.45 (0.1 to 1.96)	0.58 (0.13 to 2.7)
26-35 years	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
36-45 years	2.49 (1.41 to 4.38)**	1.96 (1.1 to 3.48)*	1.82 (1.01 to 3.27)*	1.83 (1.01 to 3.3)*
46-55 years	5.26 (3.02 to 9.15)**	2.73 (1.51 to 4.95)**	2.55 (1.38 to 4.72)**	2.2 (1.21 to 4.01)*
56-65 years	7.52 (4.14 to 13.67)**	3.17 (1.63 to 6.14)**	3.49 (1.77 to 6.89)**	2.26 (1.13 to 4.54)*
≥66 years	11.89 (6.26 to 22.60)**	4.06 (1.92 to 8.58)**	3.3 (1.42 to 7.72)**	2.41 (1.12 to 5.19)*
Sex, n (%)				
Female	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
Male	4.92 (3.05 to 7.95)**	3.44 (2.07 to 5.73)**	3.61 (2.05 to 6.34)**	3.79 (2.25 to 6.36)**
Townsend Deprivation Quintile, n (%)				
First (least deprived)	2.06 (1.09 to 3.9)*	1.87 (0.96 to 3.65)	1.85 (0.89 to 3.82)	1.84 (0.92 to 3.67)
Second	1.15 (0.59 to 2.24)	1.1 (0.54 to 2.25)	1.07 (0.49 to 2.32)	1.16 (0.55 to 2.43)
Third	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
Fourth	1.11 (0.64 to 1.93)	1.27 (0.72 to 2.24)	1.11 (0.59 to 2.07)	1.23 (0.68 to 2.24)
Fifth (most deprived)	1.34 (0.82 to 2.19)	1.69 (1.01 to 2.84)*	1.66 (0.96 to 2.88)	1.63 (0.95 to 2.8)
Ethnicity, n (%)				
White	1.00 (ref)	1.00 (ref)	1.00 (ref)	
Black African	0.74 (0.48 to 1.14)	0.99 (0.61 to 1.61)	1.07 (0.64 to 1.8)	1.00 (ref)
Bangladeshi	1.29 (0.67 to 2.47)	1.05 (0.52 to 2.14)	1.14 (0.55 to 2.37)	1.09 (0.53 to 2.22)
Caribbean	2.23 (1.01 to 4.89)*	3.32 (1.43 to 7.71)**	4.26 (1.73 to 10.52)**	1.2 (0.43 to 3.34)
Chinese	0.75 (0.44 to 1.28)	0.85 (0.47 to 1.52)	0.97 (0.53 to 1.77)	0.59 (0.32 to 1.09)
Indian	1.67 (0.82 to 3.38)	1.83 (0.84 to 3.98)	1.91 (0.83 to 4.41)	1.29 (0.53 to 3.12)
Other	0.73 (0.42 to 1.28)	1.06 (0.59 to 1.89)	0.75 (0.37 to 1.52)	0.76 (0.41 to 1.41)
Other Asian	0.59 (0.25 to 1.37)	0.87 (0.36 to 2.08)	1.03 (0.42 to 2.52)	0.61 (0.27 to 1.39)

Pakistani	0.47 (0.2 to 1.1)	0.67 (0.27 to 1.66)	0.76 (0.29 to 1.98)	0.68 (0.3 to 1.51)
Cigarette consumption, n (%)				
Non-smoker	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
Ex-smoker	1.1 (0.68 to 1.78)	0.77 (0.46 to 1.27)	0.89 (0.52 to 1.52)	0.63 (0.37 to 1.07)
Light smoker (1-9 cigarettes/day)	1.73 (1.19 to 2.53)**	1.41 (0.94 to 2.12)	1.47 (0.94 to 2.28)	1.18 (0.76 to 1.82)
Moderate to heavy smoker (≥10 cigarettes per day)	1.34 (0.67 to 2.65)	1.05 (0.51 to 2.18)	1.3 (0.61 to 2.78)	1.82 (0.95 to 3.49)
Alcohol consumption, n (%)				
Non-drinker	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
Trivial drinker (<1 unit per day)	1.24 (0.78 to 1.98)	1.4 (0.86 to 2.28)	1.23 (0.71 to 2.14)	1.02 (0.57 to 1.82)
Light drinker (1-2 units per day)	0.81 (0.36 to 1.84)	0.73 (0.32 to 1.69)	0.81 (0.35 to 1.9)	0.54 (0.2 to 1.46)
Moderate to heavy drinker (≥3 units per day)	1.26 (0.52 to 3.08)	0.87 (0.34 to 2.23)	1.22 (0.48 to 3.09)	0.63 (0.23 to 1.69)
BMI, n (%)				
Underweight (<18.5 kg/m^2)	1.3 (0.56 to 3)	1.31 (0.55 to 3.15)	1.63 (0.64 to 4.14)	0.65 (0.19 to 2.21)
Normal weight (18.5-24.9 kg/m^2)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
Overweight (25.0-29.9 kg/m^2)	1.07 (0.75 to 1.52)	1.06 (0.74 to 1.52)	1 (0.68 to 1.48)	0.86 (0.59 to 1.25)
Obese (≥30 kg/m^2)	1.06 (0.68 to 1.63)	1.02 (0.64 to 1.62)	1 (0.6 to 1.64)	0.79 (0.49 to 1.28)
Antiviral initiation, n (%)				
No initiation	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
Before CHB diagnosis	3.67 (1.71 to 7.9)**	1.85 (0.79 to 4.32)	2.15 (0.88 to 5.24)	1.39 (0.59 to 3.3)
Within 1 year of CHB diagnosis	4.93 (2.95 to 8.24)**	1.63 (0.92 to 2.9)	1.48 (0.8 to 2.76)	1.66 (0.91 to 3.03)
Within 2 years of CHB diagnosis	3.42 (1.59 to 7.35)**	1.24 (0.55 to 2.81)	1.64 (0.73 to 3.72)	1.41 (0.61 to 3.23)
Within 3 years of CHB diagnosis	4.32 (2.01 to 9.29)**	1.89 (0.83 to 4.28)	1.91 (0.78 to 4.7)	2.06 (0.88 to 4.87)
≥4 years after CHB diagnosis	1.55 (0.95 to 2.55)	0.76 (0.45 to 1.27)	0.62 (0.34 to 1.13)	0.49 (0.28 to 0.84)
Type 2 DM, n (%)	2.57 (1.79 to 3.7)**	1.57 (0.72 to 3.41)	1.07 (0.43 to 2.68)	1.46 (0.63 to 3.37)
Congestive heart failure, n (%)	2.1 (0.78 to 5.68)	0.77 (0.27 to 2.15)	1.03 (0.36 to 2.96)	0.7 (0.24 to 2.07)

Hypertension, n (%)	2.32 (1.68 to 3.22)**	1.19 (0.81 to 1.74)	1.15 (0.76 to 1.76)	0.94 (0.63 to 1.4)
Chronic kidney disease, n (%)	3.83 (2.5 to 5.88)**	1.26 (0.76 to 2.08)	1.31 (0.74 to 2.32)	1.01 (0.61 to 1.68)
Alcohol-related liver disease, n (%)	5.82 (3.06 to 11.05)**	0.95 (0.45 to 2.01)	1.17 (0.54 to 2.49)	1 (0.45 to 2.22)
Ascites, n (%)	13.61 (8.42 to 21.99)**	1.85 (1.02 to 3.36)*	2.08 (1.07 to 4.05)*	1.68 (0.93 to 3.07)
Cerebrovascular disease, n (%)	2.31 (1.4 to 3.83)**	0.64 (0.3 to 1.36)	0.72 (0.32 to 1.64)	0.65 (0.3 to 1.41)
Cirrhosis, n (%)	12.65 (9.22 to 17.36)**	6.52 (4.54 to 9.37)**	7.52 (5.07 to 11.14)**	5.11 (3.5 to 7.47)**
End-stage liver disease, n (%)	7.66 (4.34 to 13.52)**	1.24 (0.62 to 2.47)	0.92 (0.41 to 2.06)	1.75 (0.88 to 3.45)
Ischaemic heart disease, n (%)	3.01 (1.82 to 4.99)**	1.9 (0.88 to 4.11)	1.69 (0.7 to 4.08)	1.67 (0.75 to 3.73)
Non-alcoholic fatty liver disease, n (%)	0.69 (0.32 to 1.48)	0.66 (0.3 to 1.46)	0.52 (0.21 to 1.31)	0.75 (0.33 to 1.68)
Non-HCC neoplasm, n (%)	2.3 (1.45 to 3.67)**	0.95 (0.57 to 1.58)	--	
Peptic ulcer, n (%)	5.54 (3.68 to 8.34)**	2.20 (1.39 to 3.49)**	2.33 (1.42 to 3.82)**	1.96 (1.22 to 3.15)
Antidiabetic drug use, n (%)	2.53 (1.75 to 3.65)**	1.83 (0.74 to 4.56)	1.98 (0.73 to 5.37)	1.09 (0.4 to 2.94)
Antihypertensive use, n (%)	2.2 (1.39 to 3.48)**	0.48 (0.21 to 1.07)	0.48 (0.19 to 1.23)	0.92 (0.4 to 2.15)
NSAID use, n (%)	1.96 (1.24 to 3.1)**	0.96 (0.48 to 1.96)	0.96 (0.44 to 2.09)	1.16 (0.54 to 2.49)
Statin use, n (%)	1.72 (1.06 to 2.78)*	0.47 (0.22 to 0.99)*	0.64 (0.26 to 1.56)	0.38 (0.17 to 0.84)
ALT quintile, n (%)				
First quintile (mean log ALT=, SD=)	--	--	--	0.95 (0.46 to 1.99)
Second quintile (mean log ALT=, SD=)	--	--	--	1.91 (1.03 to 3.56)*
Third quintile (mean log ALT=, SD=)	--	--	--	1.00 (ref)
Fourth quintile (mean log ALT=, SD=)	--	--	--	1.93 (1.08 to 3.46)*
Fifth quintile (mean log ALT=, SD=)	--	--	--	2.34 (1.35 to 4.06)**
AST quintile, n (%)				

First quintile (mean log AST=, SD=)	--	--	--	0.6 (0.33 to 1.07)
Second quintile (mean log AST=, SD=)	--	--	--	0.99 (0.56 to 1.73)
Third quintile (mean log AST=, SD=)	--	--	--	1.00 (ref)
Fourth quintile (mean log AST=, SD=)	--	--	--	0.84 (0.5 to 1.4)
Fifth quintile (mean log AST=, SD=)	--	--	--	0.97 (0.6 to 1.57)
PL quintile, m (%)				
First quintile (mean log PL=, SD=)	--	--	--	1.25 (0.75 to 2.08)
Second quintile (mean log PL=, SD=)	--	--	--	0.67 (0.36 to 1.23)
Third quintile (mean log PL=, SD=)	--	--	--	1.00 (ref)
Fourth quintile (mean log PL=, SD=)	--	--	--	0.79 (0.41 to 1.52)
Fifth quintile (mean log PL=, SD=)	--	--	--	1.01 (0.56 to 1.84)

HR, hazards ratio; CHB, chronic hepatitis B virus infection; ICD, International classification of disease; SNOMED, Systemised Nomenclature of Medicine; SD, standard deviation; BMI, body mass index; DM, diabetes mellitus; HCC, hepatocellular carcinoma; NSAID, non-steroidal anti-inflammatory drug.

* $P < 0.05$

** $P < 0.01$

† Patients with history of any non-hepatocellular carcinoma neoplasm excluded.

‡ AST, ALT and Plt added to model.

Hazards of HCC were increased in males (adjusted hazards ratio (aHR) 3.44, 95% CI 2.07 to 5.73), with increasing age (aHR for 36-45 years 1.96, 95% CI 1.10 to 3.48; aHR for 46-55 years 2.73, 95% CI 1.51 to 4.95; aHR for 56-65 years 3.17, 95% CI 1.63 to 6.14; aHR for ≥66 years 4.06, 95% CI 1.92 to 8.58, as compared to 26-35 years reference group) and in the fifth deprivation quintile as compared to the third quintile (aHR 1.69, 95% CI 1.01 to 2.84). Hazards of HCC in the Caribbean ethnicity group were higher than those in the White reference group (aHR 3.32, 95% CI 1.43 to 7.71) but did not differ in any other ethnic category. There were no associations between alcohol consumption, cigarette smoking, or BMI with hazards of HCC. As expected, increased hazards of HCC were associated with evidence of advanced liver disease, cirrhosis (aHR 6.52, 95% CI 4.54 to 9.37) and ascites (aHR 1.85, 95% CI 1.02 to 3.36). Interestingly, peptic ulcer disease was also associated with increased

HCC hazards (aHR 2.20, 95% CI 1.39 to 3.49). Statin use was associated with reduced hazards of HCC (aHR 0.47, 95% CI 0.22 to 0.99). No other medicines, including antiviral treatment, associated with hazards of HCC.

Hazards ratios did not change materially in strength or direction upon sensitivity analysis excluding non-HCC neoplasms or including AST, ALT and Plt at baseline (**Table 3.12**).

Main model results are robust to complete-case sensitivity analysis

I undertook sensitivity analysis restricted to the subgroup of patients for whom complete data were available (n=3648 patients in whom 68 cases of HCC occurred (**Table 3.13**)). Confidence intervals for all aHRs were more precise in the model from the imputed dataset compared to the complete case analysis due to larger sample size in the former. Hazard ratios did not change materially in strength or direction in sensitivity analysis undertaken to exclude patients with history of non-HCC neoplasms (**Table 3.13**).

Table 3.13. Cox proportional hazards complete-case model conducted on adults with hepatitis B virus infection for whom data was complete in the original QResearch dataset (n = 3648, HCC cases = 68).

Characteristic	Univariable HR (95% CI)	Multivariable HR (95% CI)	Multivariable HR (95% CI) †
Age group, n (%)			
18-25 years	0.85 (0.19 to 3.88)	1.7 (0.35 to 8.26)	1.98 (0.4 to 9.91)
26-35 years	1.00 (ref)	1.00 (ref)	1.00 (ref)
36-45 years	1.59 (0.72 to 3.51)	1.44 (0.61 to 3.37)	1.46 (0.61 to 3.48)
46-55 years	3.14 (1.43 to 6.85)**	2.4 (0.99 to 5.82)	1.91 (0.75 to 4.85)
56-65 years	6.84 (3.13 to 14.97)**	3.24 (1.23 to 8.53)**	3.89 (1.43 to 10.61)**
≥66 years	5.47 (1.98 to 15.14)**	2.41 (0.7 to 8.37)	1.7 (0.37 to 7.73)
Sex, n (%)			
Female	1.00 (ref)	1.00 (ref)	1.00 (ref)
Male	5.50 (2.63 to 11.50)**	3.35 (1.42 to 7.88)**	3.49 (1.35 to 8.99)**
Townsend Deprivation Quintile, n (%)			
First (least deprived)	3.12 (0.95 to 10.24)	3.6 (1 to 12.99)	3.95 (0.9 to 17.35)
Second	1.91 (0.55 to 6.61)	2.47 (0.63 to 9.74)	1.32 (0.27 to 6.54)
Third	1.00 (ref)	1.00 (ref)	1.00 (ref)
Fourth	1.74 (0.63 to 4.79)	1.69 (0.57 to 5.02)	1.78 (0.55 to 5.74)
Fifth (most deprived)	2.18 (0.86 to 5.56)	2.34 (0.84 to 6.53)	2.75 (0.92 to 8.16)

Ethnicity, n (%)			
White	1.00 (ref)	1.00 (ref)	1.00 (ref)
Black African	1.14 (0.58 to 2.24)	2.37 (0.96 to 5.81)	2.95 (1.13 to 7.73)*
Bangladeshi	2.16 (0.92 to 5.05)	1.77 (0.63 to 4.96)	1.88 (0.64 to 5.55)
Caribbean	1.91 (0.44 to 8.32)	3.32 (0.59 to 18.6)	8.31 (1.59 to 43.59)*
Chinese	0.68 (0.25 to 1.85)	1.3 (0.42 to 4.02)	2 (0.6 to 6.6)
Indian	2.52 (0.92 to 6.88)	5.19 (1.57 to 17.09)**	7.23 (1.98 to 26.49)*
Other	1.14 (0.49 to 2.66)	2.18 (0.8 to 5.92)	1.41 (0.43 to 4.67)
Other Asian	0.6 (0.14 to 2.63)	1.41 (0.28 to 7.03)	1.49 (0.28 to 8.1)
Pakistani	0.86 (0.29 to 2.57)	1.87 (0.53 to 6.61)	1.82 (0.48 to 6.84)
Cigarette consumption, n (%)			
Non-smoker	1.00 (ref)	1.00 (ref)	1.00 (ref)
Ex-smoker	1.57 (0.77 to 3.17)	1.1 (0.51 to 2.39)	1.56 (0.68 to 3.56)
Light smoker (1-9 cigarettes/day)	2.91 (1.68 to 5.06)**	2.29 (1.15 to 4.57)*	3.28 (1.54 to 6.97)**
Moderate to heavy smoker (≥10 cigarettes per day)	1.96 (0.70 to 5.53)	2.92 (0.92 to 9.29)	3.52 (1.06 to 11.67)*
Alcohol consumption, n (%)			
Non-drinker	1.00 (ref)	1.00 (ref)	1.00 (ref)
Trivial drinker (<1 unit per day)	1.17 (0.56 to 2.45)	1.79 (0.79 to 4.02)	1.48 (0.59 to 3.71)
Light drinker (1-2 units per day)	0.66 (0.16 to 2.71)	0.45 (0.1 to 2.13)	0.6 (0.13 to 2.78)
Moderate to heavy drinker (≥3 units per day)	1.25 (0.31 to 5.14)	0.46 (0.09 to 2.3)	1.15 (0.22 to 5.93)
BMI, n (%)			
Underweight (<18.5 kg/m^2)	0.42 (0.06 to 3.11)	0.95 (0.12 to 7.78)	1.74 (0.21 to 14.22)
Normal weight (18.5-24.9 kg/m^2)	1.00 (ref)	1.00 (ref)	1.00 (ref)
Overweight (25.0-29.9 kg/m^2)	0.92 (0.55 to 1.55)	1 (0.57 to 1.77)	1.17 (0.62 to 2.18)
Obese (≥30 kg/m^2)	0.54 (0.26 to 1.15)	0.68 (0.29 to 1.63)	0.9 (0.36 to 2.25)
Antiviral initiation, n (%)			
No initiation	1.00 (ref)	1.00 (ref)	1.00 (ref)
Before CHB diagnosis	3.29 (1.01 to 10.68)*	1.78 (0.45 to 6.97)	2.87 (0.71 to 11.65)
Within 1 year of CHB diagnosis	5.11 (2.45 to 10.66)**	2.32 (1 to 5.38)	1.84 (0.74 to 4.59)
Within 2 years of CHB diagnosis	4.04 (1.57 to 10.39)**	1.46 (0.5 to 4.26)	1.51 (0.51 to 4.44)
Within 3 years of CHB diagnosis	6.03 (2.36 to 15.37)**	2.36 (0.82 to 6.81)	2.26 (0.7 to 7.35)
≥4 years after CHB diagnosis	1.95 (0.90 to 4.23)	0.92 (0.4 to 2.1)	0.67 (0.24 to 1.87)
Type 2 DM, n (%)	3.26 (1.96 to 5.42)**	2.5 (0.75 to 8.34)	1.74 (0.42 to 7.32)
Congestive heart failure, n (%)	1.06 (0.15 to 7.70)	0.22 (0.03 to 1.92)	0.19 (0.02 to 1.82)
Hypertension, n (%)	2.29 (1.41 to 3.73)	1.05 (0.56 to 1.96)	1.11 (0.56 to 2.21)

Chronic kidney disease, n (%)	5.48 (3.14 to 9.53)**	3.32 (1.65 to 6.70)**	3.11 (1.43 to 6.79)**
Alcohol-related liver disease, n (%)	5.79 (2.50 to 13.42)**	1.51 (0.49 to 4.64)	1.18 (0.37 to 3.72)
Ascites, n (%)	18.11 (8.63 to 37.99)**	1.81 (0.69 to 4.74)	1.51 (0.55 to 4.17)
Cerebrovascular disease, n (%)	2.35 (1.16 to 4.76)*	1.08 (0.37 to 3.20)**	1.48 (0.46 to 4.72)
Cirrhosis, n (%)	17.12 (10.37 to 28.25)**	8.79 (4.85 to 15.93)**	11.54 (5.95 to 22.39)**
End-stage liver disease, n (%)	9.28 (3.71 to 23.20)**	1.52 (0.46 to 5.09)	1.73 (0.46 to 6.51)
Ischaemic heart disease, n (%)	1.76 (0.75 to 4.10)	0.37 (0.1 to 1.4)	0.31 (0.07 to 1.33)
Non-alcoholic fatty liver disease, n (%)	0.60 (0.19 to 1.91)	0.56 (0.16 to 1.93)	0.63 (0.18 to 2.2)
Non-HCC neoplasm, n (%)	2.49 (1.19 to 5.22)*	1.23 (0.5 to 3.03)	--
Peptic ulcer, n (%)	4.21 (2.18 to 8.13)**	2.20 (1.03 to 4.70)*	3.07 (1.35 to 7)**
Antidiabetic drug use, n (%)	3.02 (1.82 to 4.99)**	1.43 (0.37 to 5.56)	1.47 (0.31 to 6.95)
Antihypertensive use, n (%)	2.56 (1.42 to 4.62)**	0.6 (0.2 to 1.78)	0.62 (0.18 to 2.17)
NSAID use, n (%)	1.98 (1.08 to 3.64)*	0.81 (0.32 to 2.09)	0.88 (0.31 to 2.53)
Statin use, n (%)	1.81 (0.97 to 3.38)	0.67 (0.25 to 1.81)	0.93 (0.29 to 2.98)

HR, hazard ratio; CHB, chronic hepatitis B virus infection; ICD, International classification of disease; SNOMED, Systemised Nomenclature of Medicine; SD, standard deviation; BMI, body mass index; DM, diabetes mellitus; HCC, hepatocellular carcinoma; NSAID, non-steroidal anti-inflammatory drug.

* $P < 0.05$

** $P < 0.01$

† Patients with history of any non-hepatocellular carcinoma neoplasm excluded

3.6. Discussion

CHB prevalence in the UK

This is the largest cohort of CHB individuals characterised in England to date, from either EHR or traditional prospective cohorts. However, prevalence estimate produced from this QResearch CHB cohort of 0.023% contrasts with current UK prevalence estimates identified in my literature search, which range from 0.27% to 0.73%. Since undertaking this literature review and sharing analysis outputs via Wellcome Open Research, an updated HBV surveillance report has been produced by UKHSA (30). This report includes a CHB prevalence estimate in the UK of 0.45% (95CI 0.35% to 0.60%), which has been derived using HBsAg laboratory surveillance data. These findings suggest that many cases of HBV infection are not diagnosed and/or not

recorded in primary care records. Therefore, substantial efforts are needed in clinical, governance and data domains to enhance diagnosis, linkage to care and information linkage, to support better individual patient care and to inform public health programmes.

HBV is a notifiable disease in the UK, and UKHSA has a CHB surveillance system, monitoring testing and diagnoses across the care pathway. This system incorporates data from diagnoses through to outcomes, (including end-stage liver disease, transplantation, liver cancer and deaths) using data feeding into the SSBBV surveillance system which captures testing in primary and secondary care settings), new laboratory diagnoses, hospital activity datasets and registries (NHS Digital hospital episode statistics, ONS cancer and deaths registries, NHS Blood and Transplant registry). SSBBV sources cover approximately 40% of the general population of England (201). The most recent HBV prevalence estimate reported by UKHSA was generated using a method developed by Schneir et al to estimate Scottish HBV prevalence (208). Briefly, this method quantified HBsAg positivity prevalence in women aged 15-44 , stratified by ethnicity, using antenatal HBV screening records. These estimates served as “starting point” estimates. Thereafter, Poission regression was undertaken to generate sex- and age group-specific rate ratios for HBsAg positivity; models were adjusted for test referral source and NHS region. From here, the “starting point” estimates were multiplied by relevant rate ratios to generate age-, sex- and ethnicity-specific CHB estimates. This method assumes that differences observed in rate ratios, generated on data collected in health service laboratory settings, are representative of the general population, which may not be true. Nonetheless, this likely gives the best estimate of diagnosed prevalence among a tested population, but because it combines acute incident infections and new diagnoses of pre-existing chronic infection, incidence and prevalence cannot be disaggregated.

The majority of diagnostic data are generated through testing individuals with risk factors for HBV infection or evidence of liver disease, and among those accepting risk-based testing (as captured in laboratory testing surveillance) likely overestimates the overall population prevalence. Alternative estimates can be generated from screening blood donors (209), but this group are self-selected low-risk individuals who have

undergone deferral discussions (210–212), and are predominantly UK born, and therefore are not representative of the general population, leading to an under-estimate of prevalence . No existing prevalence estimates factor in the undiagnosed burden, which represents the majority of people living with HBV infection (the WHO estimates that only 10.5% of people with CHB are aware of their infection status (1)). Furthermore, the highest prevalence of CHB is in groups for whom provision of healthcare is inadequate, and/or access to healthcare is challenging (including migrants, sex-workers, prisoners, and people experiencing homelessness), so overall there are still many gaps in the data, and it is most likely that estimates using primary care datasets considerably underestimate the true burden (16). There is a clear need for improved characterization of the CHB population in the UK, particularly to factor in the undiagnosed sub-population and to also sufficiently describe marginalized CHB subgroups.

This QResearch cohort identifies an increased burden of HBV infection in young, male, socioeconomically deprived individuals belonging to ethnic minorities within the UK (47). These individuals comprise a larger percentage of the London city population as compared to other regions in the UK (43), and therefore regional differences must be considered when deriving a population-wide prevalence estimate, which can be achieved through statistical models.

Less biased estimates are achieved by non-targeted testing programmes such as universal antenatal screening. Opt-out screening of blood drawn from patients admitted to a selection of Emergency Departments (EDs) for HBV, HCV and HIV is being piloted across 34 sites across the UK, with new HBV diagnoses outnumbering HIV or HCV by >2:1 in the first 100 days (213). Over time, expansion of this programme should provide an additional valuable source of epidemiological data.

While UKHSA surveillance data may include some demographic characteristics (age, sex, postcode for deprivation), unless linked to other healthcare datasets, they typically lack more detailed clinical and demographic indicators (for example, measures of deprivation, lifestyle factors, assessment of liver disease, HBV treatment coverage and burdens of other comorbidities) which are needed to characterise the infected population. In contrast, EHR databases (such as CPRD and QResearch) have the advantage of collecting relevant demographic and clinical metadata which

are not captured by UKHSA. However, linkage across data sources is disaggregated, and thereby each EHR-based estimate misses a portion of the infected population. For example, primary care data may not reflect testing conducted in secondary care (71), blood safety (transfusion/transplantation) and laboratory data generated by other services, while secondary care data are typically only reliable for the sub-population enrolled in consistent hospital follow-up. Poor data flow between diagnostic testing and EHR reflect a low clinical follow-up rate following a positive HBsAg test. This lack of data linkage to care reflects how services may be underserving the CHB population, with gaps in referral pathways, inadequate communication and education (including translation services), and failures to deliver services to marginalised communities. UKHSA surveillance systems such as SGSS and SSBBV undertake a systematic effort to harmonise and synthesize HBV testing and diagnostic data sources in order to monitor the population-wide HBV burden. This is not the purpose of EHR databases, and they have not been created to monitor BBVs on a population-wide scale. Therefore, EHR databases offer the potential to characterise a subset of those infected with HBV, but do not currently generate a picture that is generalisable to the wider infected population and cannot solely be used to estimate prevalence.

Prevalence estimates for Hepatitis C virus (HCV) (214) and human immunodeficiency virus (HIV) (215) have recently been generated using multi-parameter evidence synthesis and back-calculation models. Similar modelling approaches to produce estimates of HBV incidence and prevalence in the UK are warranted. Both modelling methods require certain inputs including previous incidence estimates and population demographic and migrant frequencies in order to produce robust prevalence estimates (214,216,217). Historical trend data on HBV rates, both overall and stratified by subgroup(s) of interest, is utilised in order to inform a model which generates a reliable prevalence estimate and projects future trends, as inputs are assumed to be representative of the population being modelled. To model the processes of transitioning to chronic infection, country of birth for migrant subgroups must be available, as the majority of individuals with CHB in the UK have moved from countries with high HBV burdens. The lack of these consistent and representative data required to generate HBV prevalence estimates in the UK presents a challenge to producing a population-wide estimate. Furthermore, it’s difficult to externally validate a model when available national data sources are used as inputs (216). Nevertheless, UKHSA aim

to produce estimate by the end of 2025. Enhanced investment is needed to support the establishment of national registries with robust centralised data linkage between sources including national laboratory surveillance systems of BBV testing and new diagnoses, and thus determine which population subgroups are bearing the majority of the HBV disease burden. This will inform prevalence modelling and provide an evidence base for delivery of appropriate resources and interventions, and to benchmark progress towards elimination targets.

Characterisation of the cohort and HCC risk factors

To make advances towards elimination goals for HBV infection in the UK, there is an urgent need to improve understanding of the epidemiology and outcomes of infection in different subgroups. In this chapter, I have harnessed diverse datasets to develop a better estimate of HBV epidemiology, and to identify gaps and challenges in existing data. A deep dive into primary care data allows us to explore groups at most risk of HCC complications, highlighting where more resources are needed to reduce morbidity and mortality of this complication.

The CHB population I identified in the QResearch database was ethnically diverse, with higher proportions of black and ethnic minority individuals than either the wider QResearch database (205) or the general English population (206). Importantly, the CHB population is disproportionately socioeconomically deprived, with substantial burdens of comorbid disease. This highlights a vulnerable population, for whom specific resources should be targeted to redress health inequities. The QResearch database has geographic coverage across England, and therefore my findings should be generalizable to the wider country, and potentially to other settings. However, a portion of individuals with CHB in the UK will invariably not be accessing primary care services due to the vulnerability of the disease population (16). Therefore, it is possible that this sample suffers from selection bias by including only a subset of the CHB population who are accessing GP services. However, this piece of work is nevertheless important in order to characterise and describe a large-scale CHB cohort and investigate HCC risk factors.

I report increased hazards of HCC associated with increasing age, male sex, socioeconomic deprivation, Caribbean ethnicity, severe liver disease (ascites and cirrhosis), and comorbid disease (peptic ulcer disease). Age, sex and T2DM have previously been found to positively associate with HCC risk in CHB (80), however this is the first cohort study to report these associations in an ethnically diverse cohort. It is likely that the present analysis is underpowered to detect associations with alcohol and cigarette consumption with HCC risk in this cohort, although smoking was identified as a risk factor on sensitivity analysis.

It is possible that the positive association of peptic ulcer disease with hazards of HCC is confounded by proton pump inhibitor (PPI) administration for peptic ulcer treatment. Information concerning PPI prescription and usage was not available in this QResearch extract, but previous observational studies have reported increased risks of HCC associated with PPIs (218,219). Pooled risk estimates from meta-analyses have been variable, with some confirming this association (220) and others failing to report associations (221,222), particularly after restricting to studies of participants with viral hepatitis (222)

Influence of HBV genotype

HBV genotype is not routinely determined in clinical practice, so is not available from EHR datasets. Genotype is associated with ethnicity (223,224), and viral genotypes have been associated with both an increased HCC risk (225) and antiviral treatment resistance (226). Therefore, associations of HCC risk with ethnicity and/or socioeconomic status may be driven by HBV genotype. However, the association may also be mediated by unmeasured population genetic factors and/or other exposures (including lifestyle factors). Increased HCC hazards observed in the most deprived Townsend quintile may similarly be confounded by genotype, however, it is more likely the observed association is mediated by socioeconomic factors whereby access to routine care/surveillance is poorer and additional lifestyle or environmental factors which increase HCC risk are more prevalent.

