BMJ Open

Epidemiology of liver disease and associated patient characteristics in Wales 2004-2022: a retrospective population-scale observational study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2024-093335
Article Type:	Original research
Date Submitted by the Author:	04-Sep-2024
Complete List of Authors:	Gao, Jingwei; Swansea University - Singleton Park Campus, Akbari, Ashley; Swansea University Medical School, Ahmed, Haroon; Cardiff University, Division of Population Medicine Davies, Aled; Cardiff University, PRIME Centre Wales Yeoman, Andrew; Royal Gwent Hospital Pembroke , Thomas ; University Hospital of Wales, Department of Gastroenterology and Hepatology
Keywords:	EPIDEMIOLOGY, Hepatology < INTERNAL MEDICINE, PUBLIC HEALTH

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Epidemiology of liver disease and associated patient characteristics in Wales 2004-2022: a retrospective population-scale observational study

Authors: Jingwei Gao [1], Ashley Akbari [1], Haroon Ahmed [2], Aled Davies [3], Andrew Yeoman [4], Thomas Peter Ignatius Pembroke [5]

Affiliations:

1 Population Data Science, Swansea University Medical School, Faculty of Medicine, Health & Life Science, Swansea University, Swansea, UK.

2 Division of Population Medicine, Cardiff University, Neuadd Meirionnydd, Heath Park, Cardiff, UK.

3 PRIME Centre Wales, Cardiff University, Neuadd Meirionnydd, Heath Park, Cardiff, UK.

4 Aneurin Bevan University Health Board, Newport, UK.

5 Department of Gastroenterology and Hepatology, University Hospital of Wales, Cardiff, UK.

Corresponding author:

Jingwei Gao

jingwei.gao@swansea.ac.uk

Population Data Science, Swansea University Medical School, Faculty of Medicine, Health & Life Science, Swansea University, Swansea, UK

ORCiD:

Jingwei Gao:	0000-0002-7722-6177
Ashley Akbari:	0000-0003-0814-0801
Haroon Ahmed:	0000-0002-0634-8548
Aled Davies:	0000-0002-7815-5155
Andrew Yeoman:	0000-0002-0739-3332
Thomas Peter Ignatius Pembroke:	0000-0002-2600-2034

Abstract

Objective

To describe the incidence and key demographic, socioeconomic and clinical characteristics of individuals with liver disease in Wales.

Design and setting

This study is designed as a retrospective observational study that linked data of anonymised identified individuals from primary, secondary care and mortality data from the Secure Anonymised Information Linkage (SAIL) Databank in Wales.

Participants

All Welsh residents who registered with a SAIL contributing general practitioner (GP) and diagnosed with liver disease from 2004 to 2022.

Primary and secondary outcome measures

Our primary outcome is the annual age-standardised incidence rate of liver disease. Secondary outcome is the numbers and frequencies of underlying aetiology and the associated comorbidities.

Results

Between 2004 and 2022, 111,098 individuals received a diagnosis of liver disease in Wales and were included in this study. The incidence of liver disease increased three-fold during the study period (97.7 per 100,000 inhabitants in 2004 to 316.2 per 100,000 inhabitants in 2022). A total of 79,992 individuals (72%) entered the cohort with the underlying aetiology of liver disease, including alcohol related liver disease (ArLD), non-alcoholic fatty liver disease (NAFLD), viral hepatitis, metabolic, hemochromatosis and autoimmune liver diseases. NAFLD has contributed to most of the change in incidence.

Conclusions

We observed increasing incidence rates of liver disease in Wales, with NAFLD showing a particularly sharp increase and frequently identified as an underlying condition. A better understanding of the

epidemiology of liver disease is the first step towards effective prevention, early detection and targeted intervention to improve patient outcomes.

Keywords Epidemiology; Hepatology; Non-alcoholic fatty liver disease hepatitis; Alcohol-related liver disease; Cirrhosis

Word count: 3613

Background

Liver disease is a significant global public health issue and a major contributor to morbidity and mortality [1]. Globally, cirrhosis and hepatocellular carcinoma account for an estimated two million deaths every year [2]. In the UK, liver disease has become the third most common cause of premature death [3], despite mortality rates for other major non-communicable diseases declining [4].

The management of chronic liver disease frequently involves lifestyle modification, including weight loss and reduced alcohol use, with the goal of reversing factors that can lead to disease progression [5]. As a result, public health policies in the UK have focussed on prevention and early detection, including the UK Government's Prevention Green Paper to promote the disease prevention [6], the National Institute for Health and Care Excellence guidance that focused on tackling obesity [7], and a series of policies to reduce alcohol-related harm [8]. In November 2022, the Welsh Government published a Quality Statement for Liver Disease [9], underlying the importance of the awareness of risk factors and early detection of liver disease, and set out its plans to promote the delivery of better quality, higher value and more accessible services for individuals with liver disease.

To effectively improve liver disease management, clinicians, researchers and policymakers must be aware of the epidemiology and clinical profile of the individuals with liver disease. However, a significant gap remains in the integration of primary, secondary, and mortality data, particularly across the different liver disease stages and aetiologies. To date, much of the epidemiological data is based on ICD-10 coding, which is largely derived from secondary care data sources and is therefore likely to underestimate the real-world incidence and prevalence of liver disease [10, 11]. This limitation currently precludes adequate prioritisation of research, targeting of interventions, and recruitment of individuals to clinical trials. Understanding the clinical and socio-demographic features associated with liver disease in a large-scale population is essential for the improvement of disease prevention and treatment.

The objective of our study is to describe the incidence of liver disease, as well as the key demographic and socioeconomic characteristics and the associated comorbidity of liver disease patients in Wales, as a first step towards improving capacity and capability for liver disease research.

Materials and method

Setting and data source

We used data from the Secure Anonymised Information Linkage (SAIL) Databank, which contains anonymised, individual-level linked electronic health record (EHR) data for Welsh population [12, 13, 14, 15]. The SAIL Databank includes complete secondary care data and primary care data covering approximately 86% of the Welsh population. These data reflect the demographic diversity of Wales

across age, sex, and levels of deprivation [16] and can be generalizable to the boarder UK population due to demographic similarities [17]. SAIL employs a split-file anonymisation process using National Health Service (NHS) number, name, sex, date of birth and postcode, ensuring confidentiality while enabling the linkage of individual-level data sources [13, 16, 18].

To provide a comprehensive overview of liver disease epidemiology, we combined linked primary care, hospital admissions and mortality data. Primary care data was accessed from the Welsh Longitudinal General Practice (WLGP) data, which currently uses the Read version 2 clinical coding system and collects event histories for people registered with a SAIL-supplying general practice in Wales. Hospital admission data, including in-patient admissions (emergency, elective and maternity) and day-care procedures, were collected from the Patient Episode Database for Wales (PEDW). Mortality data came from the Annual District Death Extract (ADDE) by the Office for National Statistics (ONS) death and contains the cause of death and contributory comorbidities. Both hospital admission and mortality data were coded using the International Classification of Diseases version 10 (ICD-10) system. We derived demographic and deprivation data from the Welsh Demographic Service Dataset (WDSD) and used the Welsh Index of Multiple Deprivation (WIMD) version 2019 quintiles to measure relative area-level deprivation based on geographical residential location from the Lower-layer Super Output Area (LSOA) version 2011.

Study population

We linked data from all individuals in the WLGP, PEDW, and ADDE data within the SAIL Databank using a unique anonymised individual identifier known as Anonymised Linkage Field (ALF). Individuals were extracted based on ALF and filtered for good data linkage status based on existing methodology [13, 18]. We excluded individuals not registered with a SAIL-contributing general practice (GP) or lacking residency information and identified those who received a liver disease diagnosis from 1st January 2004 to 31st December 2022. The first liver disease diagnosis is considered to be the index liver disease event. Individuals were required to be residents in Wales at the time of cohort entry. GP registration was required if the index event was from WLGP. All individuals were followed until the earliest of: GP de-registration, moving out of Wales, death, or the end of study period (2022.12.31).

Measurements

Definition of stages of liver disease and time of cohort entry

We applied a hierarchy of 3 tiers of potential aetiological diagnoses and 5 discrete stages of chronic liver disease based on perceived clinical importance and natural history of liver disease progression as

was described in our previous study [19]. The first stage is the underlying aetiological conditions without the presentation of cirrhosis and was divided into 3 tiers. Tier 1 aetiologies of liver disease include: alcohol-related liver disease (ArLD), non-alcoholic fatty liver disease (NAFLD), metabolic liver disease, hepatitis B virus (HBV), hepatitis C virus (HCV), autoimmune liver disease, haemochromatosis; tier 2 aetiologies include: unspecified hepatitis, congestive hepatopathy, and toxic liver disease, and tier 3 aetiologies were other miscellaneous diagnoses. As the disease progresses, various stages of liver disease were defined as follows: hepatic fibrosis and cirrhosis were categorised as stage 2, portal hypertension as stage 3, hepatic decompensation as stage 4, and hepatocellular carcinoma (HCC), intrahepatic cholangio carcinoma (ICC), and other primary liver cancers were classified under stage 5. Acute liver diseases were analysed separately from chronic liver diseases as they carry different challenges for primary and secondary care. Acute liver diseases were defined as conditions including acute viral hepatitis, Budd-Chiari syndrome, acute liver failure, infections and infarctions, and other unspecified acute liver injuries. A list of ICD-10 and Read v2 codes to identify individuals with liver disease can be found in Supplemental Table 1.

We defined the time of the first chronic or acute liver disease diagnosis as the index date. To assess the proportion of individuals with liver disease who presented late, the diagnoses were sequenced according to the natural history of the disease: aetiological diagnoses (stage 1), followed by cirrhosis, portal hypertension, decompensation and then HCC formation [19] (Supplemental Fig 1). To avoid duplicate counting, we divided the full cohort into four distinct groups: PEDW-only group representing hospital admission data, WLGP-only representing primary care data, ADDE-only group representing death records data, and those included in two or more data sources.

Comorbidities and drug prescription history

We collected data on demographic characteristics (age, sex, and WIMD 2019 quintile), with age divided by birth year in 10-year intervals, except for the 0-17 age group and 18-29 age group. Comorbidities including cardiovascular disease (CVD) related conditions (heart failure, transient ischaemic, other ischaemic disease, atrial fibrillation, peripheral vascular disease, angina, and stroke), diabetes, hypertension, and antihypertensive usage were recorded. These data were collected in time frames of 1 year, 1-3 years, 3-5 years, 5-10 years, and over 10 years prior to cohort entry. A list of ICD-10 and Read v2 codes to identify the associated conditions are available in our GitHub repository (URL: https://github.com/SwanseaUniversityDataScience/1492-LDCP).

Incidence of liver disease

The incidence rate was calculated as the number of incident cases per Welsh residents. An incident case was defined as an individual having a first liver-related diagnosis during the study period (01.01.2004-31.12.2022) with no prior liver disease history from 1st January 1994 to 31st December 2003. To determine the most clinically significant aetiology, we only considered the most advanced stage of diagnosis on incident event date. The Welsh population data was obtained from the ONS population estimates [20]. The incidence rate was directly standardised using the European Union (EU) standard population 2013 and were reported per 100,000 inhabitants in each calendar year from 2004 to 2022, with corresponding 95% confidence intervals (CI). Analysis was repeated for different data sources, liver disease aetiologies, and stages.

Study findings were reported in accordance with applicable reporting guidelines for observational studies using administrative data (Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) [21] and REporting of studies Conducted using Observational Routinely-collected Data (RECORD) [22]. Data cleaning, cohort assembly and statistical analyses were performed using Structured Query Language (IBM Db2 V.11.1) and R (V.4.1.0–V.4.1.3) within the SAIL Databank privacy-protecting trusted research environment (TRE).

Patient and public involvement

The Liver Research Cymru Patient Advisory group was established to provide PPI input into the development of research in Wales. The research questions and methodology addressed in this body of work were developed following discussion of their priorities based upon their lived experience of liver disease. The results were shared with this group and their feedback was sought in the interpretation and presentation of these data.

Results

Demographic and socioeconomic characteristics

We identified a total of 111,098 individuals with liver disease from PEDW, WLGP, and ADDE data sources from 2004.01.01 to 2022.12.31, contributing 441,885 person-years of follow-up (Fig 1A). Of the eligible individuals, 57,491 (51.7 %) were male. The median (interquartile range [IQR]) age was 59 (47-72). Approximately 50% of the individuals (55,787) entered the cohort after 2016 and 27,178 (24.5 %) individuals came from the most deprived areas. Of the 111,098 individuals, 86,941 (78.2 %) individuals were identified from PEDW, 42,783 (38.5%) from WLGP, and 18,122 (16.2%) from ADDE. Furthermore, 56,710 (51%) can only be accessed from PEDW, 20,319 (18.3%) exclusively from WLGP, 3459 (3.1%) exclusively from ADDE, and 30,610 (27.5 %) could be accessed in multiple data sources

(Fig 1B). The demographic variation is notable with a higher percentage of male and elderly individuals, and a higher degree of socioeconomic deprivation in the ADDE-only group (Table 1).

Incidence of chronic liver disease

We observed an increasing trend of liver disease from 2004 to 2022 in the Welsh population. The age-sex standardised incidence rate increased by 3 times during the 18 years of follow-up (97.7 per 100,000 inhabitants in 2004; 316.2 per 100,000 inhabitants in 2022, Table 2). This was significantly contributed by a 6.5 times increase in the incidence rate of the WLGP-only group, followed by 4.4 times increase in the PEDW-only group. The proportion of PEDW-only and WLGP-only incident cases increased by 14-16% during the past 2 decades (Fig 2A – 2B, Supplemental Table 2-1).

The incidence of liver disease has risen across all stages, with stage 2 (cirrhosis) and stage 3 diagnoses (portal hypertension) experiencing the most notable increase from 2004 to 2022 (stage 2 by 4 times; stage 3 by 3.9 times). However, the increases in the higher-stage liver disease cases were less pronounced, particularly for stage 4 (1.5 times). Consequently, the proportion of more advanced liver disease diagnoses decreased (stage 4 and stage 5: 8.6% in 2004, 5.3% in 2022, decreased by 1.62 times since 2004, Fig 3A – 3B, Supplemental Table 2-2).

We further analysed the stage 1 aetiologies by each diagnosis. We observed the most notable increase in the incidence of NAFLD (10.8 per 100,000 inhabitants in 2004, 124.5 per 100,000 inhabitants in 2022, increased 11 times since 2004), resulting a 3.4 time increase in the proportion and ultimately accounting for over half of the cases by the end of our follow-up. Conversely, the proportion of ArLD cases decreased by 71%, given that the incidence rate remained unchanged during the past 2 decades (Fig 4A – 4B, Supplemental Table 2-3). A breakdown of NAFLD incidence by data sources demonstrated the massive increases in both the PEDW-only and WLGP-only groups (PEDW-only by 17.7 times, WLGP-only by 14 times, Supplemental Table 3-1). In contrast, the increase in the incidence of ArLD was notably larger in the WLGP-only group compared to the PEDW-only group (PEDW-only increased by 1.4 times; WLGP-only increased by 3.2, Supplemental Table 3-2).

We did not observe an increasing trend for HBV and HCV. Both conditions fluctuated during the past 2 decades, with HBV cases increasing by 4.3 times and HCV cases 2.3 times from 2004 to 2017, followed by a decrease from 2018 to 2022 (HBV by 1.7 times; HCV by 2.6 times). Consequently, the proportion of HBV and HCV cases decreased from 2004 to 2022 (HBV by 2.3 times; HCV by 8.6 times, Fig 4A – 4B, Supplemental Table 2-3).

Underlying liver disease conditions

Of the 111,098 individuals with liver disease, 79,992 (72.0 %) entered the cohort due to a tier 1 aetiology. This was mainly contributed by the high prevalence of NAFLD (33,655 [30.3%]). Most NAFLD cases were identified in WLGP-only group (13,565 [66.8 %]), and the least NAFLD cases in ADDE-only group (727 [21.0 %]). This aligns with the distribution of individuals with tier 1 aetiology, which's proportion was the highest in WLGP-only group (18,453 [90.8 %]), and the lowest in ADDE-only group (1,728 [50.0 %]). In contrast, tier 2 and tier 3 diagnoses were higher in PEDW-only group. The proportion of advanced liver disease stages was lower in WLGP-only group (stage 3: 166 [0.8 %], stage 4: 53 [0.3 %], stage 5: 121 [0.6 %]), but higher in ADDE-only group (stage 4: 276 [8.0 %], stage 5: 473 [13.7 %]) (Table 3).

Comorbidities associated with liver disease

Amongst individuals who had liver disease diagnoses records within 1 year prior to cohort entry, 1,972 (1.8 %) individuals had onset hypertension or initiated an antihypertensive medication, followed by 707 (0.6 %) individuals who had onset diabetes, and 679 (0.6 %) individuals who had onset cardiovascular disease (CVD) related conditions. These proportions of comorbidities increased as we expand the time frame to 1-3 years (hypertension/antihypertensive: 2,924 [2.7 %]; diabetes: 766 [0.7 %]; CVD related conditions: 727 [0.7 %]), 3-5 years (hypertension/antihypertensive: 2,849 [2.7 %]; diabetes: 694 [0.7 %]; CVD related conditions: 684 [0.7 %]), 5-10 years (hypertension/antihypertensive: 7.830 [7.4 %]; diabetes: 1,680 [1.6 %]; CVD related conditions: 1,835 [1.7 %]), and over 10 years (hypertension/antihypertensive: 24,852 [23.7%]; diabetes: 3,811 [3.6 %]; CVD related conditions: 4,998 [4.8 %]) before cohort entry (Table 4).

Discussion

In this large-scale population-based study, we observed that the incidence of liver disease had increased dramatically in Wales during the past two decades. Notably, NAFLD played a significant role in this rise, with its incidence and proportion increasing sharply.

The rapid increase in the incidence rate of liver disease observed in our study is similar to our previous findings in Wales [19], where a 3.6-fold increase was observed in the in-patient chronic liver disease cases between 2001 and 2019, aligning with the trends presented in our PEDW data. This was mainly driven by the 11-fold increase in NAFLD incidence. Other studies, such as P Nasr's et al work in Sweden [23], reported a 2-fold increase in the incidence rate of NAFLD during 2005-2019. Similarly, H Tian et al [24] used data from the Global Burden of Disease study 2019 and showed an increase of 95.4% of NAFLD globally from 1990 to 2019. A systematic review that included 578 studies demonstrated a 13% higher prevalence of NAFLD during the year 2011-2021 compared with year 2000-2010 [25]. A

community study in the USA reported a 5-fold increase in the NAFLD incidence from 1997 to 2014 [26]. However, none of these findings compare to our observation in Wales. This substantial difference may be partly attributed to our inclusion of primary care data, which experienced more rapid increase than inpatient data and thus boosted the NAFLD incidence across the study population. The high prevalence of obesity [27] and rapidly increasing incidence of diabetes [28] in Wales in the recent decades may also have attributed to the sharp increase of NAFLD, as these metabolic factors including obesity [29, 30] and insulin resistance were considered to be strongly linked to NAFLD due to metabolic dysfunctions [31, 32]. This observation is further supported by our observation of autoimmune liver disease, which is not linked to metabolic factors and remained unchanged during our study period, highlighting the lifestyle-related nature of the increase in NAFLD. Another interesting observation related to NAFLD is the high proportion of those with comorbidities. NALFD evolves in those with metabolic syndrome, and these comorbidities are used to define the more recently adopted term metabolic associated fatty liver disease [33]. In keeping with this definition of NAFLD, we observed a high proportion of individuals having onset comorbidities such as hypertension, diabetes, CVD related conditions up to 10 years prior to their first liver disease diagnosis. The multimorbidity amongst people living with liver disease is anticipated considering the increasing incidence of NAFLD and the high prevalence of obesity in Wales, and this will require engagement with primary care practitioners to address their complex multidisciplinary healthcare needs and to identify those at risk of significant liver disease early in the disease trajectory [34].

Contrary to previous European studies showing a decline in ArLD incidence [19, 35], our research observed a resurgence after 2015 following a slight decrease from 2004 to 2014. We believe that changes in drinking habits [36, 37] and government actions on alcohol pricing and taxation [38] likely influenced the declining trend before 2015, while the resurgence after 2015 was contributed by the inclusion of outpatient data from the WLGP-only group. The addition of WLGP patients could explain the discrepancy in our results compared to the prior studies that focused solely on in-patients, filling the gap by providing a broader perspective by including outpatient data [19, 35]. Similar to ArLD, the incidence of HBV increased until 2017, then sharply dropped after the year of 2018. The observed decrease in the incidence of HBV and HCV post-2020 could be attributed to the initiatives outlined in the Welsh Health Circular, which seeks to meet the goal of eradicating HBV and HCV as a major public health concern [39]. This trend may represent the effort in better case finding and eradication, which led to reduced transmission for HCV, and enhanced vaccination and viral suppression for HBV [40]. Additionally, disruptions caused by the COVID-19 pandemic may also have played a role in this decline [41].

It is worth to mention our observation on the increasing number of incident cases identified from primary care data, particularly after 2015. This observation, together with the rapid increase in NAFLD and ArLD incidence, indicated an increasing detection rate of liver disease in primary care at an early

stage. This change might be attributed to the implementation of the Wales Liver Disease Delivery Plan, which emphasised the importance of early detection and ensure that 'excellent care' is accessible when necessary [42]. Another notable trend observed during our study period is the decreasing proportion of severe late-stage presentations, such as hepatic decompensation. Additionally, we observed a declining trend in the proportion of individuals who died from liver disease without receiving a prior diagnosis. This is to say, despite the increasing incidence of liver disease, there is a promising indication that more individuals were identified as the result of early detection. This provides the healthcare system with more opportunities for early-stage intervention. However, it also challenges the system to identify those at higher risk of developing further liver-related complications, given the increasing absolute number of individuals requiring secondary care [43].

The strengths of our study are the national-scale setting and population-based cohort with long-term follow-up, which allowed us to maximise the generalisability of our findings to the wider population. By including both in-patient and outpatient data, we were able to expand our sample size by at least 20% and conduct a comprehensive analysis of the incident rate of liver disease, fostering a deeper insight into its clinical implications. Our inclusion of outpatients within the primary care dataset is essential for providing a more comprehensive scope in liver disease epidemiology, as these individuals may present different clinical characteristics compared to those admitted to secondary care services. This approach also helps address the under-recognition of the disease in other reports, offering a broader and more thorough understanding of the disease's incidence. Another strength lies in the reproducibility and the use of the SAIL Databank as a research-ready data asset (RRDA). By incorporating reproducible research pipelines, we standardized and documented the cohort curation process. This enhances the study's reliability and reproducibility, enabling other researchers to replicate our findings and explore further research questions within the same framework. However, our study is vulnerable to several limitations. Regulatory laws associated with the SAIL Databank prohibited the inclusion of conditions classified as sensitive in our analysis. For instance, the sensitive Read codes for HBV and HCV could not be extracted along with the rest of the cohort, resulting in potentially incomplete data for these conditions, affecting the completeness and precision of our data analysis. Our study exclusively included individuals who had a history of GP registration and residency information in Wales. Consequently, we likely overlooked individuals were unable to register with a GP, a subpopulation that represents a group at significant risk for liver disease. Furthermore, due to the observational nature and design of our study, we were unable to establish causal relationships between the onset of liver disease and other comorbidities.

Conclusion

Our study observed a significant rise in the incidence of liver disease in Wales over the past two decades, primarily driven by the increase in NAFLD. The high prevalence of comorbid conditions among liver disease patients and the increased role of primary care in disease identification highlight the need for integrated healthcare approaches to address this growing public health concern. It is therefore crucial to focus more attention on early detection, lifestyle interventions, and comprehensive management of both liver disease and its associated comorbidities.

Acknowledgements

The Liver Disease Cymru Partnership (LDCP) received a grant from the National Institute for Health Research (NIHR154876). This project was partly funded by an unrestricted grant from the Liver Disease Implementation Group, Welsh Government (LDIG-22-19).

We wish to acknowledge all members of the LDCP who have contributed to the collaborations: Prof Kerry Hood, Prof Deborah Fitzsimmons and Prof Katherine Cullen for advice developing the group and the Liver Research Cymru patient advisory group for their perspective on the research to be undertaken.

This study uses anonymised data held in the Secure Anonymised Information Linkage (SAIL) Databank, which is part of the national population data research infrastructure for Wales. We would like to acknowledge all the data providers who make anonymised data available for research.

Author Contributions

This project was a collaborative effort among multiple team members, each contributing significantly to various aspects of the research. TP, AA, AD, AY and HA acquired the funding. TP, AA, AD, AY, HA and JG contributed to the design of the methodology. TP, AA, AD, AY and JG supervised the study and validated the research. JG and AA were responsible for project administration and data management. JG curated the data, performed data analysis, created visual representations of the data and wrote the original draft of the paper. All authors reviewed and edited the manuscript.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Funding Sources

The Liver Disease Cymru Partnership (LDCP) receives a grant from the National Institute for Health Research (NIHR154876). This project was partly funded by an unrestricted grant from the Liver Disease Implementation Group, Welsh Government (LDIG-22-19).

Strengths and limitations

The national-scale population-based cohort and long follow-up maximized the generalizability of our finding.

By incorporating both in-patient and outpatient data, the study not only expanded the sample size but also enabled a thorough analysis of liver disease incidence, enhancing understanding of its clinical implications.

With the usage of SAIL Databank as the research-ready data asset and the incorporation of reproducible research pipelines, this study strengthened its reliability and reproducibility.

Sensitive Read codes for conditions like HBV and HCV could not be included due to regulatory laws, impacting the completeness and precision of the data analysis.

The study only included individuals with a history of GP registration and residency information in Wales, likely overlooking a subpopulation at significant risk for liver disease

Data Access and Ethics Statement

The routine data used in this study are available in the SAIL Databank at Swansea University, Swansea, UK. All proposals to use SAIL data are subject to review by an Information Governance Review Panel (IGRP). Before any data can be accessed, approval must be given by the IGRP. The IGRP gives careful consideration to each project to ensure proper and appropriate use of SAIL data. When access has been approved, it is gained through a privacy-protecting trusted research environment (TRE). SAIL has established an application process to be followed by anyone who would like to access data via SAIL https://www.saildatabank.com/application-process . This study has been approved by the IGRP as project 1492. The research adhered to ethical guidelines and Data Protection Act 2018 to ensure the privacy and confidentiality of all data subjects involved.

The reproducible SQL and R code, and code lists to identify study individuals are available on Github: https://github.com/SwanseaUniversityDataScience/1492-LDCP

References

[1] Paik JM, Golabi P, Younossi Y, Mishra A, Younossi ZM. Changes in the Global Burden of Chronic Liver Diseases From 2012 to 2017: The Growing Impact of NAFLD. Hepatology. 2020 Oct 27;72(5):1605–16. https://doi.org/10.1002/hep.31173

[2] Wang H, Naghavi M, Allen C, Barber RM, Bhutta ZA, Carter A, et al. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet. 2016 Oct;388(10053):1459–544. https://doi.org/10.1016/S0140-6736(16)31012-1.

[3] Deaths registered in England and Wales (series DR): 2017. Office for National Statistics. 2017. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deathsregisteredinenglandandwalesseriesdr/2017. [Accessed 5 Jun 2024].

[4] Williams R, Aspinall R, Bellis M, Camps-Walsh G, Cramp M, Dhawan A, et al. Addressing liver disease in the UK: a blueprint for attaining excellence in health care and reducing premature mortality from lifestyle issues of excess consumption of alcohol, obesity, and viral hepatitis. Lancet. 2014 Nov 29;384(9958):1953–97. https://doi.org/10.1016/S0140-6736(14)61838-9.

[5] Nobili V, Carter-Kent C, Feldstein AE. The role of lifestyle changes in the management of chronic liver disease. BMC Medicine. 2011 Jun 6;9:70. https://doi.org/10.1186/1741-7015-9-70.

[6] Department of Health and Social Care. Advancing our health: prevention in the 2020s – consultation document. GOV.UK. 2019. https://www.gov.uk/government/consultations/advancing-our-health-prevention-in-the-2020s/advancing-our-health-prevention-in-the-2020s-consultation-document. [Accessed 5 Jun 2024].

[7] Diet, nutrition and obesity. NICE. https://www.nice.org.uk/guidance/lifestyle-and-wellbeing/diet--nutrition-and-obesity. [Accessed 5 Jun 2024].

[8] Burton R, Henn C, Lavoie D, O'Connor R, Perkins C, Sweeney K, et al. A rapid evidence review of the effectiveness and cost-effectiveness of alcohol control policies: an English perspective. The Lancet. 2017 Apr;389(10078):1558–80. https://doi.org/10.1016/S0140-6736(16)32420-5.

[9] The quality statement for liver disease: the quality statement describes what good quality liver disease services should look like. Welsh Government. 2022. https://www.gov.wales/sites/default/files/pdf-versions/2023/3/1/1679326108/quality-statement-liver-disease.pdf. [Accessed 5 Jun 2024].

[10] Nam YH, Mendelsohn AB, Panozzo CA, Maro JC, Brown JS. Health outcomes coding trends in the US Food and Drug Administration's Sentinel System during transition to International Classification of Diseases-10 coding system: A brief review. Pharmacoepidemiology and Drug Safety. 2021 Mar 17;30(7):838–42. https://doi.org/10.1002/pds.5216.

[11] Ratib S, West J, Fleming KM. Liver cirrhosis in England—an observational study: are we measuring its burden occurrence correctly? BMJ Open. 2017 Jul;7(7):e013752. https://doi.org/10.1136/bmjopen-2016-013752.

[12] Ford DV, Jones KH, Verplancke JP, Lyons RA, John G, Brown G, et al. The SAIL Databank: building a national architecture for e-health research and evaluation. BMC Health Services Research. 2009 Sep 4;9:157. https://doi.org/10.1186/1472-6963-9-157.

[13] Lyons RA, Jones KH, John G, Brooks CJ, Verplancke JP, Ford DV, et al. The SAIL databank: linking multiple health and social care datasets. BMC Medical Informatics and Decision Making. 2009 Jan 16;9:3. https://doi.org/10.1186/1472-6947-9-3.

[14] Jones KH, Ford DV, Ellwood-Thompson S, Lyons RA. The UK Secure eResearch Platform for public health research: a case study. The Lancet. 2016 Nov;388:S62. https://doi.org/10.1016/S0140-6736(16)32298-X.

[15] Wilkinson T, Schnier C, Bush K, Rannikmäe K, Lyons RA, McTaggart S, et al. Drug prescriptions and dementia incidence: a medication-wide association study of 17000 dementia cases among half a million participants. Journal of Epidemiology and Community Health. 2021 Oct 27;76(3):223–9. https://doi.org/10.1136/jech-2021-217090.

[16] Schnier C, Wilkinson T, Akbari A, Orton C, Sleegers K, Gallacher J, et al. The Secure Anonymised Information Linkage databank Dementia e-cohort (SAIL-DeC). International Journal of Population Data Science. 2020 Feb 25;5(1):1121. https://doi.org/10.23889/ijpds.v5i1.1121

[17] Population estimates for the UK, England, Wales, Scotland, and Northern Ireland: mid-2022. Office for National Statistics. 2024. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2022. [Accessed 5 Jun 2024].

[18] Ford DV, Jones KH, Verplancke JP, Lyons RA, John G, Brown G, et al. The SAIL Databank: building a national architecture for e-health research and evaluation. BMC Health Services Research. 2009 Sep 4;9:157. https://doi.org/10.1186/1472-6963-9-157.

[19] Peter T, John GR, Puyk B, Howkins K, Clarke R, Yousuf F, et al. Rising incidence, progression and changing patterns of liver disease in Wales 1999-2019. World Journal of Hepatology. 2023 Jan 27; 15(1):89–106. https://doi.org/10.4254/wjh.v15.i1.89.

[20] Estimate of the population for England and Wales. Office for National Statistics. 2023. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/estimatesofthepopulationforenglandandwales. [Accessed 5 Jun 2024].

[21] von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. PLoS Medicine. 2007 Oct 16;4(10):e296. https://doi.org/10.1016/j.jclinepi.2007.11.008.

[22] Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. PLOS Medicine. 2015 Oct 6;12(10):e1001885. https://doi.org/10.1371/journal.pmed.1001885

[23] Patrik Nasr, Erik von Seth, Mayerhofer R, Ndegwa N, Ludvigsson JF, Hannes Hagström. Incidence, prevalence and mortality of chronic liver diseases in Sweden between 2005 and 2019. European Journal of Epidemiology. 2023 Jul 25;38(9):973–84. https://doi.org/10.1007/s10654-023-01028-x.

[24] Tian H, Zhang K, Hui Z, Ren F, Ma Y, Han F, et al. Global burden of non-alcoholic fatty liver disease in 204 countries and territories from 1990 to 2019. Clinics and Research in Hepatology and Gastroenterology. 2023 Jan;47(1):102068. https://doi.org/10.1016/j.clinre.2022.102068.

[25] Liu J, Tian Y, Fu X, Mu C, Yao M, Ni Y, et al. Estimating global prevalence, incidence, and outcomes of non-alcoholic fatty liver disease from 2000 to 2021: systematic review and meta-analysis. Chinese Medical Journal. 2022 Jul 20;135(14):1682–91. https://doi.org/10.1097/CM9.0000000000002277.

[26] Allen AM, Therneau TM, Larson JJ, Coward A, Somers VK, Kamath PS. Nonalcoholic fatty liver disease incidence and impact on metabolic burden and death: A 20 year-community study. Hepatology. 2018 Mar 23;67(5):1726–36. https://doi.org/10.1002/hep.29546.

[27] Keaver L, Xu B, Jaccard A, Webber L. Morbid obesity in the UK: A modelling projection study to 2035. Scandinavian Journal of Public Health. 2018 Aug 30;48(4):422–7. https://doi.org/10.1177/1403494818794814

[28] Powell R. Diabetes prevalence – trends, risk factors, and 10-year projection [Internet]. Public Health Wales. 2023. https://phw.nhs.wales/services-and-teams/observatory/data-and-analysis/diabetes-prevalence-trends-risk-factors-and-10-year-projection/. [Accessed 5 Jun 2024].

[29] Younossi Z, Anstee QM, Marietti M, Hardy T, Henry L, Eslam M, et al. Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. Nature Reviews Gastroenterology & Hepatology. 2017 Sep 20;15(1):11–20. https://doi.org/10.1038/nrgastro.2017.109.

[30] Loomis AK, Kabadi S, Preiss D, Hyde C, Bonato V, St. Louis M, et al. Body Mass Index and Risk of Nonalcoholic Fatty Liver Disease: Two Electronic Health Record Prospective Studies. The Journal of Clinical Endocrinology & Metabolism. 2016 Mar 1;101(3):945–52. https://doi.org/10.1210/jc.2015-3444.

[31] Caussy C, Aubin A, Loomba R. The Relationship Between Type 2 Diabetes, NAFLD, and Cardiovascular Risk. Current Diabetes Reports. 2021 Mar 19;21(5):15. https://doi.org/10.1007/s11892-021-01383-7.

[32] Tanase DM, Gosav EM, Costea CF, Ciocoiu M, Lacatusu CM, Maranduca MA, et al. The Intricate Relationship between Type 2 Diabetes Mellitus (T2DM), Insulin Resistance (IR), and Nonalcoholic Fatty Liver Disease (NAFLD). Journal of Diabetes Research. 2020 Jul 31:2020:3920196. https://doi.org/10.1155/2020/3920196.

[33] Eslam M, Newsome PN, Sarin SK, Anstee QM, Targher G, Romero-Gomez M, et al. A new definition for metabolic dysfunction-associated fatty liver disease: An international expert consensus statement. Journal of Hepatology. 2020 Jul 1;73(1):202–9. https://doi.org/10.1016/j.jhep.2020.03.039.

[34] Pryke R, Guha IN. Time to focus on chronic liver diseases in the community: A review of primary care hepatology tools, pathways of care and reimbursement mechanisms. Journal of Hepatology. 2023 Mar 1 [cited 2024 May 13];78(3):663–71. https://doi.org/10.1016/j.jhep.2022.10.010.

[35] Deleuran T, Vilstrup H, Becker U, Jepsen P. Epidemiology of Alcoholic Liver Disease in Denmark 2006–2011: A Population-Based Study. Alcohol and Alcoholism. 2015 Feb 13;50(3):352–7. https://doi.org/10.1093/alcalc/agv003.

[36] Adult drinking habits in Great Britain: 2017. Office for National Statistics. 2018. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/drugusealcoholandsmoking/bulletins/opinionsandlifestylesurveyadultdrinkinghabitsingreatbritain/2017. [Accessed 5 Jun 2024].

[37] Adult lifestyles by year, 2016-17 to 2019-20. StatsWales. Available from: https://statswales.gov.wales/Catalogue/National-Survey-for-Wales/Population-Health/Adult-Lifestyles/adultlifestyles-by-year. [Accessed 5 Jun 2024].

[38] Home Office. Alcohol strategy. GOV.UK. 2012. https://www.gov.uk/government/publications/alcohol-strategy. [Accessed 5 Jun 2024].

[39] WELSH HEALTH CIRCULAR. Attaining the WHO targets for eliminating hepatitis (b and C) as a significant threat to public health. Welsh Government. 2017. https://www.gov.wales/sites/default/files/publications/2019-07/attaining-the-who-targets-for-eliminating-hepatitis-b-and-c-as-a-significant-threat-to-public-health.pdf. [Accessed 5 Jun 2024].

[40] Prevention, diagnosis and treatment of blood borne viruses in Wales: Hepatitis B, hepatitis C and HIV Annual report 2023 (Data to end 2022). Public Health Wales. 2023. https://phw.nhs.wales/publications/publications1/prevention-diagnosis-and-treatment-of-blood-borne-viruses-in-wales-hepatitis-b-hepatitis-c-and-hiv/. [Accessed 5 Jun 2024].

[41] Pley CM, McNaughton AL, Matthews PC, Lourenço J. The global impact of the COVID-19 pandemic on the prevention, diagnosis and treatment of hepatitis B virus (HBV) infection. BMJ Global Health. 2021 Jan;6(1):e004275. https://doi.org/10.1136/bmjgh-2020-004275.

[42] Together for Health -Liver Disease Delivery Plan: A Delivery Plan for NHS Wales and its Partners to 2020. Welsh Government. 2015. https://www.gov.wales/sites/default/files/publications/2018-12/liver-disease-delivery-plan-2015-to-2020.pdf. [Accessed 5 Jun 2024].

[43] Fowell A, Kirsty Fancey, Gamble KL, Bicknell K, Dowman J, Howden PJ, et al. Evaluation of a primary to secondary care referral pathway and novel nurse-led one-stop clinic for patients with suspected non-alcoholic fatty liver disease. Frontline Gastroenterology. 2020 Apr 7;12(2):102–7. https://doi.org/10.1136/flgastro-2019-101304.

Figure Legends

Fig. 1 Flowchart of study population selection (Fig.1A) and Venn diagram of study cohort composition (Fig.1B)

Abbreviations: PEDW: Patient Episode Database for Wales; WLGP: Welsh Longitudinal General Practice; ADDE: Annual District Death Extract

Fig. 2 Changes in Standardised Incidence per 100,000 Inhabitants by Data Source (2004-2022)
Bar chart presenting the changes in annual European age-sex standardised incidence rate and percentages by data sources. The x-axis represents time in years, and the y-axis represents incidence rate in Fig.2A, percentages in Fig.2B

Abbreviations: PEDW: Patient Episode Database for Wales; WLGP: Welsh Longitudinal General Practice; ADDE: Annual District Death Extract

Fig. 3 Changes in Standardised Incidence per 100,000 Inhabitants by Stages (2004-2022)
Bar chart presenting the changes in annual European age-sex standardised incidence rate and percentages by stages. The x-axis represents time in years, and the y-axis represents incidence rate in Fig.3A, percentages in Fig.3B

Abbreviations: PEDW: Patient Episode Database for Wales; WLGP: Welsh Longitudinal General Practice; ADDE: Annual District Death Extract

Fig. 4 Changes in Standardised Incidence per 100,000 Inhabitants by Liver Disease Aetiologies (2004-2022)
Bar chart presenting the changes in annual European age-sex standardised incidence rate and percentages by aetiologies. The x-axis represents time in years, and the y-axis represents incidence rate in Fig.4A, percentages in Fig.4B

Abbreviations: PEDW: Patient Episode Database for Wales; WLGP: Welsh Longitudinal General Practice; ADDE: Annual District Death Extract; ArLD: alcohol related liver disease; NAFLD: non-alcoholic fatty liver disease; HBV: hepatitis B virus; HCV: hepatitis C virus

* ArLD overlap: if an individual had 2 or more diagnoses from liver disease aetiologies (ArLD, NAFLD, HBV, HCV, metabolic, hemochromatosis and autoimmune liver diseases) and one of them was ArLD; Non-ArLD overlap: if an individual had 2 or more diagnoses from liver disease aetiologies but none of them was ArLD.

