BMC Medicine

Real world data reveal a diagnostic gap in non-alcoholic fatty liver disease

--Manuscript Draft--

Manuscript Number:	BMED-D-18-00536R3
Full Title:	Real world data reveal a diagnostic gap in non-alcoholic fatty liver disease
Article Type:	Research article
Section/Category:	Endocrinology and Metabolism
Funding Information:	FP7 Ideas: European Research Council (115372) Not applicable
Abstract:	Background: Non-alcoholic fatty liver disease (NAFLD) is the most common cause of liver disease worldwide. It affects an estimated 20% of the general population, based on cohort studies of varying size and heterogeneous selection. However, the prevalence and incidence of recorded NAFLD diagnoses in unselected 'real world' healthcare records is unknown. We harmonised health records from four major European territories and assessed age and sex-specific point prevalence and incidence of NAFLD over the past decade. Methods: Data were extracted from The Health Improvement Network (UK), Health Search Database (Italy), Information System for Research in Primary Care (Spain) and Integrated Primary Care Information (Netherlands). Each database uses a different coding system. Prevalence and incidence estimates were pooled across databases by random-effect meta-analysis after log-transformation. Results: Data were available for 17,669,973 adults. 176,114 had a recorded diagnosis of NAFLD. Pooled prevalence trebled from 0.60% in 2007 (95% confidence interval: 0.41-0.79) to 1.85% (0.91-2.79) in 2014. Incidence doubled from 1.32 (0.83-1.82) to 2.35 (1.29-3.40) per 1,000 person years. The Fib-4 non-invasive estimate of liver fibrosis could be calculated in 40.6% of patients, of whom 29.6-35.7% had indeterminate or high-risk scores. Conclusion: In the largest primary care record study of its kind to date, rates of recorded NAFLD are much lower than expected suggesting under-diagnosis and under-recording. Despite this, we have identified rising incidence and prevalence of the diagnosis. Improved recognition of NAFLD may identify people who will benefit from risk factor modification or emerging therapies to prevent progression to cardiometabolic and hepatic complications.
Corresponding Author:	William Alazawi UNITED KINGDOM
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	
Corresponding Author's Secondary Institution:	
First Author:	Myriam Alexander
First Author Secondary Information:	
Order of Authors:	Myriam Alexander
	Katrina Loomis
	Jolyon Fairburn-Beech
	Johan van der Lei
	Talita Duarte-Salles
	Daniel Prieto-Alhambra
	David Ansell

	Alessandro Pasqua
	Francesco Lapi
	Peter Rijnbeek
	Mees Mosseveld
	Paul Avillach
	Peter Egger
	Stuart Kendrick
	Dawn Waterworth
	Naveed Sattar
	William Alazawi
Order of Authors Secondary Information:	
Response to Reviewers:	Thank you for the positive reviews and for the opportunity to make the necessary changes. We have included a Declarations section at the end of the manuscript as requested.

Real world data reveal a diagnostic gap in non-alcoholic fatty liver disease

Myriam Alexander[1], A. Katrina Loomis[2], Jolyon Fairburn-Beech[1], Johan van der Lei[3], Talita Duarte-Salles[4], Daniel Prieto-Alhambra[5], David Ansell[6], Alessandro Pasqua[7], Francesco Lapi[7], Peter Rijnbeek[3], Mees Mosseveld[3], Paul Avillach[8], Peter Egger[1], Stuart Kendrick[1], Dawn M. Waterworth[1], Naveed Sattar¶, William Alazawi*¶

*Correspondence to:

Dr William Alazawi

Barts Liver Centre

Blizard Institute

Queen Mary, University of London

w.alazawi@qmul.ac.uk

¶ These authors contributed equally to the study

[1]GlaxoSmithKline, UK

[2]Worldwide Research and Development, Pfizer, USA

[3]Erasmus Universitair Medisch Centrum Rotterdam,The Netherlands

[4] Fundació Institut Universitari per a la Recerca a l'Atenció Primària de Salut Jordi Gol i Gurina, Spain.

[5]Centre for Statistics in Medicine, NDORMS, University of Oxford, UK.

[6]Quintile IMS, UK.

[7]Health Search, Italian College of General Practitioners and Primary Care, Italy.

[8] Harvard Medical School, Harvard, Boston, Massachusetts, United States.