Drug treatment and HCC development

I report a protective association of statin use with HCC hazards. This association has been reported in previous CHB cohorts (227,228), in individuals with predisposing HCC risk factors including cirrhosis and T2DM (229,230) and in a general patient population (228–230). Further analysis, including potential mediation analysis where data availability allows, is warranted to investigate the mechanisms of this association.

I did not find an association of antiviral treatment with reduced HCC risk, although there is evidence from previous observational and randomised interventional studies that CHB treatment with antiviral nucleoside analogue (NA) agents reduces HCC risk (231–233). In this sample, 10.3% of individuals initiated antiviral treatment during follow-up. However, a substantial proportion of treatment data may be missing from primary care records, as this information is typically captured in secondary/tertiary care records where HBV prescribing is based (71). For example, 16% of CHB patients in secondary care were receiving antiviral treatment, based on the NIHR Health Informatics Collaborative (HIC) database (234). My primary care cohort may systematically differ to samples in clinical trials and hospital-based cohorts whereby only a small proportion of individuals are treatment-eligible in this cohort. Additionally, receipt of antiviral treatment may target those with more severe progression/prognosis compared to untreated individuals at baseline, thus negating any signal for a positive effect. Clinical guidelines recommend criteria for treatment, but it is not possible for us to determine the extent to which the 89.7% of patients with no treatment recorded are treatment ineligible or are receiving treatment in a secondary/tertiary care setting that is not captured in QResearch data. Therefore, the association of antiviral treatment and HCC risk cannot be confidently determined from this dataset.

Application of data to HCC risk stratification

HCC risk scores (including PAGE-B (72), REACH-B (73,74), GAG-HCC (75,76) and CU-HCC (77–79), utilise various characteristics, including age, sex, and laboratory parameters to predict HCC risk. The utility of existing risk scores in homogenous patient subgroups has been demonstrated (235). The presented results highlight the

need for validation (and potentially modification) of such scores in heterogenous samples which are ethnically and clinically diverse. These scores however require quantified laboratory parameters such as platelets, HBsAg and VL, and validation of such scores in QResearch is restricted due to the absence of such laboratory data in primary care electronic health systems. Identifying groups who carry a significantly higher risk of developing HCC may inform enhanced screening and support early diagnosis. Determination of modifiable risk factors which influence HCC development should inform interventions, for example a risk associated with smoking should inform specific smoking cessation advice and support among people living with HBV infection.

Caveats of primary care EHR analysis – missing data

I highlight substantial data missingness due to poor primary care coverage and HBV representation in such data sources. Missingness likely associated with characteristics not measured in this dataset including thereby excluding marginalised subgroups (including undocumented migrants, highly mobile population subgroups, and people who do not speak English) (16), which may further increase the proportion who fall into the most socioeconomically deprived quintiles.

Most patients were identified by coding, with a minority (3.0%) having confirmatory laboratory tests accessible in QResearch. This is unsurprising given that specialist referral is recommended following a positive test result (8) thus second confirmatory HBV lab data (VL and/or HBsAg) would typically be performed in a secondary or tertiary care centre, and these results are not reliably transferred into primary care EHR systems. Efforts to enhance linkage of data between primary and secondary care would be of significant benefit to patient care, and potentially increase the chance of timely diagnosis of liver complications, including HCC, in patients presenting to their primary care practitioner with early symptoms.

Length of follow-up is shorter than anticipated for a chronic disease despite a 20-year study period from 1999 to 2019. This is not uncommon in EHR databases in which there may be time lags between notification of a patient characteristic/disease and input into electronic record systems. Additionally, given the neglected nature of CHB,

it is likely that individuals present to primary care late in the infection course and/or diagnoses are primary managed in secondary/tertiary services without linkage to primary care records. Follow-up was differential between individuals who did and did not develop HCC, with shorter median follow-up duration in HCC cases. This may be due to HCC cases only presenting to a GP practice when their liver disease has progressed to a symptomatic, more advanced stage.

Analytical corrections for missing data

I have undertaken imputation of missing baseline data, in line with previous investigations in QResearch (195–197), but was unable to impute important HBV biomarkers (including VL and HBsAg) as >90% of participants were missing any measurement at baseline. I therefore excluded these variables from analysis. High missingness was observed for additional biomarkers, specifically AST, ALT and Plt, which serve as indicators of liver function and can be used to derive proxy scores for staging fibrosis and cirrhosis. Replication of the present analysis in secondary care datasets is a future opportunity to investigate the utility of these biomarkers as markers of disease endpoints. Similarly, I was unable to time-update models for changes in alcohol and cigarette consumption throughout follow-up due to lack of repeated measurements (>90% of the cohort had a one-off record for alcohol and smoking). Due to missing data, I was also unable to validate and/or modify such scores. Future investigation in data sources where variables are more complete and remeasurement is more comprehensive is warranted to enable these analytical improvements and allow for more precise estimation of effect sizes.

Characterisation of comorbid or end-organ disease in primary care EHR data sources is also limited due to missing information. For example, cirrhosis is inferred via biopsy, hepatic imaging or laboratory scores (236). However, these data are usually collected in secondary/tertiary care and are not reliably transferred into primary care EHRs. Although I identified a subset of individuals with a diagnostic code indicative of cirrhosis, cirrhotic prevalence in CHB was likely underestimated, and due to this missingness I was unable to investigate risk factors for cirrhosis as an endpoint. Similarly, the prevalence of comorbidities, including renal and cerebrovascular disease, is likely underestimated for this reason, and I am underpowered to detect

associations with HCC risk. A high proportion of data on alcohol consumption is missing, and where alcohol intake is recorded, this may not be reliable. Although alcohol may be a relevant risk factor for HCC development, I am unable to report this association in this study.

Conclusions

The CHB population in the England is ethnically diverse and socioeconomically deprived. I have identified risk factors for HCC, and validated associations observed previously in different CHB populations, in the largest CHB cohort characterised in the UK to date. I have also used the cohort to generate an estimate of CHB population prevalence, and also summarised and appraised previous estimates via a literature review. Future analysis warrants validation and/or modification of existing CHB-HCC risk scores in diverse patient samples. This may require establishment of a prospective cohort study with long-term follow-up of CHB-individuals sampled from a diverse population. Missingness in various data fields an important observation: it is a limitation both for identification of individuals with CHB, and for robust description of those individuals whose diagnosis is recorded. Enhanced capture of such data is crucial to provide an evidence base for interventions, including diagnostic screening, treatment and surveillance, modification of risk factors for HCC, and monitoring progress towards elimination targets.

4. Chronic HBV cohort in the National Institute for Health Research Health Informatics Collaborative secondary care database: antiviral treatment coverage and eligibility, and the impact of the COVID-19 pandemic on routine surveillance

4.1. Availability of content

I herein present findings from two separate analyses undertaken in the NIHR HIC secondary care EHR database. The investigation of the impact of the coronavirus disease 2019 (COVID-19) pandemic on CHB surveillance is currently published (awaiting peer review) on *Wellcome Open Research* (234). I am the sole first author on this manuscript, having conducted the analysis and written the first draft of the manuscript. Results from the investigation of HBV treatment eligibility and coverage have been prepared into a manuscript which will be submitted to *Gut* For this study I am also the single first author on this manuscript, as I conducted the analysis and wrote the first draft of the manuscript independently.

4.2. Background

4.2.1. HBV and secondary/tertiary care in the UK

In the UK, CHB is a condition that is primarily managed in secondary and tertiary healthcare settings (8,176), although HBsAg-positive individuals may be initially diagnosed in primary care. If an individual with HBsAg positivity presents to their general practitioner, NICE guidelines recommend that multiple initial tests be undertaken, including: HBeAg status, HBV viral load (VL) quantification, HCV and HIV co-infection status, along with testing for HCC via hepatic ultrasound and alpha-fetoprotein (AFP) assessment. Referral to a specialist (hepatologist, gastroenterologist or infectious disease specialist with an interest in hepatology) is then recommended with the results of these initial tests (8).

After referral to a specialist clinic, individuals with CHB typically undergo assessment for liver disease severity and antiviral treatment eligibility. Elastography measures liver

stiffness, which is associated with inflammation or fibrosis/cirrhosis (8). Disease severity and risk of progression is assessed via various clinical investigations including transient elastography, and assessment of serum concentrations of liver enzymes (including ALT), HBV VL, and either binary or quantitative assessment of viral antigens HBsAg and HBeAg. A biopsy may occasionally be undertaken if other results are equivocal or other factors contributing to liver disease need to be assessed. These parameters are used to inform long-term surveillance and eligibility for initiation of antiviral treatment.

CHB antiviral treatment is prescribed and managed in secondary/tertiary care outpatient clinics (8). According to UK and European treatment guidelines (8,237), eligibility is determined according to a combination of factors including age, sex, the above panel of laboratory markers, and liver disease severity. Male sex, older age, high HBV VL (>2000 IU/ml), ALT derangement above the upper limit of normal (ULN, ≥30 IU/L in males and ≥19 IU/L in females) and cirrhosis or decompensated liver disease are indicators for treatment initiation. The mainstay of modern antiviral treatment is administration of a nucleoside analogue (NA) agent, which supresses viral replication and thereby decreases liver inflammation and fibrosis (8); in the UK this is typically tenofovir disoproxil fumarate (TDF) or entecavir (ETV).

In order to monitor treatment eligibility (as well as progression of liver disease), treatment indicators must be measured repeatedly over time to detect potential changes. Therefore, secondary and tertiary care electronic health systems contain a wealth of longitudinal clinical data. Parameters available include not only those relating to HBV infection and treatment, but any additional clinical and demographic characteristics including other serum biomarkers, height, weight, date of birth and ethnicity. Data spanning multiple secondary and tertiary care centres across the country can be harnessed to characterise the CHB population access specialist care, and to determine which patients are receiving antiviral treatment.

Strict antiviral treatment criteria result in only a minority of patients being eligible for treatment. It is possible that a larger percentage of the CHB population would benefit from initiation of NA treatment, as these agents are efficacious, cheap, and well tolerated with minimal side effects (8). It is anticipated that the WHO will soon release

updated CHB guidelines with less stringent criteria for treatment initiation, and this change has already been made in China (238,239). Therefore, data from secondary care EHRs are invaluable to assess which patients meet existing treatment criteria, the proportion of these who are receiving NA agents, and to determine the impact of changing guidelines on the population eligible for treatment.

The COVID-19 pandemic impacted routine clinical services across a range of specialities/departments, both within and outside the UK (240), with quantified impacts reported for cancer (241–243) and cardiovascular disease (244). This is likely also true for CHB management and surveillance, however fewer data sources are available to investigate this impact due to the clinical care pathway of CHB. Secondary care EHRs are an optimal data source in which to investigate the impact of the pandemic on routine clinical practice and care in CHB.

4.2.2. NIHR HIC secondary/tertiary care EHR database

In 2014, the NIHR HIC was established in response to Chief Medical Officer Dame Sally Davies' challenge to NHS sites across the UK to harness the power of routinely collected EHR data for translational research (68). Individual sites, supported by their respective university partners via NIHR Biomedical Research Centres (BRCs), formed a unified network to make multicentre individual-level anonymised EHR available for collaborative research purposes. The Viral Hepatitis theme, managed by Professor Eleanor Barnes and Professor Philippa Matthews, is a framework for collection of viral hepatitis data from secondary care centres in England (68,71).

The NIHR Viral Hepatitis HIC database contains such data from individuals with CHB accessing secondary care health services. Data are collected longitudinally, with lengths of follow-up varying both by individual and by respective NHS site (due to variable clinical review and data collection processes). Parameters include laboratory tests, demographics, information regarding antiviral treatment, elastography scores, imaging, liver biopsy reports and potential risk factors for liver disease. Data elements are defined according to the standardised NHS Data Dictionary (245). Given the

complexity and magnitude of available information, analysis of HIC data requires large amounts of wrangling and processing for analysis.

Data storage and entry processes are unique to individual NHS sites, with different sites using a range of EHR interfaces (including Cerner Millennium and Epic). Therefore, the Viral Hepatitis theme provides an informatics infrastructure via which data from all sites may be harmonised and hosted on a central server. Oxford University Hospitals NHS Foundation Trust hosts data for the Viral Hepatitis theme, and researchers access data remotely via a Trusted Research Environment.

Viral Hepatitis HIC data are currently collected from ten participating NHS sites, which have geographical coverage across England Data from 6 centres contributing antiviral treatment information to the HIC were included in analyses for this chapter (**Figure 4.1**). A comprehensive data governance framework has been developed to oversee data collection across the sites which use heterogenous EHR processes and environments. The NIHR HIC Viral Hepatitis theme database was approved by South Central—Oxford C Research Ethics Committee (REF Number: 21/SC/0060). Individuals are able to opt out from having their data included in the HIC via the National Data Opt-out (246). The requirement for written informed consent was waived by South Central—Oxford C Research Ethics Committee, because data have been anonymized before transmission to the theme central data repository.

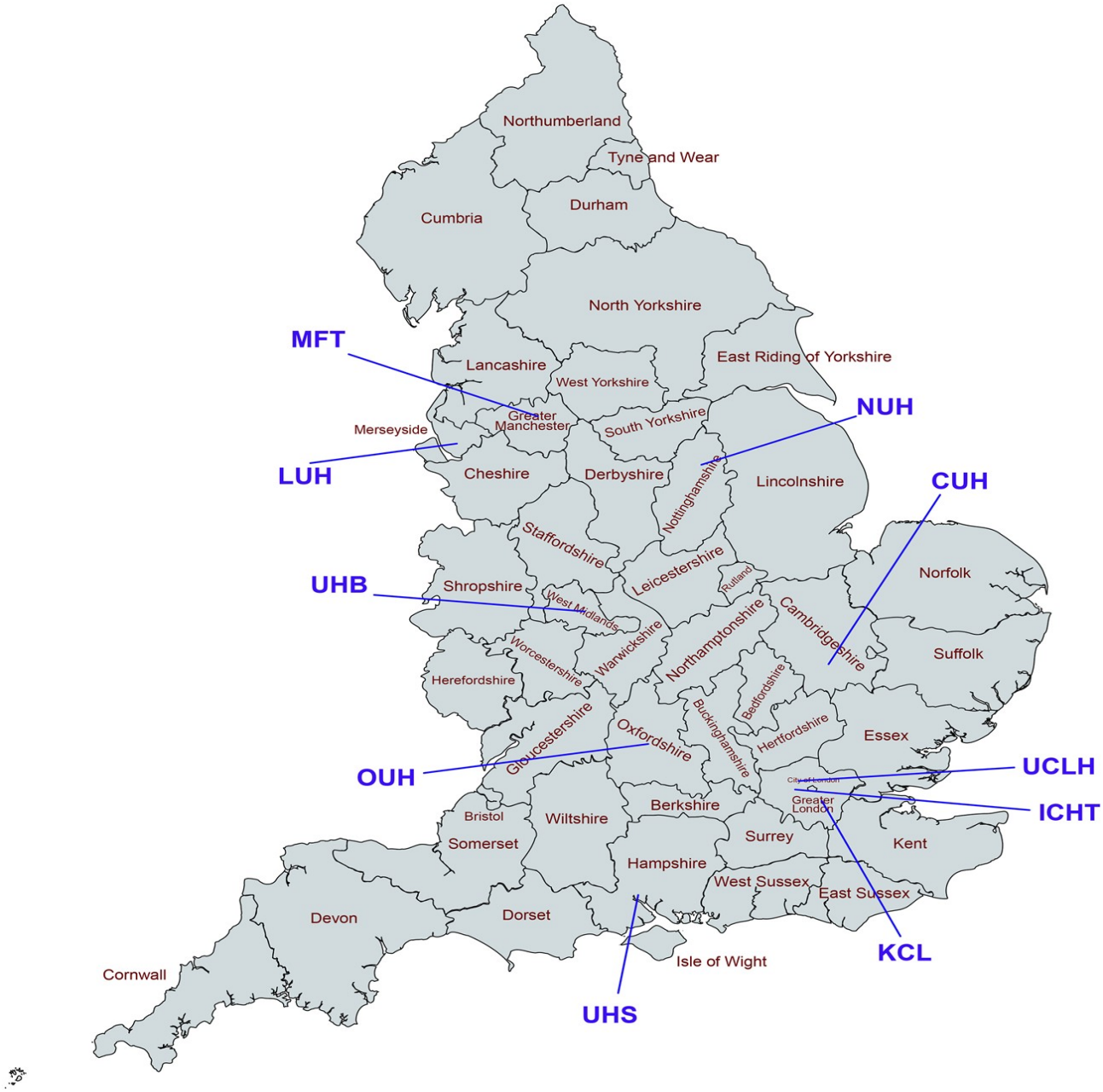


Figure 4.1. Locations of the 10 National Health Service (NHS) Trusts participating in the National Institute for Health Research (NIHR) Health Informatics Collaborative (HIC) viral hepatitis theme up to February 2022.
CUH, Cambridge University Hospitals NHS Foundation Trust; ICHT, Imperial College Healthcare NHS Trust; KCL, King's College Hospital NHS Foundation Trust; LUH, Liverpool University Hospitals NHS Foundation Trust; MFT, Manchester University NHS Foundation Trust; NUH, Nottingham University Hospitals NHS Trust; OUH, Oxford University Hospitals NHS Foundation Trust; UCLH, University College London Hospitals NHS Foundation Trust; UHB, University Hospitals Birmingham NHS Foundation Trust; UHS, University Hospital Southampton NHS Foundation Trust. CUH, ICHT, KCL, OUH and UCLH were the five NHS Trusts initially included in the NIHR HIC viral hepatitis theme, with LUH, MFT, NUH, UHB and UHS joining more recently. NB this figure has been reproduced from *Wang et al 2022* (71).

4.2.3. Aims and objectives

Multiple evidence gaps exist concerning CHB treatment eligibility and coverage in the UK, whereby the percentage of CHB individuals receiving antiviral treatment requires quantification, and factors associated with treatment initiation warrant exploration. Additionally, the impact of the COVID-19 pandemic on routine CHB care has yet to be robustly quantified. Therefore I set out to investigate two aims in the NIHR HIC: (i) to quantify the coverage of antiviral treatment in a cohort of CHB individuals from a secondary care data source, and (ii) to quantify the impact of the COVID-19 pandemic on routine CHB surveillance.

4.3. Research questions

4.3.1. Research question 1: Antiviral treatment eligibility and coverage across individuals with CHB in the NIHR HIC secondary EHR database; who is appropriately receiving treatment, who is not receiving treatment, and how would modified criteria impact treatment coverage?

4.3.2. Research question 2: What was the impact of the COVID-19 pandemic on routine CHB surveillance in secondary care settings?

4.4. Methods

4.4.1. Methods – Research question 1

Study design and follow-up

I retrospectively identified individuals with CHB in the NIHR HIC database, who were aged ≥18 years at time of earliest CHB record, and who originated from NHS sites contributing antiviral treatment data to the HIC. CHB was defined as the presence of a positive HBsAg or VL measurement on at least two recordings ≥6 months apart. This is concordant with the definition used to identify CHB in the QResearch primary care EHR database.

Participants contributed follow up time to the study from their earliest CHB record to either the initiation of antiviral treatment (outcome of interest) or censoring, whichever occurred first (**Figure 4.2**). The cohort is dynamic, whereby the earliest record of CHB varies by individual, and the cohort baseline is therefore specific to patient-level diagnosis date. A patient with an early CHB diagnosis date (relative to other individuals) might have initiated antiviral therapy sometime after CHB diagnosis (Scenario A), after which they can be said to have "experienced" the outcome of interest and no longer contribute follow-up time to the analysis. Alternatively in Scenario B, a patient might have a relatively late index date, and also exit the cohort before the cohort end date (which is the last possible date at which individual NHS sites could upload data to the HIC). Patients who don't develop the outcome of interest will be censored at the end date for the analysis, not having experienced the outcome of interest; this is visualised in Scenario C.

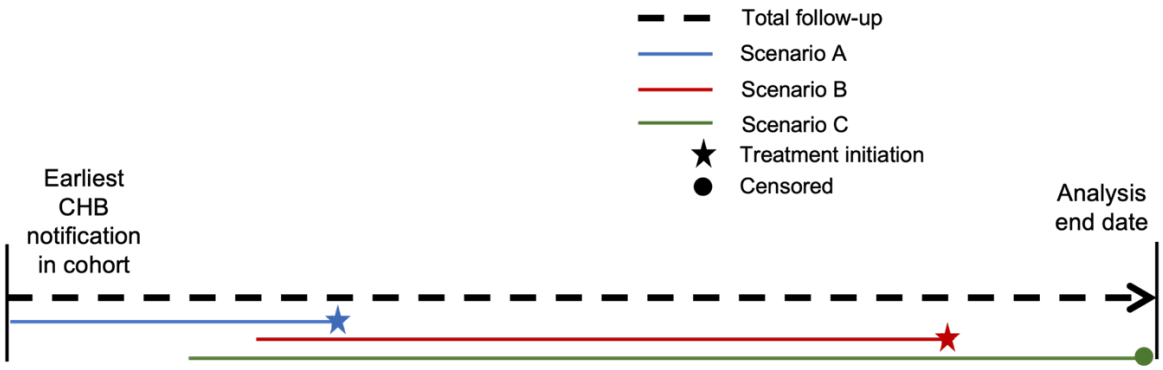


Figure 4.2. Timeline visualising different follow-up scenarios A, B and C for the analysis.
In Scenario A (blue), the patient initiates antiviral therapy sometime after CHB diagnosis, after which they have experienced the outcome of interest (treatment initiation) and no longer contribute follow-up time. In Scenario B (red) the individual enters the cohort relatively late as compared to other individuals and contributes follow up time until they experience the outcome of interest (treatment initiation). In Scenario C (green), the patient does not experience the outcome of interest and is thereby not censored until the analysis end date.

Variable and outcome ascertainment

Clinical, demographic and laboratory parameters were available for all individuals across across six NHS Trusts which contributed antiviral treatment data in the HIC at time of analysis. Quantified HBsAg was available in a subset of 5 NHS Trusts, and imaging data in a subset of 3. The primary outcome of interest in this study was time

to initiation of antiviral treatment. Due to the structure of prescribing data records in secondary care electronic systems, some patients report multiple various “start” dates for treatment due to the renewal of prescription and/or switching from one NA to another. Therefore, earliest recorded treatment start date was used. Once a patient had initiated treatment, they were assumed to continue treatment indefinitely, as recommended by NICE guidelines (8) and because it was not possible to infer treatment end date from available data. To investigate treatment eligibility across untreated individuals, proportions of individuals who demonstrated abnormal ALT measurement (≥30 IU/L in males and ≥19 IU/L in females) or VL levels indicating NA treatment (> 2000 IU/ml) according to NICE guidelines (8) were quantified. To investigate how many additional patients would be eligible for treatment under relaxed CHB treatment criteria which have been adopted in China (which recommend NA therapy in CHB individuals with abnormal ALT and detectable VL), the number of untreated patients meeting these criteria was reported (238,239).

Statistical analysis

Statistical analyses were carried out in R (version 4.1.0). Baseline characteristics were summarised for all CHB patients using descriptive statistics. Means and standard deviations (SDs) or medians and interquartile ranges (IQRs) were presented for continuous measures and were compared using *t* or Wilcoxon rank-sum tests, respectively. Patient counts and percentages were presented for continuous and binary variables; percentages were compared using chi-squared or Fisher’s exact tests.

Univariable and multivariable Cox proportional hazards models were used to investigate factors associated with antiviral treatment initiation. Variables were included in the multivariable model based on significance of univariable associations (where $P \leq 0.1$) and/or based on clinical guidelines (8,237). The main Cox model has been time-updated to account for longitudinal changes in ALT, HBeAg, HBsAg and HBV VL. Hazard ratios and 95% confidence intervals (95% CI) were reported for model outputs. The proportional hazards assumption was assessed using visual assessment of Kaplan-Meier plots to ensure survival curves remained relatively parallel over time across strata. Due to the absence of imaging data from half of the

NHS Trusts, associations of imaging history/type with treatment initiation were not investigated.

Sensitivity analyses

To test robustness of my main model, sensitivity analyses were performed (**Table 4.1**), whereby patients experiencing the outcome of interest within the first year of follow up were excluded to control for reverse causality bias. This was performed to establish temporality and scrutinise how longitudinal changes in variables associated with treatment initiation. Individuals who are clinically compromised may present acutely in a secondary care setting whereby CHB diagnosis and treatment initiation may occur within a relatively short time span of one another. Exclusion of such clinical scenarios from analysis enables an investigation which represents a more typical CHB clinical care pathway.

Table 4.1. Details of main and sensitivity analysis performed to investigate factors associated with antiviral treatment imitation for individuals with chronic hepatitis B virus infection in the NIHR Health Informatics Collaborative database.

Analysis	Main or sensitivity	Variables included	Corresponding table(s)/figure
Cox proportional hazards model for which model has been time-updated to account for longitudinal changes in ALT, HBeAg, HBsAg and HBV VL	Main	Abnormal ALT, Age, Ethnicity, HCV coinfection, HIV coinfection, Positive HBeAg Positive HBV VL, Sex,	Figure 3, Table 4
Cox proportional hazards model on cohort subset whereby participants experiencing the outcome of interest within the first year of follow up are excluded	Sensitivity	Abnormal ALT, Age, Ethnicity, HCV coinfection, HIV coinfection, Positive HBeAg	Table 5

		Positive HBV VL, Sex,	

ALT, alanine transaminase; HCV, hepatitis C virus; CT, computerised tomography; HIV, human immunodeficiency virus; HBeAg, Hepatitis B virus E antigen; HBsAg, Hepatitis B virus surface antigen; HBV VL, hepatitis B virus DNA viral load.

4.4.2. Methods – Research question 2

I undertook longitudinal and cross-sectional analyses using individual-level secondary care data collected across six NHS Trusts which contributed antiviral treatment data in the HIC at time of analysis. We investigated how ALT and VL surveillance varied between pre-COVID (years 2016-2019) to COVID (years 2020-2021), comparing surveillance metrics in patients on and off antiviral therapy.

We calculated cumulative probabilities of undergoing ≥1 ALT and HBV VL measurement each year (as a minimum standard for routine monitoring), using the Kaplan-Meier method, comparing probabilities across years using the log-rank test. Where median time to 50% of the cohort achieving ≥1 ALT or HBV VL measurement could not be computed, due to <50% of participants achieving this for one or more calendar years, Cox proportional hazards modelling was used to investigate how hazards of ≥1 ALT or HBV VL measurement varied across calendar years. We quantified the mean number of ALT and HBV VL measurements per 100 patients, during pre-COVID and COVID years, with 95% CIs calculated using the normal distribution (whereby standard error (SE) is estimated by $\frac{s}{\sqrt{n}}$ with n denoting sample size and s denoting standard deviation).

4.5. Results

4.5.1. Research question 1

Cohort characteristics and imaging

In total, 4771 individuals were eligible for inclusion in analysis. Mean follow-up was 5.4 years (SD 3.9) years (**Table 4.2**). A minority (n = 1305, 27.4%) of patients initiated treatment throughout follow-up. Nearly half (n = 2005, 42.0%) of individuals were of White ethnicity, as compared to 11.8% (n = 562) Asian, 18.9% (n = 902) Black, 2.8% (n = 132) mixed and 24.5% of other ethnicities (n = 1170). A small percentage of individuals had HIV (1.5%, n = 70) or HCV (1.0%, n = 48) coinfection.

In general, patients who initiated treatment were older, with a larger percentage of males compared to individuals who remained untreated (**Table 4.2**). Ethnicity breakdowns varied between treated and untreated individuals, whereby smaller percentages of White and Black individuals initiated treatment, compared to larger percentages of Asian and mixed ethnicity individuals ($P < 0.001$). Although minorities of patients in both groups had evidence of HIV or HCV coinfection, proportions of co-infected individuals were larger in those who initiated antiviral treatment (as expected, as HIV treatment overlaps with HBV treatment).

Table 4.2. Baseline characteristics of individuals included in this NIHR HIC analysis investigating factors associated with antiviral treatment initiation in chronic HBV infection.
Characteristics are presented overall for the total study sample (n = 4771) and stratified by those who initiated antiviral treatment (n = 1305) and did not initiate treatment (n = 3466) throughout follow-up. Percentages are expressed out of the column total.

Characteristic	Total, n (%)	Untreated, n (%)	Treated, n (%)
N	4771	3466	1305
Follow-up time (years), median (IQR)	5.40 (3.9)	6.12 (3.73)	3.46 (3.58)
Age group, n (%)			
<25 years	385 (8.1)	301 (8.7)	84 (6.4)

Characteristic	Total, n (%)	Untreated, n (%)	Treated, n (%)
25-34 years	1545 (32.4)	1201 (34.7)	344 (26.4)
35-44 years	1387 (29.1)	1032 (29.8)	355 (27.2)
45-54 years	821 (17.2)	551 (15.9)	270 (20.7)
55-64 years	436 (9.1)	258 (7.4)	178 (13.6)
≥65 years	196 (4.1)	122 (3.5)	74 (5.7)
Sex, n (%)			
Female	2218 (46.5)	1713 (49.4)	505 (38.7)
Male	2553 (53.5)	1753 (50.6)	800 (61.3)
Ethnicity, n (%)			
White	2005 (42.0)	1520 (43.9)	485 (37.2)
Asian	562 (11.8)	402 (11.6)	160 (12.3)
Black	902 (18.9)	696 (20.1)	206 (15.8)
Mixed	132 (2.8)	92 (2.7)	40 (3.1)
Other	1170 (24.5)	756 (21.8)	414 (31.7)
HIV coinfection, n (%)	70 (1.5)	28 (0.8)	42 (3.2)
HCV coinfection, n (%)	48 (1.0)	28 (0.8)	20 (1.5)
Abnormal ALT†	3139 (65.8)	2228 (64.3)	912 (69.9)
HBV VL > 2000 IU/mL	424 (8.9)	336 (9.7)	88 (6.8)
Treatment eligible under relaxed criteria‡	239 (5)	216 (6.8)	1305 (100)

IQR, interquartile range; HIV, human immunodeficiency virus; HCV, hepatitis C virus; HBeAg, hepatitis B virus E antigen; HBsAg, hepatitis B virus surface antigen; HBV VL, hepatitis B virus viral load; ALT, alanine aminotransferase; N/A, not applicable.

† ≥30 IU/L in males and ≥19 IU/L in females

‡ Relaxed treatment criteria whereby individuals with chronic HBV infection are treatment eligible if they have record of both a positive viral load measurement and an abnormal ALT measurement.

A smaller percentage of untreated individuals had an abnormal ALT reported (n = 2228, 64.3%) as compared to individuals who initiated treatment throughout follow-up (n = 912, 69.9%) ($P < 0.01$) (**Table 4.2**).

Influence of changing guidelines on proportion of population reaching treatment eligibility criteria

Amongst individuals who did not initiate treatment, 9,7% (n = 336) had at least one VL measurement > 2000 IU/ml, which is an indicator for NA therapy in CHB individuals according to NICE guidelines. Under relaxed treatment initiation criteria for ALT and VL, which have been adopted in China, a further 216 individuals (representing 6.8% of those untreated) would have been eligible to initiate antiviral treatment.

Across the 3 NHS Trusts contributing imaging data, the most common form of imaging undertaken was ultrasound (**Table 4.3**). A larger percentage of patients who initiated treatment underwent ultrasound throughout follow-up (77.7%, n = 327) as compared to those who did not initiate treatment (62.8%, n = 835). Based on the available data, elastography scores did not differ between patients who did and did not initiate treatment (however, there is a high level of missingness as elastography scores are not robustly captured by EHR systems).