Tables

Table 1. Demographic and socioeconomic characteristics of liver disease patients

Demographic characteristics	Full group, N = 111,098[1]	PEDW only group, N = 56,710[1]	WLGP only group, N = 20,319[1]	ADDE only group, N = 3,459[1]	Two or more data sources, N = 30,610[1]
Sex					
Male	57,491(51.7%)	27,558(48.6%)	10,478(51.6%)	2,019(58.4%)	17,436(57.0%)
Female	53,607(48.3%)	29,152(51.4%)	9,841(48.4%)	1,440(41.6%)	13,174(43.0%)
Age					
0-17	1,445(1.3%)	814(1.4%)	342(1.7%)	<10 (<0.3%)	<290 (<1.0%)
18-29	5,862(5.3%)	3,486(6.1%)	1,447(7.1%)	<40 (<1.2%)	<900 (<0%)
30-39	10,109(9.1%)	5,099(9.0%)	2,520(12.4%)	143(4.1%)	2,347(7.7%)
40-49	15,370(13.8%)	6,316(11.1%)	3,744(18.4%)	421(12.2%)	4,889(16.0%)
50-59	22,799(20.5%)	9,332(16.5%)	5,356(26.4%)	705(20.4%)	7,406(24.2%)
60-69	23,264(20.9%)	11,026(19.4%)	4,271(21.0%)	758(21.9%)	7,209(23.6%)
70-79	19,189(17.3%)	11,335(20.0%)	2,117(10.4%)	693(20.0%)	5,044(16.5%)
80+	13,060(11.8%)	9,302(16.4%)	522(2.6%)	702(20.3%)	2,534(8.3%)
Cohort entry year					

Demographic characteristics	Full group, N = 111,098[1]	PEDW only group, N = 56,710[1]	WLGP only group, N = 20,319[1]	ADDE only group, N = 3,459[1]	Two or more data sources, N = 30,610[1]
2004-2007	17,099(15.4%)	7,345(13.0%)	1,949(9.6%)	617(17.8%)	7,188(23.5%)
2008-2011	17,554(15.8%)	8,484(15.0%)	2,513(12.4%)	705(20.4%)	5,852(19.1%)
2012-2015	20,658(18.6%)	10,682(18.8%)	3,109(15.3%)	731(21.1%)	6,136(20.0%)
2016-2019	30,448(27.4%)	15,521(27.4%)	6,683(32.9%)	749(21.7%)	7,495(24.5%)
2020-2022	25,339(22.8%)	14,678(25.9%)	6,065(29.8%)	657(19.0%)	3,939(12.9%)
WIMD 2019 quintiles					
1, most deprived	27,178(24.5%)	14,012(24.7%)	4,417(21.7%)	928(26.8%)	7,821(25.6%)
2	24,391(22.0%)	12,411(21.9%)	4,203(20.7%)	858(24.8%)	6,919(22.6%)
3	21,066(19.0%)	10,802(19.0%)	3,828(18.8%)	647(18.7%)	5,789(18.9%)
4	19,619(17.7%)	9,855(17.4%)	3,949(19.4%)	545(15.8%)	5,270(17.2%)
5, least deprived	18,844(17.0%)	9,630(17.0%)	3,922(19.3%)	481(13.9%)	4,811(15.7%)

[1] n(%)

Abbreviations: PEDW: Patient Episodes Dataset for Wales, WLGP: Welsh Longitudinal General Practice dataset, ADDE: Annual District Deaths Extract, WIMD: Welsh Index of Multiple Deprivation

Table 2. Crude and standardised incidence of chronic liver disease from 2004 to 2022

Year	Wales population	Incident cases	Crude incidence	Crude 95% CI	STD incidence	STD 95% CI
2004	3,131,640	3,060	97.7	94.3 to 101.2	97.7	94.2 to 101.3
2005	3,105,633	3,403	109.6	105.9 to 113.3	109.4	105.7 to 113.2
2006	3,104,483	3,670	118.2	114.4 to 122.1	118.7	114.9 to 122.6
2007	3,087,732	3,537	114.6	110.8 to 118.4	115.1	111.3 to 119.0
2008	3,083,840	3,867	125.4	121.5 to 129.4	126.3	122.3 to 130.4
2009	3,081,366	4,008	130.1	126.1 to 134.2	131.7	127.7 to 135.9
2010	3,077,165	4,268	138.7	134.6 to 142.9	140.6	136.4 to 144.9
2011	3,072,739	4,519	147.1	142.8 to 151.4	149.9	145.5 to 154.4
2012	3,073,788	4,546	147.9	143.6 to 152.3	150.6	146.3 to 155.1
2013	3,071,058	4,733	154.1	149.8 to 158.6	158.0	153.5 to 162.6
2014	3,070,928	5,145	167.5	163.0 to 172.2	172.5	167.7 to 177.3
2015	3,063,758	5,659	184.7	179.9 to 189.6	192.2	187.2 to 197.4
2016	3,049,971	6,548	214.7	209.5 to 220.0	225.7	220.2 to 231.3
2017	3,038,872	6,981	229.7	224.4 to 235.2	243.2	237.4 to 249.1

Year	Wales population	Incident cases	Crude incidence	Crude 95% CI	STD incidence	STD 95% CI
2018	3,025,867	7,953	262.8	257.1 to 268.7	279.1	272.9 to 285.4
2019	3,006,299	8,481	282.1	276.1 to 288.2	301.2	294.8 to 307.8
2020	2,985,668	7,327	245.4	239.8 to 251.1	264.0	257.8 to 270.2
2021	2,969,309	9,136	307.7	301.4 to 314.1	332.0	325.0 to 338.9
2022	2,957,422	8,596	290.7	284.5 to 296.9	316.2	309.4 to 323.1

Abbreviation: STD: standardised; CI: confidence interval

Table 3. The underlying liver disease conditions at cohort entry

Phenotype	Full group, N = 111,098[1]	PEDW-only group, N = 56,710[1]	WLGP-only group, N = 20,319[1]	ADDE-only group, N = 3,459[1]	Two or more data sources, N = 30,610[1]
Chronic liver disease					
Stage 1 - tier 1	**79,992(72.0%)**	**35,102(61.9%)**	**18,453(90.8%)**	**1,728(50.0%)**	**24,709(80.7%)**
- ArLD	19,760(17.8%)	4,566(8.1%)	2,116(10.4%)	887(25.6%)	12,191(39.8%)
- NAFLD	33,655(30.3%)	13,444(23.7%)	13,565(66.8%)	727(21.0%)	5,919(19.3%)
- Metabolic liver disease	5,469(4.9%)	3,887(6.9%)	782(3.8%)	69(2.0%)	731(2.4%)
- HBV	1,063(1.0%)	905(1.6%)	0(0.0%)	0(0.0%)	158(0.5%)
- HCV	3,539(3.2%)	2,697(4.8%)	0(0.0%)	13(0.4%)	829(2.7%)
- Autoimmune liver disease	13,582(12.2%)	9,060(16.0%)	815(4.0%)	199(5.8%)	3,508(11.5%)
- Haemochromatosis	4,111(3.7%)	<940(<1.7%)	1,209(6.0%)	<10(<0.2%)	1,958(6.4%)
- ArLD overlap	799(0.7%)	158(0.3%)	15(0.1%)	173(5.0%)	453(1.5%)
- Non ArLD overlap	371(0.3%)	228(0.4%)	<20(<0.1%)	<10(<0.2%)	123(0.4%)
Stage 1 - tier 2	**6,111(5.5%)**	**4,906(8.7%)**	**81(0.4%)**	**127(3.7%)**	**997(3.3%)**

Phenotype	Full group, N = 111,098[1]	PEDW-only group, N = 56,710[1]	WLGP-only group, N = 20,319[1]	ADDE-only group, N = 3,459[1]	Two or more data sources, N = 30,610[1]
- Hepatitis not specified	4,783(4.3%)	3,867(6.8%)	11(0.1%)	102(2.9%)	803(2.6%)
- Congestive hepatopathy	574(0.5%)	514(0.9%)	<10(<0.1%)	<10(<0.2%)	<60(<0.2%)
- Toxic liver disease	757(0.7%)	526(0.9%)	68(0.3%)	21(0.6%)	142(0.5%)
Stage 1 - tier 3	**8,426(7.6%)**	**8,132(14.3%)**	**36(0.2%)**	**37(1.1%)**	**221(0.7%)**
- Miscellaneous	8,426(7.6%)	8,132(14.3%)	36(0.2%)	37(1.1%)	221(0.7%)
Stage 2	**4,562(4.1%)**	**1,664(2.9%)**	**506(2.5%)**	**690(19.9%)**	**1,702(5.6%)**
- Hepatic fibrosis	261(0.2%)	135(0.2%)	64(0.3%)	<10(<0.2%)	<60(<0.2%)
- Cirrhosis	4,301(3.9%)	1,529(2.7%)	442(2.2%)	682(19.7%)	1,648(5.4%)
Stage 3	**3,040(2.7%)**	**2,171(3.8%)**	**166(0.8%)**	**43(1.2%)**	**660(2.2%)**
- Portal hypertension	3,040(2.7%)	2,171(3.8%)	166(0.8%)	43(1.2%)	660(2.2%)
Stage 4	**3,221(2.9%)**	**2,517(4.4%)**	**53(0.3%)**	**276(8.0%)**	**375(1.2%)**
- Hepatic decompensation	3,221(2.9%)	2,517(4.4%)	53(0.3%)	276(8.0%)	375(1.2%)
Stage 5	**2,233(2.0%)**	**341(0.6%)**	**121(0.6%)**	**473(13.7%)**	**1,298(4.2%)**

Phenotype	Full group, N = 111,098[1]	PEDW-only group, N = 56,710[1]	WLGP-only group, N = 20,319[1]	ADDE-only group, N = 3,459[1]	Two or more data sources, N = 30,610[1]
- Hepatocellular carcinoma	748(0.7%)	83(0.1%)	71(0.3%)	86(2.5%)	508(1.7%)
- Intrahepatic cholangio carcinoma	1,455(1.3%)	232(0.4%)	47(0.2%)	387(11.2%)	789(2.6%)
- Other primary liver cancer	39(0.0%)	26(0.0%)	<10(<0.1%)	<10(<0.2%)	<10(<0.1%)
Acute liver disease					
Acute liver disease	**3,513(3.2%)**	**1,877(3.3%)**	**903(4.4%)**	**85(2.5%)**	**648(2.1%)**
- Acute viral hepatitis	1,898(1.7%)	962(1.7%)	761(3.7%)	<10(<0.2%)	<180(<0.6%)
- Budd-Chiari	184(0.2%)	144(0.3%)	<10(<0.1%)	<20(<0.1%)	20(0.1%)
- Acute liver failure	27(0.0%)	0(0.0%)	<30(<0.2%)	0(0.0%)	<10(<0.1%)
- Infection/infarction	1,343(1.2%)	757(1.3%)	62(0.3%)	67(1.9%)	457(1.5%)
- Other non-specified acute liver injuries	71(0.1%)	<20(<0.1%)	45(0.2%)	0(0.0%)	<10(<0.1%)

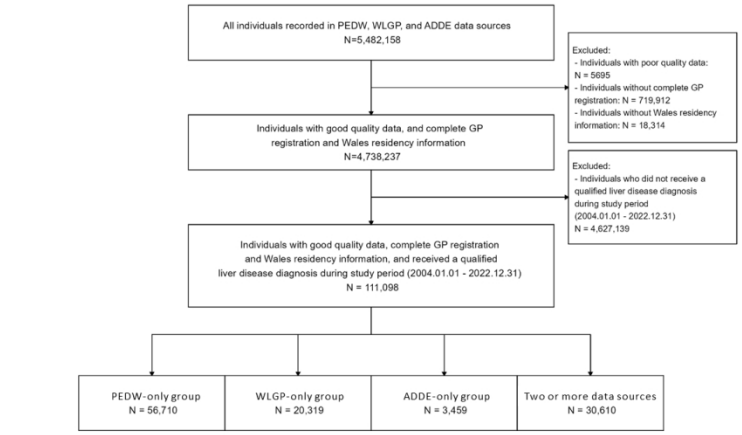
[1] n(%)

Abbreviation: ArLD: alcohol-related liver disease; NAFLD: non-alcohol fatty liver disease; HBV: hepatitis B virus; HCV: hepatitis C virus

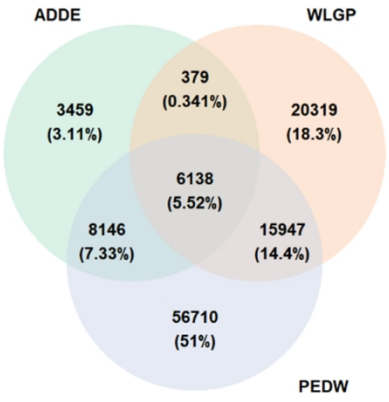
Table 4. Comorbidities associated with liver disease

Phenotype	Within 1 year[1]	In 1-3 years[1]	In 3-5 years[1]	In 5-10 years[1]	Over 10 years[1]
Number of patients who had diagnoses records	**111,089**	**106,817**	**104,625**	**106,361**	**104,817**
CVD related conditions	679(0.6%)	727(0.7%)	684(0.7%)	1,835(1.7%)	4,998(4.8%)
Diabetes	707(0.6%)	766(0.7%)	694(0.7%)	1,680(1.6%)	3,811(3.6%)
Hypertension/antihypertensive	1,972(1.8%)	2,924(2.7%)	2,849(2.7%)	7,830(7.4%)	24,852(23.7%)

Abbreviation: CVD: cardiovascular disease; COPD: chronic obstructive pulmonary disease



A



B

Fig. 1 Flowchart of study population selection (Fig.1A) and Venn diagram of study cohort composition (Fig.1B)

400x500mm (96 x 96 DPI)

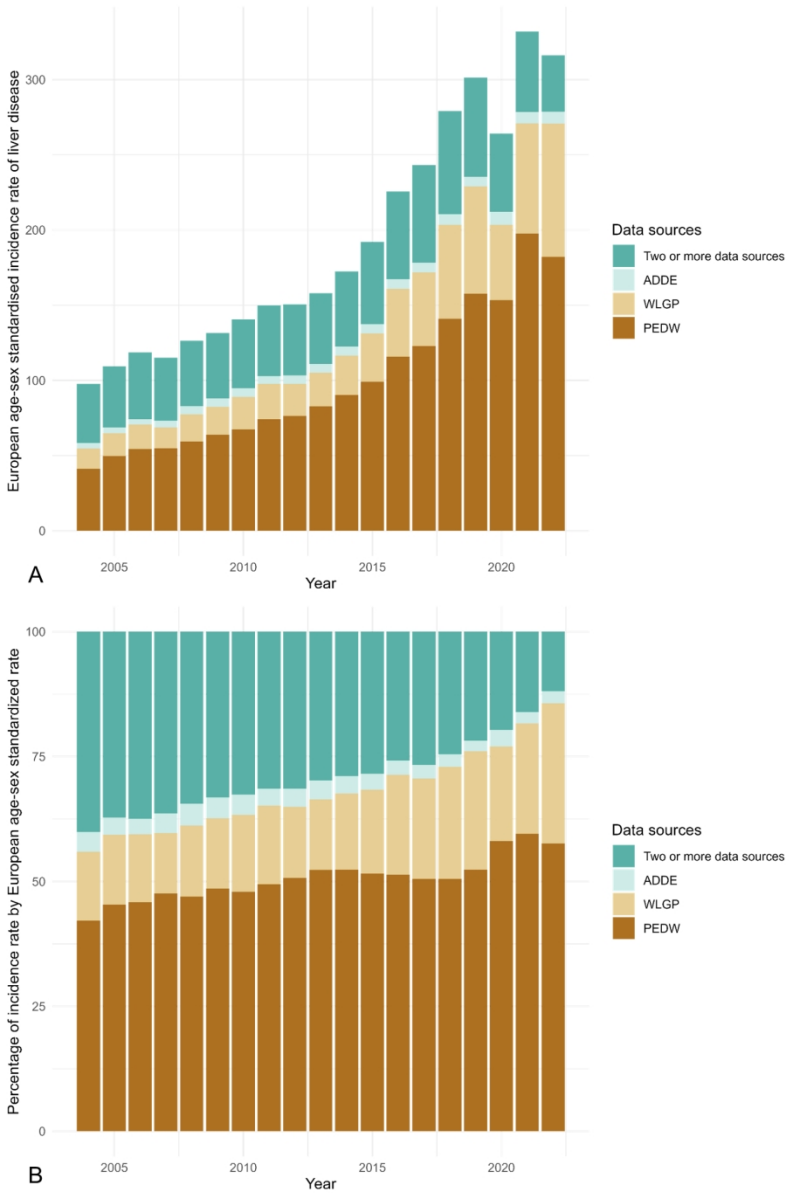


Fig. 2 Changes in Standardised Incidence per 100,000 Inhabitants by Data Source (2004-2022)
Bar chart presenting the changes in annual European age-sex standardised incidence rate and percentages by data sources. The x-axis represents time in years, and the y-axis represents incidence rate in Fig.2A, percentages in Fig.2B

400x600mm (96 x 96 DPI)

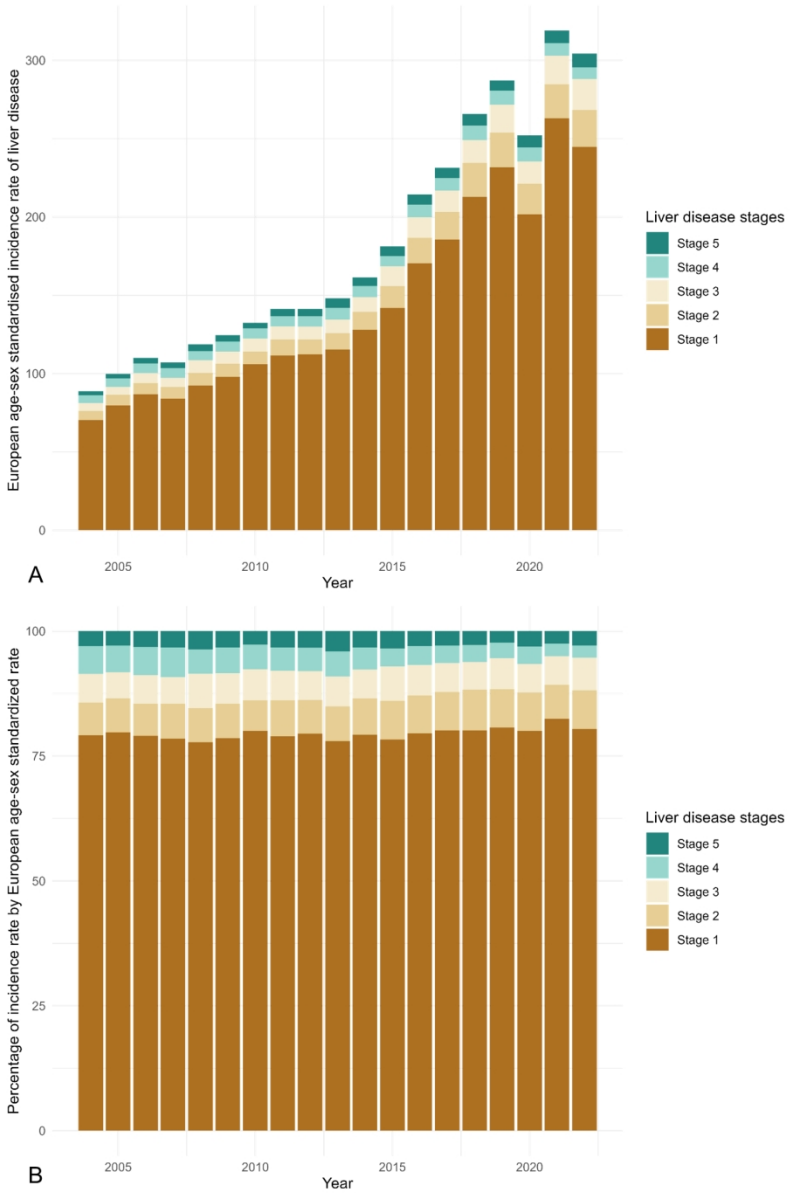


Fig. 3 Changes in Standardised Incidence per 100,000 Inhabitants by Stages (2004-2022)
Bar chart presenting the changes in annual European age-sex standardised incidence rate and percentages by stages. The x-axis represents time in years, and the y-axis represents incidence rate in Fig.3A, percentages in Fig.3B

400x600mm (96 x 96 DPI)

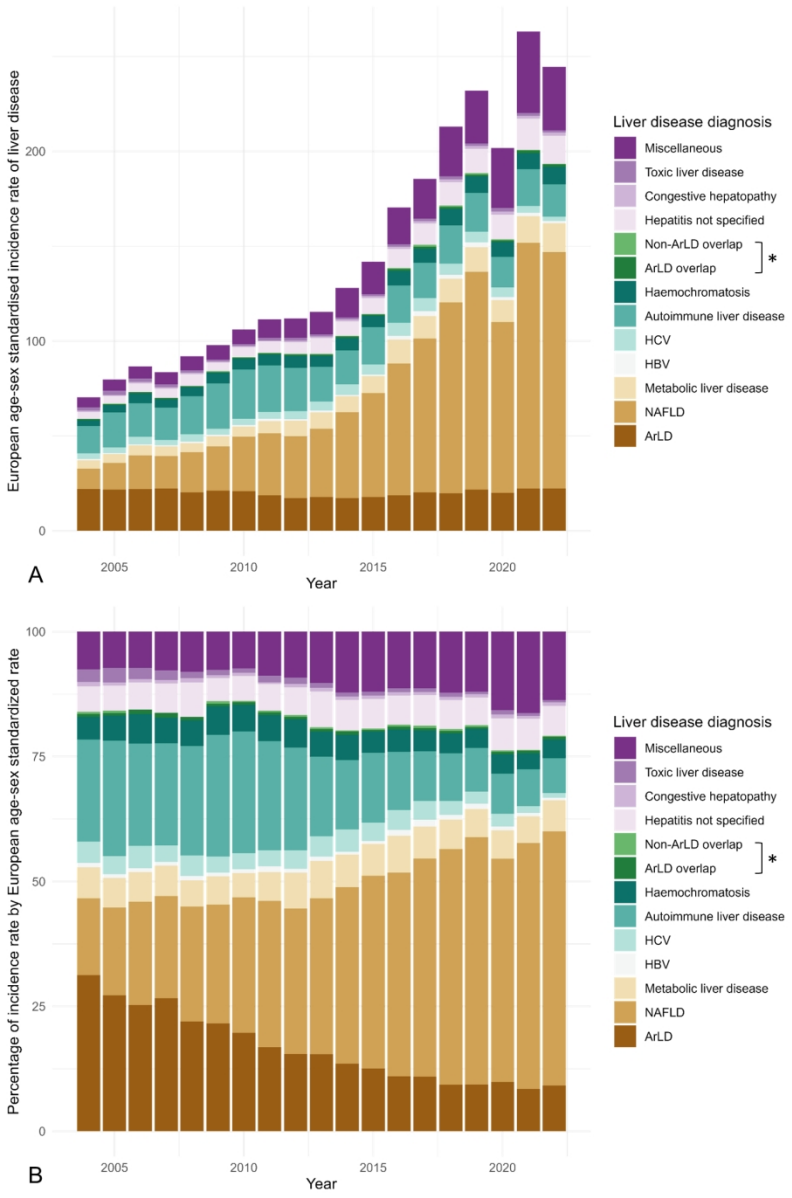


Fig. 4 Changes in Standardised Incidence per 100,000 Inhabitants by Liver Disease Aetiologies (2004-2022)
Bar chart presenting the changes in annual European age-sex standardised incidence rate and percentages by aetiologies. The x-axis represents time in years, and the y-axis represents incidence rate in Fig.4A, percentages in Fig.4B

400x600mm (96 x 96 DPI)

Supplemental Materials

Supplemental Fig. 1 Flowchart for Identifying Index Events and Determining Cohort Entry Date.

This flowchart illustrates the decision rules for identifying the index event and determining the cohort entry date for study participants. The index date is defined as the date of the first diagnosis between 2004 and 2022. The aetiological diagnoses throughout the disease progression were identified as the index event entering the cohort. If no aetiological diagnosis is present, the index event is defined based on the following sequence: stage 2 diagnoses, stage 3 diagnoses, stage 4 diagnoses, and stage 5 diagnoses.

The decision flow proceeds as follows:

Tier 1 Diagnosis: If present at any point during the disease progression, the tier 1 diagnosis is recorded as the index event, and the date of the cohort entry is the index day. Tier 1 diagnoses include ArLD, NAFLD, metabolic liver disease, HBV, HCV, autoimmune liver disease, and haemochromatosis.

Tier 2 Diagnosis: If there is no tier 1 diagnosis, the presence of a tier 2 diagnosis is checked. If present, it is recorded as the index event, and the date of cohort entry is the index day. Tier 2 diagnoses include hepatitis not specified, congestive hepatopathy, and toxic liver disease.

Tier 3 Diagnosis: In the absence of tier 1 and tier 2 diagnoses, the presence of a tier 3 diagnosis is checked. If found, it is recorded as the index event. Tier 3 diagnoses include miscellaneous conditions.

Stage 2 Diagnosis: If none of the Stage 1 diagnoses are present, the presence of a stage 2 diagnosis is considered. If found, it is recorded as the index event. Stage 2 diagnoses include hepatic fibrosis and cirrhosis.

Stage 3 Diagnosis: If no stage 1 or stage 2 diagnoses are present, a stage 3 diagnosis is checked next. If found, it is recorded as the index event. Stage 3 diagnoses include portal hypertension.

Stage 4 Diagnosis: If the stage 1, stage 2, and stage 3 diagnoses are absent, a stage 4 diagnosis is checked. If present, it is recorded as the index event. Stage 4 diagnoses include hepatic decompensation.

Stage 5 Diagnosis: If no above diagnoses are found, a stage 5 diagnosis is considered. If found, it is recorded as the index event. Stage 5 diagnoses include hepatocellular carcinoma, intrahepatic cholangiocarcinoma, other primary liver cancer.

Acute Liver Diagnosis: If none of the chronic diagnoses are present, an acute liver diagnosis is checked. If found, it is recorded as the index event. Acute liver diagnoses include acute viral hepatitis, Budd-Chiari, acute liver failure, infections/sepsis, and other non-specified acute liver injuries.

Supplemental Tables

Supplemental Table 1 Code list for identifying liver disease

Phenotype	ICD10 codes	Read codes
Acute viral hepatitis	B15, B19, B16, B17 (B171 excluded), B172, B178, B179, B180, B159, B169, B199	A70z1, AyuB0, XE2u., Q4090, A700., A701., A7052, A70.., A706., A708., A709., A70z., A70G., A704., AyuB3
Acute liver failure	K720	J6000
Budd-Chiari	I820	G820.
Infection/infarction	K750, K763, K751	J62.., J620., J6200, J6201, J6202, J6203, J6204, J620z, A053., J634., J621.
Other non-specified acute liver injuries	K752	J63y1
Autoimmune liver disease	K754, K743, K831, K753	J63B., J6141, J6160, J6617, J63X.
Haemochromatosis	E831	C3500
Metabolic liver disease	E880, E830	C3762, C3761, C3510
HBV *	B181, B180	
HCV *	B182	A70E., A70F.
Alcohol-related liver disease	K70	J613., J6130, J612., J6120, J610., J617., J6170, J611.
Non-alcoholic fatty liver disease	K760, K7581	J61y1, J61y8
Hepatitis not specified	K769, K7589, K73	Jyu72, J614y
Congestive hepatopathy	K761, K762, K765	J630., J636., J637.
Toxic liver disease	K71	J635., J6350, J6351, J63252, J6353, J6354, J6355, J6356, J6357, J635X
Miscellaneous	K764, K768, K77	J638., Jyu73, J63yz, Jyu75
Hepatic fibrosis	K740, K741, K742	J61y4, J61y6, J61y5
Cirrhosis	K703, K744, K745, K746, K749	J6161, J616z, J615z
Portal hypertension	K766, I81, I859, I982, I85	J623., G81.., G8523, G852., G8521, G8522, G852z

Hepatic decompensation	K721, K767, I850, K72, C220	J624., SP143, G850., G8520, J625., B1503, BB5D7, BB5D5, BB5D8
Hepatocellular carcinoma (HCC)	C220	B1503, BB5D7, BB5D5, BB5D8
Intrahepatic cholangio carcinoma (ICC)	C221	B150.
Other primary liver cancers	C222, C223, C224, C225, C226. C227	B808.

*We identified Read codes(A7071, A7073, ZV02B, Q4091, 43B4., A7070, A7051, A7072, A70z0, A70A., A70B., A70C., A70D., A70E.,A70F., ZV02C) and ICD-10 codes (B180, B181, B182) for HBV and HCV. However, in order to complain with Data Protection Act 2018 and the UK General Data Protection Regulation, we could not include Read codes (A7071, A7073, ZV02B, Q4091, 43B4., A7070, A7051, A7072, A70z0, A70A., A70B., A70C., A70D., ZV02C) and ICD-10 codes (B171) as these were flagged as sensitive in the latest version of known sensitive code list of SAIL Databank.

Supplemental Table 2-1. Standardised incidence rate of liver disease by data sources (2004 to 2022)

Year	PEDW only group		WLGP only group		ADDE only group		Two or more data sources	
	STD incidence	STD 95% CI	STD incidence	STD 95% CI	STD incidence	STD 95% CI	STD incidence	STD 95% CI
2004	41.2	(38.9 to 43.5)	13.5	(12.2 to 14.9)	3.8	(3.1 to 4.5)	39.2	(37.1 to 41.5)
2005	49.6	(47.2 to 52.1)	15.3	(13.9 to 16.8)	3.8	(3.1 to 4.5)	40.7	(38.5 to 43)
2006	54.4	(51.9 to 57.1)	16.1	(14.7 to 17.6)	3.7	(3.1 to 4.5)	44.4	(42.1 to 46.9)
2007	54.8	(52.2 to 57.5)	13.9	(12.6 to 15.3)	4.5	(3.7 to 5.3)	41.9	(39.6 to 44.3)
2008	59.3	(56.6 to 62.1)	18.0	(16.5 to 19.6)	5.5	(4.7 to 6.4)	43.5	(41.2 to 45.9)
2009	63.9	(61.1 to 66.9)	18.5	(17 to 20.1)	5.5	(4.7 to 6.5)	43.7	(41.4 to 46.2)
2010	67.4	(64.5 to 70.4)	21.7	(20.1 to 23.4)	5.6	(4.7 to 6.5)	45.9	(43.5 to 48.4)
2011	74.1	(71.1 to 77.3)	23.6	(21.8 to 25.4)	5.1	(4.3 to 6)	47.1	(44.7 to 49.6)
2012	76.3	(73.2 to 79.5)	21.5	(19.9 to 23.2)	5.5	(4.7 to 6.4)	47.3	(44.9 to 49.9)
2013	82.7	(79.4 to 86)	22.3	(20.6 to 24.1)	6.0	(5.1 to 6.9)	47.1	(44.6 to 49.6)
2014	90.3	(86.9 to 93.8)	26.2	(24.4 to 28.1)	6.1	(5.3 to 7.1)	49.8	(47.3 to 52.4)
2015	99.1	(95.5 to 102.8)	32.2	(30.2 to 34.3)	6.2	(5.3 to 7.2)	54.7	(52.1 to 57.5)
2016	115.9	(112 to 120)	45.1	(42.7 to 47.6)	6.4	(5.5 to 7.4)	58.3	(55.5 to 61.1)
2017	122.9	(118.8 to 127.2)	48.9	(46.4 to 51.5)	6.5	(5.5 to 7.5)	64.9	(62 to 67.9)
2018	141.0	(136.5 to 145.5)	62.5	(59.7 to 65.5)	7.1	(6.1 to 8.2)	68.5	(65.5 to 71.7)
2019	157.6	(152.9 to 162.4)	71.4	(68.4 to 74.6)	6.4	(5.4 to 7.4)	65.8	(62.8 to 68.9)
2020	153.3	(148.5 to 158.1)	50.2	(47.6 to 52.8)	8.7	(7.6 to 9.8)	51.9	(49.2 to 54.7)
2021	197.5	(192.1 to 203)	73.4	(70.2 to 76.6)	7.5	(6.5 to 8.6)	53.5	(50.8 to 56.4)
2022	182.1	(176.9 to 187.4)	88.6	(85.1 to 92.1)	7.9	(6.9 to 9.1)	37.6	(35.2 to 40.1)

Abbreviation: STD: standardized; CI: confidence interval

Supplemental Table 2-2 Standardised incidence rate of liver disease by disease stages (2004 to 2022)

Year	Stage 1		Stage 2		Stage 3		Stage 4		Stage 5	
	STD incidence	STD 95% CI	STD incidence	STD 95% CI	STD incidence	STD 95% CI	STD incidence	STD 95% CI	STD incidence	STD 95% CI
2004	70.3	(67.4 to 73.3)	5.8	(5 to 6.7)	5.1	(4.3 to 6)	4.9	(4.2 to 5.8)	2.7	(2.1 to 3.3)
2005	79.6	(76.4 to 82.8)	6.8	(5.9 to 7.8)	5.2	(4.4 to 6)	5.3	(4.5 to 6.2)	2.9	(2.3 to 3.6)
2006	86.9	(83.6 to 90.3)	7.1	(6.2 to 8.1)	6.3	(5.5 to 7.3)	6.2	(5.3 to 7.2)	3.5	(2.9 to 4.2)
2007	84.0	(80.8 to 87.4)	7.5	(6.5 to 8.5)	5.7	(4.9 to 6.6)	6.4	(5.5 to 7.3)	3.5	(2.9 to 4.2)
2008	92.3	(88.9 to 95.9)	8.1	(7.1 to 9.2)	8.2	(7.3 to 9.3)	5.7	(4.9 to 6.6)	4.4	(3.7 to 5.2)
2009	97.8	(94.3 to 101.4)	8.6	(7.6 to 9.7)	7.6	(6.6 to 8.6)	6.4	(5.5 to 7.4)	4.1	(3.4 to 4.9)
2010	106.0	(102.3 to 109.7)	8.1	(7.1 to 9.1)	8.3	(7.3 to 9.4)	6.5	(5.6 to 7.5)	3.6	(2.9 to 4.3)
2011	111.6	(107.8 to 115.5)	10.1	(9 to 11.3)	8.4	(7.4 to 9.5)	6.6	(5.7 to 7.6)	4.6	(3.9 to 5.5)
2012	112.3	(108.5 to 116.1)	9.5	(8.4 to 10.7)	8.2	(7.2 to 9.3)	6.6	(5.7 to 7.6)	4.7	(4 to 5.6)
2013	115.4	(111.6 to 119.4)	10.3	(9.1 to 11.5)	8.9	(7.9 to 10.1)	7.4	(6.4 to 8.5)	6.0	(5.2 to 7)
2014	127.9	(123.9 to 132.1)	11.6	(10.4 to 12.9)	9.4	(8.4 to 10.6)	7.1	(6.2 to 8.2)	5.3	(4.5 to 6.2)
2015	141.9	(137.6 to 146.3)	14.0	(12.7 to 15.5)	12.5	(11.2 to 13.8)	6.5	(5.6 to 7.5)	6.3	(5.4 to 7.3)
2016	170.5	(165.7 to 175.3)	16.2	(14.7 to 17.7)	13.2	(11.9 to 14.7)	8.1	(7.1 to 9.2)	6.4	(5.5 to 7.5)
2017	185.5	(180.5 to 190.6)	17.8	(16.2 to 19.4)	13.5	(12.2 to 14.9)	8.0	(7 to 9.2)	6.7	(5.7 to 7.8)
2018	212.9	(207.5 to 218.4)	21.6	(19.9 to 23.4)	14.7	(13.3 to 16.2)	9.2	(8.1 to 10.3)	7.3	(6.3 to 8.5)
2019	231.8	(226.2 to 237.6)	22.1	(20.4 to 24)	17.8	(16.2 to 19.5)	8.9	(7.8 to 10.1)	6.6	(5.7 to 7.6)
2020	201.7	(196.3 to 207.1)	19.5	(17.8 to 21.2)	14.3	(12.9 to 15.8)	8.9	(7.8 to 10.2)	7.7	(6.6 to 8.9)
2021	263.0	(256.9 to 269.3)	21.6	(19.9 to 23.5)	18.3	(16.7 to 20.1)	8.2	(7.2 to 9.4)	7.9	(6.8 to 9.1)
2022	244.7	(238.7 to 250.7)	23.5	(21.6 to 25.5)	19.9	(18.2 to 21.8)	7.5	(6.4 to 8.8)	8.7	(7.6 to 10)

Abbreviation: STD: standardized; CI: confidence interval

Supplemental Table 2-3, Standardised incidence rate of liver disease by aetiologies (2004 to 2022)

Year	ArLD	NAFLD	Metabolic liver disease	HBV	HCV	Autoimmune liver disease	Haemochromatosis	ArLD overlap	Non-ArLD overlap	Hepatitis of unspecified origin	Congestive hepatopathy	Toxic liver disease	Miscellaneous
	STD incidence (95% CI)	STD incidence (95% CI)	STD incidence (95% CI)	STD incidence (95% CI)	STD incidence (95% CI)	STD incidence (95% CI)	STD incidence (95% CI)	STD incidence (95% CI)	STD incidence (95% CI)	STD incidence (95% CI)	STD incidence (95% CI)	STD incidence (95% CI)	STD incidence (95% CI)
2004	22.0 (20.3 to 23.7)	10.8 (9.6 to 12)	4.4 (3.7 to 5.2)	0.6 (0.4 to 0.9)	3.0 (2.4 to 3.7)	14.4 (13.2 to 15.8)	3.2 (2.6 to 3.9)	0.3 (0.2 to 0.6)	0.4 (0.2 to 0.7)	3.6 (2.2 to 5.1)	0.6 (0.4 to 0.9)	1.8 (1.3 to 2.3)	5.3 (4.5 to 6.1)
2005	21.7 (20.1 to 23.5)	14.0 (12.7 to 15.4)	4.7 (3.9 to 5.5)	0.6 (0.3 to 1)	2.9 (2.3 to 3.5)	18.4 (16.9 to 19.9)	4.0 (3.3 to 4.8)	0.4 (0.2 to 0.7)	0.4 (0.2 to 0.7)	4.0 (3.1 to 4.8)	0.5 (0.3 to 0.8)	2.3 (1.8 to 2.9)	5.8 (5 to 6.7)
2006	21.9 (20.2 to 23.6)	17.9 (16.4 to 19.5)	5.1 (4.4 to 6)	0.7 (0.4 to 1.1)	3.9 (3.2 to 4.7)	17.7 (16.2 to 19.2)	5.1 (4.3 to 6)	--	0.8 (0.5 to 1.2)	4.4 (3.3 to 5.6)	0.6 (0.3 to 0.9)	1.9 (1.4 to 2.5)	6.3 (5.4 to 7.2)
2007	22.3 (20.6 to 24)	17.1 (15.7 to 18.6)	5.1 (4.4 to 6)	0.6 (0.4 to 1)	2.8 (2.2 to 3.5)	17.1 (15.7 to 18.6)	4.3 (3.6 to 5.1)	--	0.8 (0.5 to 1.2)	4.9 (4.2 to 5.8)	0.6 (0.3 to 0.9)	1.6 (1.2 to 2.1)	6.5 (5.6 to 7.5)
2008	20.2 (18.6 to 21.9)	21.2 (19.6 to 22.9)	4.8 (4 to 5.6)	0.8 (0.5 to 1.2)	3.8 (3.1 to 4.6)	20.1 (18.6 to 21.8)	4.8 (4 to 5.6)	--	0.6 (0.4 to 1)	6.3 (5.4 to 7.2)	0.8 (0.5 to 1.2)	1.2 (0.9 to 1.7)	7.4 (6.5 to 8.4)
2009	21.1 (19.5 to 22.8)	23.3 (21.6 to 25.1)	5.5 (4.7 to 6.4)	0.7 (0.4 to 1.1)	3.2 (2.6 to 4)	23.8 (22.1 to 25.6)	5.6 (4.8 to 6.6)	0.5 (0.2 to 0.8)	0.5 (0.3 to 0.9)	4.5 (3.8 to 5.3)	0.6 (0.3 to 1)	1.0 (0.7 to 1.4)	7.5 (6.6 to 8.6)
2010	20.9 (19.2 to 22.6)	28.8 (26.9 to 30.8)	5.2 (4.4 to 6)	0.8 (0.5 to 1.1)	3.4 (2.7 to 4.1)	25.9 (24.1 to 27.7)	5.7 (4.9 to 6.6)	0.3 (0.2 to 0.6)	0.5 (0.3 to 0.8)	5.3 (4.5 to 6.2)	0.7 (0.4 to 1.1)	0.9 (0.6 to 1.3)	7.8 (6.8 to 8.8)
2011	18.8 (17.3 to 20.4)	32.6 (30.6 to 34.7)	6.4 (5.5 to 7.4)	1.3 (0.9 to 1.8)	3.6 (3 to 4.4)	24.4 (22.7 to 26.3)	5.8 (5 to 6.7)	0.4 (0.2 to 0.7)	0.6 (0.4 to 1)	5.9 (5 to 6.8)	0.4 (0.2 to 0.7)	1.5 (1.1 to 2)	9.8 (8.7 to 11)
2012	17.3 (15.9 to 18.9)	32.6 (30.6 to 34.7)	8.1 (7.1 to 9.2)	0.8 (0.5 to 1.2)	4.2 (3.5 to 5.1)	23.0 (21.3 to 24.8)	6.3 (5.5 to 7.3)	0.4 (0.2 to 0.7)	0.6 (0.4 to 1)	6.2 (5.3 to 7.1)	0.8 (0.5 to 1.2)	1.4 (1 to 1.9)	10.3 (9.2 to 11.6)
2013	17.8 (16.3 to 19.4)	36.0 (33.9 to 38.2)	8.6 (7.5 to 9.7)	1.0 (0.7 to 1.5)	4.7 (4 to 5.6)	18.4 (16.9 to 20)	5.8 (5 to 6.7)	0.4 (0.2 to 0.7)	0.6 (0.4 to 1)	8.3 (7.3 to 9.4)	0.8 (0.5 to 1.2)	1.1 (0.7 to 1.5)	11.9 (10.7 to 13.2)
2014	17.3 (15.9 to 18.9)	45.2 (42.8 to 47.7)	8.4 (7.4 to 9.5)	0.7 (0.4 to 1.1)	5.7 (4.9 to 6.6)	17.8 (16.3 to 19.3)	6.5 (5.6 to 7.4)	0.3 (0.2 to 0.6)	0.8 (0.5 to 1.1)	7.8 (6.8 to 8.9)	0.9 (0.6 to 1.3)	1.0 (0.6 to 1.4)	15.6 (14.2 to 17.1)

2015	17.8 (16.3 to 19.4)	54.7 (52.1 to 57.5)	9.0 (7.9 to 10.2)	0.9 (0.6 to 1.4)	5.2 (4.4 to 6.1)	19.8 (18.2 to 21.5)	6.0 (5.2 to 7)	0.4 (0.2 to 0.7)	0.5 (0.2 to 0.8)	8.4 (7.4 to 9.6)	1.0 (0.6 to 1.4)	1.1 (0.7 to 1.6)	17.0 (15.6 to 18.6)
2016	18.7 (17.2 to 20.3)	69.5 (66.6 to 72.6)	12.6 (11.3 to 14.1)	1.9 (1.4 to 2.5)	6.9 (6 to 8)	19.8 (18.2 to 21.6)	7.6 (6.6 to 8.7)	0.6 (0.4 to 1)	1.0 (0.7 to 1.4)	10.2 (8.8 to 11.8)	1.2 (0.8 to 1.7)	1.2 (0.8 to 1.6)	19.4 (17.7 to 21.1)
2017	20.3 (18.6 to 22)	81 (77.7 to 84.3)	11.8 (10.5 to 13.2)	2.6 (2 to 3.2)	7.0 (6.1 to 8.1)	18.5 (16.9 to 20.2)	7.6 (6.6 to 8.7)	0.9 (0.6 to 1.3)	1.0 (0.7 to 1.5)	11.8 (10.2 to 13.6)	1.2 (0.9 to 1.7)	1.3 (0.9 to 1.7)	21.1 (19.4 to 22.9)
2018	19.8 (18.2 to 21.5)	100.5 (96.9 to 104.3)	12.6 (11.3 to 14.1)	2.0 (1.5 to 2.5)	5.9 (5 to 6.9)	20.3 (18.6 to 22.1)	8.5 (7.4 to 9.6)	0.8 (0.5 to 1.2)	1.0 (0.7 to 1.4)	12.2 (10.6 to 13.9)	1.4 (0.9 to 1.9)	1.7 (1.2 to 2.2)	26.1 (24.2 to 28.1)
2019	21.7 (20 to 23.5)	114.9 (110.9 to 118.9)	13.0 (11.6 to 14.5)	2.5 (2 to 3.2)	5.6 (4.8 to 6.6)	20.3 (18.6 to 22.1)	8.9 (7.8 to 10.1)	0.8 (0.5 to 1.2)	0.8 (0.5 to 1.2)	12.2 (10.6 to 13.9)	1.5 (1 to 2)	1.2 (0.8 to 1.6)	27.9 (25.9 to 30)
2020	19.9 (18.3 to 21.6)	90.1 (86.6 to 93.7)	11.5 (10.2 to 12.9)	1.6 (1.2 to 2.2)	5.1 (4.3 to 6)	16.2 (14.7 to 17.9)	7.8 (6.8 to 8.9)	0.6 (0.4 to 1)	0.9 (0.5 to 1.3)	13.3 (11.5 to 15.3)	1.6 (1.2 to 2.2)	1.8 (1.3 to 2.4)	31.6 (29.4 to 33.9)
2021	22.2 (20.4 to 24)	129.6 (125.4 to 133.9)	14.1 (12.7 to 15.7)	1.7 (1.3 to 2.3)	3.6 (2.9 to 4.3)	19.4 (17.7 to 21.2)	8.4 (7.4 to 9.6)	0.5 (0.2 to 0.8)	1.4 (1 to 1.9)	16.5 (14.5 to 18)	1.4 (0.9 to 2)	1.5 (1.1 to 2.1)	42.9 (40.4 to 45.5)
2022	22.4 (20.7 to 24.2)	124.5 (120.4 to 128.8)	15.2 (13.7 to 16.9)	1.2 (0.9 to 1.7)	2.3 (1.8 to 3)	17.0 (15.5 to 18.7)	9.2 (8 to 10.4)	0.4 (0.2 to 0.8)	1.3 (0.9 to 1.8)	14.4 (13 to 16.3)	1.7 (1.2 to 2.4)	1.2 (0.8 to 1.7)	33.5 (31.2 to 35.9)