Table 4.3. Types of hepatic imaging performed in individuals included in this NIHR HIC analysis investigating factors associated with antiviral treatment initiation in chronic HBV infection.
Data are presented from 3 NHS trusts contributing imaging data both overall (n = 1329) and stratified by those who initiated antiviral treatment (n = 421) and did not initiate treatment (908) throughout follow-up. Percentages are expressed out of the column total.

Characteristic	Total, n (%)	Untreated, n (%)	Treated, n (%)
N	1329	908	421
Elastography, n (%)	60 (4.5)	38 (4.2)	22 (5.2)
Liver stiffness score, mean (SD)	5.99 (3.06)	5.92 (3.39)	6.11 (2.44)

Characteristic	Total, n (%)	Untreated, n (%)	Treated, n (%)
MRI, n (%)	55 (4.1)	22 (4.2)	33 (7.8)
CT, n (%)	75 (5.6)	39 (4.3)	36 (8.6)
Ultrasound, n (%)	835 (62.8)	508 (5)	327 (77.7)

SD, standard deviation; MRI, magnetic resonance imaging; CT, computed tomography.

Factors associated with initiation of antiviral treatment

Mean time from patient baseline to treatment initiation in 1305 patients was 3.5 years (SD 3.6). Results from univariable investigations of how CHB laboratory measurements, demographics and record of hepatic imaging associated with antiviral treatment initiation are reported in **Table 4.4**. Univariably, HBeAg positivity, HBV VL positivity and abnormal ALT were associated with treatment initiation, as expected in line with NICE clinical guidance. Increasing age, male sex, HIV coinfection and record of CT and ultrasound imaging were also associated with antiviral treatment initiation.

Table 4.4. Cox proportional hazards univariable and main multivariable model investigating factors which are associated with antiviral treatment initiation in adults with CHB.
The main multivariable model has been time-updated to account for longitudinal changes in ALT, HBeAg and HBV VL.

Characteristic	Univariable HR (95% CI)	Multivariable HR (95% CI)
Age group, n (%)		
<25 years	0.90 (0.69 to 1.18)	0.77 (0.52 to 1.13)
25-34 years	1.00 (ref)	1.00 (ref)
35-44 years	1.19 (1/01 to 1.39)*	1.04 (0.77 to 1.42)
45-54 years	1.55 (1.30 to 1.84)***	1.48 (1.05 to 2.10)*
55-64 years	2.11 (1.74 to 2.56)***	2.08 (1.43 to 3.03)***
≥65 years	2.32 (1.78 to 3.03)***	1.62 (1.03 to 2.53)*
Sex, n (%)		
Female	1.00 (ref)	1.00 (ref)

Characteristic	Univariable HR (95% CI)	Multivariable HR (95% CI)
Male	1.51 (1.33 to 1.69)***	1.37 (1.06 to 1.77)*
Ethnicity, n (%)		
White	1.00 (ref)	1.00 (ref)
Asian	1.06 (0.88 to 1.28)	1.51 (1.10 to 2.06)*
Black	0.84 (0.71 to 0.99)*	1.30 (0.91 to 1.84)
Mixed	0.99 (69 to 1.41)	1.27 (0.83 to 1.93)
Other	1.43 (1.23 to 1.65)***	1.31 (1.03 to 1.66)*
HIV coinfection, n (%)	2.13 (1.53 to 2.97)***	1.58 (1.01 to 2.46)*
HCV coinfection, n (%)	1.46 (0.90 to 2.36)	--
Serum ALT concentration	1.03 (1.02 to 1.03)***	1.01 (1.00 to 1.02)***
Abnormal ALT†	1.27 (1.11 to 1.44)***	--
HBV VL positive	2.53 (2.21 to 2.90)***	1.60 (1.29 to 1.99)***
HBeAg positive	2.36 (2.07 to 2.71)***	2.28 (1.92 to 2.72)***

HIV, human immunodeficiency virus; HCV, hepatitis C virus; CT, computed tomography; ALT, alanine aminotransferase; HBV VL, hepatitis B virus viral load; HBeAg, hepatitis B virus E antigen.
* $P < 0.05$
** $P < 0.01$
** $P < 0.001$
† ≥30 IU/L in males and ≥19 IU/L in females

In the multivariable model, (**Table 4.4, Figure 4.3**), ethnicity remained associated with hazards of antiviral treatment initiation after adjustment for age, sex and treatment indicators outlined in clinical guidelines (including HBV VL, HBeAg and abnormal ALT). Specifically, individuals of Asian and Other ethnicities demonstrated 51% (aHR 1.51, 95% CI 1.10 to 2.06) and 31% (aHR 1.31, 95% CI 1.03 to 1.66) higher hazards of treatment initiation as compared to White individuals. An association of HIV coinfection with antiviral treatment initiation was also observed in the analysis, due to the use of common NA agents in both CHB and HIV treatment. I note that sociodemographic characteristics, which would have been relevant confounders requiring consideration, were not available in this data source.

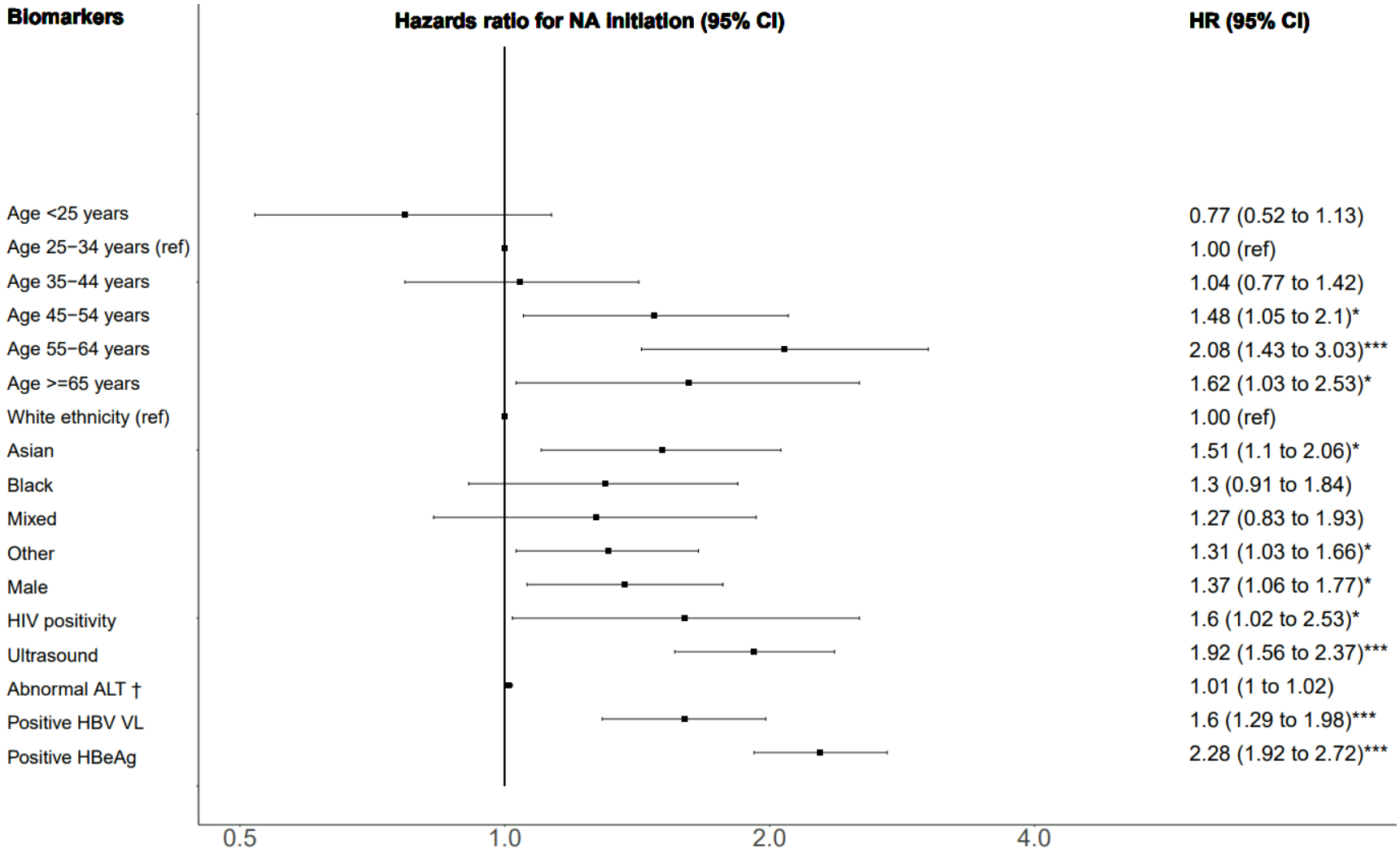


Figure 4.3. Forest plot for main multivariable Cox proportional hazards model from analysis investigating factors associated with antiviral treatment initiation in chronic Hepatitis B virus infected individuals (n = 4771) undertaken in the HIC database.
ALT, alanine aminotransferase; HBV VL, hepatitis B virus viral load; HBeAg, hepatitis B virus E antigen.

Sensitivity analyses

Sensitivity analysis was performed to test the robustness of the main model (**Table 4.5**). Individuals initiating treatment within the first year of follow-up were excluded. No hazards ratios changed in direction, or materially in magnitude.

Table 4.5. Results from sensitivity analysis undertaken to test robustness of the main model, with patients experiencing the outcome of interest within the first year of follow up excluded. This was undertaken to control for reverse causality bias.

Characteristic	Multivariable HR sensitivity analysis I (95% CI) †
Age group, n (%)	
<25 years	0.83 (0.60 to 1.14)
25-34 years	1.00 (ref)
35-44 years	1.26 (1.03 to 1.54)*
45-54 years	1.65 (1.34 to 2.05)***
55-64 years	2.06 (1.61 to 2.64)***
≥65 years	2.43 (1.71 to 3.45)***
Sex, n (%)	
Female	1.00 (ref)
Male	1.12 (0.95 to 1.32)***
Ethnicity, n (%)	
White	1.00 (ref)
Asian	1.34 (1.08 to 1.66)*
Black	0.95 (0.76 to 1.18)
Mixed	1.23 (0.82 to 1.85)
Other	1.33 (1.11 to 1.60)***
HIV coinfection, n (%)	1.06 (0.66 to 1.72)
Serum ALT concentration	1.02 (1.01 to 1.03)***
HBV VL positive	2.37 (1.99 to 2.84)***
HBeAg positive	1.84 (1.53 to 2.20)***

HIV, human immunodeficiency virus; HCV, hepatitis C virus; CT, computed tomography; ALT, alanine aminotransferase; HBV VL, hepatitis B virus viral load; HBeAg, hepatitis B virus E antigen; HBsAg, hepatitis B virus surface antigen.
* $P < 0.05$
** $P < 0.01$
** $P < 0.001$
† Patients initiating treatment within the first year of follow up excluded.
‡ Main results model fit to subset of patients with HBsAg tested (n = 3425, amongst whom 824 individuals initiated treatment throughout follow up).

4.5.2. Research question 2

Coverage of ALT and VL testing (proportion tested at least once a year)

Data were available from 3768 patients for the year 2016; this increased to 5450 by the year 2021. In 2016, 74.1% (95% CI 72.7-75.5%) of individuals with CHB had ALT measured on ≥1 occasion throughout the year. This significantly decreased to 56.0% (95% CI 54.6-57.3%) in 2020 and 47.7% (95% CI 46.4-49.0%) in 2021, suggesting a negative impact of the COVID-19 pandemic on routine clinical? surveillance (**Figure 4.4A**). Similarly, throughout 2016, 68.4% (95% CI 66.9-69.9%) of patients had VL measured at least once, whilst this significantly decreased to 51.6% (95% CI 50.3-53.0%) in 2020 and 44.7% (95% CI 43.4-46.0%) in 2021 (**Figure 4.4B**). From 2016 to 2019, median time to ALT/VL surveillance ranged between 138 and 175 days (out of 365 total in one calendar year). Median time to surveillance significantly differed in 2020, with durations of 308 and 345 days in 2020 for ALT and VL, respectively (Log-rank $P < 0.0001$). In 2021, 47.7% and 44.7% of individuals achieved ALT and VL surveillance by the end of the calendar year, respectively.

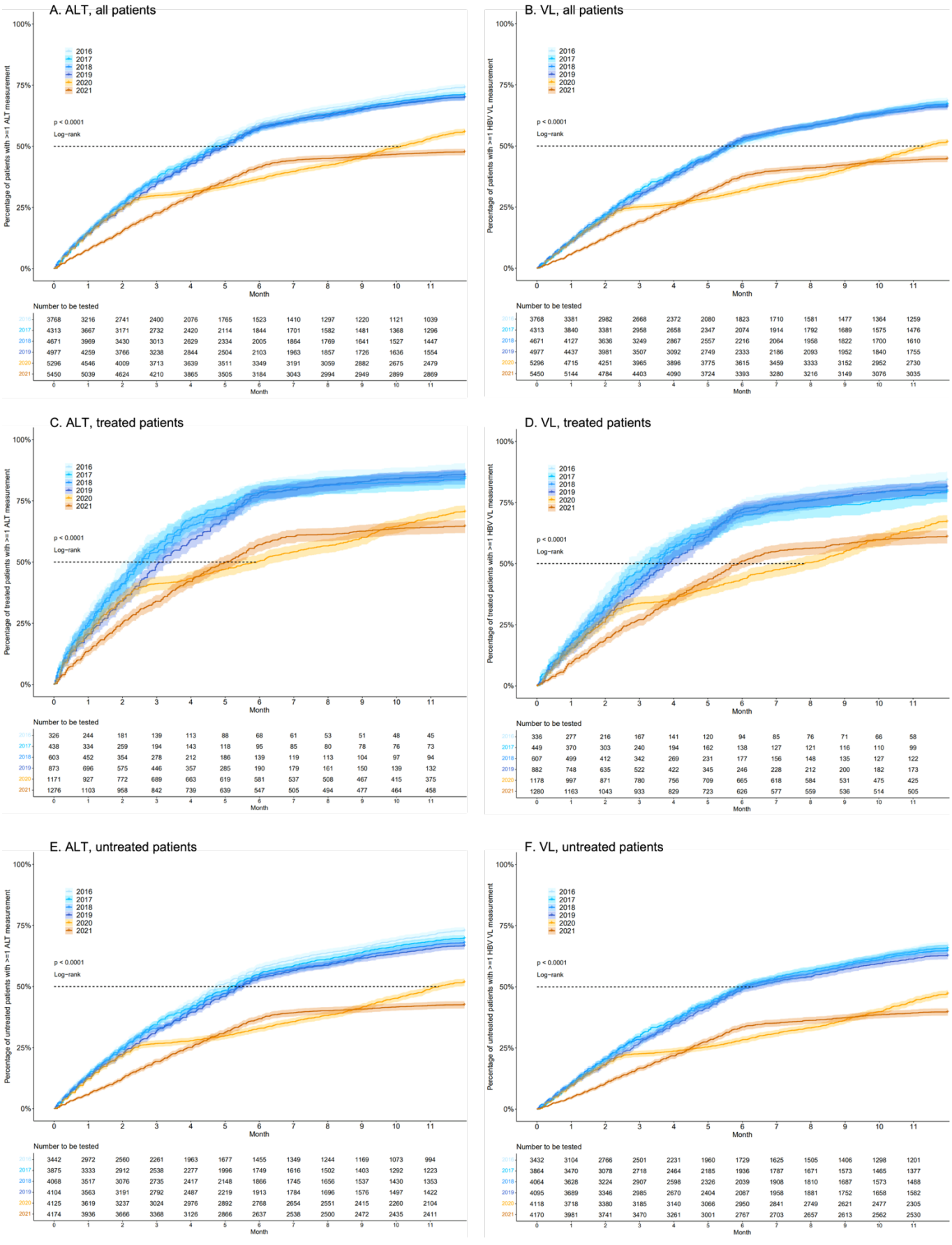


Figure 4.4. Kaplan-Meier plots demonstrating the cumulative proportion of adults with chronic Hepatitis B virus (HBV) infection undergoing routine laboratory surveillance.
Plots show patients who have had ≥1 alanine transferase (ALT) (panels A, C, E) and ≥1 HBV DNA viral load (VL) (panels B, D, F) measurement each year for pre-COVID-19 (2016–19) and COVID-19 (2020 and part of 2021) years. For both ALT and VL, plots are stratified by treated (panels C and D, respectively) and untreated (panels E and F, respectively)

patients. Dashed lines depict median time for 50% of the cohort to have the laboratory assessment undertaken. Cumulative probabilities for each year were calculated using the Kaplan-Meier method, comparing probabilities across years using the log-rank test, with 2016 serving as the reference/baseline year.

Coverage of ALT and VL testing stratified by treatment status

A subset of patients had treatment data available for years 2016 to 2021. Among treated patients, the proportion with ≥1 ALT measurement decreased significantly from 87.4% (95% CI 83.8-91.0%) in 2016 to 70.6% (95% CI 68.0-73.2%) in 2020, and 64.6% (95% CI 62.0-67.2%) in 2021 (**Figure 4.4C**). Untreated patients demonstrated a reduction from 72.9% (95% CI 71.4-74.4%) in 2016 to 51.8% (95% CI 50.3-53.3%) in 2020 and 42.5% (95% CI 41.0-44.0%) in 2021 (**Figure 4.4E**). In treated patients, median time to ALT surveillance ranged between 74 and 97 days from 2016-2019, and increased to 180 and 153 days in 2020 and 2021, respectively (Log-rank $P < 0.0001$). Differences in median time to surveillance were also observed in untreated patients, with median times to ALT surveillance ranging from 145 to 167 days during pre-COVID years and reaching 342 days in 2020 and with <50% individuals undergoing surveillance in 2021.

The proportion of treated patients with VL measured decreased from 84.2% (95% CI 80.3-88.1%) in 2016 to 67.2% (95% CI 64.6-69.9%) in 2020 and 61.0% (95% CI 58.3-63.7%) in 2021 (**Figure 4.4D**). A decrease was also observed in untreated patients from 66.8% (95% CI 65.2-68.4%) in 2016 to 47.2% (95% CI 45.7-48.7%) in 2020 and 39.7% (95% CI 38.3-41.2%) in 2021 (**Figure 4.4F**). In both treated and untreated individuals, cumulative probabilities comparing time to ALT and HBV VL measurement significantly differed across calendar years (Log-rank $P < 0.0001$). Median time to VL surveillance in treated patients ranged between 91 and 115 days from 2016-2019, increased to 238 and 153 days in 2020 and accordingly reduced to 180 days in 2021 (Log-rank $P < 0.0001$). In untreated patients, median time to VL surveillance ranged from 183 to 192 days during pre-COVID years, whilst in both 2020 and 2021 <50% of individuals underwent VL surveillance (Log-rank $P < 0.0001$).

Frequency of ALT and VL testing

Pre-COVID, the mean number of ALT measurements varied from 19 (95% CI 19-19) to 24 (95% CI 24-24) per 100 patients per month, with annual dips in August and December suggesting an impact of holiday periods when clinic activity may be reduced

(**Figure 4.5A**). Between March 2020 and December 2021, this range reduced to between 4 (95% CI 4-4) and 15 (95% CI 15-15), mean number of measurements per 100 patients per month remaining lower in these years as compared to pre-COVID years. The same trend was observed for VL (**Figure 4.5B**), with mean numbers of measurements ranging from 9 (95% CI 9-9) to 15 (95% CI 15-15) per 100 patients pre-COVID, and from 3 (95% CI 3-3) to 11 (95% CI 11-12) during the COVID period, with nadirs of 3 measurements per 100 patients (95% CI 3-3) in April 2020 and December 2021.

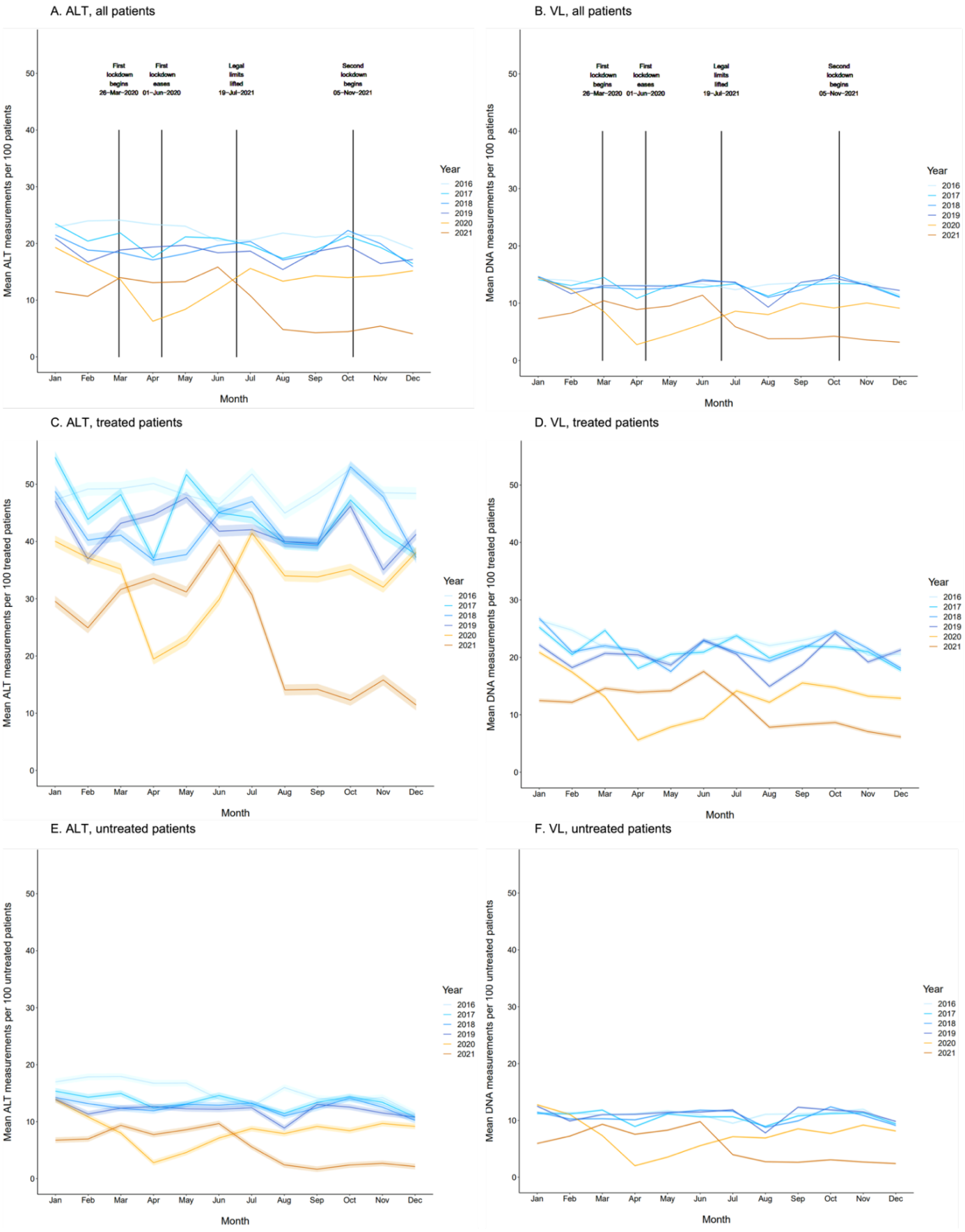


Figure 4.5. Mean numbers of ALT and HBV DNA VL measurements per 100 patients per month during pre-COVID (2016–19) and COVID-19 (2020-2021) periods.
Data are shown overall and stratified by treatment status for both viral load (VL, panels A, C, E) and alanine transferase (ALT, panels B, D, F). Dates of UK national COVID-19 lockdowns are denoted in overall plots. 95% CIs were calculated using the normal distribution (whereby standard error (SE) is estimated by $s\sqrt{n}$ with n denoting sample size and s denoting standard deviation).

Frequency of testing by HBV treatment status

VL and ALT measurements were more frequent in treated (**Figure 4.5 C, D**) compared to untreated (**Figure 4.5 E, F**) patients throughout the study period. In treated patients pre-COVID, mean number of ALT measurements fluctuated between and within years but remained >35 measurements per 100 patients per month (**Figure 4.5C**); hence, treated patients had ALT measured on average once every 4 months (in keeping with the routine interval for clinic visits in patients on treatment).

Mean numbers of ALT monthly tests fluctuated between 1 (95% CI 10-13) and 42 (95% CI 40-43) per 100 patients during the COVID-19 period but were generally significantly lower compared to pre-COVID years. This pattern was also observed for VL, with mean numbers of measurements ranging from 15 (95% CI 15-15) to 27 (95% CI 26-27) during pre-COVID years and 6 (95% CI 5-6) to 18 (95% CI 17-18) during the COVID period (**Figure 4.5D**).

In untreated patients pre-COVID, mean number of ALT measurements per 100 patients per month ranged from 64 (95% CI 64-65) to 99 (95% CI 98-99), with a drop in the COVID period to between 23 (95% CI 22-23) and 74 (95% CI 73-74) (**Figure 4.5E**). This pattern was also observed for VL, with pre-COVID and COVID measurements ranging from 8 (95% CI 8-8) to 12 (95% CI 12-13) and from 2 (95% CI 2-2) to 13 (95% CI 13-13) measurements per 100 patients, respectively (**Figure 4.5F**).

4.6 Discussion

Summary of main findings

In the treatment analysis of individuals with CHB attending specialist outpatient hepatology/infectious disease services across 6 NHS Trusts, I quantified the percentage of individuals initiating antiviral treatment throughout follow-up, and compared demographic and clinical characteristics between those who did and did not initiate treatment. Increased eligibility for antiviral treatment was observed in older males of Asian and Other ethnicities, and in those with HIV coinfection. The

association between treatment eligibility and ethnicity remained after adjustment for clinical indicators of treatment initiation, including older age, male sex, abnormal ALT, HBV VL positivity and HBeAg positivity (8,237). This observation was robust to time-updating of the Cox model to account for longitudinal changes in ALT, HBV VL and HBeAg status throughout follow-up. It was also robust to sensitivity analyses designed to establish temporality and restrict the multivariable model further for an additional HBV biomarker, quantified HBsAg.

In analysis investigating the impact of the COVID-19 pandemic on routine CHB surveillance, I demonstrated how clinical service disruptions led to significantly reduced proportions of CHB individuals having important HBV biomarkers measured. These impacts were quantified for CHB individuals overall, and also stratified by antiviral treatment status. The implications of the pandemic were also observed across both 2020 and 2021, and rates of routine clinical surveillance had not recovered to pre-COVID levels by the end of 2021. UK governmental investigations by PHE (247) and UKHSA (248) have reported concordant findings, with reductions in testing, diagnosis and treatment initiation for HBV (and BBVs generally) observed throughout the pandemic. Ongoing scrutiny will be applied to data collected during 2022, when these become available. Sharing such results with clinical services will allow them to review their practice and implement additional catch-up surveillance if warranted as face-to-face services are reinstated.

Antiviral treatment in CHB – coverage across the UK CHB population attending secondary/tertiary care services

In this analysis of individuals with CHB accessing specialist secondary/tertiary care services across 6 NHS Trusts, a minority of patients initiated antiviral treatment throughout longitudinal follow-up. This reflects current recommendations in both British/European (8,237) and global (249) treatment guidelines, according to which a minority (12-24%) of CHB individuals are eligible for antiviral treatment (8,237,249). However, strict treatment criteria may not maximise the potential benefit of antiviral treatment to individuals living with CHB, particularly treatment with NA agents which are cheap, well tolerated, and for which there are minimal side effects (8).

Recent studies have investigated the impacts of expanding treatment coverage in CHB, both short-term NA therapy administered in specific clinical scenarios (250) as well as expanding long-term treatment in populations where CHB is endemic (239). These investigations have demonstrated that expansion of treatment coverage would benefit CHB individuals in terms of clinical prognosis and overall reduction of risk of progression to end-stage disease. This benefit is reported despite potential concerns regarding NA adherence and importantly economic cost, as cost-effectiveness analysis has demonstrated that treatment expansion, specifically treating a larger percentage of CHB individuals and initiating treatment at an earlier stage in the disease pathway, is optimal (239). Importantly, expansion would enable more rapid progression towards global 2030 elimination targets (24,53). Public health benefits can also be gained through viraemic suppression which reduces the risk of transmission events.

I report an association of ethnicity with antiviral treatment initiation, whereby hazards of initiation were higher in those of Asian and Other ethnicities as compared to White ethnicity. This association is likely confounded by unmeasured characteristics, including both host and viral genetic factors. Viral genotypes may result in varying clinical presentations of CHB, with genotypes concentrating in different populations according to country of birth. There may be additional unmeasured parameters which associate with the likelihood of an individual having access to antiviral treatment. Specifically, I have previously described the CHB population in the UK to be disproportionately socioeconomically deprived, with a larger percentage of individuals of ethnic minorities experiencing deprivation. Therefore, it is possible that socioeconomic status associates with both ethnicity and access to secondary/tertiary care services in which NA agents are prescribed. Further characterisation of the social and economic characteristics of the UK CHB population, particularly the subset attending specialist care services, is warranted in order to further elucidate this potential association, and to investigate how clinical resources may be more effectively distributed.

Impact of the COVID-19 pandemic on CHB surveillance

My analysis demonstrates the negative impact of the COVID-19 pandemic on routine clinical surveillance for patients with CHB infection in the UK. Reduction in rates of

surveillance track periods of population lock-down and the introduction and relaxation of legislative restrictions (251,252). During lockdown periods, much clinical care was switched to telemedicine (253,254), intended to maintain contact between patients and healthcare providers, but at the detriment of laboratory monitoring. Although patients may have an option to attend phlebotomy appointments, services have been stretched, and clinicians and/or patients may have elected to defer routine blood tests. Some patients may have attended different services for blood tests closer to home, with results then not incorporated in the EHR. Telephone appointments are impractical for certain patients, and contact may have been lost altogether – for example, for those who do not communicate confidently in English. Furthermore, CHB patients who left the UK early in the course of the pandemic may have been unable to return due to travel restrictions, thereby foregoing routine surveillance.

The impact of the pandemic was seen across treated and untreated individuals. However, the impact on ALT measurement was greater in those on treatment during lockdowns. I was unable to investigate how surveillance disruptions might impact risk of CHB end-stage disease or complications, due to inadequate long-term follow-up and also the lack of definitive HCC identification in the current NIHR HIC Data source. As complications of CHB evolve over the long-term (months to years), it will become possible to quantify how service disruptions may have impacted incidence of HCC and other adverse endpoints. Surveillance disruption may be associated with an increased risk of CHB progression and delays in starting antiviral therapy. Further prospective data are required to determine the impact on long-term outcomes.

These data represent only six large, urban centres, in the South East of England; trends may differ across settings and additional data will be required to determine the extent to which our observations apply in other regions. HIC data collection is being expanded (71), providing enhanced opportunities to incorporate a wider view of the UK over time. Effort and resources are required to refine telemedicine services and optimise access to laboratory surveillance (255), to re-establish sufficient face-to-face services, and to catch-up on monitoring and interventions. Future analysis is warranted to investigate how imaging surveillance and telemedicine service uptake have been impacted when these data become available in the NIHR HIC database. Continued longitudinal follow-up is required to ascertain how COVID-attributable

disruptions will impact the incidence of HCC and other relevant endpoints, and to ensure activity is re-aligned to support progress to elimination goals.

Strengths and limitations

The NIHR HIC secondary care EHR database offers strengths which are unique to this data source, being one of the largest EHR databases in the UK, and the largest secondary care data source containing longitudinal laboratory parameters. Therefore, characterisation of large-scale longitudinal CHB data is possible. This provides statistical power for analysis and permits sensitivity analyses to maximise robustness.