Abbreviation: ArLD: alcohol-related liver disease; NAFLD: non-alcohol fatty liver disease; HBV: hepatitis B virus; HCV: hepatitis C virus; STD: standardised; CI: confidence interval

Supplemental Table 3-1, Standardised incidence rate of NAFLD by data sources (2004 to 2022)

Year	PEDW-only		WLGP-only		ADDE-only		Two or more data sources	
	STD incidence	STD 95% CI	STD incidence	STD 95% CI	STD incidence	STD 95% CI	STD incidence	STD 95% CI
2004	3.4	(2.8 to 4.1)	4.2	(3.5 to 5)	0.5	(0.3 to 0.8)	2.7	(2.1 to 3.3)
2005	4.6	(3.9 to 5.4)	5.2	(4.4 to 6.1)	0.3	(0.2 to 0.6)	3.9	(3.2 to 4.6)
2006	6.0	(5.2 to 7)	7.2	(6.3 to 8.3)	--	--	4.4	(3.7 to 5.3)
2007	6.1	(5.2 to 7)	6.7	(5.8 to 7.7)	0.3	(0.2 to 0.6)	4.0	(3.3 to 4.8)
2008	6.0	(5.1 to 6.9)	10.4	(9.2 to 11.6)	0.6	(0.4 to 0.9)	4.2	(3.5 to 5)
2009	6.6	(5.7 to 7.6)	10.7	(9.6 to 12)	0.6	(0.3 to 0.9)	5.4	(4.6 to 6.3)
2010	6.8	(5.9 to 7.8)	14.9	(13.5 to 16.3)	0.7	(0.4 to 1.1)	6.4	(5.6 to 7.4)
2011	9.8	(8.7 to 11)	15.4	(14 to 16.9)	0.4	(0.2 to 0.7)	7.0	(6.1 to 8)
2012	11.6	(10.4 to 12.8)	13.8	(12.5 to 15.2)	0.4	(0.2 to 0.7)	6.9	(6 to 7.9)
2013	13.0	(11.8 to 14.4)	15.2	(13.8 to 16.7)	1.0	(0.7 to 1.4)	6.7	(5.8 to 7.7)
2014	17.6	(16.2 to 19.2)	18.2	(16.7 to 19.8)	1.1	(0.7 to 1.5)	8.2	(7.2 to 9.3)
2015	21.0	(19.4 to 22.8)	23.3	(21.6 to 25.1)	1.0	(0.7 to 1.4)	9.4	(8.3 to 10.6)
2016	24.2	(22.4 to 26)	32.5	(30.5 to 34.7)	1.2	(0.8 to 1.6)	11.7	(10.5 to 13)
2017	29.7	(27.7 to 31.7)	35.7	(33.6 to 38)	1.3	(0.9 to 1.8)	14.2	(12.9 to 15.7)
2018	38.9	(36.7 to 41.3)	45.3	(42.8 to 47.8)	1.6	(1.1 to 2.1)	14.7	(13.4 to 16.2)
2019	45.1	(42.7 to 47.7)	53.4	(50.8 to 56.2)	1.6	(1.2 to 2.1)	14.7	(13.3 to 16.2)
2020	45.0	(42.5 to 47.6)	34.0	(31.9 to 36.2)	1.9	(1.4 to 2.5)	9.1	(8.1 to 10.3)
2021	65.2	(62.2 to 68.3)	52.9	(50.2 to 55.6)	1.5	(1.1 to 2.1)	10.0	(8.8 to 11.2)
2022	60.5	(57.6 to 63.5)	58.7	(55.9 to 61.6)	1.5	(1.1 to 2)	3.9	(3.2 to 4.7)

Abbreviation: STD: standardized; CI: confidence interval

Supplemental Table 3-2, Standardised incidence rate of ArLD by data sources (2004 to 2022)

Year	PEDW-only		WLGP-only		ADDE-only		Two or more data sources	
	STD incidence	STD 95% CI	STD incidence	STD 95% CI	STD incidence	STD 95% CI	STD incidence	STD 95% CI
2004	4.3	(3.6 to 5.1)	3.0	(2.4 to 3.7)	--	--	14.3	(13 to 15.8)
2005	4.9	(4.2 to 5.8)	2.7	(2.2 to 3.4)	0.6	(0.3 to 1)	13.4	(12.1 to 14.8)
2006	5.0	(4.3 to 5.9)	2.7	(2.1 to 3.3)	0.3	(0.2 to 0.6)	13.8	(12.5 to 15.2)
2007	5.4	(4.6 to 6.3)	2.3	(1.8 to 2.9)	0.6	(0.4 to 1)	14.0	(12.7 to 15.4)
2008	4.8	(4.1 to 5.7)	2.0	(1.5 to 2.6)	0.7	(0.4 to 1.1)	12.7	(11.4 to 14)
2009	5.7	(4.8 to 6.6)	2.1	(1.6 to 2.6)	0.9	(0.6 to 1.3)	12.5	(11.2 to 13.8)
2010	5.1	(4.3 to 6)	2.5	(1.9 to 3.1)	0.9	(0.6 to 1.3)	12.5	(11.2 to 13.8)
2011	4.5	(3.8 to 5.3)	2.5	(1.9 to 3.1)	0.7	(0.4 to 1.1)	11.1	(9.9 to 12.4)
2012	4.7	(3.9 to 5.5)	1.8	(1.4 to 2.4)	0.7	(0.4 to 1.1)	10.1	(9 to 11.3)
2013	4.8	(4 to 5.6)	2.0	(1.5 to 2.6)	0.6	(0.4 to 1)	10.4	(9.2 to 11.6)
2014	4.7	(3.9 to 5.5)	2.2	(1.7 to 2.8)	0.6	(0.4 to 1)	9.9	(8.8 to 11.1)
2015	4.4	(3.7 to 5.3)	2.3	(1.8 to 3)	0.7	(0.4 to 1.1)	10.4	(9.2 to 11.6)
2016	4.7	(3.9 to 5.5)	3.9	(3.2 to 4.7)	0.9	(0.6 to 1.3)	9.3	(8.2 to 10.4)
2017	4.3	(3.6 to 5.1)	4.5	(3.7 to 5.3)	0.6	(0.4 to 1)	10.9	(9.7 to 12.1)
2018	4.3	(3.5 to 5.1)	5.2	(4.4 to 6.1)	0.7	(0.4 to 1)	9.7	(8.6 to 10.9)
2019	4.8	(4.1 to 5.7)	5.6	(4.8 to 6.6)	1.0	(0.6 to 1.4)	10.3	(9.2 to 11.6)
2020	5.4	(4.6 to 6.4)	5.0	(4.2 to 5.9)	1.1	(0.7 to 1.5)	8.4	(7.3 to 9.5)
2021	4.8	(4 to 5.7)	7.5	(6.5 to 8.6)	1.3	(0.9 to 1.8)	8.5	(7.5 to 9.7)
2022	6.1	(5.2 to 7.1)	9.7	(8.5 to 10.9)	1.1	(0.8 to 1.6)	5.5	(4.7 to 6.5)

Abbreviation: STD: standardized; CI: confidence interval

SAIL project number	1492
SAIL project title	Liver Disease Cymru Partnership (LDCP)
Output title (draft)	Epidemiology of liver disease and associated patient characteristics in Wales from 2004 to 2022: a retrospective population-scale observational study
Lead author	Jingwei Gao
Co-author(s)	Haroon Ahmed, Ashley Akbari, Katherine Cullen, Aled Davies, Deborah Fitzsimmons, Kerry Hood, Tom Pembroke, Andrew Yeoman
Objective	We aim to describe incidence rate and the key clinical and sociodemographic features associated with the cohort of individuals diagnosed with liver disease in the population of Wales. ➢ We will describe the creation and methodology of the Liver Disease Cymru Partnership (LDCP) cohort and document the process and methodology. ➢ We will report the standardised incidence rate of liver disease in the population of Wales ➢ We will describe the underlying liver disease condition, demographic and socioeconomic characteristics, and clinical features of the individuals diagnosed with liver disease
Study design	The study is designed as retrospective observational study. The study cohort will include all individuals who are residents in Wales and registered with a SAIL providing general practice (GP), and were diagnosed with liver disease within the study period. A list of phenotypes to identify liver disease diagnoses is available in the **Supplementary Materials.** All missing key identifiers and linkage fields will be cleaned, and any rows/records will be removed, with other missing data reported as unknown.
Coverage:	The study period is between 2004-01-01 to the most recent available data (currently 2022-12-31). For all individuals identified within the cohort study, we will extract all available data back to 1994-01-01 to track for prevalent conditions and services received in all data sources.
Data sources	The following data sources have been approved for access by this project from the SAIL IGRP: • **ADDE** - Annual District Death Extract • **CCDS** - Critical Care Dataset • **CNIS** - Cancer Network Information System Cymru (CANISC) • **CVSP** - COVID-19 Shielded People list • **CVVD** - COVID Vaccine Data • **EDDS** - Emergency Department Dataset • **ICNC** - Intensive Care National Audit & Research Centre (ICNARC) • **OPDW** - NHS Hospital Outpatients • **PATD** - COVID-19 Pathology Test results • **PEDW** - Patient Episode Database for Wales

	• **RTDS** - Radiotherapy Data Set • **SACT** - Systemic Anti-Cancer Therapy Dataset • **WCSU** - WCISU (Welsh Cancer Intelligence Surveillance Unit) • **WDSD** - Welsh Demographic Service • **WLGP** - GP Primary Care – Audit • **WRRS** - Wales Results Reporting Service
Exposure of interest	Liver disease diagnoses during study period (2004.01.01 – 2022.12.31)
Methodology	**Liver disease aetiologies** We applied a hierarchy of 3 tiers of potential aetiological diagnoses and 5 discrete stages of chronic liver disease based on perceived clinical importance and natural history of liver disease progression as was described in our previous study. A list of ICD-10 and Read v2 codes to identify liver disease in different stages and tiers can be found in supplemental table 1.
Outcome of interest	**Incidence rate of liver disease** Our primary outcome is the incidence rate of liver disease. The incidence rate of liver disease will be reported in each calendar years by different data sources, liver disease aetiologies, and liver disease stages. **Underlying condition** We defined the first chronic or acute liver disease diagnosis during study period as the index event. However, some individuals may present earlier stage or aetiological diagnoses late due to delay of test results. Therefore, we define the lowest tier/stage of liver disease diagnosis throughout the disease progression as the underlying liver disease condition. The decision rules of identifying underlying conditions are available in supplemental fig 1. **Demographic characteristics and clinical features** We collected data on demographic characteristics such as age, sex, and WIMD 2019 quintile. The collected clinical features include cardiovascular disease (CVD) related conditions (heart failure, transient ischaemic, other ischaemic disease, atrial fibrillation, peripheral vascular disease, angina, and stroke), diabetes, hypertension, and drug prescription such as antihypertensive usage. The code list for identifying associated comorbidities can be found in our GitHub repository (URL: https://github.com/SwanseaUniversityDataScience/1492-LDCP).
Statistical analysis	**Descriptive statistics**: We will apply descriptive analysis to report the number and frequency of demographic characteristics, underlying liver disease conditions, and associated comorbidities. **Incidence of liver disease:** The incidence rate was calculated as the number of incident cases divided by the number of Welsh residents. An incident case was defined as an individual having a first liver-related diagnosis during the study period (01.01.2004-31.12.2022) whilst having no prior liver disease history from 1st January 1994 to 31st December 2003. The incidence rate was directly standardised using the European Union (EU) standard population 2013 and age categorised into five-year age bands. The crude and standardised incidence rates were reported per 100,000 inhabitants in each calendar year from 2004 to 2022 with corresponding 95% confidence intervals (CI).

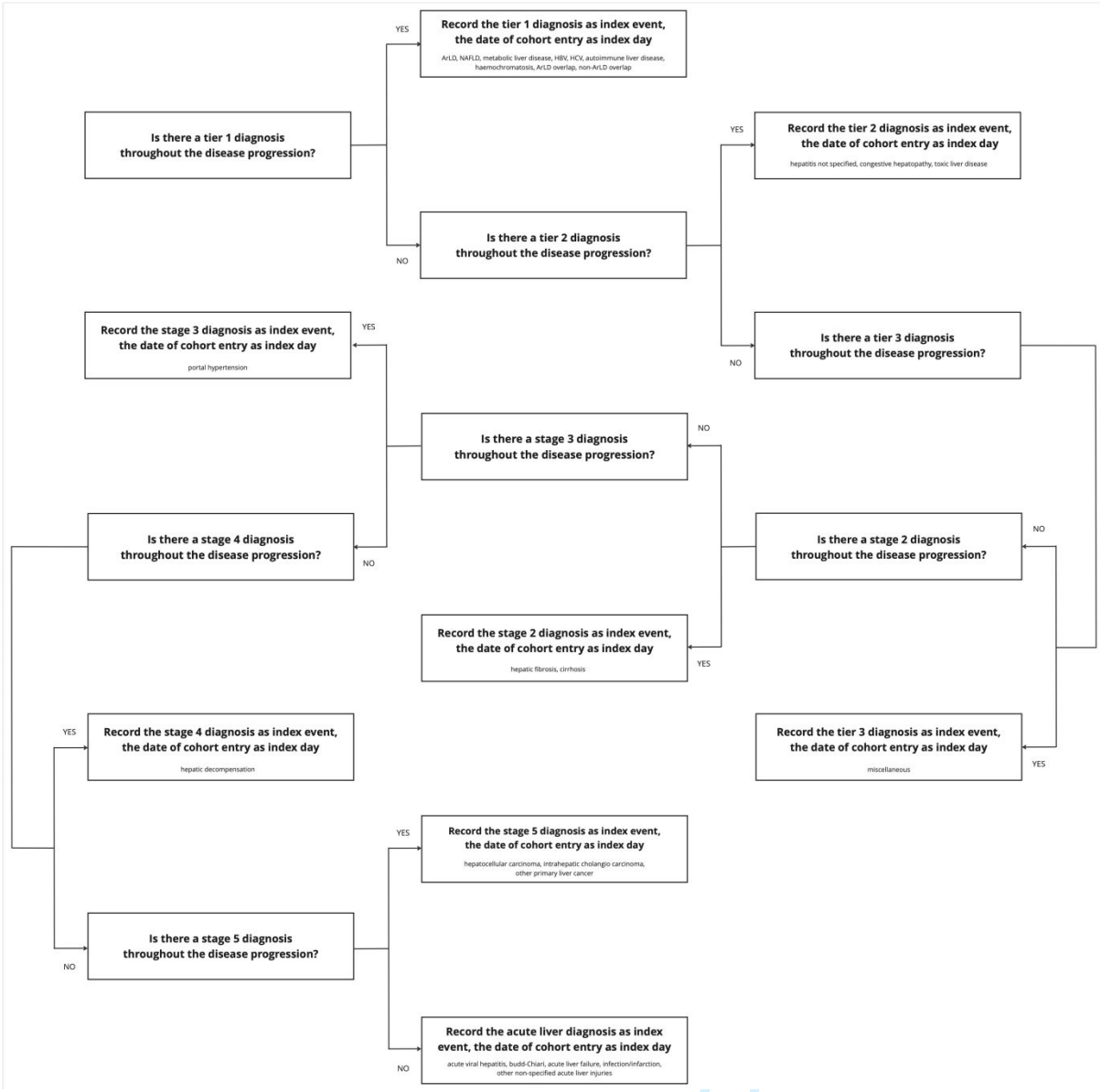
Output plan:	This protocol and specification will lead to an academic output submitted to a peer-reviewed journal (the decision on uploading a pre-print is to be decided by the group after completion of the work and at the time of submission). This document can be shared with all project members and other suitable invited stakeholders, which the project team feels is appropriate to provide their input into planning the work and resulting output. All draft reports while the work is in iterative development will be shared with the project team, along with the final output for input and opt-in for authorship before submission. Results will be reported according to the Strengthening the Reporting of Observational studies in Epidemiology (STROBE) guidelines [2].
References	[1] Pembroke TPI, John G, Puyk B, Howkins K, Clarke R, Yousuf F, Czajkowski M, Godkin A, Salmon J, Yeoman A. Rising incidence, progression and changing patterns of liver disease in Wales 1999-2019. World J Hepatol. 2023 Jan 27;15(1):89-106. [2] von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. Int J Surg. 2014 Dec;12(12):1495-9. [3] Hagström H, Adams LA, Allen AM, Byrne CD, Chang Y, Grønbaek H, Ismail M, Jepsen P, Kanwal F, Kramer J, Lazarus JV, Long MT, Loomba R, Newsome PN, Rowe IA, Ryu S, Schattenberg JM, Serper M, Sheron N, Simon TG, Tapper EB, Wild S, Wong VW, Yilmaz Y, Zelber-Sagi S, Åberg F. Administrative Coding in Electronic Health Care Record-Based Research of NAFLD: An Expert Panel Consensus Statement. Hepatology. 2021 Jul;74(1):474-482.

Supplemental Material

Supplemental Table 1 Code list for identifying liver disease

Stage	Tiers	Conditions	ICD-10 codes	Read code
Stage 1	Tier 1	Autoimmune liver disease	K754, K743, K831, K753	J63B., J6141, J6160, J6617, J63X.
		Haemochromatosis	E831	C3500
		Metabolic liver disease	E880, E830	C3762, C3761, C3510
		HBV	B181, B180	A7071, A7073, ZV02B, Q4091, 43B4., A7070, A7051
		HCV	B182	A70E., A70F., A7072, A70z0, A70A., A70B., A70C., A70D., ZV02C
		Alcohol-related liver disease (ArLD)	K70	J613., J6130, J612., J6120, J610., J617., J6170, J611.
		Non-alcoholic fatty liver disease (NAFLD)	K760, K7581	J61y1, J61y8
	Tier 2	Hepatitis not specified	K769, K7589, K73	Jyu72, J614y
		Congestive hepatopathy	K761, K762, K765	J630., J636., J637.
		Toxic liver disease	K71	J635., J6350, J6351, J63252, J6353, J6354, J6355, J6356, J6357, J635X
	Tier 3	Miscellaneous	K764, K768, K77	J638., Jyu73, J63yz, Jyu75
Stage 2		Cirrhosis	K703, K744, K745, K746, K749	J6161, J616z, J615z

		Hepatic fibrosis	K740, K741, K742	J61y4, J61y6, J61y5
Stage 3		Portal hypertension	K766, I81, I859, I982, I85	J623., G81.., G8523, G852., G8521, G8522, G852z
Stage 4		Hepatic decompensation	K721, K767, I850, K72	J624., SP143, G850., G8520, J625.
Stage 5		Hepatocellular carcinoma (HCC)	C220	B1503, BB5D7, BB5D5, BB5D8
		Intrahepatic cholangio carcinoma (ICC)	C221	B150.
		Other primary liver cancers	C222, C223, C224, C225, C226. C227	B808.
Acute liver disease		Acute viral hepatitis	B15, B19, B16, B17, B172, B178, B179, B180, B159, B169, B199	A70z1, AyuB0, XE2u., Q4090, A700., 43B5., A701., A7052, A70.., A706., A708., A709., A70z., A70G., A704., AyuB3
		Budd-Chiari syndrome	I820	G820.
		Acute liver failure	K720	J6000
		Infections and infarctions	K750, K763, K751	J62.., J620., J6200, J6201, J6202, J6203, J6204, J620z, A053., J634., J621.
		Other unspecified acute liver injuries	K752	J63y1



Supplemental Fig. 1 Flowchart for Identifying Index Events and Determining Cohort Entry Date.

This flowchart illustrates the decision rules for identifying the index event and determining the cohort entry date for study participants. The index date is defined as the date of the first diagnosis between 2004 and 2022. The aetiological diagnoses throughout the disease progression were identified as the index event entering the cohort. If no aetiological diagnosis is present, the index event is defined based on the following sequence: stage 2 diagnoses, stage 3 diagnoses, stage 4 diagnoses, and stage 5 diagnoses.

The decision flow proceeds as follows:

Tier 1 Diagnosis: If present at any point during the disease progression, the tier 1 diagnosis is recorded as the index event, and the date of the cohort entry is the index day. Tier 1 diagnoses include ArLD, NAFLD, metabolic liver disease, HBV, HCV, autoimmune liver disease, and haemochromatosis.

Tier 2 Diagnosis: If there is no tier 1 diagnosis, the presence of a tier 2 diagnosis is checked. If present, it is recorded as the index event, and the date of cohort entry is the index day. Tier 2 diagnoses include hepatitis not specified, congestive hepatopathy, and toxic liver disease.

Tier 3 Diagnosis: In the absence of tier 1 and tier 2 diagnoses, the presence of a tier 3 diagnosis is checked. If found, it is recorded as the index event. Tier 3 diagnoses include miscellaneous conditions.

Stage 2 Diagnosis: If none of the Stage 1 diagnoses are present, the presence of a stage 2 diagnosis is considered. If found, it is recorded as the index event. Stage 2 diagnoses include hepatic fibrosis and cirrhosis.

Stage 3 Diagnosis: If no stage 1 or stage 2 diagnoses are present, a stage 3 diagnosis is checked next. If found, it is recorded as the index event. Stage 3 diagnoses include portal hypertension.

Stage 4 Diagnosis: If the stage 1, stage 2, and stage 3 diagnoses are absent, a stage 4 diagnosis is checked. If present, it is recorded as the index event. Stage 4 diagnoses include hepatic decompensation.

Stage 5 Diagnosis: If no above diagnoses are found, a stage 5 diagnosis is considered. If found, it is recorded as the index event. Stage 5 diagnoses include hepatocellular carcinoma, intrahepatic cholangiocarcinoma, other primary liver cancer.

Acute Liver Diagnosis: If none of the chronic diagnoses are present, an acute liver diagnosis is checked. If found, it is recorded as the index event. Acute liver diagnoses include acute viral hepatitis, Budd-Chiari, acute liver failure, infections/sepsis, and other non-specified acute liver injuries.

BMJ Open

Incidence rate and associated patient characteristics of liver disease in Wales 2004-2022: a retrospective population-scale observational study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2024-093335.R1
Article Type:	Original research
Date Submitted by the Author:	28-Nov-2024
Complete List of Authors:	Gao, Jingwei; Swansea University - Singleton Park Campus, Akbari, Ashley; Swansea University Medical School, Ahmed, Haroon; Cardiff University, Division of Population Medicine Davies, Aled; Cardiff University, PRIME Centre Wales Yeoman, Andrew; Royal Gwent Hospital Pembroke , Thomas ; University Hospital of Wales, Department of Gastroenterology and Hepatology
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Gastroenterology and hepatology
Keywords:	EPIDEMIOLOGY, Hepatology < INTERNAL MEDICINE, PUBLIC HEALTH

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Incidence rate and associated patient characteristics of liver disease in Wales 2004-2022: a retrospective population-scale observational study

Authors: Jingwei Gao [1], Ashley Akbari [1], Haroon Ahmed [2], Aled Davies [3], Andrew Yeoman [4], Thomas Peter Ignatius Pembroke [5]

Affiliations:

1 Population Data Science, Swansea University Medical School, Faculty of Medicine, Health & Life Science, Swansea University, Swansea, UK.

2 Division of Population Medicine, Cardiff University, Neuadd Meirionnydd, Heath Park, Cardiff, UK.

3 PRIME Centre Wales, Cardiff University, Neuadd Meirionnydd, Heath Park, Cardiff, UK.

4 Aneurin Bevan University Health Board, Newport, UK.

5 Department of Gastroenterology and Hepatology, University Hospital of Wales, Cardiff, UK.

Corresponding author:

Jingwei Gao

jingwei.gao@swansea.ac.uk

Population Data Science, Swansea University Medical School, Faculty of Medicine, Health & Life Science, Swansea University, Swansea, UK

ORCiD:

Jingwei Gao:	0000-0002-7722-6177
Ashley Akbari:	0000-0003-0814-0801
Haroon Ahmed:	0000-0002-0634-8548
Aled Davies:	0000-0002-7815-5155
Andrew Yeoman:	0000-0002-0739-3332
Thomas Peter Ignatius Pembroke:	0000-0002-2600-2034

Abstract

Objective

To describe the incidence and key demographic, socioeconomic and clinical characteristics of individuals with liver disease in Wales.

Design and setting

This study is designed as a retrospective observational study that linked data of anonymised identified individuals from primary, secondary care and mortality data from the Secure Anonymised Information Linkage (SAIL) Databank in Wales.

Participants

All Welsh residents who registered with a SAIL contributing general practitioner (GP) and diagnosed with liver disease from 2004 to 2022.

Primary and secondary outcome measures

Our primary outcome is the annual age-standardised incidence rate of liver disease. Secondary outcome is the numbers and frequencies of underlying aetiology and the associated comorbidities.

Results

Between 2004 and 2022, 111,098 individuals received a diagnosis of liver disease in Wales and were included in this study. The incidence of liver disease increased three-fold during the study period (97.7 per 100,000 inhabitants in 2004 to 316.2 per 100,000 inhabitants in 2022). A total of 79,992 individuals (72%) entered the cohort with the underlying aetiology of liver disease, including alcohol related liver disease (ArLD), non-alcoholic fatty liver disease (NAFLD), viral hepatitis, metabolic, hemochromatosis and autoimmune liver diseases. NAFLD has contributed to most of the change in incidence.

Conclusions

We observed increasing incidence rates of liver disease in Wales, with NAFLD showing a particularly sharp increase and frequently identified as an underlying condition. A better understanding of the

incidence of liver disease is the first step towards effective prevention, early detection and targeted intervention to improve patient outcomes.

Keywords Epidemiology; Hepatology; Non-alcoholic fatty liver disease hepatitis; Alcohol-related liver disease; Cirrhosis

Word count: 3894

Strengths and limitations of this study

- The national-scale population-based cohort and long follow-up maximized the generalizability of our finding.
- By incorporating both in-patient and outpatient data, the study not only expanded the sample size but also enabled a thorough analysis of liver disease incidence, enhancing understanding of its clinical implications.
- With the usage of SAIL Databank as the research-ready data asset and the incorporation of reproducible research pipelines, this study strengthened its reliability and reproducibility.
- Sensitive Read codes for conditions like HBV and HCV could not be included due to regulatory laws, impacting the completeness and precision of the data analysis.
- The study only included individuals with a history of GP registration and residency information in Wales, likely overlooking a subpopulation at significant risk for liver disease

Background

Liver disease is a significant global public health issue and a major contributor to morbidity and mortality [1]. Globally, cirrhosis and hepatocellular carcinoma account for an estimated two million deaths every year [2]. In the UK, liver disease has become the third most common cause of premature death [3], despite mortality rates for other major non-communicable diseases declining [4].

The management of chronic liver disease frequently involves lifestyle modification, including weight loss and reduced alcohol use, with the goal of reversing factors that can lead to disease progression [5]. As a result, public health policies in the UK have focussed on prevention and early detection, including the UK Government's Prevention Green Paper to promote the disease prevention [6], the National Institute for Health and Care Excellence guidance that focused on tackling obesity [7], and a series of policies to reduce alcohol-related harm [8]. In November 2022, the Welsh Government published a Quality Statement for Liver Disease [9], underlying the importance of the awareness of risk factors and early detection of liver disease, and set out its plans to promote the delivery of better quality, higher value and more accessible services for individuals with liver disease.

To effectively improve liver disease management, clinicians, researchers and policymakers must be aware of the epidemiology and clinical profile of the individuals with liver disease. However, a significant gap remains in the integration of primary, secondary, and mortality data, particularly across the different liver disease stages and aetiologies. To date, much of the epidemiological data is based on ICD-10 coding, which is largely derived from secondary care data sources and therefore likely to underestimate the real-world incidence and prevalence of liver disease [10, 11]. This limitation currently precludes adequate prioritisation of research, targeting of interventions, and recruitment of individuals to clinical trials. Understanding the clinical and socio-demographic features associated with liver disease in a large-scale population is essential for the improvement of disease prevention and treatment. Therefore, we aim to fill this gap by integrating primary, secondary, and mortality data, providing a more comprehensive and accurate depiction of liver disease across different stages and aetiologies.

The objective of our study is to describe the incidence of liver disease, as well as the key demographic and socioeconomic characteristics and the associated comorbidity of liver disease patients in Wales, as a first step towards improving capacity and capability for liver disease research.

Materials and method

Setting and data source

This study is designed as retrospective cohort study. We used data from the Secure Anonymised Information Linkage (SAIL) Databank, which contains anonymised, individual-level linked electronic

health record (EHR) data for Welsh population [12, 13, 14, 15]. The SAIL Databank includes complete secondary care data and primary care data covering approximately 86% of the Welsh population. These data reflect the demographic diversity of Wales across age, sex, and levels of deprivation [16] and can be generalisable to the broader UK population due to demographic similarities [17]. SAIL employs a split-file anonymisation process using National Health Service (NHS) number, name, sex, date of birth and postcode, ensuring confidentiality while enabling the linkage of individual-level data sources [13, 16, 18].

To provide a comprehensive overview of liver disease epidemiology, we combined linked primary care, hospital admissions and mortality data. Primary care data were accessed from the Welsh Longitudinal General Practice (WLGP) data, which currently uses the Read version 2 clinical coding system and collects event histories for people registered with a SAIL-supplying general practice in Wales. Hospital admission data, including in-patient admissions (emergency, elective and maternity) and day-care procedures, were collected from the Patient Episode Database for Wales (PEDW). Mortality data came from the Annual District Death Extract (ADDE) by the Office for National Statistics (ONS) death and contains the cause of death and contributory comorbidities. Both hospital admission and mortality data were coded using the International Classification of Diseases version 10 (ICD-10) system. We derived demographic and deprivation data from the Welsh Demographic Service Dataset (WDSD) and used the Welsh Index of Multiple Deprivation (WIMD) version 2019 quintiles to measure relative area-level deprivation based on geographical residential location from the Lower-layer Super Output Area (LSOA) version 2011.

Study population

We linked data from all individuals in the WLGP, PEDW, and ADDE data within the SAIL Databank using a unique anonymised individual identifier known as Anonymised Linkage Field (ALF). Individuals were extracted based on ALF and filtered for good data linkage status based on existing methodology [13, 18]. We excluded individuals not registered with a SAIL-contributing general practice (GP) or lacking residency information and identified those who received a liver disease diagnosis from 1st January 2004 to 31st December 2022. The first liver disease diagnosis is considered to be the index liver disease event. Individuals were required to be residents in Wales at the time of cohort entry. GP registration was required if the index event was from WLGP. All individuals were followed until the earliest of: GP de-registration, moving out of Wales, death, or the end of study period (2022.12.31). We divided the full cohort into four distinct groups: individuals identified in WLGP were assigned to the primary care group; those identified in PEDW were assigned to the secondary care group; individuals identified in ADDE were assigned to the mortality group; and individuals identified across multiple data sources (two or more) were assigned to a multi-source group.

Measurements

Definition of stages of liver disease and time of cohort entry

We applied a hierarchy of 3 tiers of potential aetiological diagnoses and 5 discrete stages of chronic liver disease based on perceived clinical importance and natural history of liver disease progression as was described in our previous study [19]. The first stage is the underlying aetiological conditions without the presentation of cirrhosis and was divided into 3 tiers. Tier 1 aetiologies of liver disease include: alcohol-related liver disease (ArLD), non-alcoholic fatty liver disease (NAFLD), metabolic liver disease, hepatitis B virus (HBV), hepatitis C virus (HCV), autoimmune liver disease, haemochromatosis; tier 2 aetiologies include: unspecified hepatitis, congestive hepatopathy, and toxic liver disease, and tier 3 aetiologies were other miscellaneous diagnoses. As the disease progresses, various stages of liver disease were defined as follows: hepatic fibrosis and cirrhosis were categorised as stage 2, portal hypertension as stage 3, hepatic decompensation as stage 4, and hepatocellular carcinoma (HCC), intrahepatic cholangio carcinoma (ICC), and other primary liver cancers were classified under stage 5. Acute liver diseases were analysed separately from chronic liver diseases as they carry different challenges for primary and secondary care. Acute liver diseases were defined as conditions including acute viral hepatitis, Budd-Chiari syndrome, acute liver failure, infections and infarctions, and other unspecified acute liver injuries. A list of ICD-10 and Read v2 codes to identify individuals with liver disease can be found in Supplemental Table 1.

We defined the time of the first chronic or acute liver disease diagnosis as the index date. To assess the proportion of individuals with liver disease who presented late, the diagnoses were sequenced according to the natural history of the disease: aetiological diagnoses (stage 1), followed by cirrhosis, portal hypertension, decompensation and then HCC formation [19] (Supplemental Fig 1).

Comorbidities and drug prescription history

We collected data on demographic characteristics (age, sex, and WIMD 2019 quintile), with age divided by birth year in 10-year intervals, except for the 0-17 age group and 18-29 age group. Comorbidities including cardiovascular disease (CVD) related conditions (heart failure, transient ischaemic attack, other ischaemic disease, atrial fibrillation, peripheral vascular disease, angina, and stroke), diabetes, hypertension, and antihypertensive usage were tracked up to ten years prior to cohort entry. Lists of ICD-10 and Read v2 codes to identify the associated conditions are available as supplemental tables (Supplemental Table 2-1, Supplemental Table 2-2).

Incidence of liver disease

The incidence rate was calculated as the number of incident cases per Welsh residents. An incident case was defined as an individual having a first liver-related diagnosis during the study period (01.01.2004-31.12.2022) with no prior liver disease history from 1st January 1994 to 31st December 2003. To determine the most clinically significant aetiology, we only considered the most advanced stage of diagnosis on incident event date. The Welsh population data was obtained from the ONS population estimates [20]. The incidence rate was directly standardised using the European Union (EU) standard population 2013 and were reported per 100,000 inhabitants in each calendar year from 2004 to 2022, with corresponding 95% confidence intervals (CI). Analysis was repeated for different data sources, liver disease aetiologies, and stages.

Study findings were reported in accordance with applicable reporting guidelines for observational studies using administrative data (Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) [21] and REporting of studies Conducted using Observational Routinely-collected Data (RECORD) [22]. Data cleaning, cohort assembly and statistical analyses were performed using Structured Query Language (IBM Db2 V.11.1) and R (V.4.1.0–V.4.1.3) within the SAIL Databank privacy-protecting trusted research environment (TRE).

Patient and public involvement

The Liver Research Cymru Patient Advisory group was established to provide PPI input into the development of research in Wales. The research questions and methodology addressed in this body of work were developed following discussion of their priorities based upon their lived experience of liver disease. The results were shared with this group and their feedback was sought in the interpretation and presentation of these data.

Results

Demographic and socioeconomic characteristics

We identified a total of 111,098 individuals with liver disease from all three data sources from 1st January 2004 to 31st December 2022, contributing 441,885 person-years of follow-up (Fig 1A). Of the eligible individuals, 57,491 (51.7 %) were male. The median (interquartile range [IQR]) age was 59 (47-72). Approximately 50% of the individuals (55,787) entered the cohort after 2016 and 27,178 (24.5 %) individuals came from the most deprived areas. Of the 111,098 individuals, 20,319 (18.3%) individuals could be tracked from WLGP (primary care data source), forming the primary care group,

56,710 (51.0 %) from PEDW (secondary data source) constituted the secondary care group, and 3,459 (3.1 %) from ADDE (mortality data source) formed mortality group. Additionally, 30,610 (27.5%) appeared in multiple sources, forming the multi-source group (Fig 1B). The demographic variation is notable with a higher percentage of male and elderly individuals, and a higher degree of socioeconomic deprivation in the mortality group (Table 1).

Incidence of chronic liver disease

We observed an increasing trend of liver disease from 2004 to 2022 in the Welsh population. The age-sex standardised incidence rate increased by 3 times during the 18 years of follow-up (97.7 per 100,000 inhabitants in 2004; 316.2 per 100,000 inhabitants in 2022, Table 2). This was significantly contributed by a 6.5 times increase in the incidence rate of the primary care group, followed by 4.4 times increase in the secondary care group. The proportion of incident cases in primary care and secondary care group increased by 14-16% during the past 2 decades (Fig 2A – 2C, Supplemental Table 3-1).

The incidence of liver disease has risen across all stages, with stage 2 (cirrhosis) and stage 3 diagnoses (portal hypertension) experiencing the most notable increase from 2004 to 2022 (stage 2 by 4 times; stage 3 by 3.9 times). However, the increases in the higher-stage liver disease cases were less pronounced, particularly for stage 4 (1.5 times). Consequently, the proportion of more advanced liver disease diagnoses decreased (stage 4 and stage 5: 8.6% in 2004, 5.3% in 2022, decreased by 1.62 times since 2004, Fig 3A – 3C, Supplemental Table 3-2).

We further analysed the stage 1 aetiologies by each diagnosis. We observed the most notable increase in the incidence of NAFLD (10.8 per 100,000 inhabitants in 2004, 124.5 per 100,000 inhabitants in 2022, increased 11 times since 2004), resulting a 3.4 time increase in the proportion and ultimately accounting for over half of the cases by the end of our follow-up. Conversely, the proportion of ArLD cases decreased by 71%, given that the incidence rate remained unchanged during the past 2 decades (Fig 4A – 4B, Supplemental Table 3-3). A breakdown of NAFLD incidence by data sources demonstrated the 14 and 18 times increases in the primary care and the secondary care group, respectively (Supplemental Table 4-1). In contrast, the increase in the incidence of ArLD in the primary care group was more than twice that of the secondary care group (Supplemental Table 4-2).

We did not observe an increasing trend for HBV and HCV. Both conditions fluctuated during the past 2 decades, with HBV cases increasing by 4.3 times and HCV cases 2.3 times from 2004 to 2017, followed by a decrease from 2018 to 2022 (HBV by 1.7 times; HCV by 2.6 times). Consequently, the proportion of HBV and HCV cases decreased from 2004 to 2022 (HBV by 2.3 times; HCV by 8.6 times, Fig 4A – 4C, Supplemental Table 3-3).

Underlying liver disease conditions

Of the 111,098 individuals with liver disease, 79,992 (72.0 %) entered the cohort due to a tier 1 aetiology. This was mainly contributed by the high prevalence of NAFLD (33,655 [30.3%]). Most NAFLD cases were identified in the primary care group (13,565 [66.8 %]), and the least NAFLD cases in the mortality group (727 [21.0 %]). This aligns with the distribution of individuals with tier 1 aetiology, which's proportion was the highest in the primary care group (18,453 [90.8 %]), and the lowest in the mortality group (1,728 [50.0 %]). In contrast, tier 2 and tier 3 diagnoses were higher in the secondary care group. The proportion of advanced liver disease stages was lower in the primary care group (stage 3: 166 [0.8 %], stage 4: 53 [0.3 %], stage 5: 121 [0.6 %]), but higher in the mortality group (stage 4: 276 [8.0 %], stage 5: 473 [13.7 %]) (Table 3).

Comorbidities associated with liver disease

Amongst all eligible individuals with liver disease, 40,427 (36.4%) individuals had onset hypertension or initiated an antihypertensive medication, followed by 8,923 (8.0%) individuals who had cardiovascular disease (CVD) related conditions, and 7,658(6.9%) who had diabetes. We observed the higher proportions of comorbidities in the mortality group (CVD: 11.2%; diabetes: 8.1%; hypertension/antihypertensive: 39.7%), and the lowest in the primary care group (CVD: 4.2%; diabetes: 6.3%; hypertension/antihypertensive: 35.7%) (Table 4). The burden of comorbidities was the lowest for individuals with stage 1 liver disease conditions, but higher for late-stage (CVD highest for stage 4: 14.2%; hypertension/antihypertensive for stage 5: 45.1%) and stage 2 individuals (diabetes: 13.4%) (Supplemental Table 5-1). Amongst the individuals who entered the cohort with a stage 1 aetiology, individuals with congestive hepatopathy had the highest rates of hypertension/antihypertensive (47.6%) and CVD conditions (23.6%), while those with NAFLD had the highest rates of diabetes (8.4%) (Supplemental Table 5-2).

Discussion

In this large-scale population-based study, we observed that the incidence of liver disease had increased dramatically in Wales during the past two decades. Notably, NAFLD played a significant role in this rise, with its incidence and proportion increasing sharply.

The rapid increase in the incidence rate of liver disease observed in our study is similar to our previous findings in Wales [19], where a 3.6-fold increase was observed in the in-patient chronic liver disease cases between 2001 and 2019, aligning with the trends presented in our secondary care data. This was mainly driven by the 11-fold increase in NAFLD incidence. Other studies, such as P Nasr's et al work in Sweden [23], reported a 2-fold increase in the incidence rate of NAFLD during 2005-2019. Similarly,

H Tian et al [24] used data from the Global Burden of Disease study 2019 and showed an increase of 95.4% of NAFLD globally from 1990 to 2019. A systematic review that included 578 studies demonstrated a 13% higher prevalence of NAFLD during the year 2011-2021 compared with year 2000-2010 [25]. A community study in the USA reported a 5-fold increase in the NAFLD incidence from 1997 to 2014 [26]. However, none of these findings compare to our observation in Wales. This substantial difference may be partly attributed to our inclusion of primary care data, which experienced more rapid increase than inpatient data and thus boosted the NAFLD incidence across the study population. The high prevalence of obesity [27] and rapidly increasing incidence of diabetes [28] in Wales in the recent decades may also have attributed to the sharp increase of NAFLD, as these metabolic factors including obesity [29, 30] and insulin resistance were considered to be strongly linked to NAFLD due to metabolic dysfunctions [31, 32]. This observation is further supported by our observation of autoimmune liver disease, which is not linked to metabolic factors and remained unchanged during our study period, highlighting the lifestyle-related nature of the increase in NAFLD. Another interesting observation related to NAFLD is the high proportion of those with comorbidities. NAFLD evolves in those with metabolic syndrome, and these comorbidities are used to define the more recently adopted term metabolic associated fatty liver disease [33]. In keeping with this definition of NAFLD, we observed a high proportion of individuals having onset comorbidities such as hypertension, diabetes, CVD related conditions up to 10 years prior to their first liver disease diagnosis. The multimorbidity amongst people living with liver disease is anticipated considering the increasing incidence of NAFLD and the high prevalence of obesity in Wales, and this will require engagement with primary care practitioners to address their complex multidisciplinary healthcare needs and to identify those at risk of significant liver disease early in the disease trajectory [34].

Contrary to previous European studies showing a decline in ArLD incidence [19, 35], our research observed a resurgence after 2015 following a slight decrease from 2004 to 2014. We believe that changes in drinking habits [36, 37] and government actions on alcohol pricing and taxation [38] likely influenced the declining trend before 2015, while the resurgence after 2015 was contributed by the inclusion of outpatient data from the primary care group. The addition of primary care data could explain the discrepancy in our results compared to the prior studies that focused solely on in-patients, filling the gap by providing a broader perspective by including outpatient data [19, 35]. Similar to ArLD, the incidence of HBV increased until 2017, then sharply dropped after the year of 2018. The observed decrease in the incidence of HBV and HCV post-2020 could be attributed to the initiatives outlined in the Welsh Health Circular, which seeks to meet the goal of eradicating HBV and HCV as a major public health concern [39]. This trend may represent the effort in better case finding and eradication, which led to reduced transmission for HCV, and enhanced vaccination and viral suppression for HBV [40]. Additionally, disruptions caused by the COVID-19 pandemic may also have played a role in this decline [41].