The HIC is also comprised of real-world clinical data, with a sampling frame of individuals who are accessing 'regular' health services at multiple NHS Trusts across the country, thus representing routine healthcare provision for CHB individuals attending secondary care services and generating representative results, which can potentially be generalised to patients not included in the database. However, a proportion of CHB individuals in the UK are not attending routine secondary care services, and may have limited access to diagnosis/clinical care, and thereby HIC data may not be representative of the total UK CHB population. Ultimately the results provide a robust characterisation of current health service provision and may be used to inform clinical practice and future guidelines.

Unfortunately, I was unable to investigate how restricted antiviral treatment coverage and COVID-19-attributable health service disruptions associated with risk of disease endpoints in CHB, due to lack of robust characterisation of relevant outcomes, including HCC and cirrhosis, in the NIHR HIC. ICD-10 coding for endpoints is currently not available, and identification of such outcomes from imaging report free-texts using Natural Language Processing is not possible across all participating sites, although it has been successfully achieved using Oxford University Hospitals HIC data (256). Furthermore, prolonged longitudinal follow-up is required to order to investigate how the COVID-19 pandemic will impact risk of end-stage disease. Both of these requirements can be fulfilled with sustained maintenance and growth of the NIHR HIC dataset, which is an invaluable resource to viral hepatitis research.

Conclusions

The NIHR HIC secondary care EHR database includes patient-level data for CHB individuals accessing specialist services across the UK. Using this large-scale real-world data source, I have quantified treatment coverage across a CHB sample, demonstrating that a minority of individuals receive antiviral treatment, in line with current clinical guidance. I have also quantified the impact of the COVID-19 pandemic on routine CHB surveillance, reporting significant reductions in surveillance rates as compared to pre-COVID years, which have persisted to the end of 2021. I am passionate about continuing to use the NIHR HIC to conduct viral hepatitis investigations and produce robust, impactful findings which can benefit clinical practice and patient benefit/wellbeing.

5. UK HBV immunisation in the Clinical Practice Research Datalink primary care electronic health record database: changes in coverage in response to policy changes and the COVID-19 pandemic

5.1 Availability of content

The data presented in this chapter originate from the proprietary Clinical Practice Research Datalink (CPRD) dataset that is licensed to GSK for internal use. As a student affiliated with and co-sponsored by GSK, I have been granted internal access to these data. An external CPRD data request is currently underway in order to obtain ethnicity and socioeconomic deprivation measures. Once these parameters are obtained, analysis will be rerun to include these variables and results will be reported in a manuscript intended for publication.

5.2 Background

5.2.1. HBV immunisation in the UK

In the UK, HBV is concentrated in certain population subgroups, including people who inject drugs, people born in countries with intermediate to high prevalence (≥2%) of CHB (257,258) and socioeconomically deprived communities (47). Risk of chronic infection is also highest in those exposed in early life (259,260). HBV infection is hence likely to lead to the highest burden of sequalae in specific vulnerable sub-populations. It is possible that rates of HBV vaccination coverage also vary across ethnicity, with socioeconomic status, and by geographical region (given that population demographics vary across regions in England).

The World Health Organization's (WHO) Global Health Sector Strategy on viral hepatitis provided key recommendations in 2016 for the elimination of viral hepatitis as a public health threat by 2030. One key recommendation was introduction of national childhood HBV vaccination programmes (24). The introduction of a universal infant HBV vaccination was discussed thoroughly by the UK's Joint Commission on Vaccination and Immunisation in 2014 after review of evidence supporting the

inclusion of an HBV component into the then pentavalent universal infant vaccine schedule. Introduction of a universal HBV vaccine had been delayed due to lack of evidence of cost-effectiveness at the time regarding addition to the multivalent compound as well as uncertainty regarding maintenance of protection via monovalent immunisation. , Even when vaccine programmes are rolled out, they may not reach the most vulnerable populations due to poor access, lack of education/information, and vaccine hesitancy (261).

The UK introduced universal infant HBV vaccination in mid 2017. Since then, infants have been routinely vaccinated against hepatitis B virus (HBV) infection during their first year of life, as a component of multivalent vaccinations given at ages 8, 12 and 16 weeks (**Figure 5.1**) in line with the WHO Expanded Programme for Immunisation (EPI) and in accordance with UK guidance in The Green Book (25). A selective neonatal programme is also in place for infants born to mothers who are infected with HBV. Infants included in the selected program receive an initial dose within 24 hours of birth, followed by 3 further doses at 4 weeks, 8 weeks and one year of age. Infants born to women considered highly infective (HBsAg positivity, viral load is ≥ $2x10^5$ and negative anti-HBe) also receive HBIG. Therefore, 3 or more vaccine doses constitute full coverage, and zero, 1 or 2 doses are regarded as an incomplete schedule. Vaccination is also offered to children aged >10 years who are at higher risk of HBV infection. In comparison, the routine Meningococcal B (MenB) vaccination is administered as a monovalent compound, with two primary doses administered at 8 and 16 weeks of age, and one booster (co-administered with Meningitis C immunisation) administered at one year of life (**Figure 5.1**).

In the UK, the routine infant HBV vaccine is administered as part of a hexavalent combination vaccine containing diphtheria / tetanus / acellular pertussis / inactivated polio vaccine / *Haemophilus influenzae* type b / hepatitis B (DTaP/IPV/Hib/HepB). However, since its introduction, no systematic large-scale investigation has been undertaken to i) describe HBV vaccine uptake across demographics/subgroups (including ethnicity, socioeconomic status and geographical region), or; ii) compare HBV vaccination uptake rates to those observed pre-2017. Additionally, given recent evidence suggesting reduced coverage on childhood vaccinations in the UK (262) and

other countries (263–267) since March of 2020, an investigation of how routine HBV vaccination has been impacted by the COVID-19 pandemic is warranted.

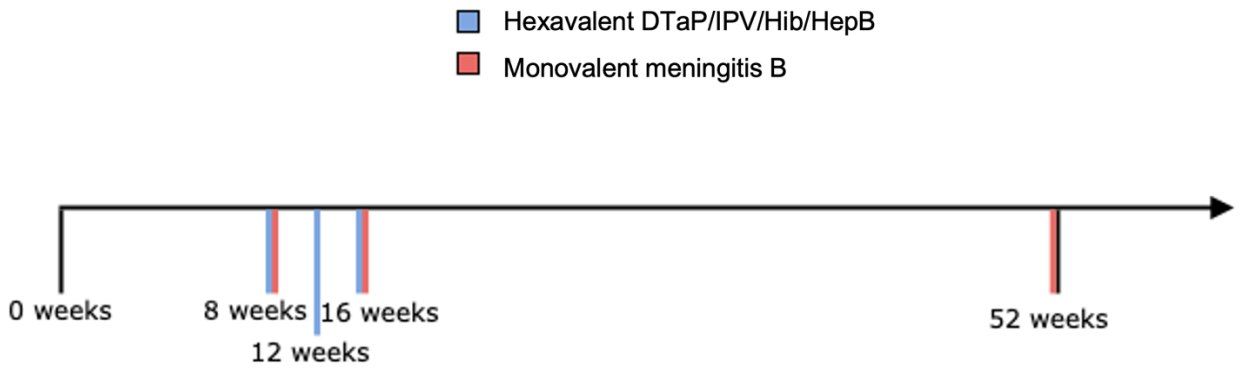


Figure 5.1 Infant vaccination timeline from 0 to 52 weeks of life depicting the vaccine schedule for the hexavalent combination vaccine and monovalent Meningitis B vaccine.
DTaP/IPV/Hib/HepB; diphtheria/tetanus/acellular pertussis/inactivated polio vaccine/Haemophilus influenzae type b/hepatitis B.

HBV vaccine coverage has recently been explored by UKHSA reports which indicate uptake of >85% of multivalent infant vaccines from years 2018 to present (26–30), based on Cover of Vaccination Evaluated Rapidly ('COVER') surveillance. COVER data for Northern Ireland, Wales and Scotland are reported by all health boards, and English data are reported by local teams at the local-authority level for general practices across all regions, and English data are collected within the NHS Digital infrastructure. Therefore, our CPRD study sample may vary to the nation-wide sampling referencing frame used to inform COVER reports, as the infants included in CPRD represent a subset of those available for inclusion in reports.

5.2.2. Clinical Practice Research Datalink (CPRD) primary care EHR database

The CPRD is a primary care EHR database including general practices located across the UK, including practices from Scotland, Wales and Northern Ireland. Established in 1987, the database contains record from practices utilising EMIS or Vision IT software (66,67). There are records from >60 million individuals in CPRD, of which 18 million are registered at general practices which currently contribute data to CPRD. Over a quarter of patients have >2 decades of follow-up. This analysis utilises a data cut from the proprietary CPRD GOLD dataset that is licensed to GSK for internal use.

The CPRD database is appropriate to address the specified objectives as it is a large-scale primary-care EHR database which is representative of the wider UK population (268) due to its size and geographical coverage. Additionally, infant vaccinations are managed in primary care and therefore relevant fields should be complete for the majority/all infants in this sample. The variables available in this dataset include year of birth, month of birth, geographical region, and record of vaccine for HBV, tetanus and MenB compounds. Linkage to additional data sources is possible for selected patients, depending on the IT system used at their respective general practice. Data may be linked to a variety of sources including Hospital Episode Statistics (HES) data (Accident and Emergency; Inpatient; Outpatient; Imaging), ONS linkage to the national death registry, national cancer registry and treatment registers, mental health services, socio-economic and ethnicity measures via local-area data linkage, mother-baby link and pregnancy register algorithms (61)

5.2.3. Aims and objectives

Novel quantification of coverage of infant HBV vaccine coverage across relevant factors, including socioeconomic status and geographic region, is warranted. In addition, it is useful to undertake this analysis in an EHR database such as CPRD in order to. validate which information is available, and the extent to which results triangulate with UKHSA reports. Therefore, I set out to undertake a cross-sectional descriptive analysis for every year from 2015 to 2020 to quantify uptake of routine infant HBV vaccination, in total and stratified by relevant subgroups. I compared HBV vaccination to two other baseline vaccinations for which schedules have been longer established, tetanus and Meningitis B, which were introduced as routine vaccinations in the 1970s and 2015, respectively (269,270). I aimed to quantify the percentage of UK infants that received 1, 2, 3 or >3 HBV vaccines within the first 14 months of life throughout the study period, and to compare yearly UK infant HBV vaccine uptake across multiple relevant birth periods (**Figure 5.2**):

A. 01/January/2015 to 01/August/2017: the period before a universal infant HBV vaccination was introduced in the UK.

B. 02/August/2017 to 01/February/2020: the period after the universal infant HBV vaccination was introduced in the UK, but before the COVID-19 pandemic had begun.
C. 02/February/2020 to latest available data entry: the period after the universal infant HBV vaccination was introduced in the UK, and during the COVID-19 pandemic.

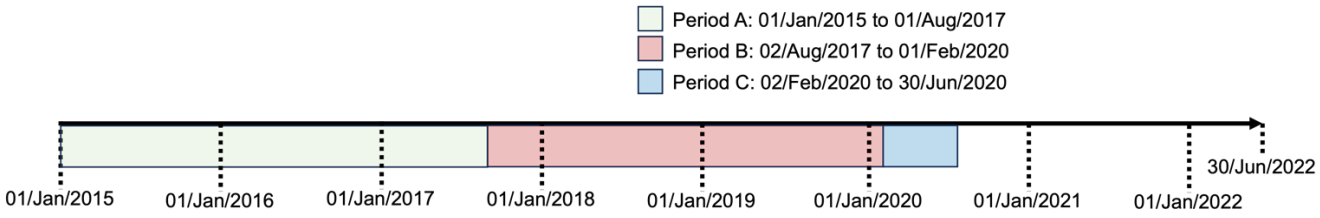


Figure 5.2. Study timeline depicting calendar years (and their midpoints) with dotted lines, and relevant study periods in shaded areas. Infants are divided into 3 study periods according to month and date of birth.
Follow up data were available to 30 June 2022, ensuring complete follow-up was achieved for infants born in any of the 3 periods. Periods are split as follows: Period A, 01/January/2015 to 01/August/2017, the period before a universal infant HBV vaccination was introduced in the UK; Period B, 02/August/2017 to 01/February/2020. the period after the universal infant HBV vaccination was introduced in the UK, but before the COVID-19 pandemic had begun and; Period C, 02/February/2020 to latest available data entry, the period after the universal infant HBV vaccination was introduced in the UK, and during the COVID-19 pandemic.

5.3. Research questions

5.3.1. Research question 1. What are the percentages of UK infants receiving 1, 2, 3 or >3 HBV vaccines within the first 14 months of life from 2015 to mid-2020?

5.3.2. Research question 2. How have the introduction of the HBV vaccine as a universal infant immunisation and the COVID-19 pandemic impacted HBV vaccination coverage in the UK, and are there differences in coverage based on geographic region or ethnic group?

5.4. Methods

5.4.1. Study design and follow-up

This is a descriptive cross-sectional study being undertaken in the CPRD primary care EHR database. Participants were categorised by month and calendar year of birth for every year from 2015 to 2020 (with calendar years split in half from January to June and from July to December) to quantify uptake (percentage of eligible infant) of routine infant HBV vaccination.

Inclusion criteria (which were minimal and included infants so long as their data quality was acceptable) for infants eligible for this study were as follows:

- Participants whose year of Current Registration Date (CRD) is between January 1st, 2015 and June 30th, 2020, with CRD defined as follows:
    - Date the individual's current period of registration with the practice began (date of the first 'permanent' record after the latest transferred out period) (271).
- Registered in a Practice already deemed to be providing "up to (research) standard" data and the "Up To Standard Date" is on or before the CRD, with relevant terms defined accordingly:
    - Up To (research) Standard: Indicator of whether practice data is deemed to be of research quality (271).
    - Up To Standard Date: Date at which the practice data is deemed to be of research quality. Derived using a CPRD algorithm that primarily looks at practice death recording and gaps in the data (271).
- Whose data quality is flagged as "acceptable" according to an internal CPRD validation process refs (272–274) where patients are unacceptable if they have disrupted/non-continuous follow up or "poor data recording that raises suspicion as to the validity of that patients record"
- Whose birth year/month was within 3 months prior to/on the CRD year/month
- Who were aged ≥14 months on earliest date of "transfer out date" or "death date" or "last collection date" (i.e., Individuals who have survived to at least 14 months of life), with the relevant terms defined accordingly:

- Transfer out date: Date the individual transferred out of the practice, if relevant (271).
- Last collection date: Date of the last collection for the practice (271).

The period of 2015 to 2020 was chosen for multiple reasons because it allowed us to sufficiently cover the period of change using recent evidence, and by including individuals born between January 2015 and June 2020 we had complete (14 months) follow-up for all eligible infants. Infants registered in CPRD within the first three months of life and followed to age 14 months were included, so as to capture those who received late vaccinations, or for whom there was a lag between vaccine administration and notification in the CPRD database.

I quantified uptake across calendar years in total and stratified by ethnicity, socioeconomic status and geographical region, comparing HBV vaccination uptake to two baseline vaccines:

I. Tetanus (given as part of the hexavalent vaccine which includes 3 doses in the first year of life administered at age 8, 12 and 16 weeks of life) (**Figure 5.1**), and;
II. Meningitis B (MenB) (not part of the hexavalent vaccine; one dose given at each of 8 and 16 weeks of life and a booster dose administered at 1 year) (**Figure 5.1**).

HBV, tetanus and MenB vaccination were identified using combinations of Read codes unique to both multivalent vaccine compounds and individual vaccine components. By including two non-HBV baseline vaccines, I aimed to elucidate whether any observed gaps in the routine schedule are HBV vaccine-specific or generalised to the wider schedule. Tetanus and MenB were chosen as comparators in order to quantify HBV vaccination uptake in comparison to i) a compound (tetanus) which is also administered in the hexavalent combination vaccine and was included in the pentavalent combination vaccine before the introduction of HBV in August 2017, and ii) a compound (MenB) which is administered as a stand-alone injection not part of a combination injection. Tetanus is part of the current hexavalent compound and is a

control comparator, so wherever an HBV vaccination is recorded as administered as part of the hexavalent combination, tetanus should have also been administered.

5.4.2. Sample size / power calculations

A Rapid Data Query (RDQ) within GSK was undertaken to assess the feasibility of this analysis by estimating the approximate number of eligible infants we could anticipate for the full dataset. Inclusion criteria were minimal whereby infants are included so long as their data quality is deemed "acceptable" according to an internal CPRD validation process refs (272–274). In short, patients are considered unacceptable if they have disrupted/non-continuous follow up or "poor data recording that raises suspicion as to the validity of the that patients record" (272). I maximised power by including all eligible infants for the years 2015 to mid 2020. This RDQ identified 144,604 infants for analysis with registered dates of birth between January 2015 - June 2020.

I calculated precision estimates using the PASS tool (275) (**Table 5.1**). Previous vaccination coverage reports from Public Health England indicate uptakes >85% of multivalent infant vaccines from years 2018 to present (26–30). With a sample size of 100,000 infants and a vaccine uptake of 85%, 95% confidence intervals will range from 84.8% to 85.2% (Figure 1). Required coverage estimates were not available for stand-alone HBV or MenB vaccinations.

Table 5.1. 95% precision estimates for vaccine uptake across a range of sample sizes and proportional vaccine uptake. Uptake estimates are based on previous vaccine uptake reports from Public Health England (26–30).

Confidence Level	Sample Size (N)	Actual Width	Proportion (P)	Lower Limit	Upper Limit	Width if *P* = 0.05
0.950	100000	0.004	0.850	0.848	0.852	0.006
0.950	100000	0.004	0.900	0.898	0.902	0.006
0.950	100000	0.003	0.950	0.949	0.951	0.006
0.950	100000	0.000	1.000	1.000	1.000	0.006
0.950	120000	0.004	0.850	0.848	0.852	0.006
0.950	120000	0.003	0.900	0.898	0.902	0.006

0.950	120000	0.002	0.950	0.949	0.951	0.006
0.950	120000	0.000	1.000	1.000	1.000	0.006
0.950	140000	0.004	0.850	0.848	0.852	0.005
0.950	140000	0.003	0.900	0.898	0.902	0.005
0.950	140000	0.002	0.950	0.949	0.951	0.005
0.950	140000	0.000	1.000	1.000	1.000	0.005
0.950	160000	0.003	0.850	0.848	0.852	0.005
0.950	160000	0.003	0.900	0.899	0.901	0.005
0.950	160000	0.002	0.950	0.949	0.951	0.005
0.950	160000	0.000	1.000	1.000	1.000	0.005
0.950	180000	0.003	0.850	0.848	0.852	0.005
0.950	180000	0.003	0.900	0.899	0.901	0.005
0.950	180000	0.002	0.950	0.949	0.951	0.005
0.950	180000	0.000	1.000	1.000	1.000	0.005

Power calculations were performed *a priori* in R using the 'pwr' package. Sample sizes were obtained via RDQ along with governmental statistics of pentavalent DTaP/IPV/Hib/HepB vaccine coverage (276,277) to perform power calculations to inform statistical comparisons of vaccine uptake (proportion of infants receiving vaccination) between years. To inform comparison of coverage between calendar years, sample sizes and estimates from years 2015 (n = 34,178) and 2019 (n = 22,414) were used to perform a power calculation. Using these parameters, analysis would have 83% power to detect a 2.5% difference in vaccine coverage between the two samples (**Figure 5.3**).

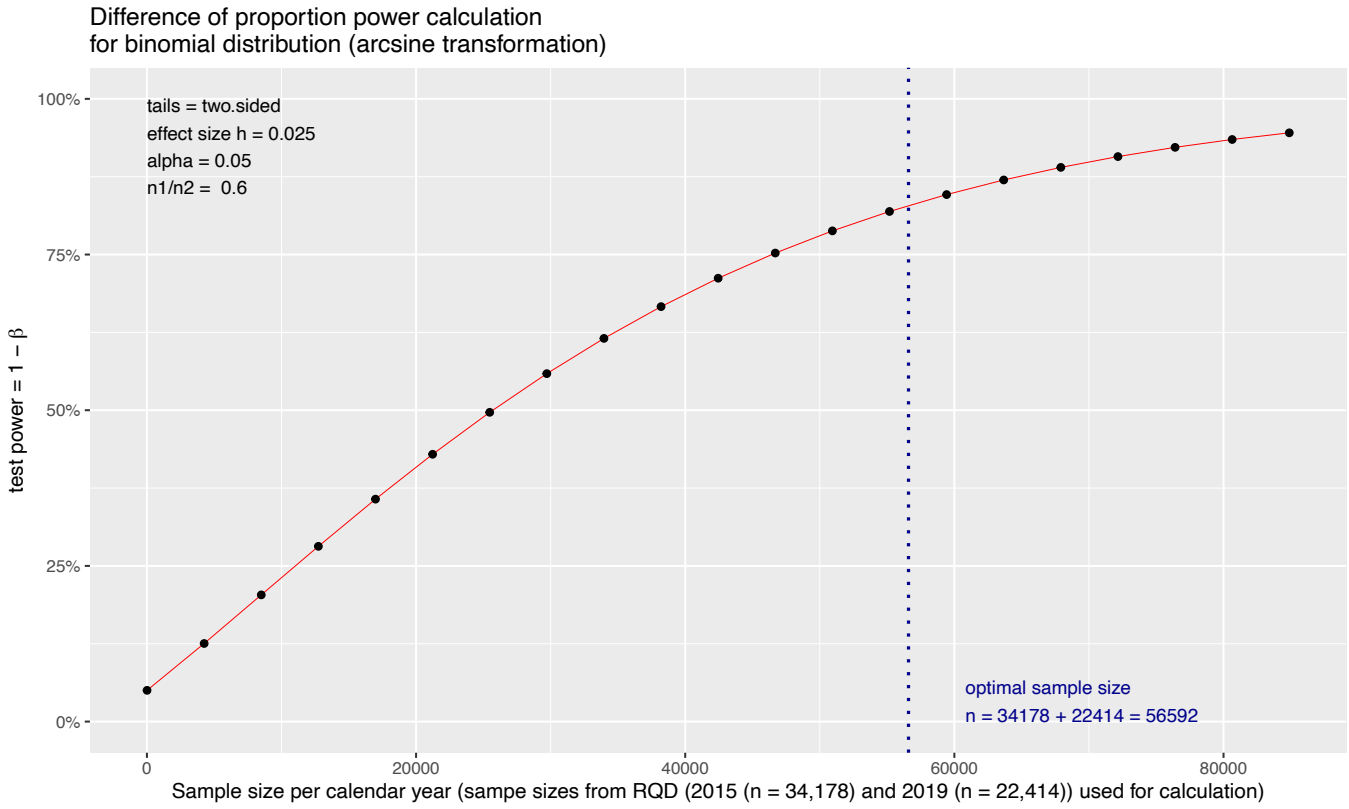


Figure 5.3. Power calculation of difference of two different sample sizes with different proportions equating to a total sample size for investigation of hexavalent DTaP/IPV/Hib/HepB vaccination coverage across calendar years.
Sample size estimates for 2015 (n = 34,178) and 2019 (n = 22,414), informed by governmental statistics of pentavalent DTaP/IPV/Hib vaccine coverage (276,277) were used. With these informed sample size estimates, a total sample of 56,592 individuals would detect a 2.5% difference in vaccine coverage between the two calendar years with 83% power (represented by the dashed vertical line).

5.4.3. Statistical analysis

I quantified the percentages of infants having received 1, 2, 3 or >3 HBV vaccine doses within the first 14 months of life, year on year, from 2015 to 2020 (with calendar years split in half from January to June and from July to December). This is because follow-up data were available to June 2022 at the time of data extract, allowing for complete follow-up of infants born from January through to June for the 2020 calendar year. Therefore, all calendar years were split into half-years in order to aid comparison and account for any season artefacts resulting from restricted data availability for infants born in 2020. A single vaccine dose is defined as a single record of relevant vaccination. For all analyses, infants were grouped by month and calendar year of registration, and vaccination uptake was compared across half-years. Descriptive

statistics (frequencies and percentages) were utilised to describe vaccine coverage across calendar years, time periods, region and ethnicity. Patient counts and percentages were presented for binary and categorical variables and were compared using chi-squared tests. Z-tests were used to test for significant differences between independent proportions of vaccine coverage. Percentage vaccination coverage per primary care practice was calculated with the numerator as the number of infants with complete vaccination status and denominator as the total number of infants registered at the practice. Primary care practices were ranked by percentage vaccine coverage and divided into the following categories: 0-10%, 11-20%, 21-40%, 41-60%, 61-80% and 81-100%.

5.5. Results

5.5.1 Description of study cohort

In total, 46,399 infants were eligible for inclusion in this study, identified from 382 general practices across 11 regions in the UK (**Table 5.2**). More infants in the sample were born in 2015 (n = 9624, 20.7%) and 2016 (n = 9786, 21.1%) as compared to later years (2018 n = 7985, 17.2%; 2019 n = 6578, 14.2%), reflecting the attrition from CPRD GOLD as practices switched from the InPS Vision GP software provider to EMIS. There was a near 50:50 division of female and male (n = 23688, 51.1%) infants across the entire cohort and study period. Nearly half (n = 22625, 48.7%) of infants originated from Scotland and Wales. Ethnicity records were missing for just over half (n = 24554, 52.9%) of infants. The largest ethnic group in the cohort was White (n = 14366, 31.0%), followed by Mixed (n = 5195, 11.2%), South Asian (n = 902, 1.9%) and Black (n = 606, 1.3%).

5.5.2 Representation by GP practice

Just over half (n = 23255, 50.1%) of all infants were registered at general practices achieving 41-60% practice-level vaccination coverage throughout the entire study period. Most remaining infants fell into practices achieving ≤40% coverage across the total study period.

Table 5.2. Table of patient characteristics including sex, regional and ethnicity tabulations.

	Number of individuals, n (%)
Total	46399
Time period of birth, n (%)	
Jan-Jun 2015	7120 (15.3)
Jul-Dec 2015	2504 (5.4)
Jan-Jun 2016	7359 (15.9)
Jul-Dec 2016	2427 (5.2)
Jan-Jun 2017	6530 (14.1)
Jul-Dec 2017	2127 (4.6)
Jan-Jun 2018	6209 (13.4)
Jul-Dec 2018	1776 (3.8)
Jan-Jun 2019	5400 (11.6)
Jul-Dec 2019	1178 (2.5)
Jan-Jun 2020	3769 (8.1)
Male sex, n (%)	23688 (51.1)
Region ‡, n (%)	
East of England	494 (1.1)
London	4499 (9.7)
North East	0 (0.00)
North West	3562 (7.7)
Northern Ireland	3156 (6.8)
Scotland	12210 (26.3)
South East	7207 (15.5)
South West	1905 (4.1)
Wales	10415 (22.4)
West Midlands	2718 (5.9)
Yorkshire & The Humber	233 (0.5)
Ethnicity ‡, n (%)	
Black	606 (1.3)
Chinese	92 (0.2)

	Number of individuals, n (%)
Indeterminant**	162 (0.3)
Missing	24554 (52.9)
Mixed	5195 (11.2)
Other	522 (1.1)
South Asian	902 (1.9)
White	14366 (31.0)

† The UK geographical regions available in CPRD are split into quintiles of coverage whereby the two regions in the first quintile have the lowest vaccine coverage and the two in the fifth quintile have the highest coverage. An individual will be considered to have full vaccine coverage if they have 3 records of HBV vaccination within the first 14 months of life.

‡ Indicates variable measured at practice-level.

** Individuals with record of two different ethnic categories.

5.5.3 Cohort coverage by geographic region

Geographical representation in this sample of infants born across English regions, Northern Ireland, Scotland and Wales varied with time (**Table 5.3**). Some regions contributed consistent numbers of infants to the cohort throughout the study period, including London, the North West, Scotland, Northern Ireland, Wales, and the West Midlands. Other regions contributed decreasing numbers of infants throughout study duration, including the South East, South West, East of England and Yorkshire & the Humber; the latter of these 2 contributed infants for only a portion of the total study duration. No infants were available for analysis from practices located in the North East.

Table 5.3. The geographical representation in the entire study sample of infants (n (%)) from across the English regions, Scotland, Wales and Northern Ireland in the cohort over time.

	East of England	London	North East	North West	Northern Ireland	Scotland	South East	South West	Wales	West Midlands	Yorkshire & The Humber
2015 Jan-Jun	136 (27.5)	638 (14.2)	0 (0.0)	400 (11.2)	560 (17.7)	1955 (16.0)	1023 (14.2)	333 (17.5)	1656 (15.9)	376 (13.8)	43 (18.5)
2015 Jul-Dec	8 (1.6)	396 (8.8)	0 (0.0)	223 (6.3)	33 (1.0)	357 (2.9)	751 (10.4)	191 (10.0)	256 (2.5)	216 (7.9)	73 (31.3)
2016 Jan-Jun	125 (25.3)	612 (13.6)	0 (0.0)	424 (11.9)	555 (17.6)	1941 (15.9)	1108 (15.4)	330 (17.3)	1759 (16.9)	445 (16.4)	60 (25.8)
2016 Jul-Dec	9 (1.8)	297 (6.6)	0 (0.0)	253 (7.1)	37 (1.2)	367 (3.0)	718 (10.0)	186 (9.8)	255 (2.4)	248 (9.1)	57 (24.5)
2017 Jan-Jun	131 (26.5)	681 (15.1)	0 (0.00)	394 (11.1)	480 (15.2)	1801 (14.8)	856 (11.9)	256 (13.4)	1588 (15.2)	343 (12.6)	0 (0.00)
2017 Jul-Dec	4 (0.8)	329 (7.3)	0 (0.0)	260 (7.3)	32 (1.0)	345 (2.8)	637 (8.8)	148 (7.8)	221 (2.1)	151 (5.6)	0 (0.0)
2018 Jan-Jun	38 (7.7)	518 (11.5)	0 (0.0)	392 (11.0)	520 (16.5)	1813 (14.8)	823 (11.4)	228 (12.0)	1522 (14.6)	355 (13.1)	0 (0.0)
2018 Jul-Dec	1 (0.2)	316 (7.0)	0 (0.0)	255 (7.2)	33 (1.0)	232 (1.9)	525 (7.3)	107 (5.6)	181 (1.7)	126 (4.6)	0 (0.0)
2019 Jan-Jun	42 (8.5)	393 (8.7)	0 (0.0)	389 (10.9)	497 (15.7)	1806 (14.8)	396 (5.5)	115 (6.0)	1471 (14.1)	291 (10.7)	0 (0.0)
2019 Jul-Dec	0 (0.0)	190 (4.2)	0 (0.0)	208 (5.8)	22 (0.7)	184 (1.5)	305 (4.2)	11 (0.6)	140 (1.3)	118 (4.3)	0 (0.0)
2020 Jan-Jun	0 (0.0)	129 (2.9)	0 (0.0)	364 (10.2)	387 (12.3)	1409 (11.5)	65 (0.9)		1366 (13.1)	49 (1.8)	0 (0.0)

5.5.4 HBV vaccine coverage over time

Overall, the percentage of infants achieving full HBV vaccine coverage (≥3 doses within the first 14 months of life) increased consistently for infants born in 2015 through to 2020 (**Table S2, Figure 5.4**). For infants born from July 2017 onwards, the majority (>85%) had records of ≥ 3 HBV vaccine doses, as compared to January 2015 to June 2017 when >95% had no reported HBV vaccination records, reflecting the change in the routine vaccination programme. Complete coverage of tetanus vaccination, which was introduced in the UK as a universal neonatal vaccine in the 1970s, remained high (>93%) across practices throughout the entire study duration (**Table S2, Figures 5.4**). Similarly, the large majority of infants received complete MenB vaccine dosing regimens from mid-2015 to mid-2020 following its introduction as a universal component in 2015. To detect a difference of ≥80% in vaccine coverage between the years 2015 (n = 9624) and 2019 (n = 6578), we had >95% power using our study sample.