It is worth to mention our observation on the increasing number of incident cases identified from primary care data, particularly after 2015. This observation, together with the rapid increase in NAFLD and ArLD incidence, indicated an increasing detection rate of liver disease in primary care at an early stage. This change might be attributed to the implementation of the Wales Liver Disease Delivery Plan, which emphasised the importance of early detection and ensure that 'excellent care' is accessible when necessary [42]. Another notable trend observed during our study period is the decreasing proportion of severe late-stage presentations, such as hepatic decompensation. Additionally, we observed a declining trend in the proportion of individuals who died from liver disease without receiving a prior diagnosis. This is to say, despite the increasing incidence of liver disease, there is a promising indication that more individuals were identified as the result of early detection. This provides the healthcare system with more opportunities for early-stage intervention. Early detection of liver disease has been linked to improved long-term outcomes, and a series of reviews and guidelines [43, 44] emphasize the importance of thorough evaluation and timely referral for patients with abnormal liver function tests (LFTs). Despite these recommendations, adherence remains inconsistent, and the factors influencing clinicians' referral practices—including potential barriers and motivating factors—are not well understood. This issue is especially pressing in the context of rising obesity rates and the emergence of new weight-loss therapies, both of which carry significant implications for liver health. Consequently, further research is essential to assess how integrated care models and early detection strategies might reduce liver disease progression and mortality.

Our study has several strengths. First, it utilised the national-scale setting and created a population-based cohort with long-term follow-up, which allowed us to maximise the generalisability of our findings to the wider population. Furthermore, by combining primary care, secondary care, and mortality data, our study uniquely expands on previous research and addresses a significant gap in understanding liver disease [10, 11]. This novel approach increased our sample size by at least 20%, allowing us to provide a more comprehensive view of liver disease incidence and characteristics. Specifically, this integrated data approach enabled us to: 1, capture a broader range of liver disease cases, reducing the likelihood of underestimating case numbers and ensuring a more accurate representation of disease incidence; and 2, include patients from diverse healthcare settings, each with unique clinical characteristics, which allowed for a detailed comparison across these settings and a deeper understanding of liver disease in varied clinical contexts. Additionally, a significant strength of the study is the consistency of coding practices in Wales, which have not changed in response to funding incentives [45]. This stability helps mitigate potential biases and misdiagnoses, providing a more reliable and objective reflection of disease incidence and management trends. Another strength lies in the reproducibility and the use of the SAIL Databank as a research-ready data asset (RRDA). By incorporating reproducible research pipelines, we standardized and documented the cohort curation

process. This enhances the study's reliability and reproducibility, enabling other researchers to replicate our findings and explore further research questions within the same framework.

Our study is also vulnerable to several limitations. Regulatory laws associated with the SAIL Databank prohibited the inclusion of conditions classified as sensitive in our analysis. For instance, the sensitive Read codes for HBV and HCV could not be extracted along with the rest of the cohort, resulting in potentially incomplete data for these conditions, affecting the completeness and precision of our data analysis. Our study exclusively included individuals who had a history of GP registration and residency information in Wales. Consequently, we likely overlooked individuals were unable to register with a GP, a subpopulation that represents a group at significant risk for liver disease. Moreover, there may be changes in diagnostic practices and healthcare-seeking behaviour over the study period that could influence the observed trends and introduce variations, potentially affecting the accuracy of our findings regarding liver disease incidence and management. Furthermore, due to the observational nature and design of our study, we were unable to establish causal relationships between the onset of liver disease and other comorbidities.

Conclusion

Our study observed a significant rise in the incidence of liver disease in Wales over the past two decades, primarily driven by the increase in NAFLD. The high prevalence of comorbid conditions among liver disease patients and the increased role of primary care in disease identification highlight the need for integrated healthcare approaches to address this growing public health concern. A better understanding of the incidence of liver disease, driven by a more comprehensive analysis of the integrated primary and secondary care data, is essential as a foundation for effective prevention, early detection, and targeted interventions to improve patient.

Acknowledgements

The Liver Disease Cymru Partnership (LDCP) received a grant from the National Institute for Health Research (NIHR154876). This project was partly funded by an unrestricted grant from the Liver Disease Implementation Group, Welsh Government (LDIG-22-19).

We wish to acknowledge all members of the LDCP who have contributed to the collaborations: Prof Kerry Hood, Prof Deborah Fitzsimmons and Prof Katherine Cullen for advice developing the group and the Liver Research Cymru patient advisory group for their perspective on the research to be undertaken.

This study uses anonymised data held in the Secure Anonymised Information Linkage (SAIL) Databank, which is part of the national population data research infrastructure for Wales. We would like to acknowledge all the data providers who make anonymised data available for research.

Author Contributions

This project was a collaborative effort among multiple team members, each contributing significantly to various aspects of the research. TP, AA, AD, AY and HA acquired the funding. TP, AA, AD, AY, HA and JG contributed to the design of the methodology. TP, AA, AD, AY and JG supervised the study and validated the research. JG and AA were responsible for project administration and data management. JG curated the data, performed data analysis, created visual representations of the data and wrote the original draft of the paper. TP acted as guarantor. All authors reviewed and edited the manuscript.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Funding Sources

The Liver Disease Cymru Partnership (LDCP) receives a grant from the National Institute for Health Research (NIHR154876). This project was partly funded by an unrestricted grant from the Liver Disease Implementation Group, Welsh Government (LDIG-22-19).

Data availability statement

The routine data used in this study are available in the SAIL Databank at Swansea University, Swansea, UK. All proposals to use SAIL data are subject to review by an Information Governance Review Panel (IGRP). Before any data can be accessed, approval must be given by the IGRP. The IGRP gives careful consideration to each project to ensure proper and appropriate use of SAIL data. When access has been approved, it is gained through a privacy-protecting trusted research environment (TRE). SAIL has established an application process to be followed by anyone who would like to access data via SAIL https://www.saildatabank.com/application-process . This study has been approved by the IGRP as project 1492. The research adhered to ethical guidelines and Data Protection Act 2018 to ensure the privacy and confidentiality of all data subjects involved.

The reproducible SQL and R code, and code lists to identify study individuals are available on Github: https://github.com/SwanseaUniversityDataScience/1492-LDCP

Ethics statements

Patient consent for publication

Not applicable.

Ethics approval

The study uses anonymised data only, and ethical approval was not required for this secondary use of data. See the following weblink: https://saildatabank.com/data/

References

[1] Paik JM, Golabi P, Younossi Y, Mishra A, Younossi ZM. Changes in the Global Burden of Chronic Liver Diseases From 2012 to 2017: The Growing Impact of NAFLD. Hepatology. 2020 Oct 27;72(5):1605–16. https://doi.org/10.1002/hep.31173

[2] Wang H, Naghavi M, Allen C, Barber RM, Bhutta ZA, Carter A, et al. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet. 2016 Oct;388(10053):1459–544. https://doi.org/10.1016/S0140-6736(16)31012-1.

[3] Deaths registered in England and Wales (series DR): 2017. Office for National Statistics. 2017. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deathsregisteredinenglandandwalesseriesdr/2017. [Accessed 5 Jun 2024].

[4] Williams R, Aspinall R, Bellis M, Camps-Walsh G, Cramp M, Dhawan A, et al. Addressing liver disease in the UK: a blueprint for attaining excellence in health care and reducing premature mortality from lifestyle issues of excess consumption of alcohol, obesity, and viral hepatitis. Lancet. 2014 Nov 29;384(9958):1953–97. https://doi.org/10.1016/S0140-6736(14)61838-9.

[5] Nobili V, Carter-Kent C, Feldstein AE. The role of lifestyle changes in the management of chronic liver disease. BMC Medicine. 2011 Jun 6;9:70. https://doi.org/10.1186/1741-7015-9-70.

[6] Department of Health and Social Care. Advancing our health: prevention in the 2020s – consultation document. GOV.UK. 2019. https://www.gov.uk/government/consultations/advancing-our-health-prevention-in-the-2020s/advancing-our-health-prevention-in-the-2020s-consultation-document. [Accessed 5 Jun 2024].

[7] Diet, nutrition and obesity. NICE. https://www.nice.org.uk/guidance/lifestyle-and-wellbeing/diet--nutrition-and-obesity. [Accessed 5 Jun 2024].

[8] Burton R, Henn C, Lavoie D, O'Connor R, Perkins C, Sweeney K, et al. A rapid evidence review of the effectiveness and cost-effectiveness of alcohol control policies: an English perspective. The Lancet. 2017 Apr;389(10078):1558–80. https://doi.org/10.1016/S0140-6736(16)32420-5.

[9] The quality statement for liver disease: the quality statement describes what good quality liver disease services should look like. Welsh Government. 2022. https://www.gov.wales/sites/default/files/pdf-versions/2023/3/1/1679326108/quality-statement-liver-disease.pdf. [Accessed 5 Jun 2024].

[10] Nam YH, Mendelsohn AB, Panozzo CA, Maro JC, Brown JS. Health outcomes coding trends in the US Food and Drug Administration's Sentinel System during transition to International Classification of Diseases-10 coding system: A brief review. Pharmacoepidemiology and Drug Safety. 2021 Mar 17;30(7):838–42. https://doi.org/10.1002/pds.5216.

[11] Ratib S, West J, Fleming KM. Liver cirrhosis in England—an observational study: are we measuring its burden occurrence correctly? BMJ Open. 2017 Jul;7(7):e013752. https://doi.org/10.1136/bmjopen-2016-013752.

[12] Ford DV, Jones KH, Verplancke JP, Lyons RA, John G, Brown G, et al. The SAIL Databank: building a national architecture for e-health research and evaluation. BMC Health Services Research. 2009 Sep 4;9:157. https://doi.org/10.1186/1472-6963-9-157.

[13] Lyons RA, Jones KH, John G, Brooks CJ, Verplancke JP, Ford DV, et al. The SAIL databank: linking multiple health and social care datasets. BMC Medical Informatics and Decision Making. 2009 Jan 16;9:3. https://doi.org/10.1186/1472-6947-9-3.

[14] Jones KH, Ford DV, Ellwood-Thompson S, Lyons RA. The UK Secure eResearch Platform for public health research: a case study. The Lancet. 2016 Nov;388:S62. https://doi.org/10.1016/S0140-6736(16)32298-X.

[15] Wilkinson T, Schnier C, Bush K, Rannikmäe K, Lyons RA, McTaggart S, et al. Drug prescriptions and dementia incidence: a medication-wide association study of 17000 dementia cases among half a million participants. Journal of Epidemiology and Community Health. 2021 Oct 27;76(3):223–9. https://doi.org/10.1136/jech-2021-217090.

[16] Schnier C, Wilkinson T, Akbari A, Orton C, Sleegers K, Gallacher J, et al. The Secure Anonymised Information Linkage databank Dementia e-cohort (SAIL-DeC). International Journal of Population Data Science. 2020 Feb 25;5(1):1121. https://doi.org/10.23889/ijpds.v5i1.1121

[17] Population estimates for the UK, England, Wales, Scotland, and Northern Ireland: mid-2022. Office for National Statistics. 2024. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2022. [Accessed 5 Jun 2024].

[18] Ford DV, Jones KH, Verplancke JP, Lyons RA, John G, Brown G, et al. The SAIL Databank: building a national architecture for e-health research and evaluation. BMC Health Services Research. 2009 Sep 4;9:157. https://doi.org/10.1186/1472-6963-9-157.

[19] Peter T, John GR, Puyk B, Howkins K, Clarke R, Yousuf F, et al. Rising incidence, progression and changing patterns of liver disease in Wales 1999-2019. World Journal of Hepatology. 2023 Jan 27; 15(1):89–106. https://doi.org/10.4254/wjh.v15.i1.89.

[20] Estimate of the population for England and Wales. Office for National Statistics. 2023. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/estimatesofthepopulationforenglandandwales. [Accessed 5 Jun 2024].

[21] von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. PLoS Medicine. 2007 Oct 16;4(10):e296. https://doi.org/10.1016/j.jclinepi.2007.11.008.

[22] Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. PLOS Medicine. 2015 Oct 6;12(10):e1001885. https://doi.org/10.1371/journal.pmed.1001885

[23] Patrik Nasr, Erik von Seth, Mayerhofer R, Ndegwa N, Ludvigsson JF, Hannes Hagström. Incidence, prevalence and mortality of chronic liver diseases in Sweden between 2005 and 2019. European Journal of Epidemiology. 2023 Jul 25;38(9):973–84. https://doi.org/10.1007/s10654-023-01028-x.

[24] Tian H, Zhang K, Hui Z, Ren F, Ma Y, Han F, et al. Global burden of non-alcoholic fatty liver disease in 204 countries and territories from 1990 to 2019. Clinics and Research in Hepatology and Gastroenterology. 2023 Jan;47(1):102068. https://doi.org/10.1016/j.clinre.2022.102068.

[25] Liu J, Tian Y, Fu X, Mu C, Yao M, Ni Y, et al. Estimating global prevalence, incidence, and outcomes of non-alcoholic fatty liver disease from 2000 to 2021: systematic review and meta-analysis. Chinese Medical Journal. 2022 Jul 20;135(14):1682–91. https://doi.org/10.1097/CM9.0000000000002277.

[26] Allen AM, Therneau TM, Larson JJ, Coward A, Somers VK, Kamath PS. Nonalcoholic fatty liver disease incidence and impact on metabolic burden and death: A 20 year-community study. Hepatology. 2018 Mar 23;67(5):1726–36. https://doi.org/10.1002/hep.29546.

[27] Keaver L, Xu B, Jaccard A, Webber L. Morbid obesity in the UK: A modelling projection study to 2035. Scandinavian Journal of Public Health. 2018 Aug 30;48(4):422–7. https://doi.org/10.1177/1403494818794814

[28] Powell R. Diabetes prevalence – trends, risk factors, and 10-year projection [Internet]. Public Health Wales. 2023. https://phw.nhs.wales/services-and-teams/observatory/data-and-analysis/diabetes-prevalence-trends-risk-factors-and-10-year-projection/. [Accessed 5 Jun 2024].

[29] Younossi Z, Anstee QM, Marietti M, Hardy T, Henry L, Eslam M, et al. Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. Nature Reviews Gastroenterology & Hepatology. 2017 Sep 20;15(1):11–20. https://doi.org/10.1038/nrgastro.2017.109.

[30] Loomis AK, Kabadi S, Preiss D, Hyde C, Bonato V, St. Louis M, et al. Body Mass Index and Risk of Nonalcoholic Fatty Liver Disease: Two Electronic Health Record Prospective Studies. The Journal of Clinical Endocrinology & Metabolism. 2016 Mar 1;101(3):945–52. https://doi.org/10.1210/jc.2015-3444.

[31] Caussy C, Aubin A, Loomba R. The Relationship Between Type 2 Diabetes, NAFLD, and Cardiovascular Risk. Current Diabetes Reports. 2021 Mar 19;21(5):15. https://doi.org/10.1007/s11892-021-01383-7.

[32] Tanase DM, Gosav EM, Costea CF, Ciocoiu M, Lacatusu CM, Maranduca MA, et al. The Intricate Relationship between Type 2 Diabetes Mellitus (T2DM), Insulin Resistance (IR), and Nonalcoholic Fatty Liver Disease (NAFLD). Journal of Diabetes Research. 2020 Jul 31:2020:3920196. https://doi.org/10.1155/2020/3920196.

[33] Eslam M, Newsome PN, Sarin SK, Anstee QM, Targher G, Romero-Gomez M, et al. A new definition for metabolic dysfunction-associated fatty liver disease: An international expert consensus statement. Journal of Hepatology. 2020 Jul 1;73(1):202–9. https://doi.org/10.1016/j.jhep.2020.03.039.

[34] Pryke R, Guha IN. Time to focus on chronic liver diseases in the community: A review of primary care hepatology tools, pathways of care and reimbursement mechanisms. Journal of Hepatology. 2023 Mar 1 [cited 2024 May 13];78(3):663–71. https://doi.org/10.1016/j.jhep.2022.10.010.

[35] Deleuran T, Vilstrup H, Becker U, Jepsen P. Epidemiology of Alcoholic Liver Disease in Denmark 2006–2011: A Population-Based Study. Alcohol and Alcoholism. 2015 Feb 13;50(3):352–7. https://doi.org/10.1093/alcalc/agv003.

[36] Adult drinking habits in Great Britain: 2017. Office for National Statistics. 2018. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/drugusealcoholandsmoking/bulletins/opinionsandlifestylesurveyadultdrinkinghabitsingreatbritain/2017. [Accessed 5 Jun 2024].

[37] Adult lifestyles by year, 2016-17 to 2019-20. StatsWales. Available from: https://statswales.gov.wales/Catalogue/National-Survey-for-Wales/Population-Health/Adult-Lifestyles/adultlifestyles-by-year. [Accessed 5 Jun 2024].

[38] Home Office. Alcohol strategy. GOV.UK. 2012. https://www.gov.uk/government/publications/alcohol-strategy. [Accessed 5 Jun 2024].

[39] WELSH HEALTH CIRCULAR. Attaining the WHO targets for eliminating hepatitis (b and C) as a significant threat to public health. Welsh Government. 2017. https://www.gov.wales/sites/default/files/publications/2019-07/attaining-the-who-targets-for-eliminating-hepatitis-b-and-c-as-a-significant-threat-to-public-health.pdf. [Accessed 5 Jun 2024].

[40] Prevention, diagnosis and treatment of blood borne viruses in Wales: Hepatitis B, hepatitis C and HIV Annual report 2023 (Data to end 2022). Public Health Wales. 2023. https://phw.nhs.wales/publications/publications1/prevention-diagnosis-and-treatment-of-blood-borne-viruses-in-wales-hepatitis-b-hepatitis-c-and-hiv/. [Accessed 5 Jun 2024].

[41] Pley CM, McNaughton AL, Matthews PC, Lourenço J. The global impact of the COVID-19 pandemic on the prevention, diagnosis and treatment of hepatitis B virus (HBV) infection. BMJ Global Health. 2021 Jan;6(1):e004275. https://doi.org/10.1136/bmjgh-2020-004275.

[42] Together for Health -Liver Disease Delivery Plan: A Delivery Plan for NHS Wales and its Partners to 2020. Welsh Government. 2015. https://www.gov.wales/sites/default/files/publications/2018-12/liver-disease-delivery-plan-2015-to-2020.pdf. [Accessed 5 Jun 2024].

[43] Eslam M, Newsome PN, Sarin SK, Anstee QM, Targher G, Romero-Gomez M, et al. A new definition for metabolic dysfunction-associated fatty liver disease: An international expert consensus statement. Journal of Hepatology. 2020 Jul;73(1):202–9. https://doi.org/10.1016/j.jhep.2020.03.039.

[44] Pratt DS, Kaplan MM. Evaluation of abnormal liver-enzyme results in asymptomatic patients. N Engl J Med. 2000 Apr 27;342(17):1266-1271. https://doi.org/10.1056/NEJM200004273421707.

[45] Welsh Government. 2021-22 Health Board and Public Health Wales NHS Trust Allocations. Cardiff: Welsh Government, 2021. https://www.gov.wales/2021-22-health-board-and-public-health-wales-nhs-trust-allocations.pdf. [Accessed 05 Nov 2024].

Figure Legends

Fig. 1 Flowchart of study population selection (Fig.1A) and Venn diagram of study cohort composition (Fig.1B)

Abbreviations: PEDW: Patient Episode Database for Wales; WLGP: Welsh Longitudinal General Practice; ADDE: Annual District Death Extract

Fig. 2 Changes in Standardised Incidence per 100,000 Inhabitants by Data Source (2004-2022)
Bar chart presenting the changes in annual European age-sex standardised incidence rate and percentages by data sources. The x-axis represents time in years, and the y-axis represents incidence rate in Fig.2A, percentages in Fig.2B.

Line chart presenting the trends in annual European age-sex standardised incidence rate. The x-axis represents time in years, and the y-axis represents incidence rate in Fig.2C.

Fig. 3 Changes in Standardised Incidence per 100,000 Inhabitants by Stages (2004-2022)
Bar chart presenting the changes in annual European age-sex standardised incidence rate and percentages by stages. The x-axis represents time in years, and the y-axis represents incidence rate in Fig.3A, percentages in Fig.3B.

Line chart presenting the trends in annual European age-sex standardised incidence rate. The x-axis represents time in years, and the y-axis represents incidence rate in Fig.3C.

Fig. 4 Changes in Standardised Incidence per 100,000 Inhabitants by Liver Disease Aetiologies (2004-2022)
Bar chart presenting the changes in annual European age-sex standardised incidence rate and percentages by aetiologies. The x-axis represents time in years, and the y-axis represents incidence rate in Fig.4A, percentages in Fig.4B.

Line chart presenting the trends in annual European age-sex standardised incidence rate. The x-axis represents time in years, and the y-axis represents incidence rate in Fig.4C.

Abbreviations: ArLD: alcohol related liver disease; NAFLD: non-alcoholic fatty liver disease; HBV: hepatitis B virus; HCV: hepatitis C virus

* ArLD overlap: if an individual had 2 or more diagnoses from liver disease aetiologies (ArLD, NAFLD, HBV, HCV, metabolic, hemochromatosis and autoimmune liver diseases) and one of them was ArLD; Non-ArLD overlap: if an individual had 2 or more diagnoses from liver disease aetiologies but none of them was ArLD.

Tables

Table 1. Demographic and socioeconomic characteristics of liver disease patients

Demographic characteristics	Full cohort, N = 111,098[1]	Primary care group, N = 20,319[1]	Secondary care group, N = 56,710[1]	Mortality group, N = 3,459[1]	Multi-source group, N = 30,610[1]
Sex					
Male	57,491(51.7%)	10,478(51.6%)	27,558(48.6%)	2,019(58.4%)	17,436(57.0%)
Female	53,607(48.3%)	9,841(48.4%)	29,152(51.4%)	1,440(41.6%)	13,174(43.0%)
Age					
0-17	1,445(1.3%)	342(1.7%)	814(1.4%)	<10 (<0.3%)	<290 (<1.0%)
18-29	5,862(5.3%)	1,447(7.1%)	3,486(6.1%)	<40 (<1.2%)	<900 (<0%)
30-39	10,109(9.1%)	2,520(12.4%)	5,099(9.0%)	143(4.1%)	2,347(7.7%)
40-49	15,370(13.8%)	3,744(18.4%)	6,316(11.1%)	421(12.2%)	4,889(16.0%)
50-59	22,799(20.5%)	5,356(26.4%)	9,332(16.5%)	705(20.4%)	7,406(24.2%)
60-69	23,264(20.9%)	4,271(21.0%)	11,026(19.4%)	758(21.9%)	7,209(23.6%)
70-79	19,189(17.3%)	2,117(10.4%)	11,335(20.0%)	693(20.0%)	5,044(16.5%)
80+	13,060(11.8%)	522(2.6%)	9,302(16.4%)	702(20.3%)	2,534(8.3%)
Cohort entry year					
2004-2007	17,099(15.4%)	1,949(9.6%)	7,345(13.0%)	617(17.8%)	7,188(23.5%)
2008-2011	17,554(15.8%)	2,513(12.4%)	8,484(15.0%)	705(20.4%)	5,852(19.1%)
2012-2015	20,658(18.6%)	3,109(15.3%)	10,682(18.8%)	731(21.1%)	6,136(20.0%)
2016-2019	30,448(27.4%)	6,683(32.9%)	15,521(27.4%)	749(21.7%)	7,495(24.5%)
2020-2022	25,339(22.8%)	6,065(29.8%)	14,678(25.9%)	657(19.0%)	3,939(12.9%)
WIMD 2019 quintiles					
1, most deprived	27,178(24.5%)	4,417(21.7%)	14,012(24.7%)	928(26.8%)	7,821(25.6%)
2	24,391(22.0%)	4,203(20.7%)	12,411(21.9%)	858(24.8%)	6,919(22.6%)
3	21,066(19.0%)	3,828(18.8%)	10,802(19.0%)	647(18.7%)	5,789(18.9%)
4	19,619(17.7%)	3,949(19.4%)	9,855(17.4%)	545(15.8%)	5,270(17.2%)
5, least deprived	18,844(17.0%)	3,922(19.3%)	9,630(17.0%)	481(13.9%)	4,811(15.7%)

[1] n(%)

Abbreviations: PEDW: Patient Episodes Dataset for Wales, WLGP: Welsh Longitudinal General Practice dataset, ADDE: Annual District Deaths Extract, WIMD: Welsh Index of Multiple Deprivation

Table 2. Crude and standardised incidence of chronic liver disease from 2004 to 2022

Year	Wales population	Incident cases	Crude incidence (95% CI)	STD incidence (95% CI)
2004	3,131,640	3,060	97.7 (94.3 - 101.2)	97.7 (94.2 - 101.3)
2005	3,105,633	3,403	109.6 (105.9 - 113.3)	109.4 (105.7 - 113.2)
2006	3,104,483	3,670	118.2 (114.4 - 122.1)	118.7 (114.9 - 122.6)
2007	3,087,732	3,537	114.6 (110.8 - 118.4)	115.1 (111.3 - 119.0)
2008	3,083,840	3,867	125.4 (121.5 - 129.4)	126.3 (122.3 - 130.4)
2009	3,081,366	4,008	130.1 (126.1 - 134.2)	131.7 (127.7 - 135.9)
2010	3,077,165	4,268	138.7 (134.6 - 142.9)	140.6 (136.4 - 144.9)
2011	3,072,739	4,519	147.1 (142.8 - 151.4)	149.9 (145.5 - 154.4)
2012	3,073,788	4,546	147.9 (143.6 - 152.3)	150.6 (146.3 - 155.1)
2013	3,071,058	4,733	154.1 (149.8 - 158.6)	158.0 (153.5 - 162.6)
2014	3,070,928	5,145	167.5 (163.0 - 172.2)	172.5 (167.7 - 177.3)
2015	3,063,758	5,659	184.7 (179.9 to 189.6)	192.2 (187.2 to 197.4)
2016	3,049,971	6,548	214.7 (209.5 to 220.0)	225.7 (220.2 to 231.3)
2017	3,038,872	6,981	229.7 (224.4 to 235.2)	243.2 (237.4 to 249.1)
2018	3,025,867	7,953	262.8 (257.1 to 268.7)	279.1 (272.9 to 285.4)
2019	3,006,299	8,481	282.1 (276.1 to 288.2)	301.2 (294.8 to 307.8)
2020	2,985,668	7,327	245.4 (239.8 to 251.1)	264.0 (257.8 to 270.2)
2021	2,969,309	9,136	307.7 (301.4 to 314.1)	332.0 (325.0 to 338.9)
2022	2,957,422	8,596	290.7 (284.5 to 296.9)	316.2 (309.4 to 323.1)

Abbreviation: STD: standardised; CI: confidence interval

Table 3. The underlying liver disease conditions at cohort entry

Phenotype	Full cohort, N = 111,098[1]	Primary care group, N = 20,319[1]	Secondary care group, N = 56,710[1]	Mortality group, N = 3,459[1]	Multi-source group, N = 30,610[1]
Chronic liver disease					
Stage 1 - tier 1	79,992(72.0%)	18,453(90.8%)	35,102(61.9%)	1,728(50.0%)	24,709(80.7%)
- ArLD	19,760(17.8%)	2,116(10.4%)	4,566(8.1%)	887(25.6%)	12,191(39.8%)
- NAFLD	33,655(30.3%)	13,565(66.8%)	13,444(23.7%)	727(21.0%)	5,919(19.3%)
- Metabolic liver disease	5,469(4.9%)	782(3.8%)	3,887(6.9%)	69(2.0%)	731(2.4%)
- HBV	1,063(1.0%)	0(0.0%)	905(1.6%)	0(0.0%)	158(0.5%)
- HCV	3,539(3.2%)	0(0.0%)	2,697(4.8%)	13(0.4%)	829(2.7%)
- Autoimmune liver disease	13,582(12.2%)	815(4.0%)	9,060(16.0%)	199(5.8%)	3,508(11.5%)
- Haemochromatosis	4,111(3.7%)	1,209(6.0%)	<940(<1.7%)	<10(<0.2%)	1,958(6.4%)
- ArLD overlap	799(0.7%)	15(0.1%)	158(0.3%)	173(5.0%)	453(1.5%)
- Non ArLD overlap	371(0.3%)	<20(<0.1%)	228(0.4%)	<10(<0.2%)	123(0.4%)
Stage 1 - tier 2	6,111(5.5%)	81(0.4%)	4,906(8.7%)	127(3.7%)	997(3.3%)
- Hepatitis not specified	4,783(4.3%)	11(0.1%)	3,867(6.8%)	102(2.9%)	803(2.6%)
- Congestive hepatopathy	574(0.5%)	<10(<0.1%)	514(0.9%)	<10(<0.2%)	<60(<0.2%)
- Toxic liver disease	757(0.7%)	68(0.3%)	526(0.9%)	21(0.6%)	142(0.5%)
Stage 1 - tier 3	8,426(7.6%)	36(0.2%)	8,132(14.3%)	37(1.1%)	221(0.7%)
- Miscellaneous	8,426(7.6%)	36(0.2%)	8,132(14.3%)	37(1.1%)	221(0.7%)
Stage 2	4,562(4.1%)	506(2.5%)	1,664(2.9%)	690(19.9%)	1,702(5.6%)
- Hepatic fibrosis	261(0.2%)	64(0.3%)	135(0.2%)	<10(<0.2%)	<60(<0.2%)
- Cirrhosis	4,301(3.9%)	442(2.2%)	1,529(2.7%)	682(19.7%)	1,648(5.4%)
Stage 3	3,040(2.7%)	166(0.8%)	2,171(3.8%)	43(1.2%)	660(2.2%)
- Portal hypertension	3,040(2.7%)	166(0.8%)	2,171(3.8%)	43(1.2%)	660(2.2%)
Stage 4	3,221(2.9%)	53(0.3%)	2,517(4.4%)	276(8.0%)	375(1.2%)
- Hepatic decompensation	3,221(2.9%)	53(0.3%)	2,517(4.4%)	276(8.0%)	375(1.2%)
Stage 5	2,233(2.0%)	121(0.6%)	341(0.6%)	473(13.7%)	1,298(4.2%)
- Hepatocellular carcinoma	748(0.7%)	71(0.3%)	83(0.1%)	86(2.5%)	508(1.7%)
- Intrahepatic cholangio carcinoma	1,455(1.3%)	47(0.2%)	232(0.4%)	387(11.2%)	789(2.6%)
- Other primary liver cancer	39(0.0%)	<10(<0.1%)	26(0.0%)	<10(<0.2%)	<10(<0.1%)
Acute liver disease					
Acute liver disease	3,513(3.2%)	903(4.4%)	1,877(3.3%)	85(2.5%)	648(2.1%)
- Budd-Chiari	184(0.2%)	<10(<0.1%)	144(0.3%)	<20(<0.1%)	20(0.1%)
- Acute liver failure	27(0.0%)	<30(<0.2%)	0(0.0%)	0(0.0%)	<10(<0.1%)
- Infection/infarction	1,343(1.2%)	62(0.3%)	757(1.3%)	67(1.9%)	457(1.5%)
- Other non-specified acute liver injuries	71(0.1%)	45(0.2%)	<20(<0.1%)	0(0.0%)	<10(<0.1%)

[1] n(%)

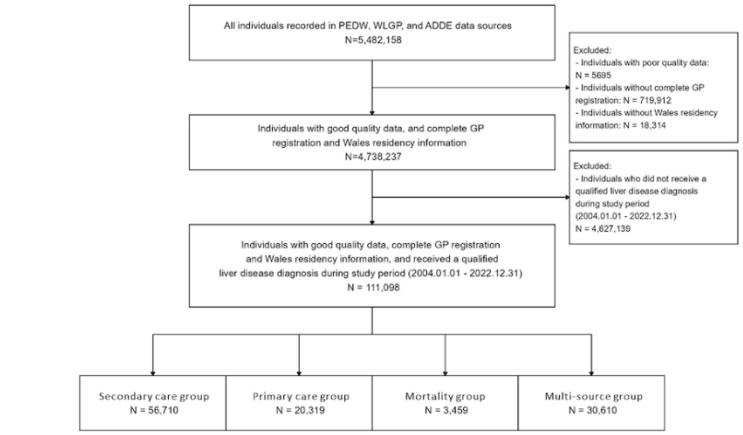
Abbreviation: ArLD: alcohol-related liver disease; NAFLD: non-alcohol fatty liver disease; HBV: hepatitis B virus; HCV: hepatitis C virus

Table 4. Comorbidities associated with liver disease

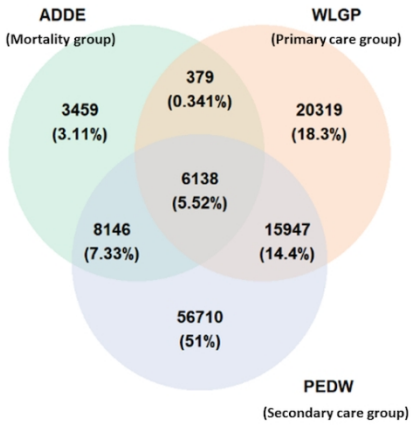
Comorbidities	Full cohort, N = 111,098[1]	Primary care group, N = 20,319[1]	Secondary care group, N = 56,710[1]	Mortality group, N = 3,459[1]	Multi-source group, N = 30,610[1]
CVD related conditions	8,923(8.0%)	852(4.2%)	5,362(9.5%)	385(11.2%)	2,324(7.6%)
Diabetes	7,658(6.9%)	1,271(6.3%)	3,577(6.3%)	279(8.1%)	2,531(8.3%)
Hypertension/antihypertensives	40,427(36.4%)	7,248(35.7%)	20,361(35.9%)	1,368(39.7%)	11,450(37.4%)

[1] n(%)

Abbreviation: CVD: cardiovascular disease; COPD: chronic obstructive pulmonary disease



A



B

Fig. 1 Flowchart of study population selection (Fig.1A) and Venn diagram of study cohort composition (Fig.1B)
Abbreviations: PEDW: Patient Episode Database for Wales; WLGP: Welsh Longitudinal General Practice; ADDE: Annual District Death Extract

99x124mm (300 x 300 DPI)

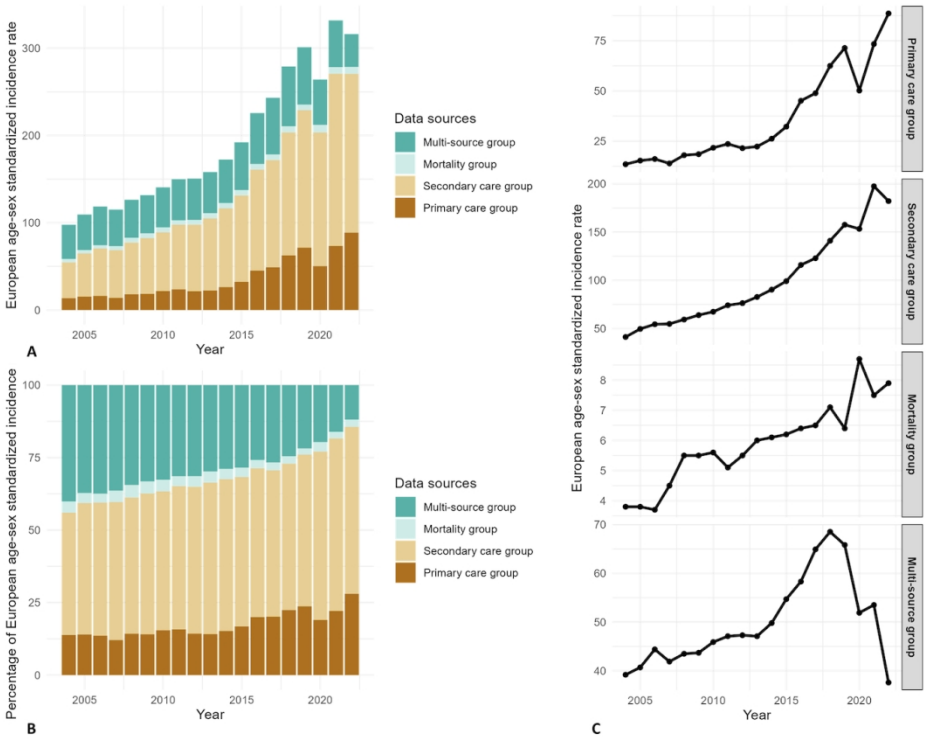


Fig. 2 Changes in Standardised Incidence per 100,000 Inhabitants by Data Source (2004-2022)
Bar chart presenting the changes in annual European age-sex standardised incidence rate and percentages by data sources. The x-axis represents time in years, and the y-axis represents incidence rate in Fig.2A, percentages in Fig.2B.
Line chart presenting the trends in annual European age-sex standardised incidence rate. The x-axis represents time in years, and the y-axis represents incidence rate in Fig.2C.

124x99mm (300 x 300 DPI)

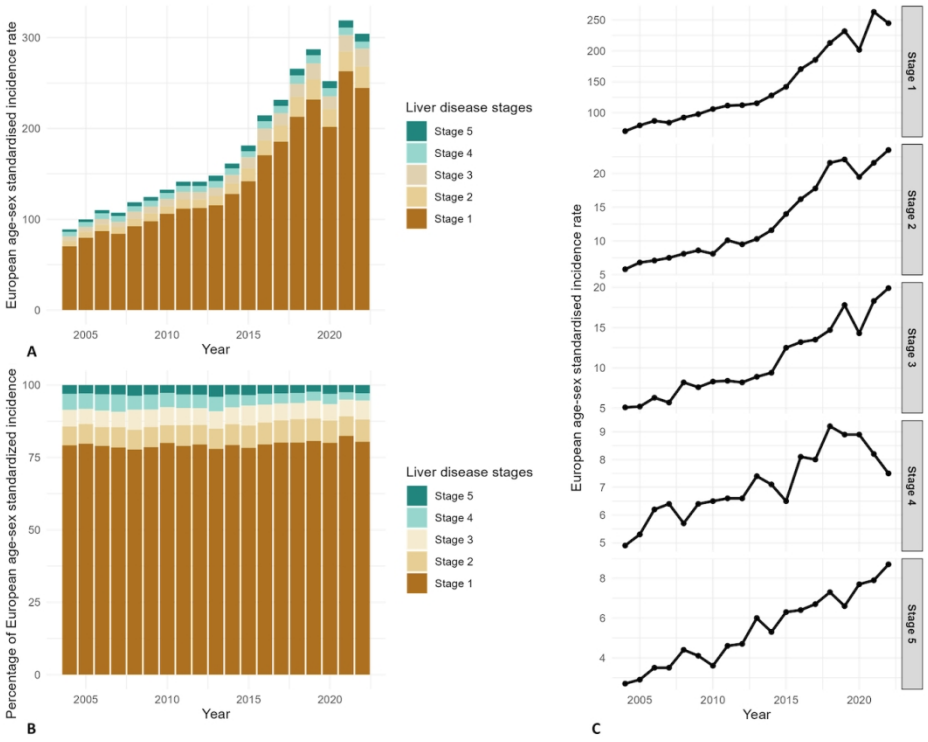


Fig. 3 Changes in Standardised Incidence per 100,000 Inhabitants by Stages (2004-2022)
Bar chart presenting the changes in annual European age-sex standardised incidence rate and percentages by stages. The x-axis represents time in years, and the y-axis represents incidence rate in Fig.3A, percentages in Fig.3B.
Line chart presenting the trends in annual European age-sex standardised incidence rate. The x-axis represents time in years, and the y-axis represents incidence rate in Fig.3C.

124x99mm (300 x 300 DPI)

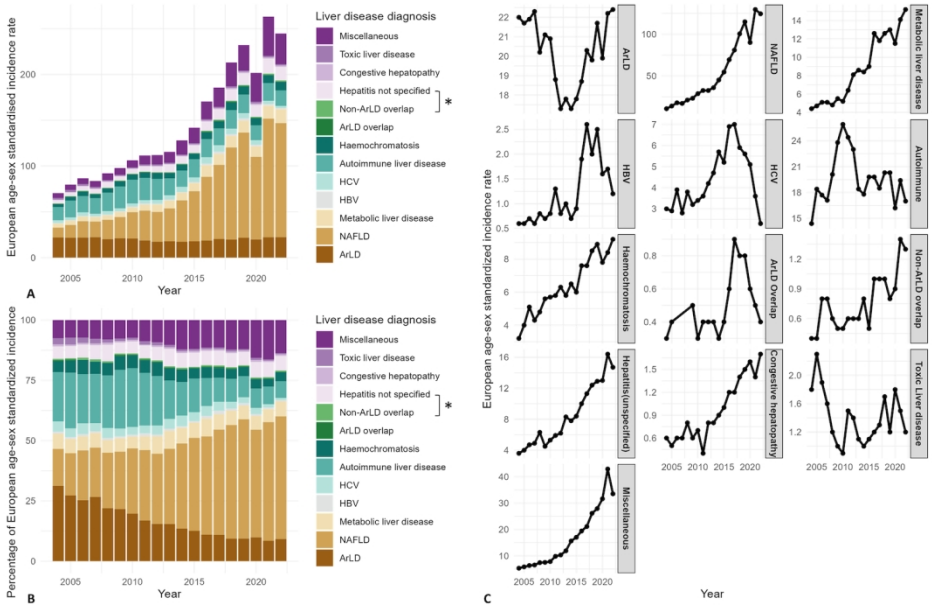
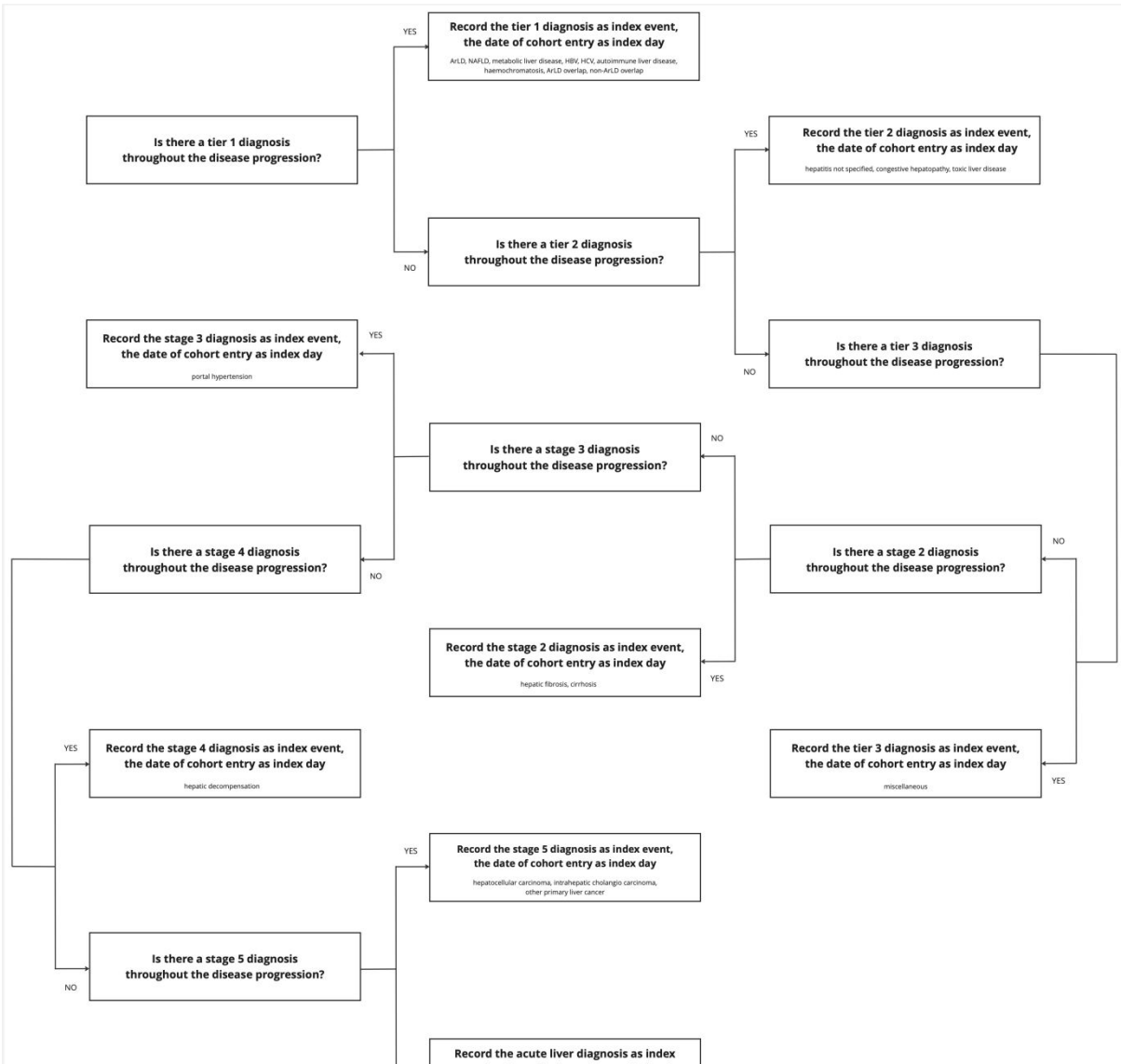


Fig. 4 Changes in Standardised Incidence per 100,000 Inhabitants by Liver Disease Aetiologies (2004-2022)
Bar chart presenting the changes in annual European age-sex standardised incidence rate and percentages by aetiologies. The x-axis represents time in years, and the y-axis represents incidence rate in Fig.4A, percentages in Fig.4B.
Line chart presenting the trends in annual European age-sex standardised incidence rate. The x-axis represents time in years, and the y-axis represents incidence rate in Fig.4C.
Abbreviations: ArLD: alcohol related liver disease; NAFLD: non-alcoholic fatty liver disease; HBV: hepatitis B virus; HCV: hepatitis C virus

* ArLD overlap: if an individual had 2 or more diagnoses from liver disease aetiologies (ArLD, NAFLD, HBV, HCV, metabolic, hemochromatosis and autoimmune liver diseases) and one of them was ArLD; Non-ArLD overlap: if an individual had 2 or more diagnoses from liver disease aetiologies but none of them was ArLD.

150x99mm (300 x 300 DPI)

Supplemental Materials



Supplemental Fig. 1 Flowchart for Identifying Index Events and Determining Cohort Entry Date.

This flowchart illustrates the decision rules for identifying the index event and determining the cohort entry date for study participants. The index date is defined as the date of the first diagnosis between 2004 and 2022. The aetiological diagnoses throughout the disease progression were identified as the index event entering the cohort. If no aetiological diagnosis is present, the index event is defined based on the following sequence: stage 2 diagnoses, stage 3 diagnoses, stage 4 diagnoses, and stage 5 diagnoses.

The decision flow proceeds as follows:

Tier 1 Diagnosis: If present at any point during the disease progression, the tier 1 diagnosis is recorded as the index event, and the date of the cohort entry is the index day. Tier 1 diagnoses include ArLD, NAFLD, metabolic liver disease, HBV, HCV, autoimmune liver disease, and haemochromatosis.

Tier 2 Diagnosis: If there is no tier 1 diagnosis, the presence of a tier 2 diagnosis is checked. If present, it is recorded as the index event, and the date of cohort entry is the index day. Tier 2 diagnoses include hepatitis not specified, congestive hepatopathy, and toxic liver disease.

Tier 3 Diagnosis: In the absence of tier 1 and tier 2 diagnoses, the presence of a tier 3 diagnosis is checked. If found, it is recorded as the index event. Tier 3 diagnoses include miscellaneous conditions.

Stage 2 Diagnosis: If none of the Stage 1 diagnoses are present, the presence of a stage 2 diagnosis is considered. If found, it is recorded as the index event. Stage 2 diagnoses include hepatic fibrosis and cirrhosis.

Stage 3 Diagnosis: If no stage 1 or stage 2 diagnoses are present, a stage 3 diagnosis is checked next. If found, it is recorded as the index event. Stage 3 diagnoses include portal hypertension.

Stage 4 Diagnosis: If the stage 1, stage 2, and stage 3 diagnoses are absent, a stage 4 diagnosis is checked. If present, it is recorded as the index event. Stage 4 diagnoses include hepatic decompensation.

Stage 5 Diagnosis: If no above diagnoses are found, a stage 5 diagnosis is considered. If found, it is recorded as the index event. Stage 5 diagnoses include hepatocellular carcinoma, intrahepatic cholangiocarcinoma, other primary liver cancer.

Acute Liver Diagnosis: If none of the chronic diagnoses are present, an acute liver diagnosis is checked. If found, it is recorded as the index event. Acute liver diagnoses include acute viral hepatitis, Budd-Chiari, acute liver failure, infections/sepsis, and other non-specified acute liver injuries.

Supplemental Tables

Supplemental Table 1 Code list for identifying liver disease

Phenotype	ICD10 codes	Read codes
Acute viral hepatitis	B15, B19, B16, B17 (B171 excluded), B172, B178, B179, B180, B159, B169, B199	A70z1, AyuB0, XE2u., Q4090, A700., A701., A7052, A70.., A706., A708., A709., A70z., A70G., A704., AyuB3
Acute liver failure	K720	J6000
Budd-Chiari	I820	G820.
Infection/infarction	K750, K763, K751	J62.., J620., J6200, J6201, J6202, J6203, J6204, J620z, A053., J634., J621.
Other non-specified acute liver injuries	K752	J63y1
Autoimmune liver disease	K754, K743, K831, K753	J63B., J6141, J6160, J6617, J63X.
Haemochromatosis	E831	C3500
Metabolic liver disease	E880, E830	C3762, C3761, C3510
HBV *	B181, B180	
HCV *	B182	A70E., A70F.
Alcohol-related liver disease	K70	J613., J6130, J612., J6120, J610., J617., J6170, J611.
Non-alcoholic fatty liver disease	K760, K7581	J61y1, J61y8
Hepatitis not specified	K769, K7589, K73	Jyu72, J614y
Congestive hepatopathy	K761, K762, K765	J630., J636., J637.
Toxic liver disease	K71	J635., J6350, J6351, J63252, J6353, J6354, J6355, J6356, J6357, J635X
Miscellaneous	K764, K768, K77	J638., Jyu73, J63yz, Jyu75
Hepatic fibrosis	K740, K741, K742	J61y4, J61y6, J61y5
Cirrhosis	K703, K744, K745, K746, K749	J6161, J616z, J615z
Portal hypertension	K766, I81, I859, I982, I85	J623., G81.., G8523, G852., G8521, G8522, G852z

Hepatic decompensation	K721, K767, I850, K72, C220	J624., SP143, G850., G8520, J625., B1503, BB5D7, BB5D5, BB5D8
Hepatocellular carcinoma (HCC)	C220	B1503, BB5D7, BB5D5, BB5D8
Intrahepatic cholangio carcinoma (ICC)	C221	B150.
Other primary liver cancers	C222, C223, C224, C225, C226. C227	B808.

*We identified Read codes(A7071, A7073, ZV02B, Q4091, 43B4., A7070, A7051, A7072, A70z0, A70A., A70B., A70C., A70D., A70E.,A70F., ZV02C) and ICD-10 codes (B180, B181, B182) for HBV and HCV. However, in order to comply with Data Protection Act 2018 and the UK General Data Protection Regulation, we could not include Read codes (A7071, A7073, ZV02B, Q4091, 43B4., A7070, A7051, A7072, A70z0, A70A., A70B., A70C., A70D., ZV02C) and ICD-10 codes (B171) as these were flagged as sensitive in the latest version of known sensitive code list of SAIL Databank.

Supplemental Table 2-1 ICD-10 code list for identifying comorbidities

Comorbidities	ICD-10 codes
Atrial fibrillation	I481, I482
Angina	I200, I201, I208, I209
Asthma	J45
Diabetes	E09, E10, E102, E103, E11, E13, K86
Heart failure	I50, I501, I502, I503, I508
Hypertension	I10, I11, I12, I13, I15
Peripheral vascular disease	E106, E116, I70
Renal disease	N18
Stroke	I691, I61, I63, I64, I60, I66, G45.0, G45.1, G45.3, G46.0, G46.2, G45.8, I65, G46.1, 45.9, G45.2, G45.4
Transient ischaemic attack	I61, I63, I66
Other ischaemic	I20, I21, I22, I23, I24, I25

Supplemental Table 2-2 Read codes list for identifying comorbidities

Comorbidities	Read codes
Atrial fibrillation	14AN., 14AR., 3272, 3273, 8CMW2, G573., G5730, G5731, G5732, G5733, G5734, G5735, G5736, G5737, G5738, G5739, G573z, 3272, 2[illegible], 793M1, 793M3
Angina	G3112, G33.., G330., G3300, G330z, G33z., G33z3, G33z7, G33zz, 662K., 662K0, 662K1, 662K2, 662Kz, 8B27., G33z1, G33z2, G33z5, G34[illegible], G34y0, Gyu30
Asthma	H33.., H330., H3300, H3301, H330z, H331., H3310, H3311, H331z, H332., H333., H334., H335., H33z., H33z0, H33z1, H33z2, H33zz,
Diabetes	66AJ., 66AJ1, 66AJz, 66An., 66Ao., 8CR2., 9OLA., 9OLA., C10.., C100., C1000, C100z, C101., C1010, C1011, C101y, C102., C1020, C1[illegible]02z, C103., C1030, C1031, C103y, C103z, C104., C1040, C1041, C104y, C104z, C105., C1050, C1051, C105y, C105z, C106., C1060, C1061, C106y, C106z, C107., C1070, C1071, C1072, C[illegible]1074, C107y, C107z, C108., C1080, C1081, C1082, C1083, C1084, C1085, C1086, C1087, C1088, C1089, C108A, C108B, C108C, C108D, C108E, C108F, C108G, C108H, C108J, C108y, C[illegible]08z, C109., C1090, C1091, C1092, C1093, C1094, C1095, C1096, C1097, C1099, C109A, C109B, C109C, C109D, C109E, C109F, C109G, C109H, C109J, C109K, C10A., C10A0, C10A1, C10A2, C10A3, C10A4, C10A5, C10A6, C10A7, C10AW, C10AX, C10B., C10B0, C10C., C10D., C10E., C10E0, C10E1, C10E2, C10E3, C10E4, C10E5, C10E6, C10E7, C10E8, C10E9, C10EA, C10EB, C10EC, C10ED, C10EE, C10EF, C10EG, C10EH, C10EJ, C10EK, C10EL, C10EM, C10EN, C10EP, C10EP, C10EQ, C10ER, C10F., C10F0, C10F1 C10F2, C10F3, C10F4, C10F5, C10F6, C10F7, C10F[illegible], C10FA, C10FB, C10FC, C10FD, C10FE, C10FF, C10FG, C10FH, C10FJ, C10FK, C10FL, C10FM, C10FN, C10FP, C10FQ, C10FR, C10FS, C10G., C10G0, C10H., C10H0, C10K., C10K0, C10M., C10M0, C10N., C10N0, C10N1, C10P., C10P0, C10P1, C10y., C10y1, C10yy, C10yz, C10z., C10z0, C10z1, C10zy, C10zz, F372., F3720, F3721, F3722, 1434, 14F4., 14P3., 9OL9.
Heart failure	G58.., G580., G5800, G5801, G5802, G5803, G5804, G581., G5810, G582., G583., G584., G58z., G232., G234., G1yz1, 1O1.., 662W., 662[illegible], 8[illegible]29., 8H2S., 9Or0., G400., G41z., G5540, G5540, G5yy9, G5yyA, R2y10, 585f., 585g., 14A6., 14AM., 1736, 1J60., 23E1., 388D., 662T., 662f., 662g., 662h., 662i., 679X., 8CL3., 8HBE., 8[illegible]Hz., 8[illegible]Hg8., 8Hk0., 9N0k., 9N2p., 9N4s., 9N4w., 9N6T., 9On.., 9On0., 9On1., 9On2., 9On3., 9On4., 9Or.., 9Or1., 9Or2., 9Or3., 9Or4., 9Or5., 9h1.., 9h11., 9h12., 9hH.., 9hH0., 9hH1., G581., H54.., H54[illegible], H5410, H541z, H54z., H584., H584z, ZRad.
Hypertension	14A2., G2..., G20.., G200., G201., G202., G203., G20z., G21.., G210., G2100, G2101, G210z, G211., G2110, G2111, G211z, G21z., G21z0, G21[illegible], G21zz, G22.., G220., G221., G222., G22z., G23.., G230., G231., G232., G233., G234., G23z., G24.., G240., G2400, G240z, G241., G2410, G241z, G244., G24z., G24z0, G24z1, G24zz, G25.., G2[illegible]0., G251., G26.., G27.., G28.., G2y.., G2z.., 6627, 6628, 662F., 662G., 662O., 662b., 662c., 662d., 662r., 7Q01., 8B26., 8BL0., 8I3N., F4042, F4213, G672., Gyu2., L122., L1220, L1221, L1223, L122z, L127., L127z, L128., L1280, L1282, Gyu21

Comorbidities	Read codes
Peripheral vascular disease	G73.., G734., G73y., G73z., G73z0, G73zz, Gyu74, 2G63., A3A0F, C107., C1070, C1071, C1073, C1074, C107z, C108G, C109F, C109F, C10EH, C10FF, G700., G702., G702z, G731., G7310, G731z, G732., G7320, G7321, G733., G73y0, G73y1, G73yz, G740., G742z, M271., M2710, M2713, R0550, R0550
Renal disease	1Z13., 1Z14., 1Z1H., 1Z1J., 1Z1K., 1Z1L., K050., K054., K055., K060., K060., K08z., K0D.., 1Z10., 1Z17., 1Z18., 1Z11., 1Z19., 1Z1A., 1Z1Y., 1Z15., 1Z16., 1Z1B., 1Z1C., 1Z1D., 1Z1E., 1Z1F., 1Z1G.,
Stroke	G6..., G61.., G610., G611., G612., G613., G614., G615., G616., G617., G618., G619., G61X., G61X0, G61X1, G61z., G63.., G630., G631., G632., G633., G634., G63y., G63y0, G63y1, G63z., G64.., G640., G6400, G641., G6410, G64z., G64z0, G64z1, G64z2, G64z3, G64z4, G66.., G660., G661., G662., G663., G664., G665., G666., G668., G67.., G670., G671., G6710, G6711, G671z, G677., G6770, G6771, G6772, G6773, G6774, G679., G67y., G67z., G6y.., G67z.
Transient ischaemic attack	G65z., G65zz, G65z1, G65y., 14AB., G65z0, Fyu55, G65.., G65.., G650., G651., G6510, G652., G653., G654., G655., G656., G657., G65y., F4236, 14AB0
Other ischaemic	G33z4, G34.., G34y., G34y0, G34y1, G34yz, G34z., G34z0, G3y.., G3z.., G31y3, G332., 6A2.., 6A4.., 8B3k., 8H2V., G3..., G31.., G3110, G31y., G31y2, G31yz, G340., G343., G344., Gyu3., Gyu32, Gyu33
Anti-hypertensive	bi1.., bi11., bi12., bi13., bi14., bi15., bi16., bi17., bi18., bi19., bi1A., bi1B., bi1C., bi1H., bi1I., bi1J., bi1K., bi1a., bi1b., bi1c., bi1d., bi1g., bi1h., bi1i., bi1j., bi1k., bi1l., bi1m., bi1n., bi1o., bi1p., bi1q., bi1r., bi1v., bi1w., bi1x., bi1y., bi1z., bi2.., bi21., bi22., bi23., bi24., bi25., bi26., bi27., bi29., bi2A., bi2B., bi2C., bi2D., bi2E., bi2F., bi2G., bi2H., bi2J., bi2K., bi2L., bi2M., bi2a., bi2t., bi2u., bi2v., bi2w., bi2x., bi2y., bi2z., bi3.., bi31., bi32., bi33., bi34., bi35., bi36., bi37., bi38., bi39., bi3a., bi3b., bi3c., bi3c., bi3d., bi3d., bi3e., bi3f., bi3g., bi3h., bi3i., bi3j., bi3k., bi3l., bi3m., bi3q., bi3r., bi3y., bi4.., bi41., bi42., bi43., bi44., bi45., bi46., bi47., bi49., bi4A., bi4B., bi4C., bi4D., bi4E., bi5.., bi51., bi52., bi53., bi54., bi57., bi58., bi6.., bi61., bi62., bi63., bi64., bi65., bi66., bi67., bi68., bi69., bi6A., bi6B., bi6C., bi6D., bi6E., bi6F., bi6G., bi6o., bi6p., bi6q., bi6r., bi6s., bi6t., bi6u., bi6v., bi6w., bi6x., bi6y., bi6z., bi7.., bi71., bi72., bi73., bi74., bi8.., bi81., bi82., bi82., bi83., bi83., bi84., bi84., bi85., bi86., bi86., bi87., bi87., bi88., bi88., bi89., bi89., bi8a., bi9.., bi91., bi92., bi93., bi94., bi94., bi95., bi95., bi96., bi96., bi97., bi98., bi99., bi9A., bi9A., bi9z., biA.., biA1., biA2., biA3., biA4., biB.., biB1., biB2., biB3., biBx., biBy., biBz., biC.., biC1., biC2., biC3., biC4., biC5., biC6., bk3.., bk31., bk32., bk33., bk34., bk37., bk38., bk3B., bk3C., bk3D., bk3E., bk3F., bk3G., bk3H., bk4.., bk41.,

Comorbidities	Read codes
	bk42., bk43., bk44., bk45., bk46., bk4A., bk4B., bk4C., bk4s., bk4t., bk4u., bk4v., bk4w., bk5.., bk51., bk52., bk53., bk54., bk55., bk56., bk7.., bk71., bk72., bk73., bk74., bk75., bk76., bk77., bk78., bk79., bk7z., bk8.., bk81., bk82., bk83., bk84., bk85., bk8z., bk9.., bk91., bk92., bk92., bk93., bk9x., bk9y., bk9z., bkB.., bkB1., bkB2., bkB3., bkB4., bkB5., bkB6., bkJ.., bkJ1., bkJ2., bkJ3., bkJ4., bkJ5., bkJ6., bd□c, bd1.., bd11., bd12., bd13., bd14., bd15., bd16., bd17., bd18., bd19., bd1A., bd1B., bd1C., bd1D., bd1E., bd1F., bd1G., bd1H., bd1J., bd1K., bd1L., bd1M., bd1N., bd1O., bd1P., bd1Q., bd1R., bd1S., bd1T., bd1U., bd1V., bd1W., bd1X., bd1Y., bd1Z., bd1a., bd1b., bd1c., bd1d., bd1e., bd1f., bd1g., bd1h., bd1i., bd1j., bd1k., bd1l., bd1l., bd1m., bd1n., bd1o., bd1p., bd1r., bd1r., bd1s., bd1t., bd1u., bd1v., bd1w., bd1x., bd1y., bd1z., bd2.., bd21., bd22., bd23., bd2w., bd2x., bd2y., bd3.., bd31., bd32., bd34., bd35., bd36., bd37., bd3a., bd3b., bd3c., bd3c., bd3d., bd3e., bd3f., bd3g., bd3h., bd3i., bd3j., bd3k., bd3l., bd3x., bd3z., bd4.., bd41., bd4z., bd5.., bd51., bd52., bd53., bd54., bd55., bd56., bd57., bd58., bd59., bd5s., bd5t., bd5u., bd5v., bd5w., bd5x., bd5y., bd6.., bd61., bd62., bd64., bd65., bd66., bd67., bd68., bd6b., bd6c., bd6d., bd6e., bd6w., bd6x., bd6z., bd7.., bd71., bd72., bd7y., bd7z., bd8.., bd81., bd82., bd83., bd84., bd85., bd86., bd87., bd88., bd89., bd8a., bd8b., bd8c., bd8d., bd8e., bd8f., bd8g., bd8h., bd8i., bd8k., bd8l., bd8m., bd8n., bd8o., bd8u., bd9.., bda.., bda1., bda2., bda3., bda4., bday., bdaz., bdb.., bdc.., bdc1., bdc1., bdc2., bdc3., bdc4., bdc5., bdcu., bdcv., bdcw., bdcx., bdd.., bdd1., bdd1., bdd2., bddz., bde.., bde1., bde2., bde3., bde4., bde5., bde6., bde7., bde8., bde9., bdeQ., bdeR., bdeS., bdeb., bdec., bded., bdee., bdef., bdeg., bdeh., bdei., bdej., bdek., bdel., bdf.., bdf1., bdf2., bdf3., bdf4., bdf5., bdf5., bdf6., bdf6., bdf7., bdf8., bdf9., bdfA., bdfB., bdfC., bdfD., bdfE., bdfF., bdfG., bdfH., bdfI., bdfJ., bdfK., bdfL., bdfM., bdfw., bdfx., bdfy., bdfz., bdg.., bdg1., bdg2., bdh.., bdh1., bdh2., bdh3., bdh4., bdi.., bdi1., bdi2., bdj.., bdj1., bdj2., bdj3., bdj4., bdj5., bdl.., bdl1., bdl1., bdl2., bdl3., bdl4., bdl5., bdl6., bdl7., bdl8., bdl8., bdm.., bdm1., bdm2., bdmy., bdmz., bdn.., bdn1., bdn2., bdn3., bdn4., bdn5., bdn6., bb3.., bb31., bb32., bb33., bb34., bb35., bb36., bb37., bb38., bb39., bb3A., bb3B., bb3C., bb3D., bb3F., bb3G., bb3H., bb3J., bb3K., bb3L., bb3M., bb3N., bb3O., bb3P., bb3Q., bb3a., bb3b., bb3d., bb3e., bb3f., bb3g., bb3h., bb3i., bb3j., bb3k., bb3l., bb3m., bb3n., bb3p., bb3q., bb3r., bb3s., bb3v., bb3w., bb3x., bb3y., bb3z., bl5.., bl51., bl52., bl53., bl54., bl55., bl56., bl57., bl58., bl59., bl5A., bl5B., bl5C., bl5D., bl5E., bl5F., bl5G., bl5H., bl5I., bl5J., bl5K., bl5L., bl5M., bl5N., bl5O., bl5P., bl5Q., bl5R., bl5S., bl5T., bl5U., bl5V., bl5V., bl5W., bl5W., bl5X., bl5Y., bl5Z., bl5a., bl5b., bl5c., bl5d., bl5e., bl5f., bl5g., bl5h., bl5j., bl5k., bl5l., bl5m., bl5n., bl5o., bl5p., bl5q., bl5r., bl5s., bl5t., bl5u., bl5v., bl5w., bl5x., bl5y., bl5z., bl7.., bl71., bl72., bl73., bl74., bl7w., bl7x., bl7y., bl7z., bl8.., bl81., bl82., bl83., bl84., bl85., bl86., bl89., bl8A., bl8B., bl8C., bl8D., bl8E., bl8F., bl8G., bl8H., bl8J., bl8K., bl8L., bl8M., bl8O., bl8P., bl8Q., bl8R., bl8S., bl8T., bl8U., bl8V., bl8W., bl8X., bl8Y., bl8Z., bl8a., bl8b., bl8c., bl8d., bl8e., bl8f., bl8g., bl8h., bl8i., bl8j., bl8k., bl8l., bl8m., bl8n., bl8o., bl8p., bl8q., bl8r., bl8s., bl8t., bl8u., bl8v., bl8w., bl8x., bl8y., bl8z., bla.., bla.., bla1., bla1., bla2., bla2., blb.., blb1., blb2., blb3., blb4., blb5., blb5., blb6., blb6., blb7., blb8., blc.., blc1., blc2., blc3., blc4., blc5., blc6., blc7., blc8., blc9., blca., blcb., blcc., blcd., blce., blcf., blcg., blch., blci., blcj., blck., blcl., blcm., blcn., blco., blcp., blcq., blcr., blcs., blct., ble.., ble1., ble2., ble3., ble4., ble5., blg.., blg1., blg2., blg3., blg4., blg5., blg6., blh.., blh1., blh2., blh3., blh4., blj.., blj1., blj2., blj3., blj4., blj5., blj6., blj7., blj8., blj9., bljA., bljB., bljC., bljD., bljE., bljF., bljG., bljH., bljJ., bljK., bljL., bljM., bljN., bljO., bljP., bljQ., bljR., bljS., bljT., bljU., bljV., bljW., bljX., bljY., bljZ., blja., bljb., bljc., bljd., blje., bljf., bll.., bll1., bll2., bll3., bll4., bll5., bll6., bll7., bll8., bll9., blla., bllb., bllc., blld., blle., bllf., bllg., bllh., blli., bllj., bllk., blll., dt1.., dt13., dt14., b2..., b21.., b211., b212., b213., b214., b215., b216., b217., b218., b219., b21A., b21B., b21a., b21b., b22.., b221., b22., b22y., b22z., b23.., b231., b232., b23y., b23z.,

Comorbidities	Read codes
	b24.., b25.., b251., b25z., b26.., b261., b262., b263., b264., b26y., b26z., b27.., b271., b27z., b28.., b281., b282., b283., b284., b285., b286., b287., b288., b289., b28z., b29.., b291., b29z., b2a.., b2a1., b2az., b2b.., b2b1., b2b2., b2b3., b2bz., b2c.., b2c1., b2cz., b2d.., b2d1., b2dz., bA1.., bA11., bA12., bA1y., bA1z., bi1D., bi1E., bi1F., bi1G., bi1H., bi1f., bi1s., bi28., bi2b., bi3n., bi3p., bi3s., bi3t., bi3u., bi3v., bi3w., bi3x., bi48., bi4F., bi55., bi56., biC7., biC8., bk35., bk36., bk39., bk3A., bk3y., bk3z., bk47., bk48., bk49., bk4x., bk4y., bk4z., bk57., bk58., bk59., bk5x., bk5y., bk5z., bk86., bk87., bk88., bk8w., bk8x., bk8y., bkC.., bkC1., bkC2., bkC3., bkCx., bkCy., bkCz., bkH.., bkH1., bkH2., bkH3., bkHx., bkHy., bkHz., bkI.., bkI1., bkI2., bkI3., bkI4., bkI5., bkL.., bkL1., bkL2., bkL3., bkL4., bkL5., bkL6., bd38., bd39., bdeA., bdeB., bdeC., bdeD., bdeE., bdeF., bdeG., bdeH., bdeJ., bdeK., bdeL., bdeM., bdeN., bdeO., bdeP., bdem., bden., bdeo., bdep., bdeq., bder., bdes., bdes., bdet., bdeu., bdev., bdew., bdex., bdey., bdez., bl5i., bh4.., bh5y., bh56., bh41., bh4x., bh5z., bh63., bh55., bh54., bh6B., bh1y., bh65., bh4z., bh14., bh4D., bh4A., bh68., bh4v., bh4B., bh6F., bh61., bh69., bh46., bh21., bh6E., bh4y., bh45., bh47., bh6H., bh42., bh6y., bh1z., bh5.., bh6C., bh6D., bh57., bh4C., bh4A., bh6G., bh66., bh1.., bh4w., bh53., bh4., bh52., bh43., bh5x., bh6z., bh13., bh51., bh64., bh67., bh49., bh48., bh11., bh2y., bh6.., bh12., bh62., bf39., bf26., bf1w., bf1x., bf35., bf3b., bf42., bf44., bf3a., bf4.., bf2.., bf2d., bf22., bf2v., bf1z., bf2., bf27., bf2j., bf23., bf24., bf12., bf3d., bf25., bf3c., bf31., bf43., bf36., bf2e., bf33., bf2b., bf2g., bf21., bf11., bf2h., bf2z., bf32., bf34., bf41., bf29., bf2a., bf2f., bf45., bf37., bf38., bf46., be3x., be3., be3.., be3z., be3y., be1.., be2x., be22., be32., be21., be31., be2.., be33.,

Supplemental Table 3-1. Standardised incidence rate of liver disease by data sources (2004 to 2022)

Year	PEDW only group		WLGP only group		ADDE only group		Two or more data sources	
	STD incidence	STD 95% CI	STD incidence	STD 95% CI	STD incidence	STD 95% CI	STD incidence	STD 95% CI
2004	41.2	(38.9 to 43.5)	13.5	(12.2 to 14.9)	3.8	(3.1 to 4.5)	39.2	(37.1 to 41.5)
2005	49.6	(47.2 to 52.1)	15.3	(13.9 to 16.8)	3.8	(3.1 to 4.5)	40.7	(38.5 to 43)
2006	54.4	(51.9 to 57.1)	16.1	(14.7 to 17.6)	3.7	(3.1 to 4.5)	44.4	(42.1 to 46.9)
2007	54.8	(52.2 to 57.5)	13.9	(12.6 to 15.3)	4.5	(3.7 to 5.3)	41.9	(39.6 to 44.3)
2008	59.3	(56.6 to 62.1)	18.0	(16.5 to 19.6)	5.5	(4.7 to 6.4)	43.5	(41.2 to 45.9)
2009	63.9	(61.1 to 66.9)	18.5	(17 to 20.1)	5.5	(4.7 to 6.5)	43.7	(41.4 to 46.2)
2010	67.4	(64.5 to 70.4)	21.7	(20.1 to 23.4)	5.6	(4.7 to 6.5)	45.9	(43.5 to 48.4)
2011	74.1	(71.1 to 77.3)	23.6	(21.8 to 25.4)	5.1	(4.3 to 6)	47.1	(44.7 to 49.6)
2012	76.3	(73.2 to 79.5)	21.5	(19.9 to 23.2)	5.5	(4.7 to 6.4)	47.3	(44.9 to 49.9)
2013	82.7	(79.4 to 86)	22.3	(20.6 to 24.1)	6.0	(5.1 to 6.9)	47.1	(44.6 to 49.6)
2014	90.3	(86.9 to 93.8)	26.2	(24.4 to 28.1)	6.1	(5.3 to 7.1)	49.8	(47.3 to 52.4)
2015	99.1	(95.5 to 102.8)	32.2	(30.2 to 34.3)	6.2	(5.3 to 7.2)	54.7	(52.1 to 57.5)
2016	115.9	(112 to 120)	45.1	(42.7 to 47.6)	6.4	(5.5 to 7.4)	58.3	(55.5 to 61.1)
2017	122.9	(118.8 to 127.2)	48.9	(46.4 to 51.5)	6.5	(5.5 to 7.5)	64.9	(62 to 67.9)
2018	141.0	(136.5 to 145.5)	62.5	(59.7 to 65.5)	7.1	(6.1 to 8.2)	68.5	(65.5 to 71.7)
2019	157.6	(152.9 to 162.4)	71.4	(68.4 to 74.6)	6.4	(5.4 to 7.4)	65.8	(62.8 to 68.9)
2020	153.3	(148.5 to 158.1)	50.2	(47.6 to 52.8)	8.7	(7.6 to 9.8)	51.9	(49.2 to 54.7)
2021	197.5	(192.1 to 203)	73.4	(70.2 to 76.6)	7.5	(6.5 to 8.6)	53.5	(50.8 to 56.4)
2022	182.1	(176.9 to 187.4)	88.6	(85.1 to 92.1)	7.9	(6.9 to 9.1)	37.6	(35.2 to 40.1)

Abbreviation: STD: standardized; CI: confidence interval

Supplemental Table 3-2 Standardised incidence rate of liver disease by disease stages (2004 to 2022)

Year	Stage 1		Stage 2		Stage 3		Stage 4		Stage 5	
	STD incidence	STD 95% CI	STD incidence	STD 95% CI	STD incidence	STD 95% CI	STD incidence	STD 95% CI	STD incidence	STD 95% CI
2004	70.3	(67.4 to 73.3)	5.8	(5 to 6.7)	5.1	(4.3 to 6)	4.9	(4.2 to 5.8)	2.7	(2.1 to 3.3)
2005	79.6	(76.4 to 82.8)	6.8	(5.9 to 7.8)	5.2	(4.4 to 6)	5.3	(4.5 to 6.2)	2.9	(2.3 to 3.6)
2006	86.9	(83.6 to 90.3)	7.1	(6.2 to 8.1)	6.3	(5.5 to 7.3)	6.2	(5.3 to 7.2)	3.5	(2.9 to 4.2)
2007	84.0	(80.8 to 87.4)	7.5	(6.5 to 8.5)	5.7	(4.9 to 6.6)	6.4	(5.5 to 7.3)	3.5	(2.9 to 4.2)
2008	92.3	(88.9 to 95.9)	8.1	(7.1 to 9.2)	8.2	(7.3 to 9.3)	5.7	(4.9 to 6.6)	4.4	(3.7 to 5.2)
2009	97.8	(94.3 to 101.4)	8.6	(7.6 to 9.7)	7.6	(6.6 to 8.6)	6.4	(5.5 to 7.4)	4.1	(3.4 to 4.9)
2010	106.0	(102.3 to 109.7)	8.1	(7.1 to 9.1)	8.3	(7.3 to 9.4)	6.5	(5.6 to 7.5)	3.6	(2.9 to 4.3)
2011	111.6	(107.8 to 115.5)	10.1	(9 to 11.3)	8.4	(7.4 to 9.5)	6.6	(5.7 to 7.6)	4.6	(3.9 to 5.5)
2012	112.3	(108.5 to 116.1)	9.5	(8.4 to 10.7)	8.2	(7.2 to 9.3)	6.6	(5.7 to 7.6)	4.7	(4 to 5.6)
2013	115.4	(111.6 to 119.4)	10.3	(9.1 to 11.5)	8.9	(7.9 to 10.1)	7.4	(6.4 to 8.5)	6.0	(5.2 to 7)
2014	127.9	(123.9 to 132.1)	11.6	(10.4 to 12.9)	9.4	(8.4 to 10.6)	7.1	(6.2 to 8.2)	5.3	(4.5 to 6.2)
2015	141.9	(137.6 to 146.3)	14.0	(12.7 to 15.5)	12.5	(11.2 to 13.8)	6.5	(5.6 to 7.5)	6.3	(5.4 to 7.3)
2016	170.5	(165.7 to 175.3)	16.2	(14.7 to 17.7)	13.2	(11.9 to 14.7)	8.1	(7.1 to 9.2)	6.4	(5.5 to 7.5)
2017	185.5	(180.5 to 190.6)	17.8	(16.2 to 19.4)	13.5	(12.2 to 14.9)	8.0	(7 to 9.2)	6.7	(5.7 to 7.8)
2018	212.9	(207.5 to 218.4)	21.6	(19.9 to 23.4)	14.7	(13.3 to 16.2)	9.2	(8.1 to 10.3)	7.3	(6.3 to 8.5)
2019	231.8	(226.2 to 237.6)	22.1	(20.4 to 24)	17.8	(16.2 to 19.5)	8.9	(7.8 to 10.1)	6.6	(5.7 to 7.6)
2020	201.7	(196.3 to 207.1)	19.5	(17.8 to 21.2)	14.3	(12.9 to 15.8)	8.9	(7.8 to 10.2)	7.7	(6.6 to 8.9)
2021	263.0	(256.9 to 269.3)	21.6	(19.9 to 23.5)	18.3	(16.7 to 20.1)	8.2	(7.2 to 9.4)	7.9	(6.8 to 9.1)
2022	244.7	(238.7 to 250.7)	23.5	(21.6 to 25.5)	19.9	(18.2 to 21.8)	7.5	(6.4 to 8.8)	8.7	(7.6 to 10)

Abbreviation: STD: standardized; CI: confidence interval

Supplemental Table 3-3, Standardised incidence rate of liver disease by aetiologies (2004 to 2022)

Year	ArLD	NAFLD	Metabolic liver disease	HBV	HCV	Autoimmune liver disease	Haemochromatosis	ArLD overlap	Non-ArLD overlap	Hepatitis of unspecified origin	Congestive hepatopathy	Toxic liver disease	Miscellaneous
	STD incidence (95% CI)	STD incidence (95% CI)	STD incidence (95% CI)	STD incidence (95% CI)	STD incidence (95% CI)	STD incidence (95% CI)	STD incidence (95% CI)	STD incidence (95% CI)	STD incidence (95% CI)	STD incidence (95% CI)	STD incidence (95% CI)	STD incidence (95% CI)	STD incidence (95% CI)
2004	22.0 (20.3 to 23.7)	10.8 (9.6 to 12)	4.4 (3.7 to 5.2)	0.6 (0.4 to 0.9)	3.0 (2.4 to 3.7)	14.4 (13.2 to 15.8)	3.2 (2.6 to 3.9)	0.3 (0.2 to 0.6)	0.4 (0.2 to 0.7)	3.6 (2.2 to 5.1)	0.6 (0.4 to 0.9)	1.8 (1.3 to 2.3)	5.3 (4.5 to 6.1)
2005	21.7 (20.1 to 23.5)	14.0 (12.7 to 15.4)	4.7 (3.9 to 5.5)	0.6 (0.3 to 1)	2.9 (2.3 to 3.5)	18.4 (16.9 to 19.9)	4.0 (3.3 to 4.8)	0.4 (0.2 to 0.7)	0.4 (0.2 to 0.7)	4.0 (2.6 to 5.6)	0.5 (0.3 to 0.8)	2.3 (1.8 to 2.9)	5.8 (5 to 6.7)
2006	21.9 (20.2 to 23.6)	17.9 (16.4 to 19.5)	5.1 (4.4 to 6)	0.7 (0.4 to 1.1)	3.9 (3.2 to 4.7)	17.7 (16.2 to 19.2)	5.1 (4.3 to 6)	--	0.8 (0.5 to 1.2)	4.4 (3 to 5.6)	0.6 (0.3 to 0.9)	1.9 (1.4 to 2.5)	6.3 (5.4 to 7.2)
2007	22.3 (20.6 to 24)	17.1 (15.7 to 18.6)	5.1 (4.4 to 6)	0.6 (0.4 to 1)	2.8 (2.2 to 3.5)	17.1 (15.7 to 18.6)	4.3 (3.6 to 5.1)	--	0.8 (0.5 to 1.2)	4.9 (4.2 to 5.8)	0.6 (0.3 to 0.9)	1.6 (1.2 to 2.1)	6.5 (5.6 to 7.5)
2008	20.2 (18.6 to 21.9)	21.2 (19.6 to 22.9)	4.8 (4 to 5.6)	0.8 (0.5 to 1.2)	3.8 (3.1 to 4.6)	20.1 (18.6 to 21.8)	4.8 (4 to 5.6)	--	0.6 (0.4 to 1)	6.3 (5.4 to 7.2)	0.8 (0.5 to 1.2)	1.2 (0.9 to 1.7)	7.4 (6.5 to 8.4)
2009	21.1 (19.5 to 22.8)	23.3 (21.6 to 25.1)	5.5 (4.7 to 6.4)	0.7 (0.4 to 1.1)	3.2 (2.6 to 4)	23.8 (22.1 to 25.6)	5.6 (4.8 to 6.6)	0.5 (0.2 to 0.8)	0.5 (0.3 to 0.9)	4.5 (3.8 to 5.3)	0.6 (0.3 to 1)	1.0 (0.7 to 1.4)	7.5 (6.6 to 8.6)
2010	20.9 (19.2 to 22.6)	28.8 (26.9 to 30.8)	5.2 (4.4 to 6)	0.8 (0.5 to 1.1)	3.4 (2.7 to 4.1)	25.9 (24.1 to 27.7)	5.7 (4.9 to 6.6)	0.3 (0.2 to 0.6)	0.5 (0.3 to 0.8)	5.3 (4.5 to 6.2)	0.7 (0.4 to 1.1)	0.9 (0.6 to 1.3)	7.8 (6.8 to 8.8)
2011	18.8 (17.3 to 20.4)	32.6 (30.6 to 34.7)	6.4 (5.5 to 7.4)	1.3 (0.9 to 1.8)	3.6 (3 to 4.4)	24.4 (22.7 to 26.3)	5.8 (5 to 6.7)	0.4 (0.2 to 0.7)	0.6 (0.4 to 1)	5.9 (5 to 6.8)	0.4 (0.2 to 0.7)	1.5 (1.1 to 2)	9.8 (8.7 to 11)
2012	17.3 (15.9 to 18.9)	32.6 (30.6 to 34.7)	8.1 (7.1 to 9.2)	0.8 (0.5 to 1.2)	4.2 (3.5 to 5.1)	23.0 (21.3 to 24.8)	6.3 (5.5 to 7.3)	0.4 (0.2 to 0.7)	0.6 (0.4 to 1)	6.2 (5.3 to 7.1)	0.8 (0.5 to 1.2)	1.4 (1 to 1.9)	10.3 (9.2 to 11.6)
2013	17.8 (16.3 to 19.4)	36.0 (33.9 to 38.2)	8.6 (7.5 to 9.7)	1.0 (0.7 to 1.5)	4.7 (4 to 5.6)	18.4 (16.9 to 20)	5.8 (5 to 6.7)	0.4 (0.2 to 0.7)	0.6 (0.4 to 1)	8.3 (7.3 to 9.4)	0.8 (0.5 to 1.2)	1.1 (0.7 to 1.5)	11.9 (10.7 to 13.2)
2014	17.3 (15.9 to 18.9)	45.2 (42.8 to 47.7)	8.4 (7.4 to 9.5)	0.7 (0.4 to 1.1)	5.7 (4.9 to 6.6)	17.8 (16.3 to 19.3)	6.5 (5.6 to 7.4)	0.3 (0.2 to 0.6)	0.8 (0.5 to 1.1)	7.8 (6.8 to 8.9)	0.9 (0.6 to 1.3)	1.0 (0.6 to 1.4)	15.6 (14.2 to 17.1)

2015	17.8 (16.3 to 19.4)	54.7 (52.1 to 57.5)	9.0 (7.9 to 10.2)	0.9 (0.6 to 1.4)	5.2 (4.4 to 6.1)	19.8 (18.2 to 21.5)	6.0 (5.2 to 7)	0.4 (0.2 to 0.7)	0.5 (0.2 to 0.8)	8.4 (7.4 to 9.6)	1.0 (0.6 to 1.4)	1.1 (0.7 to 1.6)	17.0 (15.6 to 18.6)
2016	18.7 (17.2 to 20.3)	69.5 (66.6 to 72.6)	12.6 (11.3 to 14.1)	1.9 (1.4 to 2.5)	6.9 (6 to 8)	19.8 (18.2 to 21.6)	7.6 (6.6 to 8.7)	0.6 (0.4 to 1)	1.0 (0.7 to 1.4)	10.2 (8.8 to 11.3)	1.2 (0.8 to 1.7)	1.2 (0.8 to 1.6)	19.4 (17.7 to 21.1)
2017	20.3 (18.6 to 22)	81 (77.7 to 84.3)	11.8 (10.5 to 13.2)	2.6 (2 to 3.2)	7.0 (6.1 to 8.1)	18.5 (16.9 to 20.2)	7.6 (6.6 to 8.7)	0.9 (0.6 to 1.3)	1.0 (0.7 to 1.5)	11.1 (9.5 to 12.4)	1.2 (0.9 to 1.7)	1.3 (0.9 to 1.7)	21.1 (19.4 to 22.9)
2018	19.8 (18.2 to 21.5)	100.5 (96.9 to 104.3)	12.6 (11.3 to 14.1)	2.0 (1.5 to 2.5)	5.9 (5 to 6.9)	20.3 (18.6 to 22.1)	8.5 (7.4 to 9.6)	0.8 (0.5 to 1.2)	1.0 (0.7 to 1.4)	12.1 (10.5 to 13.5)	1.4 (0.9 to 1.9)	1.7 (1.2 to 2.2)	26.1 (24.2 to 28.1)
2019	21.7 (20 to 23.5)	114.9 (110.9 to 118.9)	13.0 (11.6 to 14.5)	2.5 (2 to 3.2)	5.6 (4.8 to 6.6)	20.3 (18.6 to 22.1)	8.9 (7.8 to 10.1)	0.8 (0.5 to 1.2)	0.8 (0.5 to 1.2)	12.3 (10.7 to 13.7)	1.5 (1 to 2)	1.2 (0.8 to 1.6)	27.9 (25.9 to 30)
2020	19.9 (18.3 to 21.6)	90.1 (86.6 to 93.7)	11.5 (10.2 to 12.9)	1.6 (1.2 to 2.2)	5.1 (4.3 to 6)	16.2 (14.7 to 17.9)	7.8 (6.8 to 8.9)	0.6 (0.4 to 1)	0.9 (0.5 to 1.3)	13.3 (11.7 to 14.7)	1.6 (1.2 to 2.2)	1.8 (1.3 to 2.4)	31.6 (29.4 to 33.9)
2021	22.2 (20.4 to 24)	129.6 (125.4 to 133.9)	14.1 (12.7 to 15.7)	1.7 (1.3 to 2.3)	3.6 (2.9 to 4.3)	19.4 (17.7 to 21.2)	8.4 (7.4 to 9.6)	0.5 (0.2 to 0.8)	1.4 (1 to 1.9)	16.2 (14.3 to 17.8)	1.4 (0.9 to 2)	1.5 (1.1 to 2.1)	42.9 (40.4 to 45.5)
2022	22.4 (20.7 to 24.2)	124.5 (120.4 to 128.8)	15.2 (13.7 to 16.9)	1.2 (0.9 to 1.7)	2.3 (1.8 to 3)	17.0 (15.5 to 18.7)	9.2 (8 to 10.4)	0.4 (0.2 to 0.8)	1.3 (0.9 to 1.8)	14.4 (13 to 16.3)	1.7 (1.2 to 2.4)	1.2 (0.8 to 1.7)	33.5 (31.2 to 35.9)