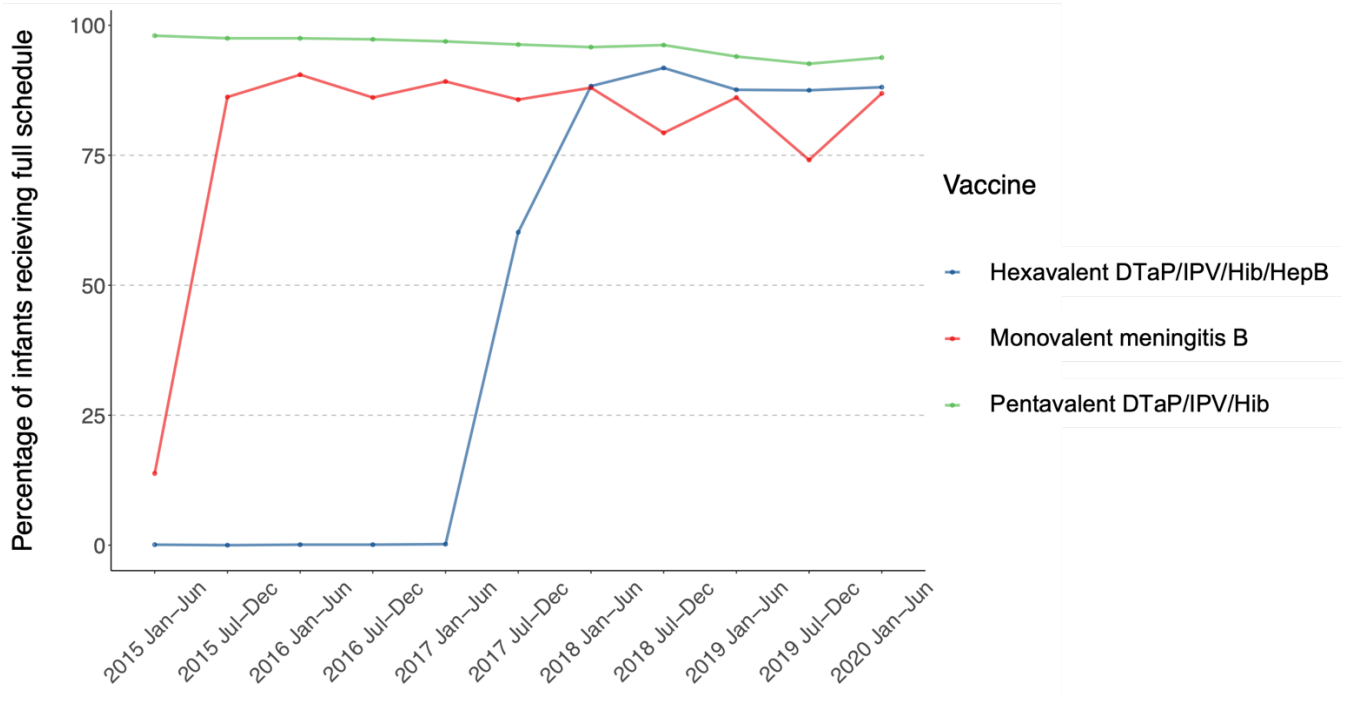


Figure 5.4. Percentages of infants receiving full vaccine schedules within the first 14 months of life for hepatitis B virus, tetanus and meningitis immunisations for infants born between January 2015 to June 2020.
A full schedule for hepatitis B virus and tetanus consists of ≥ 3 doses, and ≥ 2 doses for meningitis B. An increase towards universal coverage is observed for the HBV vaccine following its introduction to the universal hexavalent compound vaccination in August 2017. A similar result is observed for meningitis B following its introduction as a universal component in 2015.

5.5.5 HBV vaccine coverage by region

As expected, an increase in HBV vaccine coverage was observed across the UK after universal introduction (**Figures 5.4 and 5.5, Table 5.3**). Coverage appeared to be higher in southern regions as compared to northern regions for HBV, tetanus and MenB vaccinations (**Tables S3-S5**). However, a disproportionately large number of infants were registered in practices located in Scotland and Wales. Therefore, it is not clear how representative this observation is of true population-wide vaccine coverage.

Based on this sample, infants in Northern Ireland, the North West, and East achieved 90-100% HBV vaccine coverage by 2019, and 80-90% in the South East, South West, West Midlands and London in 2019 (**Figure 5.5**). In 2019, coverage in Scotland was 73.2%. Data from the North East, East Midlands and Yorkshire and the Humber were not available for 2017 and 2019.

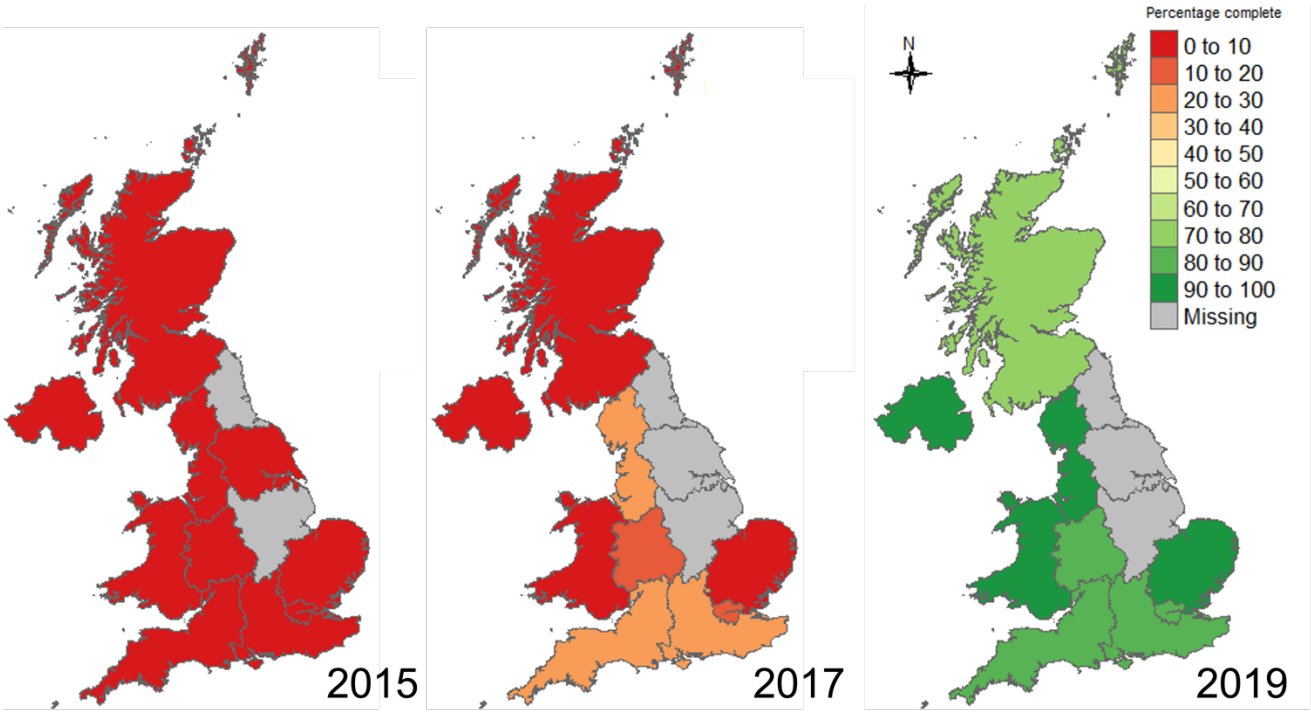


Figure 5.5. Percentages of infants receiving full HBV vaccine schedules within the first 14 months of life for years 2015, 2017 and 2019 throughout the United Kingdom.
An increase towards universal coverage is observed across the country following its introduction to the universal hexavalent compound vaccination in August 2017.

5.5.6 HBV vaccine coverage in response to health policy change and the beginning of the COVID-19 pandemic

Coverage of HBV vaccination was quantified across three time periods to observe how both the introduction of the immunisation as a universal component and the COVID-19 pandemic potentially influenced uptake (**Table 5.4, Figure 5.6**). After introduction of the vaccine universally in August 2017, complete coverage amongst infants was quantified at 88.0% for period B. Compared to infants born during the period following universal introduction but preceding the COVID-19 pandemic (Period B, 88.0%), coverage did not differ significantly in infants born during the pandemic (period C, 87.7%). Similarly, coverage of complete tetanus and MenB vaccinations did not drop during the pandemic (period C) (**Table 5.4**).

Table 5.4. Hexavalent DTaP/IPV/Hib/HepB, pentavalent DTaP/IPV/Hib and monovalent meningitis B vaccination uptake across three birth periods.

	Overall	**Period A: Pre-routine introduction and pre-COVID**	**Period B: Post-routine introduction and pre-COVID**	**Period C: Post-routine introduction and post-COVID**
Infants (N,%)	46399	26626	16031	3742
Hexavalent DTaP/IPV/Hib/HepB vaccinations (N, %)				
0 records	26165 (57.3)	25556 (97.2)	483 (3.1)	126 (3.4)
1 record	950 (2.1)	517 (2.0)	351 (2.2)	82 (2.2)
2 records	1387 (3.0)	168 (0.6)	998 (6.4)	221 (6.0)
3 records	17064 (37.4)	46 (0.2)	13809 (88.0)	3209 (87.7)
>3 records	79 (0.2)	0 (0.0)	58 (0.4)	21 (0.6)
Pentavalent DTaP/IPV/Hib vaccinations (N, %)				
0 records	43 (0.1)	15 (0.1)	22 (0.1)	6 (0.2)
1 record	482 (1.1)	219 (0.8)	224 (1.4)	39 (1.1)

2 records	1157 (2.5)	445 (1.7)	533 (3.4)	179 (4.9)
3 records	25961 (56.9)	10712 (40.8)	11841 (75.4)	3408 (93.1)
>3 records	18002 (39.4)	14896 (56.7)	3079 (19.6)	27 (0.7)
Monovalent meningitis B vaccinations (N, %)				
0 records	4118 (9.0)	4076 (15.5)	36 (0.2)	6 (0.2)
1 record	1328 (2.9)	849 (3.2)	382 (2.4)	97 (2.7)
2 records	5584 (12.2)	3301 (12.6)	1915 (12.2)	368 (10.1)
>2 records	34615 (75.8)	18061 (68.7)	13366 (85.1)	3188 (87.1)

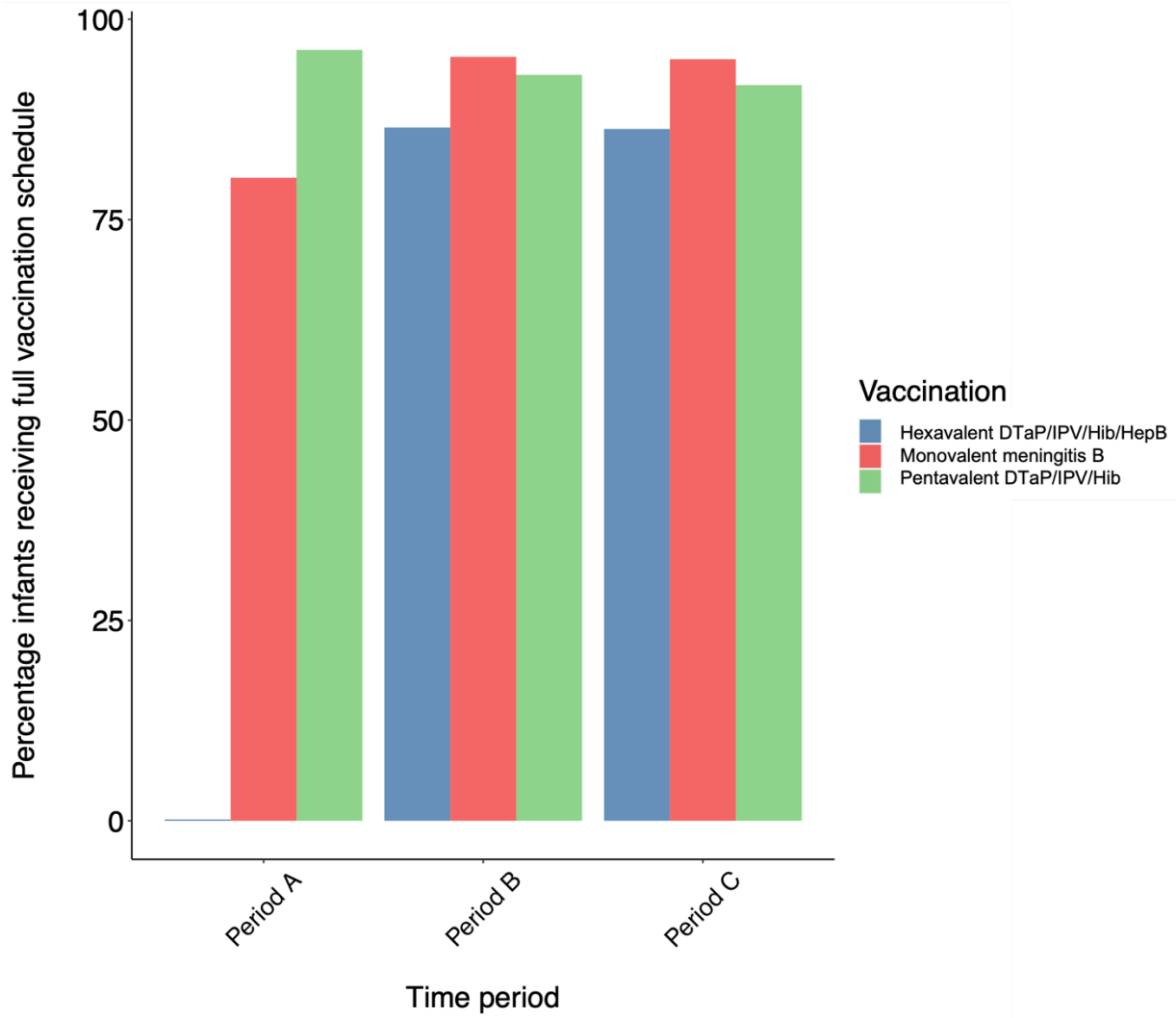


Figure 5.6. Percentages of infants receiving full vaccinations schedules across 3 separate birth periods:
Period A, pre-routine introduction and pre-COVID; Period B, post-routine introduction and pre-COVID, and; Period C, post-routine introduction and post-COVID. Coverage is visualised for hepatitis B virus, meningitis B and tetanus vaccines. HBV coverage increased from period A to B following its

introduction to the universal infant vaccine schedule. A full schedule for hepatitis B virus and tetanus consists of ≥ 3 doses, and ≥ 2 doses for meningitis B. Coverage for hepatitis B virus, meningitis B and tetanus remained high for infants born both before (period B) and during (period C) the COVID-19 pandemic.

Trends in HBV vaccine coverage across the 3 time periods did not differ by geographical region, apart from an increase in coverage of 7.7% in London between periods B and C, however this increase was non-significant (**Table S6**). Longitudinal trends in tetanus and MenB vaccinations were consistent across geographical regions (**Tables S7 and S8**).

5.5.7 HBV vaccine coverage by ethnicity

Most (52.9%) study participants were missing ethnicity information, and of those with complete ethnicity data only small minorities (<3%) were of non-White ethnicities including Black, Chinese, South Asian and other. The majority (>70%) of infants born after the introduction of HBV as a universal vaccine received full vaccinations irrespective of ethnicity (**Table S3**). No significant differences were observed across ethnicities, within-ethnicity coverage of the HBV vaccine varied widely for infants born between 2018 and 2020. This general finding was also observed for tetanus and MenB vaccinations (**Table S4 and S5**). Substantial increases in HBV, tetanus and MenB vaccine coverage between periods B and C were observed in Chinese and South Asian ethnic groups (**Tables S6-S8**). However, these increases were non-significant and should be interpreted with caution due to the small numbers of infants in these ethnicity strata.

5.6. Discussion

In this analysis, I quantify the coverage of HBV infant vaccinations across 5.5 years covering the period when universal infant vaccine was introduced (mid-2017), and the early part of the COVID-19 pandemic beginning in March 2020. I document a rapid increase in full HBV vaccine coverage in infants born after the immunisation introduction to the universal schedule and no detrimental impact during the first few months of the pandemic. My findings are presented alongside results from two baseline vaccines, tetanus and MenB, for which coverage remained high throughout the study period. Birth dose of HBV vaccination (i.e., a first dose within 24h of birth) is recommended by the WHO but is not deemed necessary in the UK due to both a low population prevalence of infection and high uptake of antenatal screening whereby pregnant women who are HBsAg positive may be offered prophylactic antiviral treatment with targeted birth dose vaccine to their neonates (25). Therefore, HBV birth dose immunisation was not investigated in the present analysis but will have been administered to a small number of babies. We did not have access to maternal HBV status.

I have undertaken this analysis in the CPRD primary care EHR database, utilising all infants with complete (14 months) follow-up born between January 2015 and June 2020. This is a large-scale database containing millions of patients, and EHR data are not publicly available. My findings suggest that increases in coverage have been achieved since the introduction of the HBV immunisation as component to the universal multivalent compound in August 2017. However, ≥10% of infants remained unvaccinated in this study sample, suggesting that vaccination is still inaccessible to some specific subgroups. Furthermore, disparities were observed across geographical regions. However, it is not certain how generalisable this finding is on a national level, given the disproportionate contributions (or lack thereof) from some regions to the study sample. Nevertheless, geographical disparities suggest inequities and vulnerabilities in specific communities, whereby enhanced intervention may be needed in order to ensure universal protection and herd immunity.

The findings I present are concordant with COVER produced by NHS Digital, who report increases across all routine childhood vaccination coverage in 2019/2020 as compared to previous years present (26–30,278–280). Specifically, increases in the hexavalent combination vaccine have been reported. NHS Digital report that most regions achieved >94% coverage in 2019/2020, with the lowest coverage in the country reported in London. This regional result differs to my findings, whereby coverage in London was amongst the highest of all regions in this CPRD sample. However, the percentage of participants originating from London in my sample was disproportionately small compared to the percentage of the UK population residing in London (280), and as such my findings may not be representative of London or other metropolitan areas in the UK. Similarly, the Scottish HBV vaccine coverage reported in this analysis is lower to that reported in UKHSA COVER reports. This suggests that the sample of infants registered at Scottish general practices contributing data to CPRD is not necessarily representative of the country-wide infant population.

It is not possible to fully quantify the impact of the COVID-19 pandemic on routine HBV vaccination in this study, as data are only available up to June 2020. The 2020/2021 report from NHS digital suggests that despite small reductions in vaccine coverage compared to 2019/2020, the percentages of children receiving complete multivalent vaccine coverage remained above targets (279). A report from UKHSA published in October 2021 supports these findings, and describes that in August 2021, 86.5% of infants had received 3 doses of the hexavalent combination vaccine within the first 6 months of life (278). This coverage is similar to that which I report in this chapter. UKHSA findings of complete vaccine coverage also align with clinical guidelines from NICE recommending that routine childhood vaccines continue throughout the pandemic providing that individuals were not displaying symptoms of COVID-19 or other infections and were not self-isolating (281).

I have used a sample of infants from CPRD to make inferences about infant HBV vaccine coverage across the UK. Overall, this analysis was well-powered, as to detect a difference of ≥80% in vaccine coverage between the years 2015 and 2019, it had >95% power using the present study sample. However, it is possible that this sample is not representative of the wider UK infant population. Specifically, there is minimal

representation from some specific regions in this sample, particularly East, North East, South West and Yorkshire & the Humber regions of England. This may be due to regional variations in electronic health data computer software usage. Representation of infants was disproportionately high from Scotland and Wales. Furthermore, over half of individuals were missing information regarding ethnicity, and in infants for whom ethnicity information was complete, only small minorities (<3%) were of non-White ethnicities. Therefore, the regional- and ethnicity-specific findings I present may not be generalisable across the UK. Additionally, small numbers of individuals in geographical and ethnic subgroups results in a lack of statistical power to detect associations. Therefore, stratified vaccine coverages should be interpreted with caution.

Future analyses warrant further investigation of HBV vaccine coverage utilising data that extend past the COVID-19 pandemic. This would serve to both triangulate findings reported in UKHSA and NHS reports, and to quantify how clinical services have persevered in the post-COVID era. Investigation of how coverage varies across socioeconomic deprivation is warranted, particularly because my previous investigations in QResearch have found that HBV prevalence is associated with both ethnicity and deprivation. Description of individuals who are receiving incomplete vaccine schedules will be particularly informative for health service provision and resource allocation with quantification across socioeconomic strata. I intend to investigate these research questions further upon receiving the external CPRD GOLD dataset which will contain linkage to these parameters.

Conclusions

I report reassuringly high percentages of infants receiving full HBV vaccination regimens following introduction of routine immunisation in the UK, with coverage maintained in the early part of the COVID-19 pandemic. Gaps in coverage require further investigation to ensure consistency and equity of coverage.

6. Discussion

6.1 Summary of work and findings

The study presented herein has broadly aimed to advance knowledge surrounding the CHB epidemiology, disease burden, management and prevention in the UK. Evidence gaps surrounding these areas had been previously identified (**Chapter 1**), and throughout my project I aimed to fill such gaps using a variety of approaches. My project therefore commenced with a systematic literature review and meta-analysis to identify observational studies characterising observational cohorts of individuals with CHB and investigating risk factors for CHB progression to HCC (**Chapter 2**). The results of this meta-analysis were used to inform subsequent analyses. Specifically, I acquired insights concerning CHB cohort size, necessary study duration required to detect a sufficient number of HCC endpoints, and *a priori* identification of relevant HCC risk factors to inform data extraction and analysis in my subsequent investigations. My findings were also of relevance to my colleagues and peers in viral hepatitis and epidemiology research, as my study has obtained 34 citations to date and the *Journal of Viral Hepatitis* awarded me with a congratulatory certificate, as the article was one of their 10 most downloaded papers.

Thereafter I went on to undertake multiple streams of work in order to address evidence gaps surrounding CHB in the UK. Firstly, I undertook an observational cohort study in order to characterise the largest cohort of individuals with CHB in the UK to date, describe CHB epidemiology and investigate risk factors for CHB progression to HCC in a socioeconomically diverse multi-ethnic Western population (**Chapter 3**), all of which had not previously been undertaken in the UK in a robust and large-scale manner. To this end I extracted cohort of individuals with CHB from the QResearch primary care EHR database. Not only did this study address the aforementioned evidence gaps, but it also served to appraise the quantity and quality of information regarding CHB epidemiology and disease burden in primary care electronic health systems, which was previously uncertain. All of my findings from this chapter are either

published or currently under review in two manuscripts at peer-reviewed scientific journals.

Further evidence gaps concerning CHB antiviral treatment and COVID-19 related disruptions to routine HBV clinical practices required investigation using real-world data. Specifically, it was not clear what proportion of individuals with CHB in the UK were receiving antiviral treatment, and how many individuals amongst those not receiving treatment were eligible for initiation of a nucleoside analogue agent. It was also not certain how many additional individuals would be treatment eligible following a relaxation of treatment criteria. Finally, an analysis of the impact of the COVID-19 pandemic on routine CHB surveillance was warranted to quantify potential disruptions to routine clinical care and inform resource allocation post-pandemic. To this end, I undertook observational analysis in the NIHR HIC secondary care EHR database (**Chapter 4)**. My analyses served to describe HBV treatment coverage overall and across relevant subgroups, investigate factors associated with antiviral treatment initiation, quantify the number of treatment-eligible individuals currently not receiving treatment, and describe how many additional individuals would become treatment-eligible following a relaxation of eligibility criteria. I also able not only to generate evidence of a COVID-19 attributable disruption to routine HBV care, but also to quantify this impact and specifically the numbers of individuals with CHB who did not undergo sufficient surveillance of ALT or VL. My findings concerning COVID-19 disruption to routine HBV care are under review in *Wellcome Open Research*. My findings regarding HBV treatment eligibility and coverage are currently being written into a manuscript for submission to a peer-reviewed scientific journal.

In August 2017, the HBV vaccine was introduced into the universal infant vaccine schedule in the UK. Specifically, it was added to the multivalent compound administered to all infants across the country, with a complete schedule consisting of 3 doses within the first 14 months of life (with a fourth booster dose sometimes additionally administered depending on clinical relevance). National COVER vaccine uptake reports from UKHSA have quantified uptake in recent years at a national level, however quantification of uptake across relevant subgroups including regions and ethnicity is informative for targeting vaccination efforts. Furthermore, investigation of how vaccine uptake has changed with both the introduction of HBV immunisation as

a universal component and the start of the COVID-19 pandemic are warranted. Therefore a descriptive study of HBV vaccination uptake in recent years was warranted to address these evidence gaps. I undertook a descriptive cross-sectional study in all infants born from 2015 to mid-2020 in the CPRD database with sufficient follow-up (14 months) to detect full HBV vaccine schedule (**Chapter 5**). Specifically I report HBV vaccine coverage across regions and ethnicities, noting regional discrepancies in vaccine uptake. Additionally I report high (>85%) levels of uptake both following the introduction of HBV immunisation as a universal component in August 2017, and following the start of the COVID-19 pandemic. These findings are currently being written into a manuscript for submission to a peer-reviewed scientific journal. I am awaiting receipt of individual-level Townsend deprivation scores for infants in this study sample. Following receipt of this parameter, I intend update the analysis to quantify vaccine uptake across Townsend deprivation quintile, and thereafter submit this manuscript to a peer-reviewed scientific journal.

I have drawn the following key conclusions from the analyses I have undertaken throughout my doctoral studies:

- The CHB disease burden in the UK is concentrated in a young, socioeconomically deprived and ethnically diverse population in individuals with CHB represented in EHR databases. This has been demonstrated in a cohort of CHB individuals identified in the QResearch primary care EHR database. This cohort is the largest cohort of CHB individuals characterised in the UK to date.
- Production of a robust population-wide estimate of CHB prevalence in the UK taking into account diverse and disparate data sources spread across primary care EHR, secondary care EHR, local authorities, health boards, national surveillance systems and other data sources is warranted to inform. This estimate must use additional data sources representative of the general population to account for undiagnosed individuals (or diagnosed individuals not captured in electronic health systems) beyond electronic health data bases. Such an estimate may inform health policy to work towards achieving elimination goals outlined in the WHO's GHSS for HBV elimination.

- Findings from meta-analysis and/or independent observational cohort study suggest increasing age, male sex, comorbid disease, and ethnicity (likely a proxy for viral genotype) are associated with increased HCC risk, and statin use is associated with a decreased risk.
- According to current NICE guidelines for CHB management and treatment, only a minority of CHB patients in the UK are eligible for antiviral treatment. Analysis using secondary/tertiary care EHRs confirms that only a minority of individuals with CHB receive treatment. Amongst those individuals not receiving antiviral treatment, a small minority are treatment eligible and not initiated on NA therapy according to available data. This broadly indicates that surveillance of CHB individuals attending secondary/tertiary care services in the UK is largely comprehensive. Relaxation of treatment guidelines would initiate a small number of individuals onto antiviral treatment and therefore would not come at an economic cost
- The COVID-19 pandemic disrupted routine CHB surveillance services, whereby significantly lower numbers of individuals underwent sufficient surveillance during years 2020 and 2021 (i.e. COVID-19 years) versus 2016 to 2019 (pre-COVID-19 years). Data from the end of 2021 did not suggest that surveillance coverage had begun to increase towards pre-COVID-19 levels. Efforts and resources are required in order to ensure surveillance services can resume normal practice.
- Uptake of infant HBV vaccination is estimated to be ≥85 % since the introduction of the universal multivalent immunisation schedule in August 2017. Coverage appears to have remained high despite potential disruption attributable to the COVID-19 pandemic. Regional discrepancies in coverage were observed, however it is not certain how generalisable the study sample from which this finding was drawn is to the whole population of the UK.

6.2 Future recommendations and next steps

Many key evidence gaps have been addressed by the analyses summarised above. These investigations have laid a robust foundation upon which next steps can be taken

in order to address remaining evidence gaps and improved CHB management both within and out with the UK. Potential next steps for CHB research are as follows:

- Validation and potential modification of existing CHB-HCC risk score using data from primary and secondary/tertiary care EHRs in the UK. Previous risk scores have largely been developed in homogenous (i.e. ethnically, socioeconomically and clinically) cohorts, and therefore further analysis is warranted to derive risk scores which have utility in diverse CHB populations.
- Development, implementation, and evaluation of novel interventions/tools to address modifiable risk factors for HCC in CHB. Such tools should be implemented into the electronic health infrastructure to directly impact patient pathways. Additionally, integration of such interventions into the electronic infrastructure allows for real-time evaluation of their impact, and also for instant modification/calibration. This is particularly relevant, as risk-scoring tools require recalibration longitudinally as the population prevalence of relevant risk factors shifts.
- Integration of primary and secondary care EHR data sources to improve characterisation of the CHB patient pathway. Specifically, linkage of disparate data sources will enable comprehensive and robust investigation risk factors for relevant CHB disease endpoints. It will also enable the identification of potential intervention opportunities along the CHB care pathway at which the effect of novel interventions may be evaluated.
- Robust cost-effective analysis to investigate of impact of expanding CHB treatment eligibility. This is warranted in order to discern whether the cost of initiating a larger number of CHB individuals on NA therapy is outweighed by the cost of treating liver complications resulting from end-stage CHB.
- Inclusion of a larger CHB cohort with longer follow-up post-COVID-19 in order to evaluate whether CHB surveillance levels have returned to those observed before the pandemic. Integration of imaging information into available data will also be vital to quantify the impact of the pandemic on HCC screening practices specifically.
- Estimation of population-wide CHB prevalence using all available data sources to integrate, to the best of abilities, parameters regarding historic HBV trend data, country of birth information, and variation among subgroups including

migrants, ethnic minorities, individuals experiencing socioeconomic deprivation and PWID, amongst other.

- Quantification of HBV vaccine coverage by ethnicity, region and socioeconomic status in national COVER reports produced by UKHSA. This would serve to enhance resource allocation and target vaccine efforts to vulnerable population subgroups.

6.3 Reflection as an electronic health record researcher

I have been very grateful to utilise electronic data from the NHS for my doctoral project. I feel it is a privilege to have undertaken my studies in a healthcare setting which enables the efficient and impactful use of routinely-collected data to address evidence gaps and benefit patient outcomes and wellbeing. Throughout my doctoral studies I have also faced various data governance and access challenges. These include heterogeneity in EHR reporting practices across NHS Trusts, high levels of missingness, lack of linkage between EHR systems/other data sources and high monetary costs associated with data access. I hope that future initiatives and systems of working can seek to not only address these hurdles, but similarly prevent additional hurdles from surfacing. Standardised data input and processing practices along with sharing of resources (such as code lists for disease and medicine categories) will help to shape this. The OpenSAFELY platform is one such initiative already in practice which aims to reduce barriers to data access and research whilst maintaining an excellent standard of data privacy and governance (63).