Abbreviation: ArLD: alcohol-related liver disease; NAFLD: non-alcohol fatty liver disease; HBV: hepatitis B virus; HCV: hepatitis C virus; STD: standardised; CI: confidence interval

Supplemental Table 4-1, Standardised incidence rate of NAFLD by data sources (2004 to 2022)

Year	PEDW-only		WLGP-only		ADDE-only		Two or more data sources	
	STD incidence	STD 95% CI	STD incidence	STD 95% CI	STD incidence	STD 95% CI	STD incidence	STD 95% CI
2004	3.4	(2.8 to 4.1)	4.2	(3.5 to 5)	0.5	(0.3 to 0.8)	2.7	(2.1 to 3.3)
2005	4.6	(3.9 to 5.4)	5.2	(4.4 to 6.1)	0.3	(0.2 to 0.6)	3.9	(3.2 to 4.6)
2006	6.0	(5.2 to 7)	7.2	(6.3 to 8.3)	--	--	4.4	(3.7 to 5.3)
2007	6.1	(5.2 to 7)	6.7	(5.8 to 7.7)	0.3	(0.2 to 0.6)	4.0	(3.3 to 4.8)
2008	6.0	(5.1 to 6.9)	10.4	(9.2 to 11.6)	0.6	(0.4 to 0.9)	4.2	(3.5 to 5)
2009	6.6	(5.7 to 7.6)	10.7	(9.6 to 12)	0.6	(0.3 to 0.9)	5.4	(4.6 to 6.3)
2010	6.8	(5.9 to 7.8)	14.9	(13.5 to 16.3)	0.7	(0.4 to 1.1)	6.4	(5.6 to 7.4)
2011	9.8	(8.7 to 11)	15.4	(14 to 16.9)	0.4	(0.2 to 0.7)	7.0	(6.1 to 8)
2012	11.6	(10.4 to 12.8)	13.8	(12.5 to 15.2)	0.4	(0.2 to 0.7)	6.9	(6 to 7.9)
2013	13.0	(11.8 to 14.4)	15.2	(13.8 to 16.7)	1.0	(0.7 to 1.4)	6.7	(5.8 to 7.7)
2014	17.6	(16.2 to 19.2)	18.2	(16.7 to 19.8)	1.1	(0.7 to 1.5)	8.2	(7.2 to 9.3)
2015	21.0	(19.4 to 22.8)	23.3	(21.6 to 25.1)	1.0	(0.7 to 1.4)	9.4	(8.3 to 10.6)
2016	24.2	(22.4 to 26)	32.5	(30.5 to 34.7)	1.2	(0.8 to 1.6)	11.7	(10.5 to 13)
2017	29.7	(27.7 to 31.7)	35.7	(33.6 to 38)	1.3	(0.9 to 1.8)	14.2	(12.9 to 15.7)
2018	38.9	(36.7 to 41.3)	45.3	(42.8 to 47.8)	1.6	(1.1 to 2.1)	14.7	(13.4 to 16.2)
2019	45.1	(42.7 to 47.7)	53.4	(50.8 to 56.2)	1.6	(1.2 to 2.1)	14.7	(13.3 to 16.2)
2020	45.0	(42.5 to 47.6)	34.0	(31.9 to 36.2)	1.9	(1.4 to 2.5)	9.1	(8.1 to 10.3)
2021	65.2	(62.2 to 68.3)	52.9	(50.2 to 55.6)	1.5	(1.1 to 2.1)	10.0	(8.8 to 11.2)
2022	60.5	(57.6 to 63.5)	58.7	(55.9 to 61.6)	1.5	(1.1 to 2)	3.9	(3.2 to 4.7)

Abbreviation: STD: standardized; CI: confidence interval

Supplemental Table 4-2, Standardised incidence rate of ArLD by data sources (2004 to 2022)

Year	PEDW-only		WLGP-only		ADDE-only		Two or more data sources	
	STD incidence	STD 95% CI	STD incidence	STD 95% CI	STD incidence	STD 95% CI	STD incidence	STD 95% CI
2004	4.3	(3.6 to 5.1)	3.0	(2.4 to 3.7)	--	--	14.3	(13 to 15.8)
2005	4.9	(4.2 to 5.8)	2.7	(2.2 to 3.4)	0.6	(0.3 to 1)	13.4	(12.1 to 14.8)
2006	5.0	(4.3 to 5.9)	2.7	(2.1 to 3.3)	0.3	(0.2 to 0.6)	13.8	(12.5 to 15.2)
2007	5.4	(4.6 to 6.3)	2.3	(1.8 to 2.9)	0.6	(0.4 to 1)	14.0	(12.7 to 15.4)
2008	4.8	(4.1 to 5.7)	2.0	(1.5 to 2.6)	0.7	(0.4 to 1.1)	12.7	(11.4 to 14)
2009	5.7	(4.8 to 6.6)	2.1	(1.6 to 2.6)	0.9	(0.6 to 1.3)	12.5	(11.2 to 13.8)
2010	5.1	(4.3 to 6)	2.5	(1.9 to 3.1)	0.9	(0.6 to 1.3)	12.5	(11.2 to 13.8)
2011	4.5	(3.8 to 5.3)	2.5	(1.9 to 3.1)	0.7	(0.4 to 1.1)	11.1	(9.9 to 12.4)
2012	4.7	(3.9 to 5.5)	1.8	(1.4 to 2.4)	0.7	(0.4 to 1.1)	10.1	(9 to 11.3)
2013	4.8	(4 to 5.6)	2.0	(1.5 to 2.6)	0.6	(0.4 to 1)	10.4	(9.2 to 11.6)
2014	4.7	(3.9 to 5.5)	2.2	(1.7 to 2.8)	0.6	(0.4 to 1)	9.9	(8.8 to 11.1)
2015	4.4	(3.7 to 5.3)	2.3	(1.8 to 3)	0.7	(0.4 to 1.1)	10.4	(9.2 to 11.6)
2016	4.7	(3.9 to 5.5)	3.9	(3.2 to 4.7)	0.9	(0.6 to 1.3)	9.3	(8.2 to 10.4)
2017	4.3	(3.6 to 5.1)	4.5	(3.7 to 5.3)	0.6	(0.4 to 1)	10.9	(9.7 to 12.1)
2018	4.3	(3.5 to 5.1)	5.2	(4.4 to 6.1)	0.7	(0.4 to 1)	9.7	(8.6 to 10.9)
2019	4.8	(4.1 to 5.7)	5.6	(4.8 to 6.6)	1.0	(0.6 to 1.4)	10.3	(9.2 to 11.6)
2020	5.4	(4.6 to 6.4)	5.0	(4.2 to 5.9)	1.1	(0.7 to 1.5)	8.4	(7.3 to 9.5)
2021	4.8	(4 to 5.7)	7.5	(6.5 to 8.6)	1.3	(0.9 to 1.8)	8.5	(7.5 to 9.7)
2022	6.1	(5.2 to 7.1)	9.7	(8.5 to 10.9)	1.1	(0.8 to 1.6)	5.5	(4.7 to 6.5)

Abbreviation: STD: standardized; CI: confidence interval

Supplemental Table 5-1 Comorbidities associated with liver disease by stages (2004 to 2022)

Comorbidities	Stage 1, N = 94,529[1]	Stage 2, N = 4,562[1]	Stage 3, N = 3,040[1]	Stage 4, N = 3,221[1]	Stage 5, N = 2,233[1]
CVD related conditions	7,018(7.4%)	587(12.9%)	310(10.2%)	457(14.2%)	264(11.8%)
Diabetes	6,135(6.5%)	612(13.4%)	303(10.0%)	234(7.3%)	200(9.0%)
Hypertension/anti-hypertensive	34,022(36.0%)	1,939(42.5%)	1,165(38.3%)	1,232(38.2%)	1,006(45.1%)

[1] n(%)

Supplemental Table 5-2 Comorbidities associated with liver disease by aetiologies (2004-2022)

Comorbidities	ArLD, N = 19,760[1]	NAFLD, N = 33,655[1]	Metablic liver disease, N = 5,469[1]	HBV, N = 1,063[1]	HCV, N = 3,539[1]	Autoimmune liver disease, N = 13,582[1]	Haemochromatosis, N = 4,111[1]	Hepatitis not specified, N = 4,283[1]	Congestive hepatopathy, N = 574[1]	Toxic liver disease, N = 757[1]	Miscellaneous, N = 8,426[1]
CVD related conditions	1,442(7.3%)	1,837(5.5%)	578(10.9%)	54(5.2%)	177(5.5%)	1,159(8.6%)	227(5.6%)	668(15.7%)	135(23.6%)	58(7.7%)	885(10.5%)
Diabetes	1,185(6.0%)	2,776(8.4%)	380(7.2%)	46(4.4%)	132(4.1%)	752(5.6%)	175(4.3%)	360(8.4%)	42(7.3%)	26(3.4%)	253(3.0%)
Hypertension/antihypertensives	6,911(35.0%)	12,455(37.5%)	1,916(36.1%)	197(19.1%)	632(19.5%)	4,670(34.8%)	1,422(35.2%)	1,678(39.2%)	273(47.6%)	203(26.9%)	3,468(41.2%)

[1] n(%)

BMJ Open

Incidence rate and associated patient characteristics of liver disease in Wales 2004-2022: a retrospective population-scale observational study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2024-093335.R2
Article Type:	Original research
Date Submitted by the Author:	21-Jan-2025
Complete List of Authors:	Gao, Jingwei; Swansea University - Singleton Park Campus, Akbari, Ashley; Swansea University Medical School, Ahmed, Haroon; Cardiff University, Division of Population Medicine Davies, Aled; Cardiff University, PRIME Centre Wales Yeoman, Andrew; Royal Gwent Hospital Pembroke , Thomas ; University Hospital of Wales, Department of Gastroenterology and Hepatology
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Gastroenterology and hepatology
Keywords:	EPIDEMIOLOGY, Hepatology < INTERNAL MEDICINE, PUBLIC HEALTH

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Incidence rate and associated patient characteristics of liver disease in Wales 2004-2022: a retrospective population-scale observational study

Authors: Jingwei Gao [1], Ashley Akbari [1], Haroon Ahmed [2], Aled Davies [3], Andrew Yeoman [4], Thomas Peter Ignatius Pembroke [5]

Affiliations:

1 Population Data Science, Swansea University Medical School, Faculty of Medicine, Health & Life Science, Swansea University, Swansea, UK.

2 Division of Population Medicine, Cardiff University, Neuadd Meirionnydd, Heath Park, Cardiff, UK.

3 PRIME Centre Wales, Cardiff University, Neuadd Meirionnydd, Heath Park, Cardiff, UK.

4 Aneurin Bevan University Health Board, Newport, UK.

5 Department of Gastroenterology and Hepatology, University Hospital of Wales, Cardiff, UK.

Corresponding author:

Jingwei Gao

jingwei.gao@swansea.ac.uk

Population Data Science, Swansea University Medical School, Faculty of Medicine, Health & Life Science, Swansea University, Swansea, UK

ORCiD:

Jingwei Gao:	0000-0002-7722-6177
Ashley Akbari:	0000-0003-0814-0801
Haroon Ahmed:	0000-0002-0634-8548
Aled Davies:	0000-0002-7815-5155
Andrew Yeoman:	0000-0002-0739-3332
Thomas Peter Ignatius Pembroke:	0000-0002-2600-2034

Abstract

Objective

To describe the incidence and key demographic, socioeconomic and clinical characteristics of individuals with liver disease in Wales.

Design and setting

This study is designed as a retrospective observational study that linked data of anonymised identified individuals from primary, secondary care and mortality data from the Secure Anonymised Information Linkage (SAIL) Databank in Wales.

Participants

All Welsh residents who registered with a SAIL contributing general practitioner (GP) and diagnosed with liver disease from 2004 to 2022.

Primary and secondary outcome measures

Our primary outcome is the annual age-standardised incidence rate of liver disease. Secondary outcome is the numbers and frequencies of underlying aetiology and the associated comorbidities.

Results

Between 2004 and 2022, 111,098 individuals received a diagnosis of liver disease in Wales and were included in this study. The incidence of liver disease increased three-fold during the study period (97.7 per 100,000 inhabitants in 2004 to 316.2 per 100,000 inhabitants in 2022). A total of 79,992 individuals (72%) entered the cohort with the underlying aetiology of liver disease, including alcohol related liver disease (ArLD), non-alcoholic fatty liver disease (NAFLD), viral hepatitis, metabolic, hemochromatosis and autoimmune liver diseases. NAFLD has contributed to most of the change in incidence.

Conclusions

We observed increasing incidence rates of liver disease in Wales, with NAFLD showing a particularly sharp increase and frequently identified as an underlying condition. A better understanding of the

incidence of liver disease is the first step towards effective prevention, early detection and targeted intervention to improve patient outcomes.

Keywords Epidemiology; Hepatology; Non-alcoholic fatty liver disease hepatitis; Alcohol-related liver disease; Cirrhosis

Word count: 3960

Strengths and limitations of this study

- The national-scale population-based cohort and long follow-up maximized the generalizability of our finding.
- By incorporating both in-patient and outpatient data, the study not only expanded the sample size but also enabled a thorough analysis of liver disease incidence, enhancing understanding of its clinical implications.
- With the usage of SAIL Databank as the research-ready data asset and the incorporation of reproducible research pipelines, this study strengthened its reliability and reproducibility.
- Sensitive Read codes for conditions like HBV and HCV could not be included due to regulatory laws, impacting the completeness and precision of the data analysis.
- The study only included individuals with a history of GP registration and residency information in Wales, likely overlooking a subpopulation at significant risk for liver disease

Background

Liver disease is a significant global public health issue and a major contributor to morbidity and mortality [1]. Globally, cirrhosis and hepatocellular carcinoma account for an estimated two million deaths every year [2]. In the UK, liver disease has become the third most common cause of premature death [3], despite mortality rates for other major non-communicable diseases declining [4].

The management of chronic liver disease frequently involves lifestyle modification, including weight loss and reduced alcohol use, with the goal of reversing factors that can lead to disease progression [5]. As a result, public health policies in the UK have focussed on prevention and early detection, including the UK Government's Prevention Green Paper to promote the disease prevention [6], the National Institute for Health and Care Excellence guidance that focused on tackling obesity [7], and a series of policies to reduce alcohol-related harm [8]. In November 2022, the Welsh Government published a Quality Statement for Liver Disease [9], underlying the importance of the awareness of risk factors and early detection of liver disease, and set out its plans to promote the delivery of better quality, higher value and more accessible services for individuals with liver disease.

To effectively improve liver disease management, clinicians, researchers and policymakers must be aware of the epidemiology and clinical profile of the individuals with liver disease. However, a significant gap remains in the integration of primary, secondary, and mortality data, particularly across the different liver disease stages and aetiologies. To date, much of the epidemiological data is based on ICD-10 coding, which is largely derived from secondary care data sources and therefore likely to underestimate the real-world incidence and prevalence of liver disease [10, 11]. This limitation currently precludes adequate prioritisation of research, targeting of interventions, and recruitment of individuals to clinical trials. Understanding the clinical and socio-demographic features associated with liver disease in a large-scale population is essential for the improvement of disease prevention and treatment. Therefore, we aim to fill this gap by integrating primary, secondary, and mortality data, providing a more comprehensive and accurate depiction of liver disease across different stages and aetiologies.

The objective of our study is to describe the incidence of liver disease, as well as the key demographic and socioeconomic characteristics and the associated comorbidity of liver disease patients in Wales, as a first step towards improving capacity and capability for liver disease research.

Materials and method

Setting and data source

This study is designed as retrospective cohort study. We used data from the Secure Anonymised Information Linkage (SAIL) Databank, which contains anonymised, individual-level linked electronic

health record (EHR) data for Welsh population [12, 13, 14, 15]. The SAIL Databank includes complete secondary care data and primary care data covering approximately 86% of the Welsh population. These data reflect the demographic diversity of Wales across age, sex, and levels of deprivation [16] and can be generalisable to the broader UK population due to demographic similarities [17]. SAIL employs a split-file anonymisation process using National Health Service (NHS) number, name, sex, date of birth and postcode, ensuring confidentiality while enabling the linkage of individual-level data sources [13, 16, 18].

To provide a comprehensive overview of liver disease epidemiology, we combined linked primary care, hospital admissions and mortality data. Primary care data were accessed from the Welsh Longitudinal General Practice (WLGP) data, which currently uses the Read version 2 clinical coding system and collects event histories for people registered with a SAIL-supplying general practice in Wales. Hospital admission data, including in-patient admissions (emergency, elective and maternity) and day-care procedures, were collected from the Patient Episode Database for Wales (PEDW). Mortality data came from the Annual District Death Extract (ADDE) by the Office for National Statistics (ONS) death and contains the cause of death and contributory comorbidities. Both hospital admission and mortality data were coded using the International Classification of Diseases version 10 (ICD-10) system. We derived demographic and deprivation data from the Welsh Demographic Service Dataset (WDSD) and used the Welsh Index of Multiple Deprivation (WIMD) version 2019 quintiles to measure relative area-level deprivation based on geographical residential location from the Lower-layer Super Output Area (LSOA) version 2011.

Study population

We linked data from all individuals in the WLGP, PEDW, and ADDE data within the SAIL Databank using a unique anonymised individual identifier known as Anonymised Linkage Field (ALF). Individuals were extracted based on ALF and filtered for good data linkage status based on existing methodology [13, 18]. We excluded individuals not registered with a SAIL-contributing general practice (GP) or lacking residency information and identified those who received a liver disease diagnosis from 1st January 2004 to 31st December 2022. The first liver disease diagnosis across all data sources is considered to be the index liver disease event. Individuals were required to be residents in Wales at the time of cohort entry. GP registration was required if the index event was from WLGP. All individuals were followed until the earliest of: GP de-registration (applicable only to individuals whose index event was identified from WLGP), moving out of Wales, death, or the end of study period (2022.12.31). GP de-registration is defined as the termination of an individual's registration with a SAIL-contributing GP, as indicated by the last end date in the registration records. We divided the full cohort into four distinct groups: individuals identified in WLGP were assigned to the primary care group;

those identified in PEDW were assigned to the secondary care group; individuals identified in ADDE were assigned to the mortality group; and individuals identified across multiple data sources (two or more) were assigned to a multi-source group.

Measurements

Definition of stages of liver disease and time of cohort entry

We applied a hierarchy of 3 tiers of potential aetiological diagnoses and 5 discrete stages of chronic liver disease based on perceived clinical importance and natural history of liver disease progression as was described in our previous study [19]. The first stage is the underlying aetiological conditions without the presentation of cirrhosis and was divided into 3 tiers. Tier 1 aetiologies of liver disease include: alcohol-related liver disease (ArLD), non-alcoholic fatty liver disease (NAFLD), metabolic liver disease, hepatitis B virus (HBV), hepatitis C virus (HCV), autoimmune liver disease, haemochromatosis; tier 2 aetiologies include: unspecified hepatitis, congestive hepatopathy, and toxic liver disease, and tier 3 aetiologies were other miscellaneous diagnoses. As the disease progresses, various stages of liver disease were defined as follows: hepatic fibrosis and cirrhosis were categorised as stage 2, portal hypertension as stage 3, hepatic decompensation as stage 4, and hepatocellular carcinoma (HCC), intrahepatic cholangio carcinoma (ICC), and other primary liver cancers were classified under stage 5. Acute liver diseases were analysed separately from chronic liver diseases as they carry different challenges for primary and secondary care. Acute liver diseases were defined as conditions including acute viral hepatitis, Budd-Chiari syndrome, acute liver failure, infections and infarctions, and other unspecified acute liver injuries. A list of ICD-10 and Read v2 codes to identify individuals with liver disease can be found in Supplemental Table 1.

We defined the time of the first chronic or acute liver disease diagnosis across all data sources as the index date. To assess the proportion of individuals with liver disease who presented late, the diagnoses were sequenced according to the natural history of the disease: aetiological diagnoses (stage 1), followed by cirrhosis, portal hypertension, decompensation and then HCC formation [19] (Supplemental Fig 1).

Comorbidities and drug prescription history

We collected data on demographic characteristics (age, sex, and WIMD 2019 quintile), with age divided by birth year in 10-year intervals, except for the 0-17 age group and 18-29 age group. Comorbidities including cardiovascular disease (CVD) related conditions (heart failure, transient ischaemic attack, other ischaemic disease, atrial fibrillation, peripheral vascular disease, angina, and stroke), diabetes,

hypertension, and antihypertensive usage were tracked up to ten years prior to cohort entry. Lists of ICD-10 and Read v2 codes to identify the associated conditions are available as supplemental tables (Supplemental Table 2-1, Supplemental Table 2-2).

Incidence of liver disease

The incidence rate was calculated as the number of incident cases per Welsh residents. An incident case was defined as an individual having a first liver-related diagnosis during the study period (01.01.2004-31.12.2022) with no prior liver disease history from 1st January 1994 to 31st December 2003. To determine the most clinically significant aetiology, we only considered the most advanced stage of diagnosis on incident event date. The Welsh population data was obtained from the ONS population estimates [20]. The incidence rate was directly standardised using the European Union (EU) standard population 2013 and were reported per 100,000 inhabitants in each calendar year from 2004 to 2022, with corresponding 95% confidence intervals (CI). Analysis was repeated for different data sources, liver disease aetiologies, and stages.

Study findings were reported in accordance with applicable reporting guidelines for observational studies using administrative data (Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) [21] and REporting of studies Conducted using Observational Routinely-collected Data (RECORD) [22]. Data cleaning, cohort assembly and statistical analyses were performed using Structured Query Language (IBM Db2 V.11.1) and R (V.4.1.0–V.4.1.3) within the SAIL Databank privacy-protecting trusted research environment (TRE).

Patient and public involvement

The Liver Research Cymru Patient Advisory group was established to provide PPI input into the development of research in Wales. The research questions and methodology addressed in this body of work were developed following discussion of their priorities based upon their lived experience of liver disease. The results were shared with this group and their feedback was sought in the interpretation and presentation of these data.

Results

Demographic and socioeconomic characteristics

We identified a total of 111,098 individuals with liver disease from all three data sources from 1st January 2004 to 31st December 2022, contributing 441,885 person-years of follow-up (Fig 1A). Of the

eligible individuals, 57,491 (51.7 %) were male. The median (interquartile range [IQR]) age was 59 (47-72). Approximately 50% of the individuals (55,787) entered the cohort after 2016 and 27,178 (24.5 %) individuals came from the most deprived areas. Of the 111,098 individuals, 20,319 (18.3%) individuals could be tracked from WLGP (primary care data source), forming the primary care group, 56,710 (51.0 %) from PEDW (secondary data source) constituted the secondary care group, and 3,459 (3.1 %) from ADDE (mortality data source) formed mortality group. Additionally, 30,610 (27.5%) appeared in multiple sources, forming the multi-source group (Fig 1B). The demographic variation is notable with a higher percentage of male and elderly individuals, and a higher degree of socioeconomic deprivation in the mortality group (Table 1).

Incidence of chronic liver disease

We observed an increasing trend of liver disease from 2004 to 2022 in the Welsh population. The age-sex standardised incidence rate increased by 3 times during the 18 years of follow-up (97.7 per 100,000 inhabitants in 2004; 316.2 per 100,000 inhabitants in 2022, Table 2). This was significantly contributed by a 6.5 times increase in the incidence rate of the primary care group, followed by 4.4 times increase in the secondary care group. The proportion of incident cases in primary care and secondary care group increased by 14-16% during the past 2 decades (Fig 2A – 2C, Supplemental Table 3-1).

The incidence of liver disease has risen across all stages, with stage 2 (cirrhosis) and stage 3 diagnoses (portal hypertension) experiencing the most notable increase from 2004 to 2022 (stage 2 by 4 times; stage 3 by 3.9 times). However, the increases in the higher-stage liver disease cases were less pronounced, particularly for stage 4 (1.5 times). Consequently, the proportion of more advanced liver disease diagnoses decreased (stage 4 and stage 5: 8.6% in 2004, 5.3% in 2022, decreased by 1.62 times since 2004, Fig 3A – 3C, Supplemental Table 3-2).

We further analysed the stage 1 aetiologies by each diagnosis. We observed the most notable increase in the incidence of NAFLD (10.8 per 100,000 inhabitants in 2004, 124.5 per 100,000 inhabitants in 2022, increased 11 times since 2004), resulting a 3.4 time increase in the proportion and ultimately accounting for over half of the cases by the end of our follow-up. Conversely, the proportion of ArLD cases decreased by 71%, given that the incidence rate remained unchanged during the past 2 decades (Fig 4A – 4B, Supplemental Table 3-3). A breakdown of NAFLD incidence by data sources demonstrated the 14 and 18 times increases in the primary care and the secondary care group, respectively (Supplemental Table 4-1). In contrast, the increase in the incidence of ArLD in the primary care group was more than twice that of the secondary care group (Supplemental Table 4-2).

We did not observe an increasing trend for HBV and HCV. Both conditions fluctuated during the past 2 decades, with HBV cases increasing by 4.3 times and HCV cases 2.3 times from 2004 to 2017, followed by a decrease from 2018 to 2022 (HBV by 1.7 times; HCV by 2.6 times). Consequently, the

proportion of HBV and HCV cases decreased from 2004 to 2022 (HBV by 2.3 times; HCV by 8.6 times, Fig 4A – 4C, Supplemental Table 3-3).

Underlying liver disease conditions

Of the 111,098 individuals with liver disease, 79,992 (72.0 %) entered the cohort due to a tier 1 aetiology. This was mainly contributed by the high prevalence of NAFLD (33,655 [30.3%]). Most NAFLD cases were identified in the primary care group (13,565 [66.8 %]), and the least NAFLD cases in the mortality group (727 [21.0 %]). This aligns with the distribution of individuals with tier 1 aetiology, which's proportion was the highest in the primary care group (18,453 [90.8 %]), and the lowest in the mortality group (1,728 [50.0 %]). In contrast, tier 2 and tier 3 diagnoses were higher in the secondary care group. The proportion of advanced liver disease stages was lower in the primary care group (stage 3: 166 [0.8 %], stage 4: 53 [0.3 %], stage 5: 121 [0.6 %]), but higher in the mortality group (stage 4: 276 [8.0 %], stage 5: 473 [13.7 %]) (Table 3).

Comorbidities associated with liver disease

Amongst all eligible individuals with liver disease, 40,427 (36.4%) individuals had onset hypertension or initiated an antihypertensive medication, followed by 8,923 (8.0%) individuals who had cardiovascular disease (CVD) related conditions, and 7,658(6.9%) who had diabetes. We observed the higher proportions of comorbidities in the mortality group (CVD: 11.2%; diabetes: 8.1%; hypertension/antihypertensive: 39.7%), and the lowest in the primary care group (CVD: 4.2%; diabetes: 6.3%; hypertension/antihypertensive: 35.7%) (Table 4). The burden of comorbidities was the lowest for individuals with stage 1 liver disease conditions, but higher for late-stage (CVD highest for stage 4: 14.2%; hypertension/antihypertensive for stage 5: 45.1%) and stage 2 individuals (diabetes: 13.4%) (Supplemental Table 5-1). Amongst the individuals who entered the cohort with a stage 1 aetiology, individuals with congestive hepatopathy had the highest rates of hypertension/antihypertensive (47.6%) and CVD conditions (23.6%), while those with NAFLD had the highest rates of diabetes (8.4%) (Supplemental Table 5-2).

Discussion

In this large-scale population-based study, we observed that the incidence of liver disease had increased dramatically in Wales during the past two decades. Notably, NAFLD played a significant role in this rise, with its incidence and proportion increasing sharply.

The rapid increase in the incidence rate of liver disease observed in our study is similar to our previous findings in Wales [19], where a 3.6-fold increase was observed in the in-patient chronic liver disease cases between 2001 and 2019, aligning with the trends presented in our secondary care data. This was mainly driven by the 11-fold increase in NAFLD incidence. Other studies, such as P Nasr's et al work in Sweden [23], reported a 2-fold increase in the incidence rate of NAFLD during 2005-2019. Similarly, H Tian et al [24] used data from the Global Burden of Disease study 2019 and showed an increase of 95.4% of NAFLD globally from 1990 to 2019. A systematic review that included 578 studies demonstrated a 13% higher prevalence of NAFLD during the year 2011-2021 compared with year 2000-2010 [25]. A community study in the USA reported a 5-fold increase in the NAFLD incidence from 1997 to 2014 [26]. However, none of these findings compare to our observation in Wales. This substantial difference may be partly attributed to our inclusion of primary care data, which experienced more rapid increase than inpatient data and thus boosted the NAFLD incidence across the study population. The high prevalence of obesity [27] and rapidly increasing incidence of diabetes [28] in Wales in the recent decades may also have attributed to the sharp increase of NAFLD, as these metabolic factors including obesity [29, 30] and insulin resistance were considered to be strongly linked to NAFLD due to metabolic dysfunctions [31, 32]. This observation is further supported by our observation of autoimmune liver disease, which is not linked to metabolic factors and remained unchanged during our study period, highlighting the lifestyle-related nature of the increase in NAFLD. Another interesting observation related to NAFLD is the high proportion of those with comorbidities. NAFLD evolves in those with metabolic syndrome, and these comorbidities are used to define the more recently adopted term metabolic associated fatty liver disease [33]. In keeping with this definition of NAFLD, we observed a high proportion of individuals having onset comorbidities such as hypertension, diabetes, CVD related conditions up to 10 years prior to their first liver disease diagnosis. The multimorbidity amongst people living with liver disease is anticipated considering the increasing incidence of NAFLD and the high prevalence of obesity in Wales, and this will require engagement with primary care practitioners to address their complex multidisciplinary healthcare needs and to identify those at risk of significant liver disease early in the disease trajectory [34].

Contrary to previous European studies showing a decline in ArLD incidence [19, 35], our research observed a resurgence after 2015 following a slight decrease from 2004 to 2014. We believe that changes in drinking habits [36, 37] and government actions on alcohol pricing and taxation [38] likely influenced the declining trend before 2015, while the resurgence after 2015 was contributed by the inclusion of outpatient data from the primary care group. The addition of primary care data could explain the discrepancy in our results compared to the prior studies that focused solely on in-patients, filling the gap by providing a broader perspective by including outpatient data [19, 35]. Similar to ArLD, the incidence of HBV increased until 2017, then sharply dropped after the year of 2018. The observed decrease in the incidence of HBV and HCV post-2020 could be attributed to the initiatives outlined in

the Welsh Health Circular, which seeks to meet the goal of eradicating HBV and HCV as a major public health concern [39]. This trend may represent the effort in better case finding and eradication, which led to reduced transmission for HCV, and enhanced vaccination and viral suppression for HBV [40]. Additionally, disruptions caused by the COVID-19 pandemic may also have played a role in this decline [41].

It is worth to mention our observation on the increasing number of incident cases identified from primary care data, particularly after 2015. This observation, together with the rapid increase in NAFLD and ArLD incidence, indicated an increasing detection rate of liver disease in primary care at an early stage. This change might be attributed to the implementation of the Wales Liver Disease Delivery Plan, which emphasised the importance of early detection and ensure that 'excellent care' is accessible when necessary [42]. Another notable trend observed during our study period is the decreasing proportion of severe late-stage presentations, such as hepatic decompensation. Additionally, we observed a declining trend in the proportion of individuals who died from liver disease without receiving a prior diagnosis. This is to say, despite the increasing incidence of liver disease, there is a promising indication that more individuals were identified as the result of early detection. This provides the healthcare system with more opportunities for early-stage intervention. Early detection of liver disease has been linked to improved long-term outcomes, and a series of reviews and guidelines [43, 44] emphasize the importance of thorough evaluation and timely referral for patients with abnormal liver function tests (LFTs). Despite these recommendations, adherence remains inconsistent, and the factors influencing clinicians' referral practices—including potential barriers and motivating factors—are not well understood. This issue is especially pressing in the context of rising obesity rates and the emergence of new weight-loss therapies, both of which carry significant implications for liver health. Consequently, further research is essential to assess how integrated care models and early detection strategies might reduce liver disease progression and mortality.

Our study has several strengths. First, it utilised the national-scale setting and created a population-based cohort with long-term follow-up, which allowed us to maximise the generalisability of our findings to the wider population. Furthermore, by combining primary care, secondary care, and mortality data, our study uniquely expands on previous research and addresses a significant gap in understanding liver disease [10, 11]. This novel approach increased our sample size by at least 20%, allowing us to provide a more comprehensive view of liver disease incidence and characteristics. Specifically, this integrated data approach enabled us to: 1, capture a broader range of liver disease cases, reducing the likelihood of underestimating case numbers and ensuring a more accurate representation of disease incidence; and 2, include patients from diverse healthcare settings, each with unique clinical characteristics, which allowed for a detailed comparison across these settings and a deeper understanding of liver disease in varied clinical contexts. Additionally, a significant strength of the study is the consistency of coding practices in Wales, which have not changed in response to funding

incentives [45]. This stability helps mitigate potential biases and misdiagnoses, providing a more reliable and objective reflection of disease incidence and management trends. Another strength lies in the reproducibility and the use of the SAIL Databank as a research-ready data asset (RRDA). By incorporating reproducible research pipelines, we standardized and documented the cohort curation process. This enhances the study's reliability and reproducibility, enabling other researchers to replicate our findings and explore further research questions within the same framework.

Our study is also vulnerable to several limitations. Regulatory laws associated with the SAIL Databank prohibited the inclusion of conditions classified as sensitive in our analysis. For instance, the sensitive Read codes for HBV and HCV could not be extracted along with the rest of the cohort, resulting in potentially incomplete data for these conditions, affecting the completeness and precision of our data analysis. Our study exclusively included individuals who had a history of GP registration and residency information in Wales. Consequently, we likely overlooked individuals were unable to register with a GP, a subpopulation that represents a group at significant risk for liver disease. The exclusion may also introduced a potential underestimation of the true incidence rate of liver disease in the primary care data. Moreover, there may be changes in diagnostic practices and healthcare-seeking behaviour over the study period that could influence the observed trends and introduce variations, potentially affecting the accuracy of our findings regarding liver disease incidence and management. Furthermore, due to the observational nature and design of our study, we were unable to establish causal relationships between the onset of liver disease and other comorbidities.

Conclusion

Our study observed a significant rise in the incidence of liver disease in Wales over the past two decades, primarily driven by the increase in NAFLD. The high prevalence of comorbid conditions among liver disease patients and the increased role of primary care in disease identification highlight the need for integrated healthcare approaches to address this growing public health concern. A better understanding of the incidence of liver disease, driven by a more comprehensive analysis of the integrated primary and secondary care data, is essential as a foundation for effective prevention, early detection, and targeted interventions to improve patient.

Acknowledgements

The Liver Disease Cymru Partnership (LDCP) received a grant from the National Institute for Health Research (NIHR154876). This project was partly funded by an unrestricted grant from the Liver Disease Implementation Group, Welsh Government (LDIG-22-19).

We wish to acknowledge all members of the LDCP who have contributed to the collaborations: Prof Kerry Hood, Prof Deborah Fitzsimmons and Prof Katherine Cullen for advice developing the group and the Liver Research Cymru patient advisory group for their perspective on the research to be undertaken.

This study uses anonymised data held in the Secure Anonymised Information Linkage (SAIL) Databank, which is part of the national population data research infrastructure for Wales. We would like to acknowledge all the data providers who make anonymised data available for research.

Author Contributions

This project was a collaborative effort among multiple team members, each contributing significantly to various aspects of the research. TP, AA, AD, AY and HA acquired the funding. TP, AA, AD, AY, HA and JG contributed to the design of the methodology. TP, AA, AD, AY and JG supervised the study and validated the research. JG and AA were responsible for project administration and data management. JG curated the data, performed data analysis, created visual representations of the data and wrote the original draft of the paper. TP acted as guarantor. All authors reviewed and edited the manuscript.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Funding Sources

The Liver Disease Cymru Partnership (LDCP) receives a grant from the National Institute for Health Research (NIHR154876). This project was partly funded by an unrestricted grant from the Liver Disease Implementation Group, Welsh Government (LDIG-22-19).

Data availability statement

The routine data used in this study are available in the SAIL Databank at Swansea University, Swansea, UK. All proposals to use SAIL data are subject to review by an Information Governance Review Panel (IGRP). Before any data can be accessed, approval must be given by the IGRP. The IGRP gives careful consideration to each project to ensure proper and appropriate use of SAIL data. When access has been approved, it is gained through a privacy-protecting trusted research environment (TRE). SAIL has established an application process to be followed by anyone who would like to access data via SAIL https://www.saildatabank.com/application-process . This study has been approved by the IGRP as project 1492. The research adhered to ethical guidelines and Data Protection Act 2018 to ensure the privacy and confidentiality of all data subjects involved.

The reproducible SQL and R code, and code lists to identify study individuals are available on Github: https://github.com/SwanseaUniversityDataScience/1492-LDCP

Ethics statements

Patient consent for publication

Not applicable.

Ethics approval

The study uses anonymised data only, and ethical approval was not required for this secondary use of data. See the following weblink: https://saildatabank.com/data/

References

[1] Paik JM, Golabi P, Younossi Y, Mishra A, Younossi ZM. Changes in the Global Burden of Chronic Liver Diseases From 2012 to 2017: The Growing Impact of NAFLD. Hepatology. 2020 Oct 27;72(5):1605–16. https://doi.org/10.1002/hep.31173

[2] Wang H, Naghavi M, Allen C, Barber RM, Bhutta ZA, Carter A, et al. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet. 2016 Oct;388(10053):1459–544. https://doi.org/10.1016/S0140-6736(16)31012-1.

[3] Deaths registered in England and Wales (series DR): 2017. Office for National Statistics. 2017. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deathsregisteredinenglandandwalesseriesdr/2017. [Accessed 5 Jun 2024].

[4] Williams R, Aspinall R, Bellis M, Camps-Walsh G, Cramp M, Dhawan A, et al. Addressing liver disease in the UK: a blueprint for attaining excellence in health care and reducing premature mortality from lifestyle issues of excess consumption of alcohol, obesity, and viral hepatitis. Lancet. 2014 Nov 29;384(9958):1953–97. https://doi.org/10.1016/S0140-6736(14)61838-9.

[5] Nobili V, Carter-Kent C, Feldstein AE. The role of lifestyle changes in the management of chronic liver disease. BMC Medicine. 2011 Jun 6;9:70. https://doi.org/10.1186/1741-7015-9-70.

[6] Department of Health and Social Care. Advancing our health: prevention in the 2020s – consultation document. GOV.UK. 2019. https://www.gov.uk/government/consultations/advancing-our-health-prevention-in-the-2020s/advancing-our-health-prevention-in-the-2020s-consultation-document. [Accessed 5 Jun 2024].

[7] Diet, nutrition and obesity. NICE. https://www.nice.org.uk/guidance/lifestyle-and-wellbeing/diet--nutrition-and-obesity. [Accessed 5 Jun 2024].

[8] Burton R, Henn C, Lavoie D, O'Connor R, Perkins C, Sweeney K, et al. A rapid evidence review of the effectiveness and cost-effectiveness of alcohol control policies: an English perspective. The Lancet. 2017 Apr;389(10078):1558–80. https://doi.org/10.1016/S0140-6736(16)32420-5.

[9] The quality statement for liver disease: the quality statement describes what good quality liver disease services should look like. Welsh Government. 2022. https://www.gov.wales/sites/default/files/pdf-versions/2023/3/1/1679326108/quality-statement-liver-disease.pdf. [Accessed 5 Jun 2024].

[10] Nam YH, Mendelsohn AB, Panozzo CA, Maro JC, Brown JS. Health outcomes coding trends in the US Food and Drug Administration's Sentinel System during transition to International Classification of Diseases-10 coding system: A brief review. Pharmacoepidemiology and Drug Safety. 2021 Mar 17;30(7):838–42. https://doi.org/10.1002/pds.5216.

[11] Ratib S, West J, Fleming KM. Liver cirrhosis in England—an observational study: are we measuring its burden occurrence correctly? BMJ Open. 2017 Jul;7(7):e013752. https://doi.org/10.1136/bmjopen-2016-013752.

[12] Ford DV, Jones KH, Verplancke JP, Lyons RA, John G, Brown G, et al. The SAIL Databank: building a national architecture for e-health research and evaluation. BMC Health Services Research. 2009 Sep 4;9:157. https://doi.org/10.1186/1472-6963-9-157.

[13] Lyons RA, Jones KH, John G, Brooks CJ, Verplancke JP, Ford DV, et al. The SAIL databank: linking multiple health and social care datasets. BMC Medical Informatics and Decision Making. 2009 Jan 16;9:3. https://doi.org/10.1186/1472-6947-9-3.

[14] Jones KH, Ford DV, Ellwood-Thompson S, Lyons RA. The UK Secure eResearch Platform for public health research: a case study. The Lancet. 2016 Nov;388:S62. https://doi.org/10.1016/S0140-6736(16)32298-X.

[15] Wilkinson T, Schnier C, Bush K, Rannikmäe K, Lyons RA, McTaggart S, et al. Drug prescriptions and dementia incidence: a medication-wide association study of 17000 dementia cases among half a million participants. Journal of Epidemiology and Community Health. 2021 Oct 27;76(3):223–9. https://doi.org/10.1136/jech-2021-217090.

[16] Schnier C, Wilkinson T, Akbari A, Orton C, Sleegers K, Gallacher J, et al. The Secure Anonymised Information Linkage databank Dementia e-cohort (SAIL-DeC). International Journal of Population Data Science. 2020 Feb 25;5(1):1121. https://doi.org/10.23889/ijpds.v5i1.1121

[17] Population estimates for the UK, England, Wales, Scotland, and Northern Ireland: mid-2022. Office for National Statistics. 2024. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2022. [Accessed 5 Jun 2024].

[18] Ford DV, Jones KH, Verplancke JP, Lyons RA, John G, Brown G, et al. The SAIL Databank: building a national architecture for e-health research and evaluation. BMC Health Services Research. 2009 Sep 4;9:157. https://doi.org/10.1186/1472-6963-9-157.

[19] Peter T, John GR, Puyk B, Howkins K, Clarke R, Yousuf F, et al. Rising incidence, progression and changing patterns of liver disease in Wales 1999-2019. World Journal of Hepatology. 2023 Jan 27; 15(1):89–106. https://doi.org/10.4254/wjh.v15.i1.89.

[20] Estimate of the population for England and Wales. Office for National Statistics. 2023. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/estimatesofthepopulationforenglandandwales. [Accessed 5 Jun 2024].

[21] von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. PLoS Medicine. 2007 Oct 16;4(10):e296. https://doi.org/10.1016/j.jclinepi.2007.11.008.

[22] Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. PLOS Medicine. 2015 Oct 6;12(10):e1001885. https://doi.org/10.1371/journal.pmed.1001885

[23] Patrik Nasr, Erik von Seth, Mayerhofer R, Ndegwa N, Ludvigsson JF, Hannes Hagström. Incidence, prevalence and mortality of chronic liver diseases in Sweden between 2005 and 2019. European Journal of Epidemiology. 2023 Jul 25;38(9):973–84. https://doi.org/10.1007/s10654-023-01028-x.

[24] Tian H, Zhang K, Hui Z, Ren F, Ma Y, Han F, et al. Global burden of non-alcoholic fatty liver disease in 204 countries and territories from 1990 to 2019. Clinics and Research in Hepatology and Gastroenterology. 2023 Jan;47(1):102068. https://doi.org/10.1016/j.clinre.2022.102068.

[25] Liu J, Tian Y, Fu X, Mu C, Yao M, Ni Y, et al. Estimating global prevalence, incidence, and outcomes of non-alcoholic fatty liver disease from 2000 to 2021: systematic review and meta-analysis. Chinese Medical Journal. 2022 Jul 20;135(14):1682–91. https://doi.org/10.1097/CM9.0000000000002277.

[26] Allen AM, Therneau TM, Larson JJ, Coward A, Somers VK, Kamath PS. Nonalcoholic fatty liver disease incidence and impact on metabolic burden and death: A 20 year-community study. Hepatology. 2018 Mar 23;67(5):1726–36. https://doi.org/10.1002/hep.29546.

[27] Keaver L, Xu B, Jaccard A, Webber L. Morbid obesity in the UK: A modelling projection study to 2035. Scandinavian Journal of Public Health. 2018 Aug 30;48(4):422–7. https://doi.org/10.1177/1403494818794814

[28] Powell R. Diabetes prevalence – trends, risk factors, and 10-year projection [Internet]. Public Health Wales. 2023. https://phw.nhs.wales/services-and-teams/observatory/data-and-analysis/diabetes-prevalence-trends-risk-factors-and-10-year-projection/. [Accessed 5 Jun 2024].

[29] Younossi Z, Anstee QM, Marietti M, Hardy T, Henry L, Eslam M, et al. Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. Nature Reviews Gastroenterology & Hepatology. 2017 Sep 20;15(1):11–20. https://doi.org/10.1038/nrgastro.2017.109.

[30] Loomis AK, Kabadi S, Preiss D, Hyde C, Bonato V, St. Louis M, et al. Body Mass Index and Risk of Nonalcoholic Fatty Liver Disease: Two Electronic Health Record Prospective Studies. The Journal of Clinical Endocrinology & Metabolism. 2016 Mar 1;101(3):945–52. https://doi.org/10.1210/jc.2015-3444.

[31] Caussy C, Aubin A, Loomba R. The Relationship Between Type 2 Diabetes, NAFLD, and Cardiovascular Risk. Current Diabetes Reports. 2021 Mar 19;21(5):15. https://doi.org/10.1007/s11892-021-01383-7.

[32] Tanase DM, Gosav EM, Costea CF, Ciocoiu M, Lacatusu CM, Maranduca MA, et al. The Intricate Relationship between Type 2 Diabetes Mellitus (T2DM), Insulin Resistance (IR), and Nonalcoholic Fatty Liver Disease (NAFLD). Journal of Diabetes Research. 2020 Jul 31:2020:3920196. https://doi.org/10.1155/2020/3920196.

[33] Eslam M, Newsome PN, Sarin SK, Anstee QM, Targher G, Romero-Gomez M, et al. A new definition for metabolic dysfunction-associated fatty liver disease: An international expert consensus statement. Journal of Hepatology. 2020 Jul 1;73(1):202–9. https://doi.org/10.1016/j.jhep.2020.03.039.

[34] Pryke R, Guha IN. Time to focus on chronic liver diseases in the community: A review of primary care hepatology tools, pathways of care and reimbursement mechanisms. Journal of Hepatology. 2023 Mar 1 [cited 2024 May 13];78(3):663–71. https://doi.org/10.1016/j.jhep.2022.10.010.

[35] Deleuran T, Vilstrup H, Becker U, Jepsen P. Epidemiology of Alcoholic Liver Disease in Denmark 2006–2011: A Population-Based Study. Alcohol and Alcoholism. 2015 Feb 13;50(3):352–7. https://doi.org/10.1093/alcalc/agv003.

[36] Adult drinking habits in Great Britain: 2017. Office for National Statistics. 2018. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/drugusealcoholandsmoking/bulletins/opinionsandlifestylesurveyadultdrinkinghabitsingreatbritain/2017. [Accessed 5 Jun 2024].

[37] Adult lifestyles by year, 2016-17 to 2019-20. StatsWales. Available from: https://statswales.gov.wales/Catalogue/National-Survey-for-Wales/Population-Health/Adult-Lifestyles/adultlifestyles-by-year. [Accessed 5 Jun 2024].

[38] Home Office. Alcohol strategy. GOV.UK. 2012. https://www.gov.uk/government/publications/alcohol-strategy. [Accessed 5 Jun 2024].

[39] WELSH HEALTH CIRCULAR. Attaining the WHO targets for eliminating hepatitis (b and C) as a significant threat to public health. Welsh Government. 2017. https://www.gov.wales/sites/default/files/publications/2019-07/attaining-the-who-targets-for-eliminating-hepatitis-b-and-c-as-a-significant-threat-to-public-health.pdf. [Accessed 5 Jun 2024].

[40] Prevention, diagnosis and treatment of blood borne viruses in Wales: Hepatitis B, hepatitis C and HIV Annual report 2023 (Data to end 2022). Public Health Wales. 2023. https://phw.nhs.wales/publications/publications1/prevention-diagnosis-and-treatment-of-blood-borne-viruses-in-wales-hepatitis-b-hepatitis-c-and-hiv/. [Accessed 5 Jun 2024].

[41] Pley CM, McNaughton AL, Matthews PC, Lourenço J. The global impact of the COVID-19 pandemic on the prevention, diagnosis and treatment of hepatitis B virus (HBV) infection. BMJ Global Health. 2021 Jan;6(1):e004275. https://doi.org/10.1136/bmjgh-2020-004275.

[42] Together for Health -Liver Disease Delivery Plan: A Delivery Plan for NHS Wales and its Partners to 2020. Welsh Government. 2015. https://www.gov.wales/sites/default/files/publications/2018-12/liver-disease-delivery-plan-2015-to-2020.pdf. [Accessed 5 Jun 2024].

[43] Eslam M, Newsome PN, Sarin SK, Anstee QM, Targher G, Romero-Gomez M, et al. A new definition for metabolic dysfunction-associated fatty liver disease: An international expert consensus statement. Journal of Hepatology. 2020 Jul;73(1):202–9. https://doi.org/10.1016/j.jhep.2020.03.039.

[44] Pratt DS, Kaplan MM. Evaluation of abnormal liver-enzyme results in asymptomatic patients. N Engl J Med. 2000 Apr 27;342(17):1266-1271. https://doi.org/10.1056/NEJM200004273421707.

[45] Welsh Government. 2021-22 Health Board and Public Health Wales NHS Trust Allocations. Cardiff: Welsh Government, 2021. https://www.gov.wales/2021-22-health-board-and-public-health-wales-nhs-trust-allocations.pdf. [Accessed 05 Nov 2024].

Figure Legends

Fig. 1 Flowchart of study population selection (Fig.1A) and Venn diagram of study cohort composition (Fig.1B)

Abbreviations: PEDW: Patient Episode Database for Wales; WLGP: Welsh Longitudinal General Practice; ADDE: Annual District Death Extract

Fig. 2 Changes in Standardised Incidence per 100,000 Inhabitants by Data Source (2004-2022)
Bar chart presenting the changes in annual European age-sex standardised incidence rate and percentages by data sources. The x-axis represents time in years, and the y-axis represents incidence rate in Fig.2A, percentages in Fig.2B.

Line chart presenting the trends in annual European age-sex standardised incidence rate. The x-axis represents time in years, and the y-axis represents incidence rate in Fig.2C.

Fig. 3 Changes in Standardised Incidence per 100,000 Inhabitants by Stages (2004-2022)
Bar chart presenting the changes in annual European age-sex standardised incidence rate and percentages by stages. The x-axis represents time in years, and the y-axis represents incidence rate in Fig.3A, percentages in Fig.3B.

Line chart presenting the trends in annual European age-sex standardised incidence rate. The x-axis represents time in years, and the y-axis represents incidence rate in Fig.3C.

Fig. 4 Changes in Standardised Incidence per 100,000 Inhabitants by Liver Disease Aetiologies (2004-2022)
Bar chart presenting the changes in annual European age-sex standardised incidence rate and percentages by aetiologies. The x-axis represents time in years, and the y-axis represents incidence rate in Fig.4A, percentages in Fig.4B.

Line chart presenting the trends in annual European age-sex standardised incidence rate. The x-axis represents time in years, and the y-axis represents incidence rate in Fig.4C.

Abbreviations: ArLD: alcohol related liver disease; NAFLD: non-alcoholic fatty liver disease; HBV: hepatitis B virus; HCV: hepatitis C virus

* ArLD overlap: if an individual had 2 or more diagnoses from liver disease aetiologies (ArLD, NAFLD, HBV, HCV, metabolic, hemochromatosis and autoimmune liver diseases) and one of them was ArLD; Non-ArLD overlap: if an individual had 2 or more diagnoses from liver disease aetiologies but none of them was ArLD.

Tables

Table 1. Demographic and socioeconomic characteristics of liver disease patients

Demographic characteristics	Full cohort, N = 111,098[1]	Primary care group, N = 20,319[1]	Secondary care group, N = 56,710[1]	Mortality group, N = 3,459[1]	Multi-source group, N = 30,610[1]
Sex					
Male	57,491(51.7%)	10,478(51.6%)	27,558(48.6%)	2,019(58.4%)	17,436(57.0%)
Female	53,607(48.3%)	9,841(48.4%)	29,152(51.4%)	1,440(41.6%)	13,174(43.0%)
Age					
0-17	1,445(1.3%)	342(1.7%)	814(1.4%)	<10 (<0.3%)	<290 (<1.0%)
18-29	5,862(5.3%)	1,447(7.1%)	3,486(6.1%)	<40 (<1.2%)	<900 (<0%)
30-39	10,109(9.1%)	2,520(12.4%)	5,099(9.0%)	143(4.1%)	2,347(7.7%)
40-49	15,370(13.8%)	3,744(18.4%)	6,316(11.1%)	421(12.2%)	4,889(16.0%)
50-59	22,799(20.5%)	5,356(26.4%)	9,332(16.5%)	705(20.4%)	7,406(24.2%)
60-69	23,264(20.9%)	4,271(21.0%)	11,026(19.4%)	758(21.9%)	7,209(23.6%)
70-79	19,189(17.3%)	2,117(10.4%)	11,335(20.0%)	693(20.0%)	5,044(16.5%)
80+	13,060(11.8%)	522(2.6%)	9,302(16.4%)	702(20.3%)	2,534(8.3%)
Cohort entry year					
2004-2007	17,099(15.4%)	1,949(9.6%)	7,345(13.0%)	617(17.8%)	7,188(23.5%)
2008-2011	17,554(15.8%)	2,513(12.4%)	8,484(15.0%)	705(20.4%)	5,852(19.1%)
2012-2015	20,658(18.6%)	3,109(15.3%)	10,682(18.8%)	731(21.1%)	6,136(20.0%)
2016-2019	30,448(27.4%)	6,683(32.9%)	15,521(27.4%)	749(21.7%)	7,495(24.5%)
2020-2022	25,339(22.8%)	6,065(29.8%)	14,678(25.9%)	657(19.0%)	3,939(12.9%)
WIMD 2019 quintiles					
1, most deprived	27,178(24.5%)	4,417(21.7%)	14,012(24.7%)	928(26.8%)	7,821(25.6%)
2	24,391(22.0%)	4,203(20.7%)	12,411(21.9%)	858(24.8%)	6,919(22.6%)
3	21,066(19.0%)	3,828(18.8%)	10,802(19.0%)	647(18.7%)	5,789(18.9%)
4	19,619(17.7%)	3,949(19.4%)	9,855(17.4%)	545(15.8%)	5,270(17.2%)
5, least deprived	18,844(17.0%)	3,922(19.3%)	9,630(17.0%)	481(13.9%)	4,811(15.7%)

[1] n(%)

Abbreviations: PEDW: Patient Episodes Dataset for Wales, WLGP: Welsh Longitudinal General Practice dataset, ADDE: Annual District Deaths Extract, WIMD: Welsh Index of Multiple Deprivation

Table 2. Crude and standardised incidence of chronic liver disease from 2004 to 2022

Year	Wales population	Incident cases	Crude incidence (95% CI)	STD incidence (95% CI)
2004	3,131,640	3,060	97.7 (94.3 - 101.2)	97.7 (94.2 - 101.3)
2005	3,105,633	3,403	109.6 (105.9 - 113.3)	109.4 (105.7 - 113.2)
2006	3,104,483	3,670	118.2 (114.4 - 122.1)	118.7 (114.9 - 122.6)
2007	3,087,732	3,537	114.6 (110.8 - 118.4)	115.1 (111.3 - 119.0)
2008	3,083,840	3,867	125.4 (121.5 - 129.4)	126.3 (122.3 - 130.4)
2009	3,081,366	4,008	130.1 (126.1 - 134.2)	131.7 (127.7 - 135.9)
2010	3,077,165	4,268	138.7 (134.6 - 142.9)	140.6 (136.4 - 144.9)
2011	3,072,739	4,519	147.1 (142.8 - 151.4)	149.9 (145.5 - 154.4)
2012	3,073,788	4,546	147.9 (143.6 - 152.3)	150.6 (146.3 - 155.1)
2013	3,071,058	4,733	154.1 (149.8 - 158.6)	158.0 (153.5 - 162.6)
2014	3,070,928	5,145	167.5 (163.0 - 172.2)	172.5 (167.7 - 177.3)
2015	3,063,758	5,659	184.7 (179.9 - 189.6)	192.2 (187.2 - 197.4)
2016	3,049,971	6,548	214.7 (209.5 - 220.0)	225.7 (220.2 - 231.3)
2017	3,038,872	6,981	229.7 (224.4 - 235.2)	243.2 (237.4 - 249.1)
2018	3,025,867	7,953	262.8 (257.1 - 268.7)	279.1 (272.9 - 285.4)
2019	3,006,299	8,481	282.1 (276.1 - 288.2)	301.2 (294.8 - 307.8)
2020	2,985,668	7,327	245.4 (239.8 - 251.1)	264.0 (257.8 - 270.2)
2021	2,969,309	9,136	307.7 (301.4 - 314.1)	332.0 (325.0 - 338.9)
2022	2,957,422	8,596	290.7 (284.5 - 296.9)	316.2 (309.4 - 323.1)

Abbreviation: STD: standardised; CI: confidence interval

Table 3. The underlying liver disease conditions at cohort entry

Phenotype	Full cohort, N = 111,098[1]	Primary care group, N = 20,319[1]	Secondary care group, N = 56,710[1]	Mortality group, N = 3,459[1]	Multi-source group, N = 30,610[1]
Chronic liver disease					
Stage 1 - tier 1	79,992(72.0%)	18,453(90.8%)	35,102(61.9%)	1,728(50.0%)	24,709(80.7%)
- ArLD	19,760(17.8%)	2,116(10.4%)	4,566(8.1%)	887(25.6%)	12,191(39.8%)
- NAFLD	33,655(30.3%)	13,565(66.8%)	13,444(23.7%)	727(21.0%)	5,919(19.3%)
- Metabolic liver disease	5,469(4.9%)	782(3.8%)	3,887(6.9%)	69(2.0%)	731(2.4%)
- HBV	1,063(1.0%)	0(0.0%)	905(1.6%)	0(0.0%)	158(0.5%)
- HCV	3,539(3.2%)	0(0.0%)	2,697(4.8%)	13(0.4%)	829(2.7%)
- Autoimmune liver disease	13,582(12.2%)	815(4.0%)	9,060(16.0%)	199(5.8%)	3,508(11.5%)
- Haemochromatosis	4,111(3.7%)	1,209(6.0%)	<940(<1.7%)	<10(<0.2%)	1,958(6.4%)
- ArLD overlap	799(0.7%)	15(0.1%)	158(0.3%)	173(5.0%)	453(1.5%)
- Non ArLD overlap	371(0.3%)	<20(<0.1%)	228(0.4%)	<10(<0.2%)	123(0.4%)
Stage 1 - tier 2	6,111(5.5%)	81(0.4%)	4,906(8.7%)	127(3.7%)	997(3.3%)
- Hepatitis not specified	4,783(4.3%)	11(0.1%)	3,867(6.8%)	102(2.9%)	803(2.6%)
- Congestive hepatopathy	574(0.5%)	<10(<0.1%)	514(0.9%)	<10(<0.2%)	<60(<0.2%)
- Toxic liver disease	757(0.7%)	68(0.3%)	526(0.9%)	21(0.6%)	142(0.5%)
Stage 1 - tier 3	8,426(7.6%)	36(0.2%)	8,132(14.3%)	37(1.1%)	221(0.7%)
- Miscellaneous	8,426(7.6%)	36(0.2%)	8,132(14.3%)	37(1.1%)	221(0.7%)
Stage 2	4,562(4.1%)	506(2.5%)	1,664(2.9%)	690(19.9%)	1,702(5.6%)
- Hepatic fibrosis	261(0.2%)	64(0.3%)	135(0.2%)	<10(<0.2%)	<60(<0.2%)
- Cirrhosis	4,301(3.9%)	442(2.2%)	1,529(2.7%)	682(19.7%)	1,648(5.4%)
Stage 3	3,040(2.7%)	166(0.8%)	2,171(3.8%)	43(1.2%)	660(2.2%)
- Portal hypertension	3,040(2.7%)	166(0.8%)	2,171(3.8%)	43(1.2%)	660(2.2%)
Stage 4	3,221(2.9%)	53(0.3%)	2,517(4.4%)	276(8.0%)	375(1.2%)
- Hepatic decompensation	3,221(2.9%)	53(0.3%)	2,517(4.4%)	276(8.0%)	375(1.2%)
Stage 5	2,233(2.0%)	121(0.6%)	341(0.6%)	473(13.7%)	1,298(4.2%)
- Hepatocellular carcinoma	748(0.7%)	71(0.3%)	83(0.1%)	86(2.5%)	508(1.7%)
- Intrahepatic cholangio carcinoma	1,455(1.3%)	47(0.2%)	232(0.4%)	387(11.2%)	789(2.6%)
- Other primary liver cancer	39(0.0%)	<10(<0.1%)	26(0.0%)	<10(<0.2%)	<10(<0.1%)
Acute liver disease					
Acute liver disease	3,513(3.2%)	903(4.4%)	1,877(3.3%)	85(2.5%)	648(2.1%)
- Budd-Chiari	184(0.2%)	<10(<0.1%)	144(0.3%)	<20(<0.1%)	20(0.1%)
- Acute liver failure	27(0.0%)	<30(<0.2%)	0(0.0%)	0(0.0%)	<10(<0.1%)
- Infection/infarction	1,343(1.2%)	62(0.3%)	757(1.3%)	67(1.9%)	457(1.5%)
- Other non-specified acute liver injuries	71(0.1%)	45(0.2%)	<20(<0.1%)	0(0.0%)	<10(<0.1%)

[1] n(%)

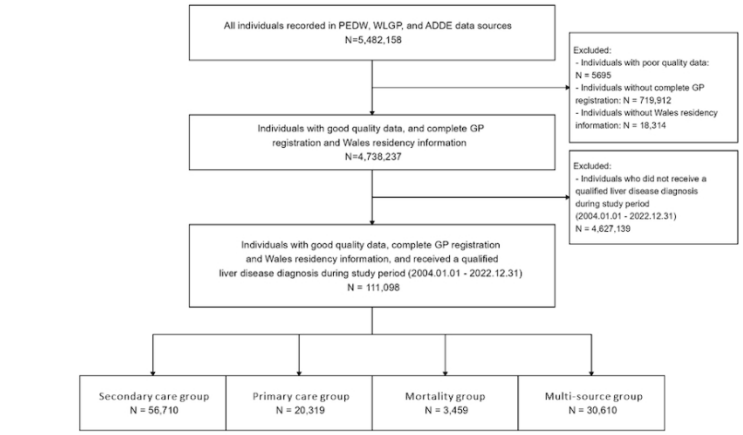
Abbreviation: ArLD: alcohol-related liver disease; NAFLD: non-alcohol fatty liver disease; HBV: hepatitis B virus; HCV: hepatitis C virus

Table 4. Comorbidities associated with liver disease

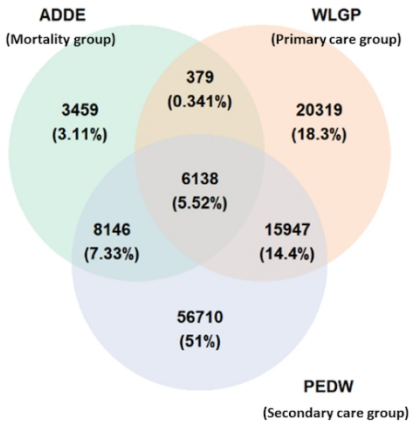
Comorbidities	Full cohort, N = 111,098[1]	Primary care group, N = 20,319[1]	Secondary care group, N = 56,710[1]	Mortality group, N = 3,459[1]	Multi-source group, N = 30,610[1]
CVD related conditions	8,923(8.0%)	852(4.2%)	5,362(9.5%)	385(11.2%)	2,324(7.6%)
Diabetes	7,658(6.9%)	1,271(6.3%)	3,577(6.3%)	279(8.1%)	2,531(8.3%)
Hypertension/antihypertensives	40,427(36.4%)	7,248(35.7%)	20,361(35.9%)	1,368(39.7%)	11,450(37.4%)

[1] n(%)

Abbreviation: CVD: cardiovascular disease; COPD: chronic obstructive pulmonary disease



A



B

Fig. 1 Flowchart of study population selection (Fig.1A) and Venn diagram of study cohort composition (Fig.1B)
Abbreviations: PEDW: Patient Episode Database for Wales; WLGP: Welsh Longitudinal General Practice; ADDE: Annual District Death Extract

99x124mm (300 x 300 DPI)

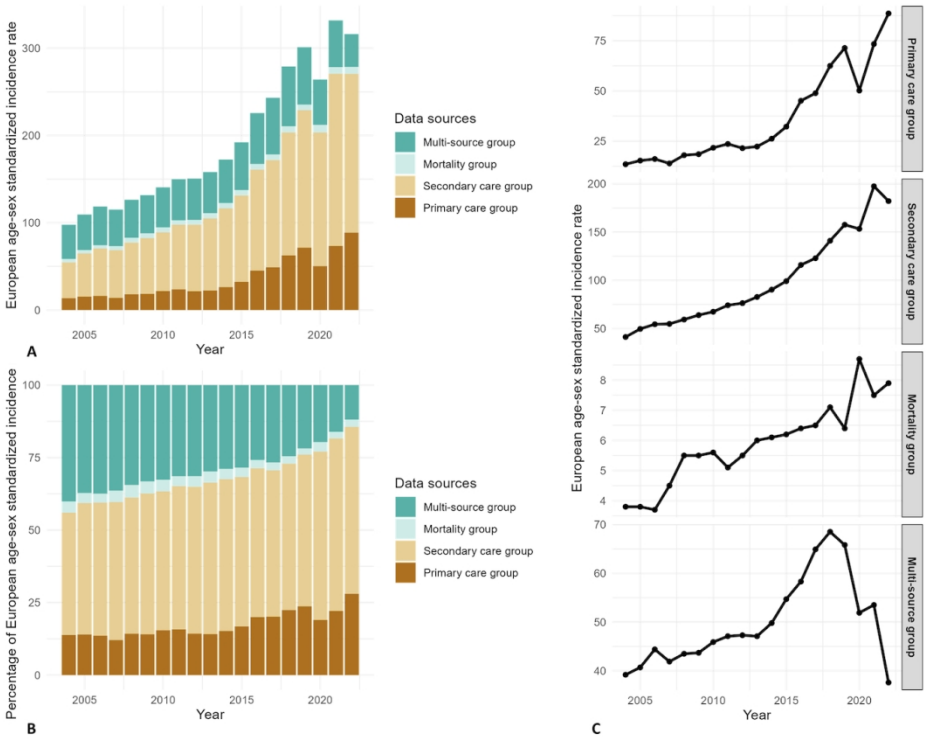


Fig. 2 Changes in Standardised Incidence per 100,000 Inhabitants by Data Source (2004-2022)
Bar chart presenting the changes in annual European age-sex standardised incidence rate and percentages by data sources. The x-axis represents time in years, and the y-axis represents incidence rate in Fig.2A, percentages in Fig.2B.
Line chart presenting the trends in annual European age-sex standardised incidence rate. The x-axis represents time in years, and the y-axis represents incidence rate in Fig.2C.

124x99mm (300 x 300 DPI)

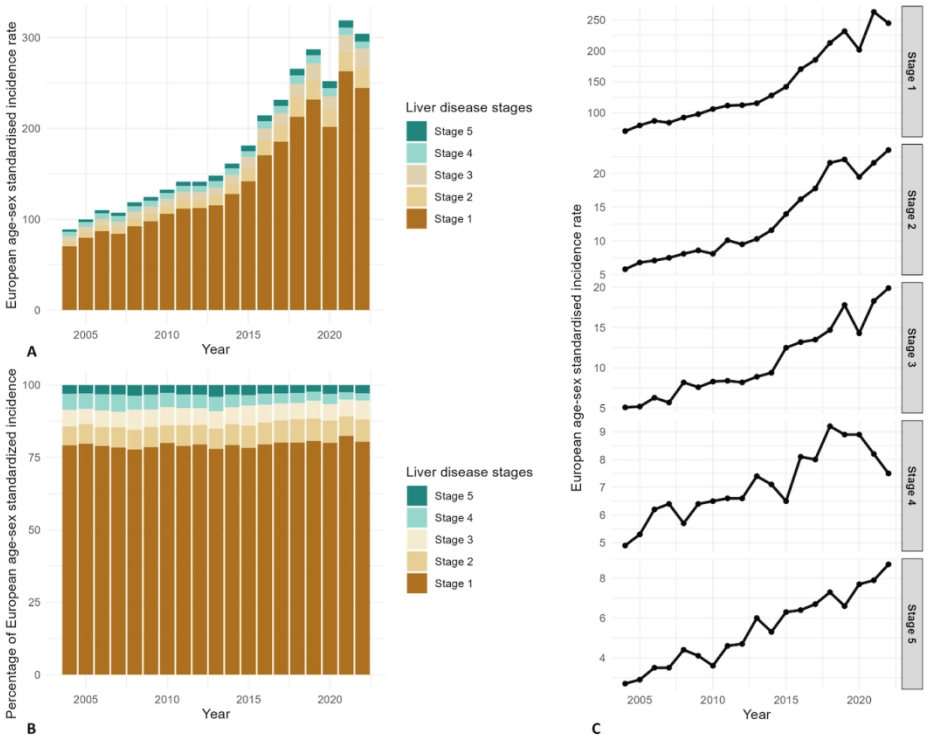


Fig. 3 Changes in Standardised Incidence per 100,000 Inhabitants by Stages (2004-2022)
Bar chart presenting the changes in annual European age-sex standardised incidence rate and percentages by stages. The x-axis represents time in years, and the y-axis represents incidence rate in Fig.3A, percentages in Fig.3B.
Line chart presenting the trends in annual European age-sex standardised incidence rate. The x-axis represents time in years, and the y-axis represents incidence rate in Fig.3C.

124x99mm (300 x 300 DPI)

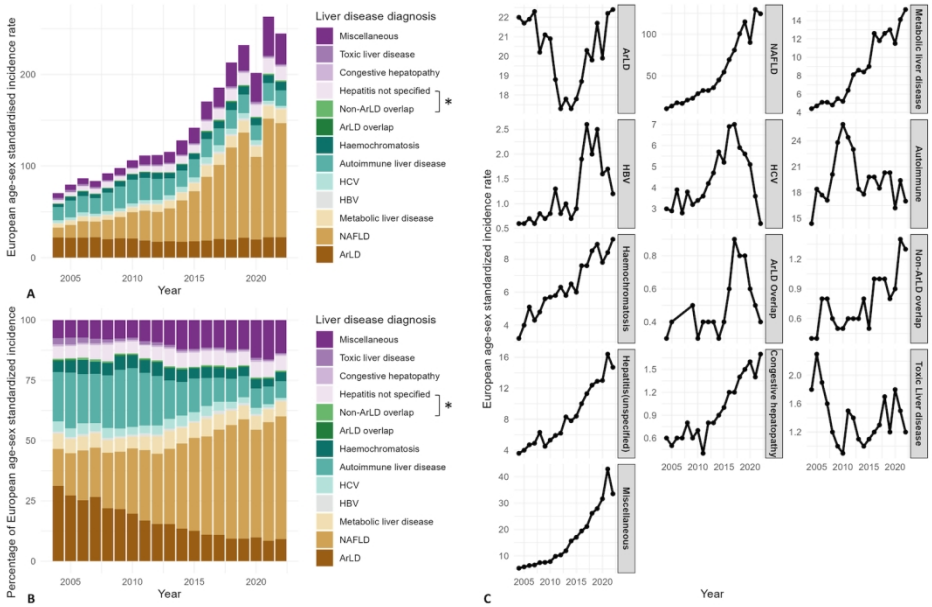
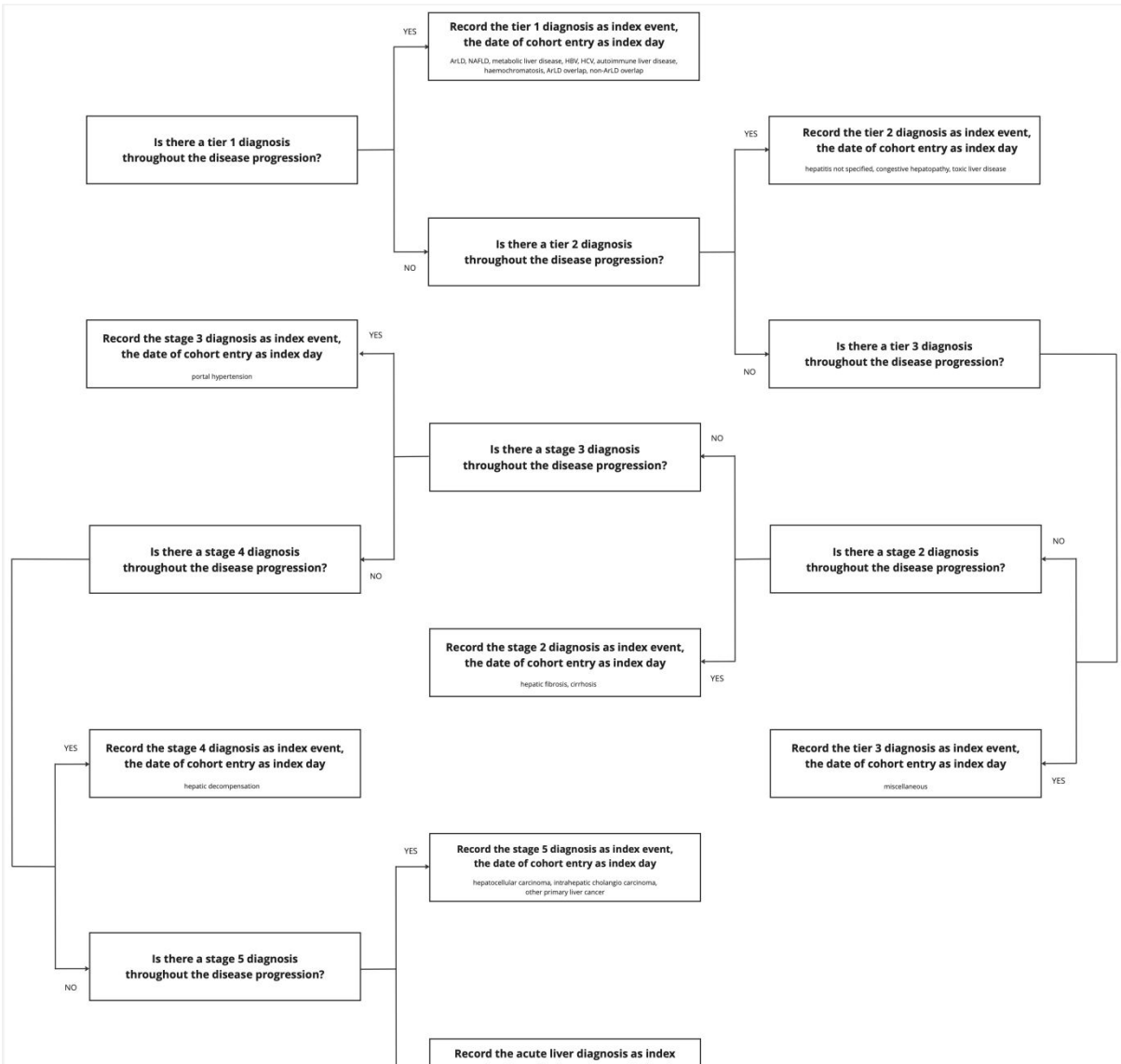


Fig. 4 Changes in Standardised Incidence per 100,000 Inhabitants by Liver Disease Aetiologies (2004-2022)
Bar chart presenting the changes in annual European age-sex standardised incidence rate and percentages by aetiologies. The x-axis represents time in years, and the y-axis represents incidence rate in Fig.4A, percentages in Fig.4B.
Line chart presenting the trends in annual European age-sex standardised incidence rate. The x-axis represents time in years, and the y-axis represents incidence rate in Fig.4C.
Abbreviations: ArLD: alcohol related liver disease; NAFLD: non-alcoholic fatty liver disease; HBV: hepatitis B virus; HCV: hepatitis C virus

* ArLD overlap: if an individual had 2 or more diagnoses from liver disease aetiologies (ArLD, NAFLD, HBV, HCV, metabolic, hemochromatosis and autoimmune liver diseases) and one of them was ArLD; Non-ArLD overlap: if an individual had 2 or more diagnoses from liver disease aetiologies but none of them was ArLD.

150x99mm (300 x 300 DPI)

Supplemental Materials



Supplemental Fig. 1 Flowchart for Identifying Index Events and Determining Cohort Entry Date.

This flowchart illustrates the decision rules for identifying the index event and determining the cohort entry date for study participants. The index date is defined as the date of the first diagnosis between 2004 and 2022. The aetiological diagnoses throughout the disease progression were identified as the index event entering the cohort. If no aetiological diagnosis is present, the index event is defined based on the following sequence: stage 2 diagnoses, stage 3 diagnoses, stage 4 diagnoses, and stage 5 diagnoses.

The decision flow proceeds as follows:

Tier 1 Diagnosis: If present at any point during the disease progression, the tier 1 diagnosis is recorded as the index event, and the date of the cohort entry is the index day. Tier 1 diagnoses include ArLD, NAFLD, metabolic liver disease, HBV, HCV, autoimmune liver disease, and haemochromatosis.

Tier 2 Diagnosis: If there is no tier 1 diagnosis, the presence of a tier 2 diagnosis is checked. If present, it is recorded as the index event, and the date of cohort entry is the index day. Tier 2 diagnoses include hepatitis not specified, congestive hepatopathy, and toxic liver disease.

Tier 3 Diagnosis: In the absence of tier 1 and tier 2 diagnoses, the presence of a tier 3 diagnosis is checked. If found, it is recorded as the index event. Tier 3 diagnoses include miscellaneous conditions.

Stage 2 Diagnosis: If none of the Stage 1 diagnoses are present, the presence of a stage 2 diagnosis is considered. If found, it is recorded as the index event. Stage 2 diagnoses include hepatic fibrosis and cirrhosis.

Stage 3 Diagnosis: If no stage 1 or stage 2 diagnoses are present, a stage 3 diagnosis is checked next. If found, it is recorded as the index event. Stage 3 diagnoses include portal hypertension.

Stage 4 Diagnosis: If the stage 1, stage 2, and stage 3 diagnoses are absent, a stage 4 diagnosis is checked. If present, it is recorded as the index event. Stage 4 diagnoses include hepatic decompensation.

Stage 5 Diagnosis: If no above diagnoses are found, a stage 5 diagnosis is considered. If found, it is recorded as the index event. Stage 5 diagnoses include hepatocellular carcinoma, intrahepatic cholangiocarcinoma, other primary liver cancer.

Acute Liver Diagnosis: If none of the chronic diagnoses are present, an acute liver diagnosis is checked. If found, it is recorded as the index event. Acute liver diagnoses include acute viral hepatitis, Budd-Chiari, acute liver failure, infections/sepsis, and other non-specified acute liver injuries.

Supplemental Tables

Supplemental Table 1 Code list for identifying liver disease

Phenotype	ICD10 codes	Read codes
Acute viral hepatitis	B15, B19, B16, B17 (B171 excluded), B172, B178, B179, B180, B159, B169, B199	A70z1, AyuB0, XE2u., Q4090, A700., A701., A7052, A70.., A706., A708., A709., A70z., A70G., A704., AyuB3
Acute liver failure	K720	J6000
Budd-Chiari	I820	G820.
Infection/infarction	K750, K763, K751	J62.., J620., J6200, J6201, J6202, J6203, J6204, J620z, A053., J634., J621.
Other non-specified acute liver injuries	K752	J63y1
Autoimmune liver disease	K754, K743, K831, K753	J63B., J6141, J6160, J6617, J63X.
Haemochromatosis	E831	C3500
Metabolic liver disease	E880, E830	C3762, C3761, C3510
HBV *	B181, B180	
HCV *	B182	A70E., A70F.
Alcohol-related liver disease	K70	J613., J6130, J612., J6120, J610., J617., J6170, J611.
Non-alcoholic fatty liver disease	K760, K7581	J61y1, J61y8
Hepatitis not specified	K769, K7589, K73	Jyu72, J614y
Congestive hepatopathy	K761, K762, K765	J630., J636., J637.
Toxic liver disease	K71	J635., J6350, J6351, J63252, J6353, J6354, J6355, J6356, J6357, J635X
Miscellaneous	K764, K768, K77	J638., Jyu73, J63yz, Jyu75
Hepatic fibrosis	K740, K741, K742	J61y4, J61y6, J61y5
Cirrhosis	K703, K744, K745, K746, K749	J6161, J616z, J615z
Portal hypertension	K766, I81, I859, I982, I85	J623., G81.., G8523, G852., G8521, G8522, G852z

Hepatic decompensation	K721, K767, I850, K72, C220	J624., SP143, G850., G8520, J625., B1503, BB5D7, BB5D5, BB5D8
Hepatocellular carcinoma (HCC)	C220	B1503, BB5D7, BB5D5, BB5D8
Intrahepatic cholangio carcinoma (ICC)	C221	B150.
Other primary liver cancers	C222, C223, C224, C225, C226. C227	B808.

*We identified Read codes(A7071, A7073, ZV02B, Q4091, 43B4., A7070, A7051, A7072, A70z0, A70A., A70B., A70C., A70D., A70E.,A70F., ZV02C) and ICD-10 codes (B180, B181, B182) for HBV and HCV. However, in order to comply with Data Protection Act 2018 and the UK General Data Protection Regulation, we could not include Read codes (A7071, A7073, ZV02B, Q4091, 43B4., A7070, A7051, A7072, A70z0, A70A., A70B., A70C., A70D., ZV02C) and ICD-10 codes (B171) as these were flagged as sensitive in the latest version of known sensitive code list of SAIL Databank.