6.4. Research output outside doctoral project

In addition to the studies mentioned throughout my results chapters, I have contributed to a variety of other studies and investigations outside of my doctoral project. These additional investigations have been wide-ranging, and I have utilised my statistical, epidemiological and programming skills to contribute to/conduct meta-analyses, rapid COVID studies, traditional longitudinal cohort data analysis (particularly the General Population Cohort (GPC) in Uganda). Below is a list of my citations which I have contributed to over the course of my DPhil studies but which out with my project scope:

- **Campbell C**, Andersson MI, Ansari MA, Moswela O, Misbah SA, Klenerman P, et al. Risk of Reactivation of Hepatitis B Virus (HBV) and Tuberculosis (TB) and Complications of Hepatitis C Virus (HCV) Following Tocilizumab Therapy: A Systematic Review to Inform Risk Assessment in the COVID-19 Era. *Frontiers in Medicine*. 2021 Aug 20;8:1346.
- Matthews PC, **Campbell** C **(joint first author)**, Săndulescu O, Matičič M, Ruta SM, Rivero-Juárez A, et al. Acute severe hepatitis outbreak in children: A perfect storm. What do we know, and what questions remain? Front Pharmacol. 2022 Nov 25;13:1062408.
- Mugisha J, **Campbell C (joint first author)**, Waddilove E, Makanga R, Wang T, Barnes E, Kimono B, Muzaale FN, Matthews PC, Newton R. Evidence suggests significant alcohol-associated mortality in a large population cohort in Uganda. Presented at EASL ILC 2023, manuscript in preparation.
- Wang T, Smith DA, **Campbell C (joint first author)**, Harris S, Salih H, Várnai KA, et al. Longitudinal Analysis of the Utility of Liver Biochemistry as Prognostic Markers in Hospitalized Patients With Corona Virus Disease 2019. *Hepatol Commun*. 2021 Jul 10.
- Downs LO, **Campbell C**, Yonga P, Ansari MA, Matthews PC, Etyang AO. A systematic review of Hepatitis B virus (HBV) prevalence and genotypes in Kenya: Data to inform clinical care and health policy. *medRxiv*. 2022 Apr 10;2022.04.08.22273611.
- Wang T, Smith DA, **Campbell C**, Freeman O, Moysova Z, Noble T, et al. Cohort Profile: The National Institute for Health Research Health Informatics Collaborative: Hepatitis B Virus (NIHR HIC HBV) research dataset. *International Journal of Epidemiology*. 2022 Jun 16;2022:1–11.
- Liao W, Coupland CAC, Innes H, Jepsen P, Matthews PC, **Campbell C**, et al. Disparities in care and outcomes for primary liver cancer in England during 2008–2018: a cohort study of 8.52 million primary care population using the QResearch database. *EClinicalMedicine*. 2023 May 1;59:101969.
- Wang T, Glampson B, Mercuri L, Papadimitriou D, Jones CR, Smith DA, Salih H, **Campbell C** et al. Identifying Hepatocellular Carcinoma from imaging reports using natural language processing to facilitate data extraction from electronic patient records. *medRxiv*. 2022 Aug 24. 2022.08.23.22279119.
- Liao W, Jepsen P, Coupland C, Innes H, Matthews PC, **Campbell C**, et al. Development and validation of personalised risk prediction models for early detection and diagnosis of primary liver cancer among the English primary care population using the QResearch® database: research protocol and statistical analysis plan. *medRxiv*. 2022 Jan 25;2022.01.24.22269755.
- Murray SM, Barbanti M, **Campbell C**, Brown A, Chen LI, Dhanapal JI, et al. Impaired humoral and cellular response to primary COVID-19 vaccination in patients less than 2 years after allogeneic bone marrow transplant. *British Journal of Haematology*. 2022 Jun 22;00:1–12.
- Wang T, Smith DA, **Campbell C**, Mokaya J, Freeman O, Salih H, et al. Hepatitis B virus (HBV) viral load, liver and renal function in adults treated with tenofovir disoproxil fumarate (TDF) vs. untreated: a retrospective longitudinal UK cohort study. *BMC Infect Dis*. 2021 June;21(1):610.
- Mulder T, **Campbell CJ**, Ruxton GD. Evaluation of disruptive camouflage of avian cup-nests. *Ibis* (Lond 1859). 2020;163(1):150– 8.

6.5. Conclusion

I am thankful to have had the opportunities throughout my doctoral project to work in multidisciplinary teams of clinicians, statisticians, engineers, computer scientists and epidemiologists in order to efficiently utilise routinely collected electronic health data to address evidence gaps in HBV. I hope future investigations work towards eliminating HBV as a public health threat by 2023, in line with targets set by the WHO. There is a huge impetus for ongoing work to tackle a growing disease burden, and I aspire to continue working towards reducing this burden in my future endeavours.

7. References

1. World Health Organization. Hepatitis B Fact Sheet. [Internet]. 2022. Available from: https://www.who.int/en/news-room/fact-sheets/detail/hepatitis-b
2. Hepatitis B - FAQs, Statistics, Data, & Guidelines | CDC [Internet]. [cited 2023 Feb 16]. Available from: https://www.cdc.gov/hepatitis/hbv/index.htm
3. Guidelines for the prevention, care and treatment of persons with chronic hepatitis B infection [Internet]. Geneva; 2015 [cited 2023 Jan 13]. Available from: https://www.who.int/publications/i/item/WHO-HIV-2015-5
4. Fitzmaurice C, Akinyemiju T, Abera S, Ahmed M, Alam N, Alemayohu MA, et al. The burden of primary liver cancer and underlying etiologies from 1990 to 2015 at the global, regional, and national level results from the global burden of disease study 2015. JAMA Oncol [Internet]. 2017 Dec 1 [cited 2021 Apr 10];3(12):1683–91. Available from: /pmc/articles/PMC5824275/
5. Trépo C, Chan HLY, Lok A. Hepatitis B virus infection. The Lancet. 2014 Dec 6;384(9959):2053–63.
6. Rehermann B, Nascimbeni M. Immunology of hepatitis B virus and hepatitis C virus infection. Nature Reviews Immunology 2005 5:3 [Internet]. 2005 Mar [cited 2023 Feb 27];5(3):215–29. Available from: https://www.nature.com/articles/nri1573

7. Jiang Y, Han Q, Zhao H, Zhang J. The Mechanisms of HBV-Induced Hepatocellular Carcinoma. J Hepatocell Carcinoma [Internet]. 2021 May [cited 2023 Jan 13];8:435. Available from: /pmc/articles/PMC8147889/
8. National Institute for Health and Care Excellence. Hepatitis B (chronic): diagnosis and management (CG165). 2017;2016(June):1–36. Available from: https://www.nice.org.uk/guidance/cg165
9. Downs LO, Smith DA, Lumley SF, Patel M, McNaughton AL, Mokaya J, et al. Electronic health informatics data to describe clearance dynamics of hepatitis B surface antigen (HBsAg) and e antigen (HBeAg) in chronic hepatitis B virus infection. mBio [Internet]. 2019 May 1 [cited 2023 Feb 27];10(3). Available from: https://journals.asm.org/doi/10.1128/mBio.00699-19
10. Chisenga CC, Musukuma K, Chilengi R, Zürcher S, Munamunungu V, Siyunda A, et al. Field performance of the Determine HBsAg point-of-care test for diagnosis of hepatitis B virus co-infection among HIV patients in Zambia. Journal of Clinical Virology. 2018 Jan 1;98:5–7.
11. Njai HF, Shimakawa Y, Sanneh B, Ferguson L, Ndow G, Mendy M, et al. Validation of Rapid Point-of-Care (POC) Tests for Detection of Hepatitis B Surface Antigen in Field and Laboratory Settings in the Gambia, Western Africa. J Clin Microbiol [Internet]. 2015 Apr 1 [cited 2023 Feb 26];53(4):1156–63. Available from: https://journals.asm.org/doi/10.1128/JCM.02980-14
12. Komas NP, Baï-Sepou S, Manirakiza A, Léal J, Béré A, le Faou A. The prevalence of hepatitis B virus markers in a cohort of students in Bangui, Central African Republic. BMC Infect Dis [Internet]. 2010 Jul 29 [cited 2023 Feb 26];10(1):1–6. Available from: https://link.springer.com/articles/10.1186/1471-2334-10-226
13. EASL. European Association for the Study of the Liver (EASL) 2017 Clinical Practice Guidelines on the management of hepatitis B virus infection. J

Hepatol [Internet]. 2017 [cited 2023 Feb 27];67(2):370–98. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28427875%0Ahttp://linkinghub.elsevier.com/retrieve/pii/S016882781730185X

14. Terrault NA, Lok ASF, McMahon BJ, Chang KM, Hwang JP, Jonas MM, et al. Update on prevention, diagnosis, and treatment of chronic hepatitis B: AASLD 2018 hepatitis B guidance. Hepatology [Internet]. 2018 [cited 2023 Feb 27];67(4):1560–99. Available from: https://aasldpubs.onlinelibrary.wiley.com/doi/10.1002/hep.29800
15. Sarin SK, Kumar • M, Lau • G K, Abbas • Z, Chan • H L Y, Chen • C J, et al. Asian-Pacific clinical practice guidelines on the management of hepatitis B: a 2015 update. Hepatol Int [Internet]. 2072 [cited 2023 Feb 27];10:1–98. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4722087/
16. Martyn E, Eisen S, Longley N, Harris P, Surey J, Norman J, et al. The forgotten people: Hepatitis B virus (HBV) infection as a priority for the inclusion health agenda. Elife. 2023 Feb 9;12.
17. Hepatitis B vaccine - NHS [Internet]. [cited 2023 Dec 20]. Available from: https://www.nhs.uk/conditions/vaccinations/hepatitis-b-vaccine/
18. Public Health England. Acute hepatitis B: guide to national enhanced surveillance [Internet]. [cited 2023 Dec 17]. Available from: https://www.gov.uk/government/publications/acute-hepatitis-b-guide-to-national-enhanced-surveillance
19. Hepatitis B: clinical and public health management [Internet]. 2019 [cited 2023 Dec 17]. Available from: https://www.gov.uk/guidance/hepatitis-b-clinical-and-public-health-management
20. UKHSA. Blood, tissue and organ donors: surveillance schemes [Internet]. 2022 [cited 2024 Jan 6]. Available from: https://www.gov.uk/guidance/blood-tissue-and-organ-donors-surveillance-schemes
21. UK Health Security Agency. NHS Infectious diseases in pregnancy screening programme handbook. [Internet]. 2023 [cited 2023 Dec 17]. Available from:

https://www.gov.uk/government/publications/infectious-diseases-in-pregnancy-screening-programme-handbook/nhs-infectious-diseases-in-pregnancy-screening-programme-handbook

22. Health Security Agency U. Integrated guidance on health clearance of healthcare workers and the management of healthcare workers living with bloodborne viruses (hepatitis B, hepatitis C and HIV). 2023.
23. BBVs in healthcare workers: health clearance and management - GOV.UK [Internet]. [cited 2023 Dec 17]. Available from: https://www.gov.uk/government/publications/bbvs-in-healthcare-workers-health-clearance-and-management
24. Global health sector strategies on, respectively, HIV, viral hepatitis and sexually transmitted infections for the period 2022-2030. . Geneva; 2022.
25. Ramsay - Health Security Agency MU. The Green Book on Immunisation - Chapter 18 Hepatitis B. 2022 [cited 2023 Feb 27]; Available from: https://www.gov.uk/government/publications/national-antenatal-infections-
26. UK Health Security Agency [Internet]. 2023 [cited 2023 Jul 6]. Quarterly vaccination coverage statistics for children aged up to 5 years in the UK (COVER programme): July to September 2022. Available from: https://www.gov.uk/government/statistics/cover-of-vaccination-evaluated-rapidly-cover-programme-2022-to-2023-quarterly-data/quarterly-vaccination-coverage-statistics-for-children-aged-up-to-5-years-in-the-uk-cover-programme-july-to-september-2022
27. Public Health England [Internet]. 2021. Quarterly vaccination coverage statistics for children aged up to five years in the UK (COVER programme): January to March 2020.
28. UK Health Security Agency [Internet]. 2022. Quarterly vaccination coverage statistics for children aged up to 5 years in the UK (COVER programme): January to March 2021.

29. Public Health England [Internet]. 2020. Quarterly vaccination coverage statistics for children aged up to five years in the UK (COVER programme): January to March 2019.
30. Hepatitis B in England: Working to eliminate hepatitis B as a public health threat. 2023.
31. Public Health England [Internet]. 2019 [cited 2024 Jan 23]. Infectious diseases in pregnancy screening programme: ISOSS. Available from: https://www.gov.uk/government/publications/integrated-screening-outcomes-surveillance-service-isoss/infectious-diseases-in-pregnancy-screening-programme-isoss
32. Guidance on the hepatitis B antenatal screening and selective neonatal immunisation pathway. [Internet]. 2023 [cited 2024 Jan 23]. Available from: https://www.gov.uk/government/publications/hepatitis-b-antenatal-screening-and-selective-neonatal-immunisation-pathway/guidance-on-the-hepatitis-b-antenatal-screening-and-selective-neonatal-immunisation-pathway--2#the-purpose-of-the-guidelines
33. UK Health Security Agency [Internet]. 2023 [cited 2024 Jan 23]. Infectious diseases in pregnancy screening standards valid for data collected from 1 April 2023 . Available from: https://www.gov.uk/government/publications/infectious-diseases-in-pregnancy-screening-programme-standards/infectious-diseases-in-pregnancy-screening-standards-valid-for-data-collected-from-1-april-2018#idps-s06-diagnosisintervention-timely-specialist-assessment-of-women-with-hepatitis-b-newly-diagnosed-or-higher-infectivity
34. UK Health Security Agency [Internet]. 2021 [cited 2024 Jan 23]. Protecting your baby against hepatitis B guide. Available from: https://www.gov.uk/government/publications/protecting-your-baby-against-hepatitis-b-leaflet/protecting-your-baby-against-hepatitis-b-with-the-hepatitis-b-vaccine

35. Public Health England [Internet]. ImmForm - manual input guidance. .
36. NHS Digital [Internet]. 2022 [cited 2024 Jan 23]. Appendix D – Data collection, validation and quality. Childhood Vaccination Coverage Statistics- England, 2021-22. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/nhs-immunisation-statistics/2021-22/appendices#appendix-d-data-collection-validation-and-quality
37. Laboratory reporting to UKHSA: A guide for diagnostic laboratories Reporting to UKHSA: a guide for diagnostic laboratories. 2023.
38. UKHSA. HPR volume 17 issue 12: news (16 October 2023) [Internet]. 2023 [cited 2024 Jan 6]. Available from: https://www.gov.uk/government/publications/health-protection-report-volume-17-2023/hpr-volume-17-issue-12-news-16-october-2023
39. UK Health Security Agency [Internet]. 2023 [cited 2024 Jan 23]. Shooting Up: infections and other injecting-related harms among people who inject drugs in the UK, data to end of 2021. Available from: https://www.gov.uk/government/publications/shooting-up-infections-among-people-who-inject-drugs-in-the-uk/shooting-up-infections-and-other-injecting-related-harms-among-people-who-inject-drugs-in-the-uk-data-to-end-of-2021
40. Sheena BS, Hiebert L, Han H, Ippolito H, Abbasi-Kangevari M, Abbasi-Kangevari Z, et al. Global, regional, and national burden of hepatitis B, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Gastroenterol Hepatol [Internet]. 2022 Sep 1 [cited 2022 Dec 6];7(9):796–829. Available from: http://www.thelancet.com/article/S2468125322001248/fulltext
41. Campbell C, Wang T, Burrow R, Mandal S, Hippisley-Cox J, Barnes E, et al. Estimating the epidemiology of chronic Hepatitis B Virus (HBV) infection in the UK: what do we know and what are we missing? Wellcome Open Res [Internet]. 2022 Jan 1 [cited 2023 Mar 19];7:203–203. Available from: https://europepmc.org/articles/PMC9975428

42. Ahmad AA, Falla AM, Duffell E, Noori T, Bechini A, Reintjes R, et al. Estimating the scale of chronic hepatitis B virus infection among migrants in EU/EEA countries. BMC Infect Dis [Internet]. 2018 Jan 11 [cited 2024 Jan 10];18(1). Available from: https://pubmed.ncbi.nlm.nih.gov/29325525/
43. Public Health England [Internet]. 2020 [cited 2024 Jan 23]. Hepatitis B virus (HBV): regional reports. Available from: https://www.gov.uk/government/publications/hepatitis-b-virus-hbv-regional-reports
44. Razavi-Shearer D, Gamkrelidze I, Nguyen MH, Chen DS, van Damme P, Abbas Z, et al. Global prevalence, treatment, and prevention of hepatitis B virus infection in 2016: a modelling study. Lancet Gastroenterol Hepatol [Internet]. 2018 Jun 1 [cited 2022 Feb 23];3(6):383–403. Available from: https://pubmed.ncbi.nlm.nih.gov/29599078/
45. British Liver Trust. Hepatitis B report [Internet]. 2017 [cited 2023 Feb 23]. Available from: https://britishlivertrust.org.uk/wp-content/uploads/Hep-B-website.pdf
46. Ferreira G, Stuurman AL, Horsmans Y, Cattaert T, Verstraeten T, Feng Y, et al. Hepatitis B virus infection and the risk of liver disease progression in type 2 diabetic patients with potential nonalcoholic fatty liver disease. Eur J Gastroenterol Hepatol [Internet]. 2020 Jan 1 [cited 2020 Jun 30];32(1):101–9. Available from: http://journals.lww.com/00042737-202001000-00017
47. Campbell C, Wang T, Gillespie I, Barnes E, Matthews PC. Analysis of primary care electronic health record data of people living with hepatitis B virus: infection and hepatocellular carcinoma risk associated with socio-economic deprivation. Public Health. 2024 Jan 1;226:215–27.
48. Trickey A, Bivegete S, Duffell E, McNaughton AL, Nerlander L, Walker JG, et al. Estimating hepatitis B virus prevalence among key population groups for European Union and European Economic Area countries and the United

Kingdom: a modelling study. BMC Infect Dis [Internet]. 2023 Dec 1 [cited 2024 Jan 9];23(1). Available from: /pmc/articles/PMC10331985/

49. By the numbers: counting the costs of infectious illness. Nature. 2021 Oct 28;598(7882):S18–9.

50. Total burden related to hepatitis B — Cause | Institute for Health Metrics and Evaluation [Internet]. [cited 2022 Dec 27]. Available from: https://www.healthdata.org/results/gbd_summaries/2019/total-burden-related-to-hepatitis-b-level-3-cause

51. Stanaway JD, Flaxman AD, Naghavi M, Fitzmaurice C, Vos T, Abubakar I, et al. The global burden of viral hepatitis from 1990 to 2013: findings from the Global Burden of Disease Study 2013. The Lancet [Internet]. 2016 Sep 10 [cited 2021 Apr 10];388(10049):1081–8. Available from: /pmc/articles/PMC5100695/

52. Liver disease profiles, October 2022 update - GOV.UK [Internet]. [cited 2023 Jan 4]. Available from: https://www.gov.uk/government/statistics/liver-disease-profiles-october-2022-update/liver-disease-profiles-october-2022-update

53. Global Health Sector Strategy on viral hepatitis 2016–2021. [Internet]. Geneva; 2016 [cited 2021 Jun 14]. Available from: https://apps.who.int/iris/bitstream/handle/10665/246177/WHO-HIV-2016.06-eng.pdf?sequence=1&isAllowed=y

54. Introduction to healthcare technology: The healthcare ecosystem [Internet]. 2021 Jul [cited 2023 Feb 27]. Available from: https://digital.nhs.uk/developer/guides-and-documentation/introduction-to-healthcare-technology/the-healthcare-ecosystem#primary-care

55. Baker C. Population estimates & GP registers: why the difference? [Internet]. [cited 2023 Jan 11]. Available from: https://commonslibrary.parliament.uk/population-estimates-gp-registers-why-the-difference/

56. Safeguarding patient data | CPRD [Internet]. [cited 2023 Jan 11]. Available from: https://cprd.com/safeguarding-patient-data
57. Choose if data from your health records is shared for research and planning - NHS [Internet]. [cited 2023 Jan 11]. Available from: https://www.nhs.uk/your-nhs-data-matters/
58. Trends and clinical characteristics of COVID-19 vaccine recipients: a federated analysis of 57.9 million patients' primary care records in situ using OpenSAFELY | Enhanced Reader.
59. Notice under Regulation 3(4) of the Health Service (Control of Patient Information) Regulations 2002 - GOV.UK [Internet]. [cited 2023 Jan 11]. Available from: https://www.gov.uk/government/publications/covid-19-notification-to-gps-and-nhs-england-to-share-information/notice-under-regulation-34-of-the-health-service-control-of-patient-information-regulations-2002
60. Read Codes - NHS Digital [Internet]. [cited 2023 Jan 11]. Available from: https://digital.nhs.uk/services/terminology-and-classifications/read-codes
61. CPRD linked data [Internet]. [cited 2023 Jan 12]. Available from: https://cprd.com/cprd-linked-data
62. Data - QResearch [Internet]. [cited 2023 Jan 11]. Available from: https://www.qresearch.org/data/
63. What is OpenSAFELY? | Bennett Institute for Applied Data Science [Internet]. [cited 2023 Jan 11]. Available from: https://www.bennett.ox.ac.uk/blog/2020/10/what-is-opensafely/
64. Hippisley-Cox J. A Description of the 4th Version of the QRESEARCH Database .
65. QResearch: Information for Researchers [Internet]. [cited 2023 Feb 26]. Available from: https://www.qresearch.org/information/information-for-researchers/

66. 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 [Internet]. 2015 Feb 1 [cited 2023 Jan 12];65(631):54–5. Available from: https://bjgp.org/content/65/631/54

67. Data | CPRD [Internet]. [cited 2023 Jan 12]. Available from: https://cprd.com/data

68. Smith DA, Wang T, Freeman O, Crichton C, Salih H, Matthews PC, et al. National Institute for Health Research Health Informatics Collaborative: development of a pipeline to collate electronic clinical data for viral hepatitis research. BMJ Health Care Inform [Internet]. 2020 Nov 1 [cited 2023 Jan 12];27(3):e100145. Available from: https://informatics.bmj.com/content/27/3/e100145

69. OPCS Classification of Interventions and Procedures [Internet]. [cited 2023 Jan 12]. Available from: https://www.datadictionary.nhs.uk/supporting_information/opcs_classification_of_interventions_and_procedures.html

70. Health Informatics Collaborative – National Institute for Health Research [Internet]. [cited 2023 Jan 12]. Available from: https://hic.nihr.ac.uk/

71. Wang T, Smith DA, Campbell C, Freeman O, Moysova Z, Noble T, et al. Cohort Profile: The National Institute for Health Research Health Informatics Collaborative: Hepatitis B Virus (NIHR HIC HBV) research dataset. Int J Epidemiol [Internet]. 2022 Jun 16 [cited 2023 Jan 12];2022:1–11. Available from: https://academic.oup.com/ije/advance-article/doi/10.1093/ije/dyac127/6609198

72. Papatheodoridis G, Dalekos G, Sypsa V, Yurdaydin C, Buti M, Goulis J, et al. PAGE-B predicts the risk of developing hepatocellular carcinoma in Caucasians with chronic hepatitis B on 5-year antiviral therapy. J Hepatol [Internet]. 2016 Apr 1 [cited 2022 Jul 22];64(4):800–6. Available from: http://www.journal-of-hepatology.eu/article/S0168827815007953/fulltext

73. Yang HI, Yuen MF, Chan HLY, Han KH, Chen PJ, Kim DY, et al. Risk estimation for hepatocellular carcinoma in chronic hepatitis B (REACH-B): development and validation of a predictive score. Lancet Oncol [Internet]. 2011 [cited 2022 Jul 22];12(6):568–74. Available from: https://pubmed.ncbi.nlm.nih.gov/21497551/

74. Lee MH, Yang HI, Liu J, Batrla-Utermann R, Jen CL, Iloeje UH, et al. Prediction models of long-term Cirrhosis and hepatocellular carcinoma risk in chronic hepatitis B patients: Risk scores integrating host and virus profiles. Hepatology [Internet]. 2013 Aug [cited 2020 Jun 8];58(2):546–54. Available from: https://aasldpubs.onlinelibrary.wiley.com/doi/full/10.1002/hep.26385

75. Wong GLH, Wong VWS. Risk prediction of hepatitis B virus-related hepatocellular carcinoma in the era of antiviral therapy. World J Gastroenterol [Internet]. 2013 Oct 21 [cited 2020 Jun 8];19(39):6515–22. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3801362/

76. Yuen MF, Tanaka Y, Fong DYT, Fung J, Wong DKH, Yuen JCH, et al. Independent risk factors and predictive score for the development of hepatocellular carcinoma in chronic hepatitis B. J Hepatol [Internet]. 2009 Jan [cited 2020 Jun 8];50(1):80–8. Available from: https://pubmed.ncbi.nlm.nih.gov/18977053/

77. Chan HLY, Hui AY, Wong ML, Tse AML, Hung LCT, Wong VWS, et al. Genotype C hepatitis B virus infection is associated with an increased risk of hepatocellular carcinoma. Gut [Internet]. 2004 Oct [cited 2020 Jun 8];53(10):1494–8. Available from: https://gut.bmj.com/content/53/10/1494

78. Chan HLY, Tse CH, Mo F, Koh J, Wong VWS, Wong GLH, et al. High viral load and hepatitis B virus subgenotype Ce are associated with increased risk of hepatocellular carcinoma. Journal of Clinical Oncology [Internet]. 2008 Jan 10 [cited 2020 Jun 8];26(2):177–82. Available from: https://pubmed.ncbi.nlm.nih.gov/18182659/

79. Wong VWS, Chan SL, Mo F, Chan TC, Loong HHF, Wong GLH, et al. Clinical scoring system to predict hepatocellular carcinoma in chronic hepatitis B carriers. Journal of Clinical Oncology [Internet]. 2010 Apr 1 [cited 2020 Jun 8];28(10):1660–5. Available from: https://pubmed.ncbi.nlm.nih.gov/20194845/

80. Campbell C, Wang T, McNaughton AL, Barnes E, Matthews PC. Risk factors for the development of hepatocellular carcinoma (HCC) in chronic hepatitis B virus (HBV) infection: a systematic review and meta-analysis. J Viral Hepat [Internet]. 2021 Mar 1 [cited 2022 Mar 3];28(3):493–507. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/jvh.13452

81. Simon TG, King LY, Chong DQ, Nguyen LH, Ma Y, VoPham T, et al. Diabetes, metabolic comorbidities, and risk of hepatocellular carcinoma: Results from two prospective cohort studies. Hepatology. 2018 May 1;67(5):1797–806.

82. Yang WS, Va P, Bray F, Gao S, Gao J, Li HL, et al. The role of pre-existing diabetes mellitus on hepatocellular carcinoma occurrence and prognosis: A meta-analysis of prospective cohort studies. PLoS One. 2011 Dec 21;6(12):e27326.

83. Hassan MM, Curley SA, Li D, Kaseb A, Davila M, Abdalla EK, et al. Association of diabetes duration and diabetes treatment with the risk of hepatocellular carcinoma. Cancer. 2010 Apr 15;116(8):1938–46.

84. Davila JA, Morgan RO, Shaib Y, McGlynn KA, El-Serag HB. Diabetes increases the risk of hepatocellular carcinoma in the United States: A population based case control study. Gut. 2005 Apr 1;54(4):533–9.

85. Turati F, Talamini R, Pelucchi C, Polesel J, Franceschi S, Crispo A, et al. Metabolic syndrome and hepatocellular carcinoma risk. Br J Cancer. 2013 Jan 15;108(1):222–8.

86. Borena W, Strohmaier S, Lukanova A, Bjørge T, Lindkvist B, Hallmans G, et al. Metabolic risk factors and primary liver cancer in a prospective study of 578,700 adults. Int J Cancer. 2012 Jul 1;131(1):193–200.

87. Global Report on Diabetes. Geneva; 2016.
88. GBD Chronic Kidney Disease Collaboration. Global, regional, and national burden of chronic kidney disease, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2020 Feb 13;0(0).
89. Forouzanfar MH, Liu P, Roth GA, Ng M, Biryukov S, Marczak L, et al. Global burden of hypertension and systolic blood pressure of at least 110 to 115mmHg, 1990-2015. JAMA - Journal of the American Medical Association. 2017 Jan 10;317(2):165–82.
90. Bansilal S, Castellano JM, Fuster V. Global burden of CVD: Focus on secondary prevention of cardiovascular disease. Int J Cardiol. 2015 Dec 1;201:S1–7.
91. Liu J, Ayada I, Zhang X, Wang L, Li Y, Wen T, et al. Estimating Global Prevalence of Metabolic Dysfunction-Associated Fatty Liver Disease in Overweight or Obese Adults. Clin Gastroenterol Hepatol [Internet]. 2022 Mar 1 [cited 2023 Jul 4];20(3):e573–82. Available from: https://pubmed.ncbi.nlm.nih.gov/33618024/
92. Teng ML, Ng CH, Huang DQ, Chan KE, Tan DJ, Lim WH, et al. Global incidence and prevalence of nonalcoholic fatty liver disease. Clin Mol Hepatol [Internet]. 2023 Feb 28 [cited 2023 Jul 4];29(Suppl):S32. Available from: /pmc/articles/PMC10029957/
93. Cheemerla S, Balakrishnan M. Global Epidemiology of Chronic Liver Disease. Clin Liver Dis (Hoboken) [Internet]. 2021 May 1 [cited 2023 Jul 4];17(5):365–70. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/cld.1061
94. Chan KE, Koh TJL, Tang ASP, Quek J, Yong JN, Tay P, et al. Global Prevalence and Clinical Characteristics of Metabolic-associated Fatty Liver Disease: A Meta-Analysis and Systematic Review of 10 739 607 Individuals. J Clin Endocrinol Metab [Internet]. 2022 Sep 1 [cited 2023 Jul 4];107(9):2691–700. Available from: https://pubmed.ncbi.nlm.nih.gov/35587339/

95. Saraei P, Asadi I, Kakar MA, Moradi-Kor N. The beneficial effects of metformin on cancer prevention and therapy: A comprehensive review of recent advances. Cancer Manag Res. 2019;11:3295–313.
96. Chae YK, Arya A, Malecek MK, Shin DS, Carneiro B, Chandra S, et al. Repurposing metformin for cancer treatment: Current clinical studies. Oncotarget. 2016;7(26):40767–80.
97. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ. 2009 Jul 21;339.
98. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: A practical tutorial. Evid Based Ment Health. 2019 Nov 1;22(4):153–60.
99. Wells G, Shea B, O'Connell D, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses [Internet]. [cited 2020 Jun 17]. Available from: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp
100. Goh MJ, Sinn DH, Kim S, Woo SY, Cho H, Kang W, et al. Statin Use and the Risk of Hepatocellular Carcinoma in Patients With Chronic Hepatitis B. Hepatology. 2020 Jun 14;71(6):2023–32.
101. Kim SU, Seo YS, Lee HA, Kim MN, Kim EH, Kim HY, et al. Validation of the CAMD Score in Patients With Chronic Hepatitis B Virus Infection Receiving Antiviral Therapy. Clinical Gastroenterology and Hepatology. 2020 Mar 1;18(3):693-699.e1.
102. Lim CT, Goh GBB, Li H, Lim TKH, Leow WQ, Wan WK, et al. Presence of Hepatic Steatosis Does Not Increase the Risk of Hepatocellular Carcinoma in Patients With Chronic Hepatitis B Over Long Follow-Up. Microbiol Insights. 2020 Jan;13:117863612091887.
103. Yang HI, Yeh ML, Wong GL et al. Real-World Effectiveness From the Asia Pacific Rim Liver Consortium for HBV Risk Score for the Prediction of

Hepatocellular Carcinoma in Chronic Hepatitis B Patients Treated With Oral Antiviral Therapy - PubMed. J Infect Dis. 2020;221(3):389–99.