Supplemental Table 2-1 ICD-10 code list for identifying comorbidities

Comorbidities	ICD-10 codes
Atrial fibrillation	I481, I482
Angina	I200, I201, I208, I209
Asthma	J45
Diabetes	E09, E10, E102, E103, E11, E13, K86
Heart failure	I50, I501, I502, I503, I508
Hypertension	I10, I11, I12, I13, I15
Peripheral vascular disease	E106, E116, I70
Renal disease	N18
Stroke	I691, I61, I63, I64, I60, I66, G45.0, G45.1, G45.3, G46.0, G46.2, G45.8, I65, G46.1, 45.9, G45.2, G45.4
Transient ischaemic attack	I61, I63, I66
Other ischaemic	I20, I21, I22, I23, I24, I25

Supplemental Table 2-2 Read codes list for identifying comorbidities

Comorbidities	Read codes
Atrial fibrillation	14AN., 14AR., 3272, 3273, 8CMW2, G573., G5730, G5731, G5732, G5733, G5734, G5735, G5736, G5737, G5738, G5739, G573z, 3272, 2[illegible], 793M1, 793M3
Angina	G3112, G33.., G330., G3300, G330z, G33z., G33z3, G33z7, G33zz, 662K., 662K0, 662K1, 662K2, 662Kz, 8B27., G33z1, G33z2, G33z5, G34[illegible], G34y0, Gyu30
Asthma	H33.., H330., H3300, H3301, H330z, H331., H3310, H3311, H331z, H332., H333., H334., H335., H33z., H33z0, H33z1, H33z2, H33zz,
Diabetes	66AJ., 66AJ1, 66AJz, 66An., 66Ao., 8CR2., 9OLA., 9OLA., C10.., C100., C1000, C100z, C101., C1010, C1011, C101y, C102., C1020, C1[illegible]02z, C103., C1030, C1031, C103y, C103z, C104., C1040, C1041, C104y, C104z, C105., C1050, C1051, C105y, C105z, C106., C1060, C1061, C106y, C106z, C107., C1070, C1071, C1072, C[illegible]1074, C107y, C107z, C108., C1080, C1081, C1082, C1083, C1084, C1085, C1086, C1087, C1088, C1089, C108A, C108B, C108C, C108D, C108E, C108F, C108G, C108H, C108J, C108y, C[illegible]08z, C109., C1090, C1091, C1092, C1093, C1094, C1095, C1096, C1097, C1099, C109A, C109B, C109C, C109D, C109E, C109F, C109G, C109H, C109J, C109K, C10A., C10A0, C10A1, C10A2, C10A3, C10A4, C10A5, C10A6, C10A7, C10AW, C10AX, C10B., C10B0, C10C., C10D., C10E., C10E0, C10E1, C10E2, C10E3, C10E4, C10E5, C10E6, C10E7, C10E8, C10E9, C10EA, C10EB, C10EC, C10ED, C10EE, C10EF, C10EG, C10EH, C10EJ, C10EK, C10EL, C10EM, C10EN, C10EP, C10EP, C10EQ, C10ER, C10F., C10F0, C10F1 C10F2, C10F3, C10F4, C10F5, C10F6, C10F7, C10F[illegible], C10FA, C10FB, C10FC, C10FD, C10FE, C10FF, C10FG, C10FH, C10FJ, C10FK, C10FL, C10FM, C10FN, C10FP, C10FQ, C10FR, C10FS, C10G., C10G0, C10H., C10H0, C10K., C10K0, C10M., C10M0, C10N., C10N0, C10N1, C10P., C10P0, C10P1, C10y., C10y1, C10yy, C10yz, C10z., C10z0, C10z1, C10zy, C10zz, F372., F3720, F3721, F3722, 1434, 14F4., 14P3., 9OL9.
Heart failure	G58.., G580., G5800, G5801, G5802, G5803, G5804, G581., G5810, G582., G583., G584., G58z., G232., G234., G1yz1, 1O1.., 662W., 662[illegible], 8[illegible]29., 8H2S., 9Or0., G400., G41z., G5540, G5540, G5yy9, G5yyA, R2y10, 585f., 585g., 14A6., 14AM., 1736, 1J60., 23E1., 388D., 662T., 662f., 662g., 662h., 662i., 679X., 8CL3., 8HBE., 8[illegible]Hz., 8[illegible]Hg8., 8Hk0., 9N0k., 9N2p., 9N4s., 9N4w., 9N6T., 9On.., 9On0., 9On1., 9On2., 9On3., 9On4., 9Or.., 9Or1., 9Or2., 9Or3., 9Or4., 9Or5., 9h1.., 9h11., 9h12., 9hH.., 9hH0., 9hH1., G581., H54.., H54[illegible], H5410, H541z, H54z., H584., H584z, ZRad.
Hypertension	14A2., G2..., G20.., G200., G201., G202., G203., G20z., G21.., G210., G2100, G2101, G210z, G211., G2110, G2111, G211z, G21z., G21z0, G21[illegible], G21zz, G22.., G220., G221., G222., G22z., G23.., G230., G231., G232., G233., G234., G23z., G24.., G240., G2400, G240z, G241., G2410, G241z, G244., G24z., G24z0, G24z1, G24zz, G25.., G2[illegible]0., G251., G26.., G27.., G28.., G2y.., G2z.., 6627, 6628, 662F., 662G., 662O., 662b., 662c., 662d., 662r., 7Q01., 8B26., 8BL0., 8I3N., F4042, F4213, G672., Gyu2., L122., L1220, L1221, L1223, L122z, L127., L127z, L128., L1280, L1282, Gyu21

Comorbidities	Read codes
Peripheral vascular disease	G73.., G734., G73y., G73z., G73z0, G73zz, Gyu74, 2G63., A3A0F, C107., C1070, C1071, C1073, C1074, C107z, C108G, C109F, C109F, C10EH, C10FF, G700., G702., G702z, G731., G7310, G731z, G732., G7320, G7321, G733., G73y0, G73y1, G73yz, G740., G742z, M271., M2710, M2713, R0550, R0550
Renal disease	1Z13., 1Z14., 1Z1H., 1Z1J., 1Z1K., 1Z1L., K050., K054., K055., K060., K060., K08z., K0D.., 1Z10., 1Z17., 1Z18., 1Z11., 1Z19., 1Z1A., 1Z1Y., 1Z15., 1Z16., 1Z1B., 1Z1C., 1Z1D., 1Z1E., 1Z1F., 1Z1G.,
Stroke	G6..., G61.., G610., G611., G612., G613., G614., G615., G616., G617., G618., G619., G61X., G61X0, G61X1, G61z., G63.., G630., G631., G632., G633., G634., G63y., G63y0, G63y1, G63z., G64.., G640., G6400, G641., G6410, G64z., G64z0, G64z1, G64z2, G64z3, G64z4, G66.., G660., G661., G662., G663., G664., G665., G666., G668., G67.., G670., G671., G6710, G6711, G671z, G677., G6770, G6771, G6772, G6773, G6774, G679., G67y., G67z., G6y.., G67z.
Transient ischaemic attack	G65z., G65zz, G65z1, G65y., 14AB., G65z0, Fyu55, G65.., G65.., G650., G651., G6510, G652., G653., G654., G655., G656., G657., G65y., F4236, 14AB0
Other ischaemic	G33z4, G34.., G34y., G34y0, G34y1, G34yz, G34z., G34z0, G3y.., G3z.., G31y3, G332., 6A2.., 6A4.., 8B3k., 8H2V., G3..., G31.., G3110, G31y., G31y2, G31yz, G340., G343., G344., Gyu3., Gyu32, Gyu33
Anti-hypertensive	bi1.., bi11., bi12., bi13., bi14., bi15., bi16., bi17., bi18., bi19., bi1A., bi1B., bi1C., bi1H., bi1I., bi1J., bi1K., bi1a., bi1b., bi1c., bi1d., bi1g., bi1h., bi1i., bi1j., bi1k., bi1l., bi1m., bi1n., bi1o., bi1p., bi1q., bi1r., bi1v., bi1w., bi1x., bi1y., bi1z., bi2.., bi21., bi22., bi23., bi24., bi25., bi26., bi27., bi29., bi2A., bi2B., bi2C., bi2D., bi2E., bi2F., bi2G., bi2H., bi2J., bi2K., bi2L., bi2M., bi2a., bi2t., bi2u., bi2v., bi2w., bi2x., bi2y., bi2z., bi3.., bi31., bi32., bi33., bi34., bi35., bi36., bi37., bi38., bi39., bi3a., bi3b., bi3c., bi3c., bi3d., bi3d., bi3e., bi3f., bi3g., bi3h., bi3i., bi3j., bi3k., bi3l., bi3m., bi3q., bi3r., bi3y., bi4.., bi41., bi42., bi43., bi44., bi45., bi46., bi47., bi49., bi4A., bi4B., bi4C., bi4D., bi4E., bi5.., bi51., bi52., bi53., bi54., bi57., bi58., bi6.., bi61., bi62., bi63., bi64., bi65., bi66., bi67., bi68., bi69., bi6A., bi6B., bi6C., bi6D., bi6E., bi6F., bi6G., bi6o., bi6p., bi6q., bi6r., bi6s., bi6t., bi6u., bi6v., bi6w., bi6x., bi6y., bi6z., bi7.., bi71., bi72., bi73., bi74., bi8.., bi81., bi82., bi82., bi83., bi83., bi84., bi84., bi85., bi86., bi86., bi87., bi87., bi88., bi88., bi89., bi89., bi8a., bi9.., bi91., bi92., bi93., bi94., bi94., bi95., bi95., bi96., bi96., bi97., bi98., bi99., bi9A., bi9A., bi9z., biA.., biA1., biA2., biA3., biA4., biB.., biB1., biB2., biB3., biBx., biBy., biBz., biC.., biC1., biC2., biC3., biC4., biC5., biC6., bk3.., bk31., bk32., bk33., bk34., bk37., bk38., bk3B., bk3C., bk3D., bk3E., bk3F., bk3G., bk3H., bk4.., bk41.,

Comorbidities	Read codes
	bk42., bk43., bk44., bk45., bk46., bk4A., bk4B., bk4C., bk4s., bk4t., bk4u., bk4v., bk4w., bk5.., bk51., bk52., bk53., bk54., bk55., bk56., bk7.., bk71., bk72., bk73., bk74., bk75., bk76., bk77., bk78., bk79., bk7z., bk8.., bk81., bk82., bk83., bk84., bk85., bk8z., bk9.., bk91., bk92., bk92., bk93., bk9x., bk9y., bk9z., bkB.., bkB1., bkB2., bkB3., bkB4., bkB5., bkB6., bkJ.., bkJ1., bkJ2., bkJ3., bkJ4., bkJ5., bkJ6., bd□c, bd1.., bd11., bd12., bd13., bd14., bd15., bd16., bd17., bd18., bd19., bd1A., bd1B., bd1C., bd1D., bd1E., bd1F., bd1G., bd1H., bd1J., bd1K., bd1L., bd1M., bd1N., bd1O., bd1P., bd1Q., bd1R., bd1S., bd1T., bd1U., bd1V., bd1W., bd1X., bd1Y., bd1Z., bd1a., bd1b., bd1c., bd1d., bd1e., bd1f., bd1g., bd1h., bd1i., bd1j., bd1k., bd1l., bd1l., bd1m., bd1n., bd1o., bd1p., bd1r., bd1r., bd1s., bd1t., bd1u., bd1v., bd1w., bd1x., bd1y., bd1z., bd2.., bd21., bd22., bd23., bd2w., bd2x., bd2y., bd3.., bd31., bd32., bd34., bd35., bd36., bd37., bd3a., bd3b., bd3c., bd3c., bd3d., bd3e., bd3f., bd3g., bd3h., bd3i., bd3j., bd3k., bd3l., bd3x., bd3z., bd4.., bd41., bd4z., bd5.., bd51., bd52., bd53., bd54., bd55., bd56., bd57., bd58., bd59., bd5s., bd5t., bd5u., bd5v., bd5w., bd5x., bd5y., bd6.., bd61., bd62., bd64., bd65., bd66., bd67., bd68., bd6b., bd6c., bd6d., bd6e., bd6w., bd6x., bd6z., bd7.., bd71., bd72., bd7y., bd7z., bd8.., bd81., bd82., bd83., bd84., bd85., bd86., bd87., bd88., bd89., bd8a., bd8b., bd8c., bd8d., bd8e., bd8f., bd8g., bd8h., bd8i., bd8k., bd8l., bd8m., bd8n., bd8o., bd8u., bd9.., bda.., bda1., bda2., bda3., bda4., bday., bdaz., bdb.., bdc.., bdc1., bdc1., bdc2., bdc3., bdc4., bdc5., bdcu., bdcv., bdcw., bdcx., bdd.., bdd1., bdd1., bdd2., bddz., bde.., bde1., bde2., bde3., bde4., bde5., bde6., bde7., bde8., bde9., bdeQ., bdeR., bdeS., bdeb., bdec., bded., bdee., bdef., bdeg., bdeh., bdei., bdej., bdek., bdel., bdf.., bdf1., bdf2., bdf3., bdf4., bdf5., bdf5., bdf6., bdf6., bdf7., bdf8., bdf9., bdfA., bdfB., bdfC., bdfD., bdfE., bdfF., bdfG., bdfH., bdfI., bdfJ., bdfK., bdfL., bdfM., bdfw., bdfx., bdfy., bdfz., bdg.., bdg1., bdg2., bdh.., bdh1., bdh2., bdh3., bdh4., bdi.., bdi1., bdi2., bdj.., bdj1., bdj2., bdj3., bdj4., bdj5., bdl.., bdl1., bdl1., bdl2., bdl3., bdl4., bdl5., bdl6., bdl7., bdl8., bdl8., bdm.., bdm1., bdm2., bdmy., bdmz., bdn.., bdn1., bdn2., bdn3., bdn4., bdn5., bdn6., bb3.., bb31., bb32., bb33., bb34., bb35., bb36., bb37., bb38., bb39., bb3A., bb3B., bb3C., bb3D., bb3F., bb3G., bb3H., bb3J., bb3K., bb3L., bb3M., bb3N., bb3O., bb3P., bb3Q., bb3a., bb3b., bb3d., bb3e., bb3f., bb3g., bb3h., bb3i., bb3j., bb3k., bb3l., bb3m., bb3n., bb3p., bb3q., bb3r., bb3s., bb3v., bb3w., bb3x., bb3y., bb3z., bl5.., bl51., bl52., bl53., bl54., bl55., bl56., bl57., bl58., bl59., bl5A., bl5B., bl5C., bl5D., bl5E., bl5F., bl5G., bl5H., bl5I., bl5J., bl5K., bl5L., bl5M., bl5N., bl5O., bl5P., bl5Q., bl5R., bl5S., bl5T., bl5U., bl5V., bl5V., bl5W., bl5W., bl5X., bl5Y., bl5Z., bl5a., bl5b., bl5c., bl5d., bl5e., bl5f., bl5g., bl5h., bl5j., bl5k., bl5l., bl5m., bl5n., bl5o., bl5p., bl5q., bl5r., bl5s., bl5t., bl5u., bl5v., bl5w., bl5x., bl5y., bl5z., bl7.., bl71., bl72., bl73., bl74., bl7w., bl7x., bl7y., bl7z., bl8.., bl81., bl82., bl83., bl84., bl85., bl86., bl89., bl8A., bl8B., bl8C., bl8D., bl8E., bl8F., bl8G., bl8H., bl8J., bl8K., bl8L., bl8M., bl8O., bl8P., bl8Q., bl8R., bl8S., bl8T., bl8U., bl8V., bl8W., bl8X., bl8Y., bl8Z., bl8a., bl8b., bl8c., bl8d., bl8e., bl8f., bl8g., bl8h., bl8i., bl8j., bl8k., bl8l., bl8m., bl8n., bl8o., bl8p., bl8q., bl8r., bl8s., bl8t., bl8u., bl8v., bl8w., bl8x., bl8y., bl8z., bla.., bla.., bla1., bla1., bla2., bla2., blb.., blb1., blb2., blb3., blb4., blb5., blb5., blb6., blb6., blb7., blb8., blc.., blc1., blc2., blc3., blc4., blc5., blc6., blc7., blc8., blc9., blca., blcb., blcc., blcd., blce., blcf., blcg., blch., blci., blcj., blck., blcl., blcm., blcn., blco., blcp., blcq., blcr., blcs., blct., ble.., ble1., ble2., ble3., ble4., ble5., blg.., blg1., blg2., blg3., blg4., blg5., blg6., blh.., blh1., blh2., blh3., blh4., blj.., blj1., blj2., blj3., blj4., blj5., blj6., blj7., blj8., blj9., bljA., bljB., bljC., bljD., bljE., bljF., bljG., bljH., bljJ., bljK., bljL., bljM., bljN., bljO., bljP., bljQ., bljR., bljS., bljT., bljU., bljV., bljW., bljX., bljY., bljZ., blja., bljb., bljc., bljd., blje., bljf., bll.., bll1., bll2., bll3., bll4., bll5., bll6., bll7., bll8., bll9., blla., bllb., bllc., blld., blle., bllf., bllg., bllh., blli., bllj., bllk., blll., dt1.., dt13., dt14., b2..., b21.., b211., b212., b213., b214., b215., b216., b217., b218., b219., b21A., b21B., b21a., b21b., b22.., b221., b22., b22y., b22z., b23.., b231., b232., b23y., b23z.,

Comorbidities	Read codes
	b24.., b25.., b251., b25z., b26.., b261., b262., b263., b264., b26y., b26z., b27.., b271., b27z., b28.., b281., b282., b283., b284., b285., b286., b287., b288., b289., b28z., b29.., b291., b29z., b2a.., b2a1., b2az., b2b.., b2b1., b2b2., b2b3., b2bz., b2c.., b2c1., b2cz., b2d.., b2d1., b2dz., bA1.., bA11., bA12., bA1y., bA1z., bi1D., bi1E., bi1F., bi1G., bi1H., bi1f., bi1s., bi28., bi2b., bi3n., bi3p., bi3s., bi3t., bi3u., bi3v., bi3w., bi3x., bi48., bi4F., bi55., bi56., biC7., biC8., bk35., bk36., bk39., bk3A., bk3y., bk3z., bk47., bk48., bk49., bk4x., bk4y., bk4z., bk57., bk58., bk59., bk5x., bk5y., bk5z., bk86., bk87., bk88., bk8w., bk8x., bk8y., bkC.., bkC1., bkC2., bkC3., bkCx., bkCy., bkCz., bkH.., bkH1., bkH2., bkH3., bkHx., bkHy., bkHz., bkI.., bkI1., bkI2., bkI3., bkI4., bkI5., bkL.., bkL1., bkL2., bkL3., bkL4., bkL5., bkL6., bd38., bd39., bdeA., bdeB., bdeC., bdeD., bdeE., bdeF., bdeG., bdeH., bdeJ., bdeK., bdeL., bdeM., bdeN., bdeO., bdeP., bdem., bden., bdeo., bdep., bdeq., bder., bdes., bdes., bdet., bdeu., bdev., bdew., bdex., bdey., bdez., bl5i., bh4.., bh5y., bh56., bh41., bh4x., bh5z., bh63., bh55., bh54., bh6B., bh1y., bh65., bh4z., bh14., bh4D., bh4A., bh68., bh4v., bh4B., bh6F., bh61., bh69., bh46., bh21., bh6E., bh4y., bh45., bh47., bh6H., bh42., bh6y., bh1z., bh5.., bh6C., bh6D., bh57., bh4C., bh4A., bh6G., bh66., bh1.., bh4w., bh53., bh4., bh52., bh43., bh5x., bh6z., bh13., bh51., bh64., bh67., bh49., bh48., bh11., bh2y., bh6.., bh12., bh62., bf39., bf26., bf1w., bf1x., bf35., bf3b., bf42., bf44., bf3a., bf4.., bf2.., bf2d., bf22., bf2v., bf1z., bf2., bf27., bf2j., bf23., bf24., bf12., bf3d., bf25., bf3c., bf31., bf43., bf36., bf2e., bf33., bf2b., bf2g., bf21., bf11., bf2h., bf2z., bf32., bf34., bf41., bf29., bf2a., bf2f., bf45., bf37., bf38., bf46., be3x., be3., be3.., be3z., be3y., be1.., be2x., be22., be32., be21., be31., be2.., be33.,

Supplemental Table 3-1. Standardised incidence rate of liver disease by data sources (2004 to 2022)

Year	PEDW only group		WLGP only group		ADDE only group		Two or more data sources	
	STD incidence	STD 95% CI	STD incidence	STD 95% CI	STD incidence	STD 95% CI	STD incidence	STD 95% CI
2004	41.2	(38.9 to 43.5)	13.5	(12.2 to 14.9)	3.8	(3.1 to 4.5)	39.2	(37.1 to 41.5)
2005	49.6	(47.2 to 52.1)	15.3	(13.9 to 16.8)	3.8	(3.1 to 4.5)	40.7	(38.5 to 43)
2006	54.4	(51.9 to 57.1)	16.1	(14.7 to 17.6)	3.7	(3.1 to 4.5)	44.4	(42.1 to 46.9)
2007	54.8	(52.2 to 57.5)	13.9	(12.6 to 15.3)	4.5	(3.7 to 5.3)	41.9	(39.6 to 44.3)
2008	59.3	(56.6 to 62.1)	18.0	(16.5 to 19.6)	5.5	(4.7 to 6.4)	43.5	(41.2 to 45.9)
2009	63.9	(61.1 to 66.9)	18.5	(17 to 20.1)	5.5	(4.7 to 6.5)	43.7	(41.4 to 46.2)
2010	67.4	(64.5 to 70.4)	21.7	(20.1 to 23.4)	5.6	(4.7 to 6.5)	45.9	(43.5 to 48.4)
2011	74.1	(71.1 to 77.3)	23.6	(21.8 to 25.4)	5.1	(4.3 to 6)	47.1	(44.7 to 49.6)
2012	76.3	(73.2 to 79.5)	21.5	(19.9 to 23.2)	5.5	(4.7 to 6.4)	47.3	(44.9 to 49.9)
2013	82.7	(79.4 to 86)	22.3	(20.6 to 24.1)	6.0	(5.1 to 6.9)	47.1	(44.6 to 49.6)
2014	90.3	(86.9 to 93.8)	26.2	(24.4 to 28.1)	6.1	(5.3 to 7.1)	49.8	(47.3 to 52.4)
2015	99.1	(95.5 to 102.8)	32.2	(30.2 to 34.3)	6.2	(5.3 to 7.2)	54.7	(52.1 to 57.5)
2016	115.9	(112 to 120)	45.1	(42.7 to 47.6)	6.4	(5.5 to 7.4)	58.3	(55.5 to 61.1)
2017	122.9	(118.8 to 127.2)	48.9	(46.4 to 51.5)	6.5	(5.5 to 7.5)	64.9	(62 to 67.9)
2018	141.0	(136.5 to 145.5)	62.5	(59.7 to 65.5)	7.1	(6.1 to 8.2)	68.5	(65.5 to 71.7)
2019	157.6	(152.9 to 162.4)	71.4	(68.4 to 74.6)	6.4	(5.4 to 7.4)	65.8	(62.8 to 68.9)
2020	153.3	(148.5 to 158.1)	50.2	(47.6 to 52.8)	8.7	(7.6 to 9.8)	51.9	(49.2 to 54.7)
2021	197.5	(192.1 to 203)	73.4	(70.2 to 76.6)	7.5	(6.5 to 8.6)	53.5	(50.8 to 56.4)
2022	182.1	(176.9 to 187.4)	88.6	(85.1 to 92.1)	7.9	(6.9 to 9.1)	37.6	(35.2 to 40.1)

Abbreviation: STD: standardized; CI: confidence interval

Supplemental Table 3-2 Standardised incidence rate of liver disease by disease stages (2004 to 2022)

Year	Stage 1		Stage 2		Stage 3		Stage 4		Stage 5	
	STD incidence	STD 95% CI	STD incidence	STD 95% CI	STD incidence	STD 95% CI	STD incidence	STD 95% CI	STD incidence	STD 95% CI
2004	70.3	(67.4 to 73.3)	5.8	(5 to 6.7)	5.1	(4.3 to 6)	4.9	(4.2 to 5.8)	2.7	(2.1 to 3.3)
2005	79.6	(76.4 to 82.8)	6.8	(5.9 to 7.8)	5.2	(4.4 to 6)	5.3	(4.5 to 6.2)	2.9	(2.3 to 3.6)
2006	86.9	(83.6 to 90.3)	7.1	(6.2 to 8.1)	6.3	(5.5 to 7.3)	6.2	(5.3 to 7.2)	3.5	(2.9 to 4.2)
2007	84.0	(80.8 to 87.4)	7.5	(6.5 to 8.5)	5.7	(4.9 to 6.6)	6.4	(5.5 to 7.3)	3.5	(2.9 to 4.2)
2008	92.3	(88.9 to 95.9)	8.1	(7.1 to 9.2)	8.2	(7.3 to 9.3)	5.7	(4.9 to 6.6)	4.4	(3.7 to 5.2)
2009	97.8	(94.3 to 101.4)	8.6	(7.6 to 9.7)	7.6	(6.6 to 8.6)	6.4	(5.5 to 7.4)	4.1	(3.4 to 4.9)
2010	106.0	(102.3 to 109.7)	8.1	(7.1 to 9.1)	8.3	(7.3 to 9.4)	6.5	(5.6 to 7.5)	3.6	(2.9 to 4.3)
2011	111.6	(107.8 to 115.5)	10.1	(9 to 11.3)	8.4	(7.4 to 9.5)	6.6	(5.7 to 7.6)	4.6	(3.9 to 5.5)
2012	112.3	(108.5 to 116.1)	9.5	(8.4 to 10.7)	8.2	(7.2 to 9.3)	6.6	(5.7 to 7.6)	4.7	(4 to 5.6)
2013	115.4	(111.6 to 119.4)	10.3	(9.1 to 11.5)	8.9	(7.9 to 10.1)	7.4	(6.4 to 8.5)	6.0	(5.2 to 7)
2014	127.9	(123.9 to 132.1)	11.6	(10.4 to 12.9)	9.4	(8.4 to 10.6)	7.1	(6.2 to 8.2)	5.3	(4.5 to 6.2)
2015	141.9	(137.6 to 146.3)	14.0	(12.7 to 15.5)	12.5	(11.2 to 13.8)	6.5	(5.6 to 7.5)	6.3	(5.4 to 7.3)
2016	170.5	(165.7 to 175.3)	16.2	(14.7 to 17.7)	13.2	(11.9 to 14.7)	8.1	(7.1 to 9.2)	6.4	(5.5 to 7.5)
2017	185.5	(180.5 to 190.6)	17.8	(16.2 to 19.4)	13.5	(12.2 to 14.9)	8.0	(7 to 9.2)	6.7	(5.7 to 7.8)
2018	212.9	(207.5 to 218.4)	21.6	(19.9 to 23.4)	14.7	(13.3 to 16.2)	9.2	(8.1 to 10.3)	7.3	(6.3 to 8.5)
2019	231.8	(226.2 to 237.6)	22.1	(20.4 to 24)	17.8	(16.2 to 19.5)	8.9	(7.8 to 10.1)	6.6	(5.7 to 7.6)
2020	201.7	(196.3 to 207.1)	19.5	(17.8 to 21.2)	14.3	(12.9 to 15.8)	8.9	(7.8 to 10.2)	7.7	(6.6 to 8.9)
2021	263.0	(256.9 to 269.3)	21.6	(19.9 to 23.5)	18.3	(16.7 to 20.1)	8.2	(7.2 to 9.4)	7.9	(6.8 to 9.1)
2022	244.7	(238.7 to 250.7)	23.5	(21.6 to 25.5)	19.9	(18.2 to 21.8)	7.5	(6.4 to 8.8)	8.7	(7.6 to 10)

Abbreviation: STD: standardized; CI: confidence interval

Supplemental Table 3-3, Standardised incidence rate of liver disease by aetiologies (2004 to 2022)

Year	ArLD	NAFLD	Metabolic liver disease	HBV	HCV	Autoimmune liver disease	Haemochromatosis	ArLD overlap	Non-ArLD overlap	Hepatitis of unspecified origin	Congestive hepatopathy	Toxic liver disease	Miscellaneous
	STD incidence (95% CI)	STD incidence (95% CI)	STD incidence (95% CI)	STD incidence (95% CI)	STD incidence (95% CI)	STD incidence (95% CI)	STD incidence (95% CI)	STD incidence (95% CI)	STD incidence (95% CI)	STD incidence (95% CI)	STD incidence (95% CI)	STD incidence (95% CI)	STD incidence (95% CI)
2004	22.0 (20.3 to 23.7)	10.8 (9.6 to 12)	4.4 (3.7 to 5.2)	0.6 (0.4 to 0.9)	3.0 (2.4 to 3.7)	14.4 (13.2 to 15.8)	3.2 (2.6 to 3.9)	0.3 (0.2 to 0.6)	0.4 (0.2 to 0.7)	3.6 (2.2 to 5.1)	0.6 (0.4 to 0.9)	1.8 (1.3 to 2.3)	5.3 (4.5 to 6.1)
2005	21.7 (20.1 to 23.5)	14.0 (12.7 to 15.4)	4.7 (3.9 to 5.5)	0.6 (0.3 to 1)	2.9 (2.3 to 3.5)	18.4 (16.9 to 19.9)	4.0 (3.3 to 4.8)	0.4 (0.2 to 0.7)	0.4 (0.2 to 0.7)	4.0 (2.6 to 5.6)	0.5 (0.3 to 0.8)	2.3 (1.8 to 2.9)	5.8 (5 to 6.7)
2006	21.9 (20.2 to 23.6)	17.9 (16.4 to 19.5)	5.1 (4.4 to 6)	0.7 (0.4 to 1.1)	3.9 (3.2 to 4.7)	17.7 (16.2 to 19.2)	5.1 (4.3 to 6)	--	0.8 (0.5 to 1.2)	4.4 (3 to 5.6)	0.6 (0.3 to 0.9)	1.9 (1.4 to 2.5)	6.3 (5.4 to 7.2)
2007	22.3 (20.6 to 24)	17.1 (15.7 to 18.6)	5.1 (4.4 to 6)	0.6 (0.4 to 1)	2.8 (2.2 to 3.5)	17.1 (15.7 to 18.6)	4.3 (3.6 to 5.1)	--	0.8 (0.5 to 1.2)	4.9 (4.2 to 5.8)	0.6 (0.3 to 0.9)	1.6 (1.2 to 2.1)	6.5 (5.6 to 7.5)
2008	20.2 (18.6 to 21.9)	21.2 (19.6 to 22.9)	4.8 (4 to 5.6)	0.8 (0.5 to 1.2)	3.8 (3.1 to 4.6)	20.1 (18.6 to 21.8)	4.8 (4 to 5.6)	--	0.6 (0.4 to 1)	6.3 (5.4 to 7.2)	0.8 (0.5 to 1.2)	1.2 (0.9 to 1.7)	7.4 (6.5 to 8.4)
2009	21.1 (19.5 to 22.8)	23.3 (21.6 to 25.1)	5.5 (4.7 to 6.4)	0.7 (0.4 to 1.1)	3.2 (2.6 to 4)	23.8 (22.1 to 25.6)	5.6 (4.8 to 6.6)	0.5 (0.2 to 0.8)	0.5 (0.3 to 0.9)	4.5 (3.8 to 5.3)	0.6 (0.3 to 1)	1.0 (0.7 to 1.4)	7.5 (6.6 to 8.6)
2010	20.9 (19.2 to 22.6)	28.8 (26.9 to 30.8)	5.2 (4.4 to 6)	0.8 (0.5 to 1.1)	3.4 (2.7 to 4.1)	25.9 (24.1 to 27.7)	5.7 (4.9 to 6.6)	0.3 (0.2 to 0.6)	0.5 (0.3 to 0.8)	5.3 (4.5 to 6.2)	0.7 (0.4 to 1.1)	0.9 (0.6 to 1.3)	7.8 (6.8 to 8.8)
2011	18.8 (17.3 to 20.4)	32.6 (30.6 to 34.7)	6.4 (5.5 to 7.4)	1.3 (0.9 to 1.8)	3.6 (3 to 4.4)	24.4 (22.7 to 26.3)	5.8 (5 to 6.7)	0.4 (0.2 to 0.7)	0.6 (0.4 to 1)	5.9 (5 to 6.8)	0.4 (0.2 to 0.7)	1.5 (1.1 to 2)	9.8 (8.7 to 11)
2012	17.3 (15.9 to 18.9)	32.6 (30.6 to 34.7)	8.1 (7.1 to 9.2)	0.8 (0.5 to 1.2)	4.2 (3.5 to 5.1)	23.0 (21.3 to 24.8)	6.3 (5.5 to 7.3)	0.4 (0.2 to 0.7)	0.6 (0.4 to 1)	6.2 (5.3 to 7.1)	0.8 (0.5 to 1.2)	1.4 (1 to 1.9)	10.3 (9.2 to 11.6)
2013	17.8 (16.3 to 19.4)	36.0 (33.9 to 38.2)	8.6 (7.5 to 9.7)	1.0 (0.7 to 1.5)	4.7 (4 to 5.6)	18.4 (16.9 to 20)	5.8 (5 to 6.7)	0.4 (0.2 to 0.7)	0.6 (0.4 to 1)	8.3 (7.3 to 9.4)	0.8 (0.5 to 1.2)	1.1 (0.7 to 1.5)	11.9 (10.7 to 13.2)
2014	17.3 (15.9 to 18.9)	45.2 (42.8 to 47.7)	8.4 (7.4 to 9.5)	0.7 (0.4 to 1.1)	5.7 (4.9 to 6.6)	17.8 (16.3 to 19.3)	6.5 (5.6 to 7.4)	0.3 (0.2 to 0.6)	0.8 (0.5 to 1.1)	7.8 (6.8 to 8.9)	0.9 (0.6 to 1.3)	1.0 (0.6 to 1.4)	15.6 (14.2 to 17.1)

2015	17.8 (16.3 to 19.4)	54.7 (52.1 to 57.5)	9.0 (7.9 to 10.2)	0.9 (0.6 to 1.4)	5.2 (4.4 to 6.1)	19.8 (18.2 to 21.5)	6.0 (5.2 to 7)	0.4 (0.2 to 0.7)	0.5 (0.2 to 0.8)	8.4 (7.4 to 9.6)	1.0 (0.6 to 1.4)	1.1 (0.7 to 1.6)	17.0 (15.6 to 18.6)
2016	18.7 (17.2 to 20.3)	69.5 (66.6 to 72.6)	12.6 (11.3 to 14.1)	1.9 (1.4 to 2.5)	6.9 (6 to 8)	19.8 (18.2 to 21.6)	7.6 (6.6 to 8.7)	0.6 (0.4 to 1)	1.0 (0.7 to 1.4)	10.2 (8.8 to 11.3)	1.2 (0.8 to 1.7)	1.2 (0.8 to 1.6)	19.4 (17.7 to 21.1)
2017	20.3 (18.6 to 22)	81 (77.7 to 84.3)	11.8 (10.5 to 13.2)	2.6 (2 to 3.2)	7.0 (6.1 to 8.1)	18.5 (16.9 to 20.2)	7.6 (6.6 to 8.7)	0.9 (0.6 to 1.3)	1.0 (0.7 to 1.5)	11.1 (9.5 to 12.4)	1.2 (0.9 to 1.7)	1.3 (0.9 to 1.7)	21.1 (19.4 to 22.9)
2018	19.8 (18.2 to 21.5)	100.5 (96.9 to 104.3)	12.6 (11.3 to 14.1)	2.0 (1.5 to 2.5)	5.9 (5 to 6.9)	20.3 (18.6 to 22.1)	8.5 (7.4 to 9.6)	0.8 (0.5 to 1.2)	1.0 (0.7 to 1.4)	12.1 (10.5 to 13.5)	1.4 (0.9 to 1.9)	1.7 (1.2 to 2.2)	26.1 (24.2 to 28.1)
2019	21.7 (20 to 23.5)	114.9 (110.9 to 118.9)	13.0 (11.6 to 14.5)	2.5 (2 to 3.2)	5.6 (4.8 to 6.6)	20.3 (18.6 to 22.1)	8.9 (7.8 to 10.1)	0.8 (0.5 to 1.2)	0.8 (0.5 to 1.2)	12.3 (10.7 to 13.7)	1.5 (1 to 2)	1.2 (0.8 to 1.6)	27.9 (25.9 to 30)
2020	19.9 (18.3 to 21.6)	90.1 (86.6 to 93.7)	11.5 (10.2 to 12.9)	1.6 (1.2 to 2.2)	5.1 (4.3 to 6)	16.2 (14.7 to 17.9)	7.8 (6.8 to 8.9)	0.6 (0.4 to 1)	0.9 (0.5 to 1.3)	13.3 (11.7 to 14.7)	1.6 (1.2 to 2.2)	1.8 (1.3 to 2.4)	31.6 (29.4 to 33.9)
2021	22.2 (20.4 to 24)	129.6 (125.4 to 133.9)	14.1 (12.7 to 15.7)	1.7 (1.3 to 2.3)	3.6 (2.9 to 4.3)	19.4 (17.7 to 21.2)	8.4 (7.4 to 9.6)	0.5 (0.2 to 0.8)	1.4 (1 to 1.9)	16.2 (14.3 to 17.8)	1.4 (0.9 to 2)	1.5 (1.1 to 2.1)	42.9 (40.4 to 45.5)
2022	22.4 (20.7 to 24.2)	124.5 (120.4 to 128.8)	15.2 (13.7 to 16.9)	1.2 (0.9 to 1.7)	2.3 (1.8 to 3)	17.0 (15.5 to 18.7)	9.2 (8 to 10.4)	0.4 (0.2 to 0.8)	1.3 (0.9 to 1.8)	14.4 (13 to 16.3)	1.7 (1.2 to 2.4)	1.2 (0.8 to 1.7)	33.5 (31.2 to 35.9)

Abbreviation: ArLD: alcohol-related liver disease; NAFLD: non-alcohol fatty liver disease; HBV: hepatitis B virus; HCV: hepatitis C virus; STD: standardised; CI: confidence interval

Supplemental Table 4-1, Standardised incidence rate of NAFLD by data sources (2004 to 2022)

Year	PEDW-only		WLGP-only		ADDE-only		Two or more data sources	
	STD incidence	STD 95% CI	STD incidence	STD 95% CI	STD incidence	STD 95% CI	STD incidence	STD 95% CI
2004	3.4	(2.8 to 4.1)	4.2	(3.5 to 5)	0.5	(0.3 to 0.8)	2.7	(2.1 to 3.3)
2005	4.6	(3.9 to 5.4)	5.2	(4.4 to 6.1)	0.3	(0.2 to 0.6)	3.9	(3.2 to 4.6)
2006	6.0	(5.2 to 7)	7.2	(6.3 to 8.3)	--	--	4.4	(3.7 to 5.3)
2007	6.1	(5.2 to 7)	6.7	(5.8 to 7.7)	0.3	(0.2 to 0.6)	4.0	(3.3 to 4.8)
2008	6.0	(5.1 to 6.9)	10.4	(9.2 to 11.6)	0.6	(0.4 to 0.9)	4.2	(3.5 to 5)
2009	6.6	(5.7 to 7.6)	10.7	(9.6 to 12)	0.6	(0.3 to 0.9)	5.4	(4.6 to 6.3)
2010	6.8	(5.9 to 7.8)	14.9	(13.5 to 16.3)	0.7	(0.4 to 1.1)	6.4	(5.6 to 7.4)
2011	9.8	(8.7 to 11)	15.4	(14 to 16.9)	0.4	(0.2 to 0.7)	7.0	(6.1 to 8)
2012	11.6	(10.4 to 12.8)	13.8	(12.5 to 15.2)	0.4	(0.2 to 0.7)	6.9	(6 to 7.9)
2013	13.0	(11.8 to 14.4)	15.2	(13.8 to 16.7)	1.0	(0.7 to 1.4)	6.7	(5.8 to 7.7)
2014	17.6	(16.2 to 19.2)	18.2	(16.7 to 19.8)	1.1	(0.7 to 1.5)	8.2	(7.2 to 9.3)
2015	21.0	(19.4 to 22.8)	23.3	(21.6 to 25.1)	1.0	(0.7 to 1.4)	9.4	(8.3 to 10.6)
2016	24.2	(22.4 to 26)	32.5	(30.5 to 34.7)	1.2	(0.8 to 1.6)	11.7	(10.5 to 13)
2017	29.7	(27.7 to 31.7)	35.7	(33.6 to 38)	1.3	(0.9 to 1.8)	14.2	(12.9 to 15.7)
2018	38.9	(36.7 to 41.3)	45.3	(42.8 to 47.8)	1.6	(1.1 to 2.1)	14.7	(13.4 to 16.2)
2019	45.1	(42.7 to 47.7)	53.4	(50.8 to 56.2)	1.6	(1.2 to 2.1)	14.7	(13.3 to 16.2)
2020	45.0	(42.5 to 47.6)	34.0	(31.9 to 36.2)	1.9	(1.4 to 2.5)	9.1	(8.1 to 10.3)
2021	65.2	(62.2 to 68.3)	52.9	(50.2 to 55.6)	1.5	(1.1 to 2.1)	10.0	(8.8 to 11.2)
2022	60.5	(57.6 to 63.5)	58.7	(55.9 to 61.6)	1.5	(1.1 to 2)	3.9	(3.2 to 4.7)

Abbreviation: STD: standardized; CI: confidence interval

Supplemental Table 4-2, Standardised incidence rate of ArLD by data sources (2004 to 2022)

Year	PEDW-only		WLGP-only		ADDE-only		Two or more data sources	
	STD incidence	STD 95% CI	STD incidence	STD 95% CI	STD incidence	STD 95% CI	STD incidence	STD 95% CI
2004	4.3	(3.6 to 5.1)	3.0	(2.4 to 3.7)	--	--	14.3	(13 to 15.8)
2005	4.9	(4.2 to 5.8)	2.7	(2.2 to 3.4)	0.6	(0.3 to 1)	13.4	(12.1 to 14.8)
2006	5.0	(4.3 to 5.9)	2.7	(2.1 to 3.3)	0.3	(0.2 to 0.6)	13.8	(12.5 to 15.2)
2007	5.4	(4.6 to 6.3)	2.3	(1.8 to 2.9)	0.6	(0.4 to 1)	14.0	(12.7 to 15.4)
2008	4.8	(4.1 to 5.7)	2.0	(1.5 to 2.6)	0.7	(0.4 to 1.1)	12.7	(11.4 to 14)
2009	5.7	(4.8 to 6.6)	2.1	(1.6 to 2.6)	0.9	(0.6 to 1.3)	12.5	(11.2 to 13.8)
2010	5.1	(4.3 to 6)	2.5	(1.9 to 3.1)	0.9	(0.6 to 1.3)	12.5	(11.2 to 13.8)
2011	4.5	(3.8 to 5.3)	2.5	(1.9 to 3.1)	0.7	(0.4 to 1.1)	11.1	(9.9 to 12.4)
2012	4.7	(3.9 to 5.5)	1.8	(1.4 to 2.4)	0.7	(0.4 to 1.1)	10.1	(9 to 11.3)
2013	4.8	(4 to 5.6)	2.0	(1.5 to 2.6)	0.6	(0.4 to 1)	10.4	(9.2 to 11.6)
2014	4.7	(3.9 to 5.5)	2.2	(1.7 to 2.8)	0.6	(0.4 to 1)	9.9	(8.8 to 11.1)
2015	4.4	(3.7 to 5.3)	2.3	(1.8 to 3)	0.7	(0.4 to 1.1)	10.4	(9.2 to 11.6)
2016	4.7	(3.9 to 5.5)	3.9	(3.2 to 4.7)	0.9	(0.6 to 1.3)	9.3	(8.2 to 10.4)
2017	4.3	(3.6 to 5.1)	4.5	(3.7 to 5.3)	0.6	(0.4 to 1)	10.9	(9.7 to 12.1)
2018	4.3	(3.5 to 5.1)	5.2	(4.4 to 6.1)	0.7	(0.4 to 1)	9.7	(8.6 to 10.9)
2019	4.8	(4.1 to 5.7)	5.6	(4.8 to 6.6)	1.0	(0.6 to 1.4)	10.3	(9.2 to 11.6)
2020	5.4	(4.6 to 6.4)	5.0	(4.2 to 5.9)	1.1	(0.7 to 1.5)	8.4	(7.3 to 9.5)
2021	4.8	(4 to 5.7)	7.5	(6.5 to 8.6)	1.3	(0.9 to 1.8)	8.5	(7.5 to 9.7)
2022	6.1	(5.2 to 7.1)	9.7	(8.5 to 10.9)	1.1	(0.8 to 1.6)	5.5	(4.7 to 6.5)

Abbreviation: STD: standardized; CI: confidence interval

Supplemental Table 5-1 Comorbidities associated with liver disease by stages (2004 to 2022)

Comorbidities	Stage 1, N = 94,529[1]	Stage 2, N = 4,562[1]	Stage 3, N = 3,040[1]	Stage 4, N = 3,221[1]	Stage 5, N = 2,233[1]
CVD related conditions	7,018(7.4%)	587(12.9%)	310(10.2%)	457(14.2%)	264(11.8%)
Diabetes	6,135(6.5%)	612(13.4%)	303(10.0%)	234(7.3%)	200(9.0%)
Hypertension/anti-hypertensive	34,022(36.0%)	1,939(42.5%)	1,165(38.3%)	1,232(38.2%)	1,006(45.1%)

[1] n(%)

Supplemental Table 5-2 Comorbidities associated with liver disease by aetiologies (2004-2022)

Comorbidities	ArLD, N = 19,760[1]	NAFLD, N = 33,655[1]	Metablic liver disease, N = 5,469[1]	HBV, N = 1,063[1]	HCV, N = 3,539[1]	Autoimmune liver disease, N = 13,582[1]	Haemochromatosis, N = 4,111[1]	Hepatitis not specified, N = 4,283[1]	Congestive hepatopathy, N = 574[1]	Toxic liver disease, N = 757[1]	Miscellaneous, N = 8,426[1]
CVD related conditions	1,442(7.3%)	1,837(5.5%)	578(10.9%)	54(5.2%)	177(5.5%)	1,159(8.6%)	227(5.6%)	668(15.7%)	135(23.6%)	58(7.7%)	885(10.5%)
Diabetes	1,185(6.0%)	2,776(8.4%)	380(7.2%)	46(4.4%)	132(4.1%)	752(5.6%)	175(4.3%)	360(8.4%)	42(7.3%)	26(3.4%)	253(3.0%)
Hypertension/antihypertensives	6,911(35.0%)	12,455(37.5%)	1,916(36.1%)	197(19.1%)	632(19.5%)	4,670(34.8%)	1,422(35.2%)	1,678(39.2%)	273(47.6%)	203(26.9%)	3,468(41.2%)

[1] n(%)