104. Cheuk-Fung Yip T, Wai-Sun Wong V, Lik-Yuen Chan H, Tse YK, Pik-Shan Kong A, Long-Yan Lam K, et al. Effects of Diabetes and Glycemic Control on Risk of Hepatocellular Carcinoma After Seroclearance of Hepatitis B Surface Antigen. Clinical Gastroenterology and Hepatology. 2018 May 1;16(5):765-773.e2.
105. Brichler S, Nahon P, Zoulim F, Layese R, Bourcier V, Audureau E, et al. Non-virological factors are drivers of hepatocellular carcinoma in virosuppressed hepatitis B cirrhosis: Results of ANRS CO12 CirVir cohort. J Viral Hepat. 2019 Mar 1;26(3):384–96.
106. Cho H, Chang Y, Lee JH, Cho YY, Nam JY, Lee Y Bin, et al. Radiologic Nonalcoholic Fatty Liver Disease Increases the Risk of Hepatocellular Carcinoma in Patients with Suppressed Chronic Hepatitis B. J Clin Gastroenterol. 2019;
107. Shyu Y, Huang T, Chien C, Yeh C, Lin C, Chien R. Diabetes poses a higher risk of hepatocellular carcinoma and mortality in patients with chronic hepatitis B: A population-based cohort study. J Viral Hepat. 2019 Jun 5;26(6):718–26.
108. Tan Y, Zhang X, Zhang W, Tang L, Yang H, Yan K, et al. The influence of metabolic syndrome on the risk of hepatocellular carcinoma in patients with chronic hepatitis B infection in mainland China. Cancer Epidemiology Biomarkers and Prevention. 2019 Dec 1;28(12):2038–46.
109. Wang HW, Lai HC, Hu TH, Su WP, Lu SN, Lin CH, et al. Stratification of hepatocellular carcinoma risk through modified FIB-4 index in chronic hepatitis B patients on entecavir therapy. J Gastroenterol Hepatol. 2019 Feb 1;34(2):442–9.
110. Hsu YC, Yip TCF, Ho HJ, Wong VWS, Huang YT, El-Serag HB, et al. Development of a scoring system to predict hepatocellular carcinoma in

Asians on antivirals for chronic hepatitis B. J Hepatol. 2018 Aug 1;69(2):278–85.

111. Hsu YC, Ho HJ, Lee TY, Huang YT, Wu MS, Lin JT, et al. Temporal trend and risk determinants of hepatocellular carcinoma in chronic hepatitis B patients on entecavir or tenofovir. J Viral Hepat. 2018 May 1;25(5):543–51.
112. Kennedy K, Graham SM, Arora N, Shuhart MC, Kim HN. Hepatocellular carcinoma among US and non- US-born patients with chronic hepatitis B: Risk factors and age at diagnosis. PLoS One. 2018 Sep 1;13(9).
113. Kim K, Choi S, Park SM. Association of fasting serum glucose level and type 2 diabetes with hepatocellular carcinoma in men with chronic hepatitis B infection: A large cohort study. Eur J Cancer. 2018 Oct 1;102:103–13.
114. Li X, Xu H, Gao P. Diabetes Mellitus is a Risk Factor for Hepatocellular Carcinoma in Patients with Chronic Hepatitis B Virus Infection in China. Med Sci Monit. 2018 Sep 24;24:6729–34.
115. Chan AWH, Wong GLH, Chan HY, Tong JHM, Yu YH, Choi PCL, et al. Concurrent fatty liver increases risk of hepatocellular carcinoma among patients with chronic hepatitis B. J Gastroenterol Hepatol. 2017 Mar;32(3):667–76.
116. Chayanupatkul M, Omino R, Mittal S, Kramer JR, Richardson P, Thrift AP, et al. Hepatocellular carcinoma in the absence of cirrhosis in patients with chronic hepatitis B virus infection. J Hepatol. 2017 Feb 1;66(2):355–62.
117. Han H, Deng H, Han T, Zhao H, Hou F, Qi X. Association between hepatocellular carcinoma and type 2 diabetes mellitus in chinese hepatitis b virus cirrhosis patients: A case-control study. Medical Science Monitor. 2017 Jul 9;23:3324–34.
118. Kim JH, Sinn DH, Gwak GY, Kang W, Paik YH, Choi MS, et al. Insulin resistance and the risk of hepatocellular carcinoma in chronic hepatitis B patients. J Gastroenterol Hepatol. 2017 May;32(5):1100–6.

119. Mallet V, Hamed K, Schwarzinger M. Prognosis of patients with chronic hepatitis B in France (2008–2013): A nationwide, observational and hospital-based study. J Hepatol. 2017 Mar 1;66(3):514–20.
120. Shim JJ, Oh CH, Kim JW, Lee CK, Kim BH. Liver cirrhosis stages and the incidence of hepatocellular carcinoma in chronic hepatitis B patients receiving antiviral therapy. Scand J Gastroenterol. 2017 Sep 2;52(9):1029–36.
121. Yu MW, Lin CL, Liu CJ, Yang SH, Tseng YL, Wu CF. Influence of Metabolic Risk Factors on Risk of Hepatocellular Carcinoma and Liver-Related Death in Men With Chronic Hepatitis B: A Large Cohort Study. Gastroenterology. 2017 Oct 1;153(4):1006-1017.e5.
122. Lee J, Yoo SH, Sohn W, Kim HW, Choi YS, Won JH, et al. Obesity and hepatocellular carcinoma in patients receiving entecavir for chronic Hepatitis B. Clin Mol Hepatol. 2016;22(3):339–49.
123. Brouwer WP, van der Meer AJ, Boonstra A, Pas SD, de Knegt RJ, de Man RA, et al. The impact of PNPLA3 (rs738409 C>G) polymorphisms on liver histology and long-term clinical outcome in chronic hepatitis B patients. Liver International. 2015 Feb 1;35(2):438–47.
124. Fu SC, Huang YW, Wang TC, Hu JT, Chen DS, Yang SS. Increased risk of hepatocellular carcinoma in chronic hepatitis B patients with new onset diabetes: a nationwide cohort study. Aliment Pharmacol Ther. 2015 Jun;41(11):1200–9.
125. Hsiang JC, Gane EJ, Bai WW, Gerred SJ. Type 2 diabetes: A risk factor for liver mortality and complications in hepatitis B cirrhosis patients. J Gastroenterol Hepatol. 2015 Mar;30(3):591–9.
126. Hsu YC, Wu CY, Lane HY, Chang CY, Tai CM, Tseng CH, et al. Determinants of hepatocellular carcinoma in cirrhotic patients treated with nucleos(t)ide analogues for chronic hepatitis B. Journal of Antimicrobial Chemotherapy. 2014;69(7):1920–7.

127. Kim SS, Hwang JC, Lim SG, Ahn SJ, Cheong JY, Cho SW. Effect of Virological Response to Entecavir on the Development of Hepatocellular Carcinoma in Hepatitis B Viral Cirrhotic Patients: Comparison Between Compensated and Decompensated Cirrhosis. American Journal of Gastroenterology. 2014 Aug;109(8):1223–33.
128. Wu CY, Lin JT, Ho HJ, Su CW, Lee TY, Wang SY, et al. Association of nucleos(T)ide analogue therapy with reduced risk of hepatocellular carcinoma in patients with chronic hepatitis B - A nationwide cohort study. Gastroenterology. 2014;147(1).
129. Chen CT, Chen JY, Wang JH, Chang KC, Tseng PL, Kee KM, et al. Diabetes mellitus, metabolic syndrome and obesity are not significant risk factors for hepatocellular carcinoma in an HBV- and HCV-endemic area of Southern Taiwan. Kaohsiung Journal of Medical Sciences. 2013 Aug 1;29(8):451–9.
130. Gao C, Fang L, Zhao HC, Li JT, Yao SK. Potential role of diabetes mellitus in the progression of cirrhosis to hepatocellular carcinoma: A cross-sectional case-control study from Chinese patients with HBV infection. Hepatobiliary and Pancreatic Diseases International. 2013;12(4):385–93.
131. Li Q, Li WW, Yang X, Fan WB, Yu JH, Xie SS, et al. Type 2 diabetes and hepatocellular carcinoma: A case-control study in patients with chronic hepatitis B. Int J Cancer. 2012 Sep 1;131(5):1197–202.
132. Chao LT, Wu CF, Sung FY, Lin CL, Liu CJ, Huang CJ, et al. Insulin, glucose and hepatocellular carcinoma risk in male hepatitis B carriers: Results from 17-year follow-up of a population-based cohort. Carcinogenesis. 2011 Jun;32(6):876–81.
133. Walter SR, Thein HH, Gidding HF, Amin J, Law MG, George J, et al. Risk factors for hepatocellular carcinoma in a cohort infected with hepatitis B or C. J Gastroenterol Hepatol. 2011 Dec 1;26(12):1757–64.
134. Wang CS, Yao WJ, Chang TT, Wang ST, Chou P. The impact of type 2 diabetes on the development of hepatocellular carcinoma in different viral

hepatitis statuses. Cancer Epidemiology Biomarkers and Prevention. 2009 Jul;18(7):2054–60.

135. Chen CL, Yang HI, Yang WS, Liu CJ, Chen PJ, You SL, et al. Metabolic Factors and Risk of Hepatocellular Carcinoma by Chronic Hepatitis B/C Infection: A Follow-up Study in Taiwan. Gastroenterology. 2008 Jul;135(1):111–21.

136. Yu MW, Shih WL, Lin CL, Liu CJ, Jian JW, Tsai KS, et al. Body-mass index and progression of hepatitis B: A population-based cohort study in men. Journal of Clinical Oncology. 2008 Dec 1;26(34):5576–82.

137. Lai MS, Hsieh MS, Chiu YH, Chen THH. Type 2 diabetes and hepatocellular carcinoma: A cohort study in high prevalence area of hepatitis virus infection. Hepatology. 2006 Jun 1;43(6):1295–302.

138. Tan Y, Wei S, Zhang W, Yang J, Yang J, Yan L. Type 2 diabetes mellitus increases the risk of hepatocellular carcinoma in subjects with chronic hepatitis B virus infection: A meta-analysis and systematic review. Vol. 11, Cancer Management and Research. Dove Medical Press Ltd; 2019. p. 705–13.

139. Lai SW, Chen PC, Liao KF, Muo CH, Lin CC, Sung FC. Risk of hepatocellular carcinoma in diabetic patients and risk reduction associated with anti-diabetic therapy: A population-based cohort study. American Journal of Gastroenterology. 2012 Jan;107(1):46–52.

140. Veldt BJ, Chen W, Heathcote EJ, Wedemeyer H, Reichen J, Hofmann WP, et al. Increased risk of hepatocellular carcinoma among patients with hepatitis C cirrhosis and diabetes mellitus. Hepatology. 2008 Jun;47(6):1856–62.

141. Antonelli A, Ferri C, Fallahi P, Pampana A, Ferrari SM, Goglia F, et al. Hepatitis C virus infection: Evidence for an association with type 2 diabetes. Diabetes Care. 2005 Oct;28(10):2548–50.

142. Grimbert S, Valensi P, Lévy-Marchal C, Perret G, Richardet JP, Raffoux C, et al. High Prevalence of Diabetes Mellitus in Patients With Chronic Hepatitis

C. A Case-Control Study - PubMed. Gastroenterol Clin Biol. 1996;20(6–7):544–8.

143. Mehta SH, Brancati FL, Sulkowski MS, Strathdee SA, Szklo M, Thomas DL. Prevalence of type 2 diabetes mellitus among persons with hepatitis C virus infection in the United States. Ann Intern Med. 2000;133(8):592–9.

144. Mehta SH, Brancati FL, Strathdee SA, Pankow JS, Netski D, Coresh J, et al. Hepatitis C virus infection and incident type 2 diabetes. Hepatology. 2003 Jul 1;38(1):50–6.

145. Donadon V, Balbi M, Mas MD, Casarin P, Zanette G. Metformin and reduced risk of hepatocellular carcinoma in diabetic patients with chronic liver disease. Liver International. 2010 May 1;30(5):750–8.

146. Mantovani A, Targher G. Type 2 diabetes mellitus and risk of hepatocellular carcinoma: Spotlight on nonalcoholic fatty liver disease. Vol. 5, Annals of Translational Medicine. AME Publishing Company; 2017.

147. Di Bisceglie AM. What Every Hepatologist Should Know about Endocrinology: Obesity, Diabetes, and Liver Disease. Vol. 126, Gastroenterology. W.B. Saunders; 2004. p. 604–6.

148. Garcia-Compean D, Jacquez-Quintana JO, Gonzalez-Gonzalez JA, Maldonado-Garza H. Liver cirrhosis and diabetes: Risk factors, pathophysiology, clinical implications and management. Vol. 15, World Journal of Gastroenterology. Baishideng Publishing Group Inc; 2009. p. 280–8.

149. Moore MA, Park CB, Tsuda H. Implications of the Hyperinsulinaemia-Diabetes-Cancer Link for Preventive Efforts - PubMed. Eur J Cancer Prev. 1998 Apr;7(2):89–107.

150. Calle EE, Kaaks R. Overweight, obesity and cancer: Epidemiological evidence and proposed mechanisms. Vol. 4, Nature Reviews Cancer. Nature Publishing Group; 2004. p. 579–91.

151. Garg SK, Maurer H, Reed K, Selagamsetty R. Diabetes and cancer: Two diseases with obesity as a common risk factor. Vol. 16, Diabetes, Obesity and Metabolism. Blackwell Publishing Ltd; 2014. p. 97–110.
152. Park EJ, Lee JH, Yu GY, He G, Ali SR, Holzer RG, et al. Dietary and Genetic Obesity Promote Liver Inflammation and Tumorigenesis by Enhancing IL-6 and TNF Expression. Cell. 2010 Jan 22;140(2):197–208.
153. Ali Kamkar MM, Ahmad R, Alsmadi O, Behbehani K. Insight into the impact of diabetes mellitus on the increased risk of hepatocellular carcinoma: Mini-review. Vol. 13, Journal of Diabetes and Metabolic Disorders. BioMed Central Ltd.; 2014. p. 57.
154. Foretz M, Guigas B, Viollet B. Understanding the glucoregulatory mechanisms of metformin in type 2 diabetes mellitus. Vol. 15, Nature Reviews Endocrinology. Nature Publishing Group; 2019. p. 569–89.
155. He X, Esteva FJ, Ensor J, Hortobagyi GN, Lee MH, Yeung SCJ. Metformin and thiazolidinediones are associated with improved breast cancer-specific survival of diabetic women with HER2+ breast cancer. Ann Oncol. 2012 Jul;23(7):1771–80.
156. Lee JH, Kim T Il, Jeon SM, Hong SP, Cheon JH, Kim WH. The effects of metformin on the survival of colorectal cancer patients with diabetes mellitus. Vol. 131, International Journal of Cancer. John Wiley & Sons, Ltd; 2012. p. 752–9.
157. Evans JMM, Donnelly LA, Emslie-Smith AM, Alessi DR, Morris AD. Metformin and reduced risk of cancer in diabetic patients. Br Med J. 2005 Jun 4;330(7503):1304–5.
158. Landman GWD, Kleefstra N, Van Hateren KJJ, Groenier KH, Gans ROB, Bilo HJG. Metformin associated with lower cancer mortality in type 2 diabetes: Zodiac-16. Diabetes Care. 2010 Feb;33(2):322–6.
159. Higurashi T, Hosono K, Takahashi H, Komiya Y, Umezawa S, Sakai E, et al. Metformin for chemoprevention of metachronous colorectal adenoma or

polyps in post-polypectomy patients without diabetes: a multicentre double-blind, placebo-controlled, randomised phase 3 trial. Lancet Oncol. 2016 Apr 1;17(4):475–83.

160. The Metformin Active Surveillance Trial (MAST) Study - Full Text View - ClinicalTrials.gov [Internet]. [cited 2020 Jun 8]. Available from: https://www.clinicaltrials.gov/ct2/show/NCT01864096?term=metformin&recrs=abdf&type=Intr&cond=cancer&phase=23&draw=2&rank=6
161. A Phase III Randomized Trial of Metformin vs Placebo in Early Stage Breast Cancer - Full Text View - ClinicalTrials.gov [Internet]. [cited 2020 Jun 8]. Available from: https://www.clinicaltrials.gov/ct2/show/NCT01101438?term=metformin&recrs=abdf&type=Intr&cond=cancer&phase=23&draw=2&rank=3
162. Neoadjuvant Chemotherapy With or Without Metformin in Early Breast Cancer. - Full Text View - ClinicalTrials.gov [Internet]. [cited 2020 Jun 8]. Available from: https://www.clinicaltrials.gov/ct2/show/NCT04387630?term=metformin&recrs=abdf&type=Intr&cond=cancer&phase=23&draw=2&rank=5
163. U.S. National Library of Medicine. [Internet]. [cited 2020 Jun 8]. Available from: https://www.clinicaltrials.gov/ct2/home
164. Global hepatitis report. Geneva; 2017.
165. Yang JD, Hainaut P, Gores GJ, Amadou A, Plymoth A, Roberts LR. A global view of hepatocellular carcinoma: trends, risk, prevention and management. Vol. 16, Nature Reviews Gastroenterology and Hepatology. Nature Publishing Group; 2019. p. 589–604.
166. 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. Lancet Gastroenterol Hepatol. 2020 Mar 1;5(3):245–66.

167. Huang DQ, Hoang JK, Leong J, Riveiro-Barciela M, Maeda M, Yang JD, et al. Differential characteristics and outcomes of Asian and non-Asian patients with HBV-related hepatocellular carcinoma. Liver Int [Internet]. 2021 Aug 1 [cited 2023 Apr 21];41(8):1922–32. Available from: https://pubmed.ncbi.nlm.nih.gov/33713386/

168. Yu MW, Lin CL, Liu CJ, Wu WJ, Hu JT, Huang YW. Metabolic-Associated Fatty Liver Disease, Hepatitis B Surface Antigen Seroclearance, and Long-Term Risk of Hepatocellular Carcinoma in Chronic Hepatitis B. Cancers (Basel) [Internet]. 2022 Dec 1 [cited 2023 Apr 21];14(23). Available from: https://pubmed.ncbi.nlm.nih.gov/36497492/

169. Shah SHBU, Alavi M, Hajarizadeh B, Matthews G V., Martinello M, Danta M, et al. Trends in decompensated cirrhosis and hepatocellular carcinoma among people with a hepatitis B notification in New South Wales. JHEP Rep [Internet]. 2022 Oct 1 [cited 2023 Apr 21];4(10). Available from: https://pubmed.ncbi.nlm.nih.gov/36119722/

170. Yun B, Ahn SH, Yoon JH, Kim BK. Clinical Indication of Aspirin Associated With Reduced Risk of Liver Cancer in Chronic Hepatitis B: A Nationwide Cohort Study. Am J Gastroenterol [Internet]. 2022 May 1 [cited 2023 Apr 21];117(5):758–68. Available from: https://pubmed.ncbi.nlm.nih.gov/35294425/

171. Mak LY, Hui RWH, Lee CH, Mao XH, Cheung KS, Wong DKH, et al. Glycemic burden and the risk of adverse hepatic outcomes in patients with chronic hepatitis B with type 2 diabetes. Hepatology [Internet]. 2023 [cited 2023 Apr 21];77(2). Available from: https://pubmed.ncbi.nlm.nih.gov/36130882/

172. Lee Y Bin, Moon H, Lee JH, Cho EJ, Yu SJ, Kim YJ, et al. Association of Metabolic Risk Factors With Risks of Cancer and All-Cause Mortality in Patients With Chronic Hepatitis B. Hepatology [Internet]. 2021 Jun 1 [cited 2023 Apr 21];73(6):2266–77. Available from: https://pubmed.ncbi.nlm.nih.gov/33140415/

173. Yun B, Ahn SH, Oh J, Yoon JH, Kim BK. Effect of metabolic dysfunction-associated fatty liver disease on liver cancer risk in a population with chronic hepatitis B virus infection: A nationwide study. Hepatol Res [Internet]. 2022 Dec 1 [cited 2023 Apr 21];52(12):975–84. Available from: https://pubmed.ncbi.nlm.nih.gov/35976670/

174. Liver cancer statistics | Cancer Research UK [Internet]. [cited 2023 Jan 19]. Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/liver-cancer#heading-One

175. Yang HI, Yuen MF, Chan HLY, Han KH, Chen PJ, Kim DY, et al. Risk estimation for hepatocellular carcinoma in chronic hepatitis B (REACH-B): Development and validation of a predictive score. Lancet Oncol [Internet]. 2011 [cited 2020 Jun 8];12(6):568–74. Available from: https://pubmed.ncbi.nlm.nih.gov/21497551/

176. Sarri G, Westby M, Bermingham S, Hill-Cawthorne G, Thomas H. Diagnosis and management of chronic hepatitis B in children, young people, and adults: summary of NICE guidance. BMJ [Internet]. 2013 [cited 2023 Apr 4];346:f3893. Available from: https://pubmed.ncbi.nlm.nih.gov/23804177/

177. Hahné S, Ramsay M, Balogun K, Edmunds WJ, Mortimer P. Incidence and routes of transmission of hepatitis B virus in England and Wales, 1995-2000: Implications for immunisation policy. Journal of Clinical Virology [Internet]. 2004 [cited 2023 Apr 4];29(4):211–20. Available from: https://pubmed.ncbi.nlm.nih.gov/15018847/

178. [ARCHIVED CONTENT] Getting ahead of the curve: a strategy for combating infectious diseases (including other aspects of health protection) : Department of Health - Publications [Internet]. [cited 2023 Feb 23]. Available from: https://webarchive.nationalarchives.gov.uk/ukgwa/+/www.dh.gov.uk/en/Publi

cationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4007697

179. Hippisley-Cox J. QResearch Protocol for an Ethical High Quality General Practice Database for Research. 2021;(9.0).

180. Characteristics of Chronic Hepatitis B associated with Cirrhosis and Cancer [Internet]. [cited 2023 Feb 23]. Available from: https://www.qresearch.org/research/approved-research-programs-and-projects/characteristics-of-chronic-hepatitis-b-associated-with-cirrhosis-and-cancer/

181. Office for National Statistics. Population estimates for the UK, England and Wales, Scotland and Northern Ireland: mid-2016 [Internet]. 2017 [cited 2022 Mar 29]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2016

182. Office for National Statistics [Internet]. 2020 [cited 2024 Jan 14]. Population estimates for the UK, England and Wales, Scotland and Northern Ireland: mid-2019. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2019estimates

183. Office for National Statistics [Internet]. 2020 [cited 2023 Feb 23]. Population estimates for the UK, England and Wales, Scotland and Northern Ireland: mid-2020. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2016

184. Lee MH, Yang HI, Liu J, Batrla-Utermann R, Jen CL, Iloeje UH, et al. Prediction models of long-term cirrhosis and hepatocellular carcinoma risk in

chronic hepatitis B patients: risk scores integrating host and virus profiles. Hepatology [Internet]. 2013 Aug [cited 2022 Jul 22];58(2):546–54. Available from: https://pubmed.ncbi.nlm.nih.gov/23504622/

185. Wong GLH, Wong VWS. Risk prediction of hepatitis B virus-related hepatocellular carcinoma in the era of antiviral therapy. World J Gastroenterol [Internet]. 2013 Oct 21 [cited 2022 Jul 22];19(39):6515–22. Available from: https://pubmed.ncbi.nlm.nih.gov/24151375/

186. Yuen MF, Tanaka Y, Fong DYT, Fung J, Wong DKH, Yuen JCH, et al. Independent risk factors and predictive score for the development of hepatocellular carcinoma in chronic hepatitis B. J Hepatol [Internet]. 2009 Jan [cited 2022 Jul 22];50(1):80–8. Available from: https://pubmed.ncbi.nlm.nih.gov/18977053/

187. Chan HLY, Hui AY, Wong ML, Tse AML, Hung LCT, Wong VWS, et al. Genotype C hepatitis B virus infection is associated with an increased risk of hepatocellular carcinoma. Gut [Internet]. 2004 Oct [cited 2022 Jul 22];53(10):1494–8. Available from: https://pubmed.ncbi.nlm.nih.gov/15361502/

188. Chan HLY, Tse CH, Mo F, Koh J, Wong VWS, Wong GLH, et al. High viral load and hepatitis B virus subgenotype ce are associated with increased risk of hepatocellular carcinoma. J Clin Oncol [Internet]. 2008 Jan 10 [cited 2022 Jul 22];26(2):177–82. Available from: https://pubmed.ncbi.nlm.nih.gov/18182659/

189. Wong VWS, Chan SL, Mo F, Chan TC, Loong HHF, Wong GLH, et al. Clinical scoring system to predict hepatocellular carcinoma in chronic hepatitis B carriers. J Clin Oncol [Internet]. 2010 Apr 1 [cited 2022 Jul 22];28(10):1660–5. Available from: https://pubmed.ncbi.nlm.nih.gov/20194845/

190. Lee YCA, Cohet C, Yang YC, Stayner L, Hashibe M, Straif K. Meta-analysis of epidemiologic studies on cigarette smoking and liver cancer. Int J Epidemiol

[Internet]. 2009 Aug 29 [cited 2023 Jul 4];38(6):1497–511. Available from: https://pubmed.ncbi.nlm.nih.gov/19720726/

191. Latest global data on cancer burden and alcohol consumption: More than 740 000 new cases of cancer in 2020 attributed to alcohol Questions and Answers (Q&A). 2021 [cited 2023 Jul 4]; Available from: https://doi.org/10.1016/S1470-2045
192. List of ethnic groups - GOV.UK [Internet]. [cited 2022 Aug 12]. Available from: https://www.ethnicity-facts-figures.service.gov.uk/style-guide/ethnic-groups#2001-census
193. A healthy lifestyle - WHO recommendations [Internet]. [cited 2022 Aug 12]. Available from: https://www.who.int/europe/news-room/fact-sheets/item/a-healthy-lifestyle---who-recommendations
194. Summary Analysis of Linked Cancer Data - QResearch [Internet]. [cited 2022 Feb 6]. Available from: https://www.qresearch.org/data/summary-analysis-of-linked-cancer-data/
195. Saatci D, Ranger TA, Garriga C, Clift AK, Zaccardi F, Tan PS, et al. Association Between Race and COVID-19 Outcomes Among 2.6 Million Children in England. JAMA Pediatr [Internet]. 2021 Sep 1 [cited 2022 Feb 6];175(9):928–38. Available from: https://jamanetwork.com/journals/jamapediatrics/fullarticle/2780966
196. Vinogradova Y, Coupland C, Hippisley-Cox J. Use of hormone replacement therapy and risk of venous thromboembolism: nested case-control studies using the QResearch and CPRD databases. BMJ [Internet]. 2019 Jan 9 [cited 2022 Feb 6];364:4810. Available from: https://www.bmj.com/content/364/bmj.k4810
197. Hippisley-Cox J, Young D, Coupland C, Channon KM, Tan PS, Harrison DA, et al. Risk of severe COVID-19 disease with ACE inhibitors and angiotensin receptor blockers: cohort study including 8.3 million people. Heart [Internet].

2020 Oct 1 [cited 2022 Feb 6];106(19):1503–11. Available from: https://heart.bmj.com/content/106/19/1503

198. Van Buuren S. Flexible Imputation of Missing Data. Second Edition. [Internet]. Boca Raton, FL: Chapman & Hall/CRC. ; 2018 [cited 2022 Aug 16]. Available from: https://stefvanbuuren.name/fimd/

199. Rubin DB. Multiple Imputation for Nonresponse in Surveys [Internet]. 1st ed. Vol. 1. Cambridge, Massachusetts: John Wiley & Sons, Ltd; 1987 [cited 2022 Feb 6]. i–xxix. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/9780470316696.fmatter

200. Marshall A, Billingham LJ, Bryan S. Can we afford to ignore missing data in cost-effectiveness analyses? Eur J Health Econ [Internet]. 2009 Feb [cited 2022 Feb 6];10(1):1–3. Available from: https://pubmed.ncbi.nlm.nih.gov/18936993/

201. Public Health England. Annual report from the sentinel surveillance of blood borne virus testing in England 2020: main report. 2021 Jul 28;15(13).

202. Public Health England. Antenatal screening standards: data report 1 April 2019 to 31 March 2020 - GOV.UK [Internet]. 2021 [cited 2023 Feb 23]. Available from: https://www.gov.uk/government/statistics/antenatal-screening-standards-data-report-2019-to-2020/antenatal-screening-standards-data-report-1-april-2019-to-31-march-2020

203. Public Health England. Hepatitis B: migrant health guide - GOV.UK [Internet]. 2017 [cited 2019 Nov 20]. Available from: https://www.gov.uk/guidance/hepatitis-b-migrant-health-guide

204. Hepatitis B and C testing: people at risk of infection (PH43) [Internet]. 2012 [cited 2023 Feb 23]. Available from: www.nice.org.uk/guidance/ph43

205. Summary of GP population characteristics - QResearch [Internet]. [cited 2020 Oct 3]. Available from: https://www.qresearch.org/data/summary-of-gp-population-characteristics/

206. Population estimates by ethnic group, England and Wales - Office for National Statistics [Internet]. [cited 2022 Jul 19]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/datasets/populationestimatesbyethnicgroupenglandandwales

207. Hippisley-Cox J, Coupland C, Brindle P. The performance of seven QPrediction risk scores in an independent external sample of patients from general practice: a validation study. BMJ Open [Internet]. 2014 Aug 1 [cited 2022 Aug 12];4(8):e005809. Available from: https://bmjopen.bmj.com/content/4/8/e005809

208. Schnier C, Wallace L, Tempelton K, Aitken C, Gunson RN, Molyneaux P, et al. Use of laboratory-based surveillance data to estimate the number of people chronically infected with hepatitis B living in Scotland. Epidemiol Infect [Internet]. 2014 [cited 2020 Oct 23];142(10):2121–30. Available from: https://pubmed.ncbi.nlm.nih.gov/24480044/

209. Harvala H, Reynolds C, Gibney Z, Derrick J, Ijaz S, Davison KL, et al. Hepatitis B infections among blood donors in England between 2009 and 2018: Is an occult hepatitis B infection a risk for blood safety? Transfusion (Paris) [Internet]. 2021 Aug 1 [cited 2023 Jan 10];61(8):2402–13. Available from: https://pubmed.ncbi.nlm.nih.gov/34114670/

210. Joint United Kingdom (UK) Blood Transfusion and Tissue Transplantation Services Professional Advisory Committee. Guidelines for the Blood Transfusion Services in the UK [Internet]. 2013 [cited 2023 Feb 20]. Available from: https://www.transfusionguidelines.org/red-book

211. Data sources and methods 2020: NHS Blood and Transplant and UK Health Security Agency. NHS Blood and Transplant [Internet]. 2021 Oct [cited 2023 Feb 20]; Available from: moz-extension://49067288-6c59-45ad-aa2a-28b5d147fe2c/enhanced-reader.html?openApp&pdf=https%3A%2F%2Fnhsbtdbe.blob.core.windows.

net%2Fumbraco-assets-corp%2F25124%2Fnhsbt-ukhsa-data-sources-and-methods-2020.pdf

212. Reynolds C, Davison KL, Brailsford SR. Safe supplies: few infections in UK blood and tissue donors. Transfusion Medicine [Internet]. 2019 Aug 1 [cited 2023 Feb 20];29(4):239–46. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/tme.12576

213. NHS England » Emergency department opt out testing for HIV, hepatitis B and hepatitis C: The first 100 days [Internet]. [cited 2023 Jan 16]. Available from: https://www.england.nhs.uk/long-read/emergency-department-opt-out-testing-for-hiv-hepatitis-b-and-hepatitis-c-the-first-100-days/

214. Harris RJ, Harris HE, Mandal S, Ramsay M, Vickerman P, Hickman M, et al. Monitoring the hepatitis C epidemic in England and evaluating intervention scale-up using routinely collected data. J Viral Hepat [Internet]. 2019 May 1 [cited 2023 Feb 23];26(5):541. Available from: https://pubmed.ncbi.nlm.nih.gov/30663179/

215. Presanis AM, Kirwan PD, Jackson CH, de Angelis D, Health England P, D Kirwan UP, et al. Trends in undiagnosed HIV prevalence in England and implications for eliminating HIV transmission by 2030: an evidence synthesis model. Lancet Public Health [Internet]. 2021 Oct 1 [cited 2023 Feb 23];6(10):e739–51. Available from: https://www.thelancet.com/journals/lanpub/article/PIIS2468-2667(21)00142-0/fulltext

216. Harris RJ, Thomas B, Griffiths J, Costella A, Chapman R, Ramsay M, et al. Increased uptake and new therapies are needed to avert rising hepatitis C-related end stage liver disease in England: Modelling the predicted impact of treatment under different scenarios. J Hepatol [Internet]. 2014 Sep 1 [cited 2024 Jan 16];61(3):530–7. Available from: http://www.journal-of-hepatology.eu/article/S0168827814003183/fulltext

217. Sweeting MJ, De Angelis D, Hickman M, Ades AE. Estimating hepatitis C prevalence in England and Wales by synthesizing evidence from multiple data sources. Assessing data conflict and model fit. Biostatistics [Internet]. 2008 Oct 1 [cited 2024 Jan 16];9(4):715–34. Available from: https://dx.doi.org/10.1093/biostatistics/kxn004

218. Shao YHJ, Chan TS, Tsai K, Wu SY. Association between proton pump inhibitors and the risk of hepatocellular carcinoma. Aliment Pharmacol Ther [Internet]. 2018 Aug 1 [cited 2022 Jul 20];48(4):460–8. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/apt.14835

219. Kim S, Jeong S, Park SJ, Chang J, Choi S, Cho Y, et al. Association between Proton Pump Inhibitor Use and Risk of Hepatocellular Carcinoma: A Korean Nationally Representative Cohort Study. Journal of Clinical Medicine 2022, Vol 11, Page 2865 [Internet]. 2022 May 19 [cited 2022 Jul 20];11(10):2865. Available from: https://www.mdpi.com/2077-0383/11/10/2865/htm

220. Song HJ, Jiang X, Henry L, Nguyen MH, Park H. Proton pump inhibitors and risk of liver cancer and mortality in patients with chronic liver disease: a systematic review and meta-analysis. European Journal of Clinical Pharmacology 2020 76:6 [Internet]. 2020 Mar 14 [cited 2022 Jul 20];76(6):851–66. Available from: https://link.springer.com/article/10.1007/s00228-020-02854-8

221. Wang J, Wu Y, Bi Q, Zheng X, Zhang J, Huang W. Adverse outcomes of proton pump inhibitors in chronic liver disease: a systematic review and meta-analysis. Hepatology International 2020 14:3 [Internet]. 2020 Jan 7 [cited 2022 Jul 20];14(3):385–98. Available from: https://link.springer.com/article/10.1007/s12072-019-10010-3

222. Kao WY, Su CW, Chia-Hui Tan E, Lee PC, Chen PH, Tang JH, et al. Proton Pump Inhibitors and Risk of Hepatocellular Carcinoma in Patients With Chronic Hepatitis B or C. Hepatology [Internet]. 2019 Mar 1 [cited 2022 Jul

20];69(3):1151–64. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/hep.30247

223. Lin CL, Kao JH. The clinical implications of hepatitis B virus genotype: Recent advances. J Gastroenterol Hepatol [Internet]. 2011 [cited 2022 Jul 22];26(SUPPL. 1):123–30. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1440-1746.2010.06541.x

224. Lin CL, Kao JH. Hepatitis B Virus Genotypes and Variants. Cold Spring Harb Perspect Med [Internet]. 2015 May 1 [cited 2022 Jul 22];5(5):a021436. Available from: http://perspectivesinmedicine.cshlp.org/content/5/5/a021436.full

225. Wong GLH, Chan HLY, Yiu KKL, Lai JWY, Chan VKK, Cheung KKC, et al. Meta-analysis: the association of hepatitis B virus genotypes and hepatocellular carcinoma. Aliment Pharmacol Ther [Internet]. 2013 Mar 1 [cited 2022 Jul 22];37(5):517–26. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/apt.12207

226. Araujo NM, Teles SA, Spitz N. Comprehensive Analysis of Clinically Significant Hepatitis B Virus Mutations in Relation to Genotype, Subgenotype and Geographic Region. Front Microbiol. 2020 Dec 14;11:3241.

227. Goh MJ, Sinn DH, Kim S, Woo SY, Cho H, Kang W, et al. Statin Use and the Risk of Hepatocellular Carcinoma in Patients With Chronic Hepatitis B. Hepatology [Internet]. 2020 Jun 1 [cited 2022 Jul 22];71(6):2023–32. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/hep.30973

228. Facciorusso A, El Aziz MAA, Singh S, Pusceddu S, Milione M, Giacomelli L, et al. Statin Use Decreases the Incidence of Hepatocellular Carcinoma: An Updated Meta-Analysis. Cancers 2020, Vol 12, Page 874 [Internet]. 2020 Apr 3 [cited 2022 Jul 22];12(4):874. Available from: https://www.mdpi.com/2072-6694/12/4/874/htm

229. Islam MM, Poly TN, Walther BA, Yang HC, Li YC. Statin Use and the Risk of Hepatocellular Carcinoma: A Meta-Analysis of Observational Studies. Cancers 2020, Vol 12, Page 671 [Internet]. 2020 Mar 13 [cited 2022 Jul 22];12(3):671. Available from: https://www.mdpi.com/2072-6694/12/3/671/htm

230. Kim G, Jang SY, Nam CM, Kang ES. Statin use and the risk of hepatocellular carcinoma in patients at high risk: A nationwide nested case-control study. J Hepatol. 2018 Mar 1;68(3):476–84.

231. Udompap P, Kim WR. Development of Hepatocellular Carcinoma in Patients With Suppressed Viral Replication: Changes in Risk Over Time. Clin Liver Dis (Hoboken) [Internet]. 2020 Feb 1 [cited 2022 Nov 7];15(2):85. Available from: /pmc/articles/PMC7098665/

232. Huang DQ, Lim SG. Hepatitis B: Who to treat? A critical review of international guidelines. Liver International [Internet]. 2020 Feb 1 [cited 2022 Nov 7];40(S1):5–14. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/liv.14365

233. Tan DJH, Ng CH, Tay PWL, Syn N, Muthiah MD, Lim WH, et al. Risk of Hepatocellular Carcinoma With Tenofovir vs Entecavir Treatment for Chronic Hepatitis B Virus: A Reconstructed Individual Patient Data Meta-analysis. JAMA Netw Open [Internet]. 2022 Jun 29 [cited 2022 Nov 7];5(6). Available from: https://pubmed.ncbi.nlm.nih.gov/35767258/

234. Campbell C, Wang T, Smith DA, Freeman O, Noble T, Várnai KA, et al. Impact of the COVID-19 pandemic on routine surveillance for adults with chronic hepatitis B virus (HBV) infection in the UK. Wellcome Open Research 2022 7:51 [Internet]. 2022 Feb 11 [cited 2022 Feb 16];7:51. Available from: https://wellcomeopenresearch.org/articles/7-51

235. Voulgaris T, Papatheodoridi M, Lampertico P, Papatheodoridis G V. Clinical utility of hepatocellular carcinoma risk scores in chronic hepatitis B. Liver

International [Internet]. 2020 Mar 1 [cited 2022 Aug 17];40(3):484–95. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/liv.14334

236. National Institute for Health and Care Excellence. Cirrhosis in over 16s: assessment and management. 2016.

237. European Association for the Study of the Liver. EASL 2017 Clinical Practice Guidelines on the management of hepatitis B virus infection. J Hepatol. 2017;67:370–98.

238. Wang G, Duan Z. Guidelines for Prevention and Treatment of Chronic Hepatitis B. J Clin Transl Hepatol [Internet]. 2021 Oct 10 [cited 2023 Jun 15];9(5):769. Available from: /pmc/articles/PMC8516840/

239. Zhang S, Wang C, Liu B, Lu Q Bin, Shang J, Zhou Y, et al. Cost-effectiveness of expanded antiviral treatment for chronic hepatitis B virus infection in China: an economic evaluation. Lancet Reg Health West Pac [Internet]. 2023 [cited 2023 May 15];0(0). Available from: http://www.thelancet.com/article/S2666606523000561/fulltext

240. Chudasama Y V., Gillies CL, Zaccardi F, Coles B, Davies MJ, Seidu S, et al. Impact of COVID-19 on routine care for chronic diseases: A global survey of views from healthcare professionals. Diabetes & Metabolic Syndrome: Clinical Research & Reviews. 2020 Sep 1;14(5):965–7.

241. Richards M, Anderson M, Carter P, Ebert BL, Mossialos E. The impact of the COVID-19 pandemic on cancer care. Nature Cancer 2020 1:6 [Internet]. 2020 May 20 [cited 2023 May 9];1(6):565–7. Available from: https://www.nature.com/articles/s43018-020-0074-y

242. Mayor S. COVID-19: impact on cancer workforce and delivery of care. Lancet Oncol [Internet]. 2020 May 1 [cited 2023 May 9];21(5):633. Available from: http://www.thelancet.com/article/S1470204520302400/fulltext

243. Sud A, Jones ME, Broggio J, Loveday C, Torr B, Garrett A, et al. Collateral damage: the impact on outcomes from cancer surgery of the COVID-19 pandemic. Annals of Oncology. 2020 Aug 1;31(8):1065–74.

244. Wosik J, Clowse MEB, Overton R, Adagarla B, Economou-Zavlanos N, Cavalier J, et al. Impact of the COVID-19 pandemic on patterns of outpatient cardiovascular care. Am Heart J. 2021 Jan 1;231:1–5.
245. NHS Data Model and Dictionary [Internet]. [cited 2023 May 9]. Available from: https://www.datadictionary.nhs.uk/
246. National data opt-out - NHS Digital [Internet]. [cited 2023 May 9]. Available from: https://digital.nhs.uk/services/national-data-opt-out
247. Public Health England [Internet]. 2020 [cited 2024 Jan 16]. COVID-19: impact on STIs, HIV and viral hepatitis. Available from: https://www.gov.uk/government/publications/covid-19-impact-on-stis-hiv-and-viral-hepatitis
248. Mitchell HD, Vilaplana TG, Mandal S, Ratna N, Glancy M, Shah A, et al. Effects of COVID-19 Pandemic Response on Service Provision for Sexually Transmitted Infections, HIV, and Viral Hepatitis, England. Emerg Infect Dis [Internet]. 2022 Mar 1 [cited 2024 Jan 16];28(3):739. Available from: /pmc/articles/PMC8888218/
249. Tan M, Bhadoria AS, Cui F, Tan A, Van Holten J, Easterbrook P, et al. Estimating the proportion of people with chronic hepatitis B virus infection eligible for hepatitis B antiviral treatment worldwide: a systematic review and meta-analysis. Lancet Gastroenterol Hepatol. 2021 Feb 1;6(2):106–19.
250. Chien RN;, Liaw YF, Bartosch B, Chien RN, Liaw YF. Current Trend in Antiviral Therapy for Chronic Hepatitis B. Viruses 2022, Vol 14, Page 434 [Internet]. 2022 Feb 21 [cited 2023 May 15];14(2):434. Available from: https://www.mdpi.com/1999-4915/14/2/434/htm
251. Timeline of UK government coronavirus lockdowns and restrictions | Institute for Government [Internet]. [cited 2023 May 15]. Available from: https://www.instituteforgovernment.org.uk/data-visualisation/timeline-coronavirus-lockdowns

252. 2 Years of COVID-19 on GOV.UK - Government Digital Service [Internet]. [cited 2023 May 15]. Available from: https://gds.blog.gov.uk/2022/07/25/2-years-of-covid-19-on-gov-uk/
253. The impact of Covid-19 on the use of digital technology in the NHS | Nuffield Trust [Internet]. [cited 2023 May 15]. Available from: https://www.nuffieldtrust.org.uk/research/the-impact-of-covid-19-on-the-use-of-digital-technology-in-the-nhs
254. Specialty guides for patient management during the coronavirus pandemic Clinical guide for the management of remote consultations and remote working in secondary care during the coronavirus pandemic. 2020;
255. Sowah L, Chiou C. Impact of Coronavirus Disease 2019 Pandemic on Viral Hepatitis Elimination: What Is the Price? AIDS Res Hum Retroviruses [Internet]. 2021 Aug 1 [cited 2023 May 15];37(8):585–8. Available from: https://pubmed.ncbi.nlm.nih.gov/33913731/
256. Wang T, Glampson B, Mercuri L, Papadimitriou D, Jones CR, Smith DA, et al. Identifying Hepatocellular Carcinoma from imaging reports using natural language processing to facilitate data extraction from electronic patient records. medRxiv [Internet]. 2022 Aug 24 [cited 2022 Sep 23];2022.08.23.22279119. Available from: https://www.medrxiv.org/content/10.1101/2022.08.23.22279119v1
257. Cochrane A, Evlampidou I, Irish C, Ingle SM, Hickman M. Hepatitis B infection prevalence by country of birth in migrant populations in a large UK city. J Clin Virol [Internet]. 2015 Jul 1 [cited 2021 Dec 14];68:79–82. Available from: https://pubmed.ncbi.nlm.nih.gov/26071342/
258. Evlampidou I, Hickman M, Irish C, Young N, Oliver I, Gillett S, et al. Low hepatitis B testing among migrants: a cross-sectional study in a UK city. British Journal of General Practice [Internet]. 2016 Jun 1 [cited 2021 Dec 14];66(647):e382–91. Available from: https://bjgp.org/content/66/647/e382

259. Mendy ME, McConkey SJ, Sande Van Der MAB, Crozier S, Kaye S, Jeffries D, et al. Changes in viral load and HBsAg and HBeAg status with age in HBV chronic carriers in The Gambia. Virol J [Internet]. 2008 [cited 2021 Dec 14];5. Available from: https://pubmed.ncbi.nlm.nih.gov/18416832/

260. Chu CM, Liaw YF. Chronic hepatitis B virus infection acquired in childhood: special emphasis on prognostic and therapeutic implication of delayed HBeAg seroconversion. J Viral Hepat [Internet]. 2007 Mar [cited 2021 Dec 14];14(3):147–52. Available from: https://pubmed.ncbi.nlm.nih.gov/17305879/

261. Glampson B, Brittain J, Kaura A, Mulla A, Mercuri LD, Brett SJ, et al. Assessing COVID-19 Vaccine Uptake and Effectiveness Through the North West London Vaccination Program: Retrospective Cohort Study. JMIR Public Health Surveill 2021;7(9):e30010 https://publichealth.jmir.org/2021/9/e30010 [Internet]. 2021 Sep 17 [cited 2023 Jun 23];7(9):e30010. Available from: https://publichealth.jmir.org/2021/9/e30010

262. McDonald HI, Tessier E, White JM, Woodruff M, Knowles C, Bates C, et al. Early impact of the coronavirus disease (COVID-19) pandemic and physical distancing measures on routine childhood vaccinations in England, January to April 2020. Eurosurveillance [Internet]. 2020 May 14 [cited 2022 Feb 14];25(19):2000848. Available from: https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2020.25.19.2000848

263. Masresha BG, Luce R, Jr, Shibeshi ME, Ntsama B, N'Diaye A, et al. The performance of routine immunization in selected African countries during the first six months of the COVID-19 pandemic. Pan Afr Med J [Internet]. 2020 [cited 2022 Feb 14];37(Suppl 1):12. Available from: /pmc/articles/PMC7733346/

264. Chandir S, Siddiqi DA, Setayesh H, Khan AJ. Impact of COVID-19 lockdown on routine immunisation in Karachi, Pakistan. Lancet Glob Health. 2020 Sep 1;8(9):e1118–20.
265. Desilva MB, Haapala J, Vazquez-Benitez G, Daley MF, Nordin JD, Klein NP, et al. Association of the COVID-19 Pandemic With Routine Childhood Vaccination Rates and Proportion Up to Date With Vaccinations Across 8 US Health Systems in the Vaccine Safety Datalink. JAMA Pediatr [Internet]. 2022 Jan 1 [cited 2022 Feb 14];176(1):68–77. Available from: https://jamanetwork.com/journals/jamapediatrics/fullarticle/2784888
266. Aizawa Y, Katsuta T, Sakiyama H, Tanaka-Taya K, Moriuchi H, Saitoh A. Changes in childhood vaccination during the coronavirus disease 2019 pandemic in Japan. Vaccine. 2021 Jun 29;39(29):4006–12.
267. Ackerson BK, Sy LS, Glenn SC, Qian L, Park CH, Riewerts RJ, et al. Pediatric vaccination during the COVID-19 pandemic. Pediatrics [Internet]. 2021 Jul 1 [cited 2022 Feb 14];148(1). Available from: /pediatrics/article/148/1/e2020047092/179954/Pediatric-Vaccination-During-the-COVID-19-Pandemic
268. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol [Internet]. 2015 Jun 1 [cited 2018 Dec 27];44(3):827–36. Available from: https://academic.oup.com/ije/article-lookup/doi/10.1093/ije/dyv098
269. Tetanus | Vaccine Knowledge [Internet]. [cited 2021 Dec 15]. Available from: https://vk.ovg.ox.ac.uk/vk/tetanus
270. MenB Vaccine (Meningococcal B Vaccine) | Vaccine Knowledge [Internet]. [cited 2021 Dec 15]. Available from: https://vk.ovg.ox.ac.uk/vk/menb-vaccine
271. Padmanabhan S. CPRD GOLD Data Specification. 2017;
272. Medicines & Healthcare products Regulatory Agency. CPRD GOLD Glossary of terms/Data definitions v2.0.

273. Denominator data | CPRD [Internet]. [cited 2023 Jul 23]. Available from: https://cprd.com/denominator-data
274. CPRD GOLD Data Specification.
275. Sample Size Software | Power Analysis Software | PASS | NCSS.com [Internet]. [cited 2022 Apr 20]. Available from: https://www.ncss.com/software/pass/
276. Vaccine uptake guidance and the latest coverage data - GOV.UK [Internet]. [cited 2022 Apr 6]. Available from: https://www.gov.uk/government/collections/vaccine-uptake
277. Coverage Statistics 2020-21 - NHS Digital [Internet]. [cited 2022 Apr 6]. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/nhs-immunisation-statistics/england---2020-21/coverage-statistics#-6-in-1-vaccine-and-5-in-1-vaccine
278. COVID-19: impact on childhood vaccinations: data to August 2021 - GOV.UK [Internet]. [cited 2023 May 30]. Available from: https://www.gov.uk/government/publications/covid-19-impact-on-childhood-vaccinations-data-to-august-2021
279. Childhood Vaccination Coverage Statistics - 2020-21 - NDRS [Internet]. [cited 2023 May 30]. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/nhs-immunisation-statistics/england---2020-21
280. Childhood Vaccination Coverage Statistics - England - 2019-20 - NDRS [Internet]. [cited 2023 May 30]. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/nhs-immunisation-statistics/england---2019-20
281. Maintaining routine immunisation programmes during COVID-19 General principles. 2020;

8. Appendix

Table S1. PRISMA checklist.

Section/topic	#	Checklist item	Reported in section
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	Abstract
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	Abstract
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	1.
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	1.
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	N/A
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	2.1
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	2.1
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	2.1, supplement
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	2.1
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	2.2
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	2.2
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	2.3
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	2.2

Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency ($e.g., I^2$) for each meta-analysis.	2.2

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	N/A
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	2.2
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	3.1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	Supplement
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	3.2, Supplement
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	3.3, 3.4
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	3.3.1
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	N/A
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	3.3.2
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	4.1
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	4.2, 4.3, 4.6
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

Table S2. Main tabulation of hepatitis B virus, tetanus and meningitis B vaccination uptake from years 2015 to mid-2020:

Vaccinations	Overall	2015 Jan-Jun/ Annual	2015 Jul-Dec/ Annual	2016 Jan-Jun/ Annual	2016 Jul-Dec/ Annual	2017 Jan-Jun/ Annual	2017 Jul-Dec/ Annual	2018 Jan-Jun/ Annual	2018 Jul-Dec/ Annual	2019 Jan-Jun/ Annual	2019 Jul-Dec/ Annual	2020 Jan-Jun/ Annual
Infants (n,%)	46399	7120	2504	7359	2427	6530	2127	6209	1776	5400	1178	3769
Hepatitis B virus vaccination (n,%)												
0 records	26165 (57.3)	6973 (99.1)	2447 (98.9)	7170 (98.6)	2352 (98.4)	6116 (95.1)	530 (25.3)	220 (3.6)	25 (1.4)	191 (3.6)	12 (1.0)	129 (3.5)
1 record	950 (2.1)	35 (0.5)	17 (0.7)	83 (1.1)	25 (1.0)	252 (3.9)	135 (6.5)	141 (2.3)	31 (1.8)	119 (2.2)	27 (2.3)	85 (2.3)
2 records	1387 (3.0)	23 (0.3)	10 (0.4)	15 (0.2)	12 (0.5)	54 (0.8)	168 (8.0)	346 (5.7)	85 (4.9)	345 (6.5)	105 (9.1)	224 (6.1)
3 records	17064 (37.4)	6 (0.1)	<5	7 (0.1)	<5	10 (0.2)	1252 (59.8)	5355 (88.0)	1587 (91.6)	4615 (87.2)	1005 (87.2)	3224 (87.5)
>3 records	79 (0.2)	<5	<5	<5	0 (0.0)	0 (0.0)	8 (0.4)	20 (0.3)	<5	21 (0.4)	<5	22 (0.6)
Tetanus vaccinations (n,%)												
0 records	43 (0.1)	<5	<5	<5	<5	7 (0.1)	<5	7 (0.1)	<5	8 (0.2)	<5	7 (0.2)

1 record	482 (1.1)	47 (0.7)	24 (1.0)	56 (0.8)	20 (0.8)	62 (1.0)	27 (1.3)	86 (1.4)	23 (1.3)	70 (1.3)	25 (2.2)	42 (1.1)
2 records	1157 (2.5)	94 (1.3)	36 (1.5)	126 (1.7)	44 (1.8)	127 (2.0)	46 (2.2)	165 (2.7)	40 (2.3)	238 (4.5)	59 (5.1)	182 (4.9)
3 records	25961 (56.9)	1982 (28.2)	1302 (52.6)	2573 (35.4)	1588 (66.4)	2749 (42.7)	1616 (77.2)	3089 (50.8)	1648 (95.2)	4927 (93.1)	1062 (92.1)	3425 (93.0)
>3 records	18002 (39.4)	4912 (69.8)	1112 (44.9)	4517 (62.1)	739 (30.9)	3487 (54.2)	400 (19.1)	2735 (45.0)	18 (1.0)	48 (0.9)	6 (0.5)	28 (0.8)
Meningitis B vaccinations (n,%)												
0 records	4118 (9.0)	3999 (56.8)	18 (0.7)	29 (0.4)	5 (0.2)	23 (0.4)	5 (0.2)	12 (0.2)	<5	15 (0.3)	2 (0.2)	7 (0.2)
1 record	1328 (2.9)	459 (6.5)	60 (2.4)	130 (1.8)	47 (2.0)	130 (2.0)	51 (2.4)	151 (2.5)	47 (2.7)	109 (2.1)	45 (3.9)	99 (2.7)
2 record	5584 (12.2)	1610 (22.9)	263 (10.6)	531 (7.3)	280 (11.7)	542 (8.4)	244 (11.7)	569 (9.4)	309 (17.8)	609 (11.5)	252 (21.9)	375 (10.2)
>2 records	34615 (75.8)	969 (13.8)	2134 (86.2)	6585 (90.5)	2059 (86.1)	5737 (89.2)	1793 (85.7)	5350 (88.0)	1373 (79.3)	4558 (86.1)	854 (74.1)	3203 (86.9)

Table S3. Tabulation of full hepatitis B virus vaccination uptake across year of birth, region and ethnicity.

	2015	**2016**	**2017**	**2018**	**2019**	**2020**
Full hepatitis b virus vaccination (%)	0.1	0.1	14.9	89.1	87.6	88.1
Region ‡ (%)						
East of England	0.0	0.0	3.0	94.9	97.6	NA
London	0.1	0.0	16.3	85.4	84.6	91.5
North East	0	0	0	0	0	0
North West	0.0	0.0	26.5	91.2	94.3	92.0
Northern Ireland	0.2	0.0	3.5	94.9	95.8	94.3
Scotland	0.1	0.1	7.9	75.4	73.2	73.0
South East	0.1	0.1	29.3	93.5	86.2	90.8
South West	0.0	0.0	22.0	91.3	89.7	NA
Wales	0.1	0.2	6.6	91.1	93.8	94.9
West Midlands	0.0	0.0	19.2	92.3	89.7	89.8
Yorkshire	0.0	0.0	NA	NA	NA	NA
Ethnicity ‡ (%)						
Black	0.0	0.0	12.7	87.4	73.1	90.0
Chinese	0.0	0.0	31.3	81.3	71.4	100.0

Indeterminant*	0.0	0.0	24.1	88.0	94.4	50.0
Missing	0.1	0.1	14.1	88.0	85.8	85.6
Mixed	0.0	0.0	15.5	91.0	93.6	94.7
Other	0.7	0.0	24.0	89.2	80.3	82.8
South Asian	0.5	0.4	15.1	95.1	83.8	96.8
White	0.0	0.0	14.9	84.1	84.1	85.2

* Individuals with record of two different ethnic categories.

Table S4. Tabulation of full tetanus vaccination uptake across year of birth, region and ethnicity.

	2015	**2016**	**2017**	**2018**	**2019**	**2020**
Full tetanus vaccination (%)	52.3	97.8	97.5	97.3	97.3	97.1
Region ‡ (%)						
East of England	97.2	94.8	94.1	94.9	97.6	NA
London	96.4	95.3	94.6	94.4	94.0	91.5
North East	0	0	0	0	0	0
North West	97.3	96.6	96.0	94.0	95.3	92.0
Northern Ireland	98.3	99.2	95.5	96.9	96.5	96.1
Scotland	95.3	94.9	92.7	90.1	85.4	85.3
South East	97.6	96.8	97.3	94.7	91.6	95.4
South West	97.5	96.9	94.8	94.3	93.7	NA
Wales	96.5	96.7	96.6	95.8	96.5	96.6
West Midlands	97.1	95.7	97.6	94.8	90.7	91.8
Yorkshire	99.1	96.6	NA	NA	NA	NA
Ethnicity ‡ (%)						
Black	97.5	96.6	92.4	92.6	86.6	90.0
Chinese	100.0	100.0	100.0	93.8	85.7	100.0

Indeterminant*	100.0	95.7	89.7	100.0	94.4	100.0
Missing	96.2	96.1	95.1	94.0	91.6	91.2
Mixed	97.4	97.0	96.5	95.9	95.2	96.3
Other	95.9	94.9	97.3	91.9	90.2	86.2
South Asian	96.2	95.3	96.4	96.7	91.9	100.0
White	97.2	96.3	95.2	92.7	91.6	91.3

* Individuals with record of two different ethnic categories.

Table S5. Tabulation of full meningitis B vaccination uptake across year of birth, region and ethnicity.

	2015	**2016**	**2017**	**2018**	**2019**	**2020**
Full meningitis B vaccination (%)	97.9	97.4	96.8	95.9	93.8	93.7
Region ‡ (%)						
East of England	47.9	94.0	94.1	94.9	97.6	NA
London	62.9	95.5	94.2	94.7	94.5	91.5
North East	NA	NA	NA	NA	NA	NA
North West	52.0	95.7	95.7	94.9	96.1	93.1
Northern Ireland	35.6	98.0	95.5	97.5	97.5	97.2
Scotland	39.8	96.6	95.1	93.8	93.2	92.2
South East	65.9	97.2	97.7	95.0	95.7	95.4
South West	66.2	97.1	95.0	96.1	95.2	NA
Wales	42.6	96.9	97.1	96.4	97.8	98.0
West Midlands	63.5	95.7	98.0	95.0	92.9	91.8
Yorkshire	81.9	98.3	NA	NA	NA	NA
Ethnicity ‡ (%)						
Black	96.0	94.1	95.8	88.1	90.0	96.0
Chinese	100.0	100.0	93.8	92.9	100.0	100.0

Indeterminant*	51.5	95.7	93.1	100.0	100.0	100.0
Missing	50.7	96.2	95.7	95.1	95.0	95.1
Mixed	55.9	97.3	96.8	96.8	97.6	97.1
Other	53.1	93.4	96.0	93.2	93.4	93.1
South Asian	59.3	96.6	97.9	98.4	94.6	100.0
White	50.9	97.2	96.3	94.6	95.7	93.7

* Individuals with record of two different ethnic categories.

Table S6. Tabulation of full hepatitis B virus vaccination uptake by region across three time periods.

	Period A: Pre-routine introduction and pre-COVID	Period B: Post-routine introduction and pre-COVID	Period C: Post-routine introduction and post-COVID
Full hepatitis B virus vaccination (n,%)	46 (0.2)	13867 (88.3)	3230 (88.3)
Region ‡, n (%)			
East of England	0.0	96.5	NA
London	0.3	83.8	91.5
North East	NA	NA	NA
North West	0.1	92.4	93.0
Northern Ireland	0.1	95.2	94.2
Scotland	0.2	74.1	73.1
South East	0.2	91.4	92.2
South West	0.2	91.3	NA
Wales	0.1	91.9	95.1
West Midlands	0.0	91.6	89.8
Yorkshire	0.0	NA	NA
Ethnicity ‡, n (%)			
Black	0.2	79.8	90.0
Chinese	0.0	80.0	100.0

Indeterminant*	1.0	84.9	50.0
Missing	0.2	86.8	85.8
Mixed	0.1	91.6	95.1
Other	0.6	86.8	82.8
South Asian	0.3	88.8	96.8
White	0.1	84.3	85.4

* Individuals with record of two different ethnic categories.

Table S7. Tabulation of full tetanus vaccination uptake by region across three time periods.

	Period A: Pre-routine introduction and pre-COVID	Period B: Post-routine introduction and pre-COVID	Period C: Post-routine introduction and post-COVID
Full tetanus vaccination (n,%)	25608 (97.4)	14920 (95.0)	3435 (93.9)
Region ‡, n (%)			
East of England	95.4	96.5	NA
London	95.3	94.6	91.5
North East	NA	NA	NA
North West	96.7	94.6	93.0
Northern Ireland	97.8	96.7	96.1
Scotland	94.6	87.7	85.3
South East	97.3	94.1	96.9
South West	96.6	94.4	NA
Wales	96.7	96.0	96.8
West Midlands	96.6	93.4	91.8
Yorkshire	97.9	NA	NA
Ethnicity ‡, n (%)			
Black	96.0	90.2	90.0

Chinese	100.0	91.4	100.0
Indeterminant*	96.0	96.2	100.0
Missing	95.8	93.0	91.4
Mixed	97.0	95.7	96.7
Other	95.6	92.1	86.2
South Asian	95.8	94.8	100.0
White	96.4	92.4	91.4

* Individuals with record of two different ethnic categories.

Table S8. Tabulation of full meningitis B vaccination uptake by region across three time periods.

	Period A: Pre-routine introduction and pre-COVID	Period B: Post-routine introduction and pre-COVID	Period C: Post-routine introduction and post-COVID
Full meningitis B vaccination (n,%)	21362 (81.3)	15281 (97.3)	3556 (97.2)
Region ‡, n (%)			
East of England	77.8	96.5	NA
London	82.6	94.9	91.5
North East	NA	NA	NA
North West	80.3	95.4	93.6
Northern Ireland	75.2	97.4	97.1
Scotland	76.2	93.5	92.2
South East	85.3	95.7	96.9
South West	84.7	95.7	NA
Wales	78.4	97.0	98.2
West Midlands	84.9	94.4	91.8
Yorkshire	90.1	NA	NA
Ethnicity ‡, n (%)			
Black	78.4	92.3	90.0

Chinese	83.0	94.3	100.0
Indeterminant*	79.8	100.0	100.0
Missing	79.9	95.1	95.2
Mixed	81.9	97.2	97.5
Other	76.2	94.1	93.1
South Asian	84.1	96.7	100.0
White	80.0	95.1	93.8

* Individuals with record of two different ethnic categories.