[9]University of Glasgow, BHF Glasgow Cardiovascular Research Centre, Glasgow, UK

Keywords

Epidemiology Population NAFLD NASH

Abbreviations

ALT	Alanine transaminase
ANOVA	Analysis of variance
AST	Aspartate transaminase
BMI	Body Mass Index
CI	Confidence interval
EHR	Electronic Health Record
EMIF	European Medical Information Framework
GP	General Practitioner
HSD	Health Search Database
IPCI	Information System for Research in Primary Care
LFT	Liver function tests
NAFLD	Non-alcoholic fatty liver disease
NASH	Non-alcoholic steatohepatitis
SIDIAP	Information System for Research in Primary Care
THIN	The Health Improvement Network
UK	United Kingdom
US	United States

Abstract

Background: Non-alcoholic fatty liver disease (NAFLD) is the most common cause of liver disease worldwide. It affects an estimated 20% of the general population, based on cohort studies of varying size and heterogeneous selection. However, the prevalence and incidence of recorded NAFLD diagnoses in unselected 'real world' healthcare records is unknown. We harmonised health records from four major European territories and assessed age and sex-specific point prevalence and incidence of NAFLD over the past decade.

Methods: Data were extracted from The Health Improvement Network (UK), Health Search Database (Italy), Information System for Research in Primary Care (Spain) and Integrated Primary Care Information (Netherlands). Each database uses a different coding system. Prevalence and incidence estimates were pooled across databases by random-effect meta-analysis after log-transformation.

Results: Data were available for 17,669,973 adults. 176,114 had a recorded diagnosis of NAFLD. Pooled prevalence trebled from 0.60% in 2007 (95% confidence interval: 0.41-0.79) to 1.85% (0.91-2.79) in 2014. Incidence doubled from 1.32 (0.83-1.82) to 2.35 (1.29-3.40) per 1,000 person years. The Fib-4 non-invasive estimate of liver fibrosis could be calculated in 40.6% of patients, of whom 29.6-35.7% had indeterminate or high-risk scores.

Conclusion: In the largest primary care record study of its kind to date, rates of recorded NAFLD are much lower than expected suggesting under-diagnosis and under-recording. Despite this, we have identified rising incidence and prevalence of the diagnosis. Improved recognition of NAFLD may identify people who will benefit from risk factor modification or emerging therapies to prevent progression to cardiometabolic and hepatic complications.

Non-alcoholic fatty liver disease (NAFLD) is rapidly becoming the most common cause of chronic liver disease worldwide [1]. NAFLD is a spectrum of diseases that encompasses uncomplicated steatosis, non-alcoholic steatohepatitis (NASH) and fibrosis, which in a small proportion can lead to complications including cirrhosis, liver failure and hepatocellular carcinoma [2]. NAFLD is a multisystem disease with a multidirectional relationship with the metabolic syndrome [3-5]. NAFLD is associated with increased risk of cardiovascular disease [5-7] and cancer [8]. Among other high risk groups [9], people with diabetes and NAFLD are at increased risk of micro- and macrovascular complications [10, 11] and these patients have a 2-fold increased risk of all-cause mortality [12].

The estimated point-prevalence of NAFLD in the general Western population is 20-30%, largely based on cohort studies with heterogeneous inclusion criteria and research methods [13]. Prevalence of NAFLD rises to 40-70% among patients with type 2 diabetes and up to 90% among patients with morbid obesity [14-16]. Moreover, as the rates of diabetes and obesity rise worldwide, it is expected that NAFLD will become even more common. NAFLD-related cirrhosis is currently the third most common indication and is anticipated to become the leading indication for liver transplantation in the USA within the next one to two decades [17].

There is much debate about whether screening programmes in the general population or in at-risk groups such as people with diabetes [9] should be implemented [18, 19]. This debate is based on our current understanding of the epidemiology and natural history of NAFLD, which, in turn, derives from cohort or cross-sectional studies [13]. These are often highly selected studies of individuals with metabolic risk factors, or they involve extensive phenotyping that would be unrealistic in routine practice.

A pragmatic approach is to focus on real-world patients in whom the diagnosis of NAFLD has been made in the course of the routine clinical care. The diagnosis of NAFLD is often made following abnormal imaging of the liver or elevated serum liver enzymes (so-called liver function tests or LFTs) and involves exclusion of other causes of liver injury such as excess alcohol consumption and viral hepatitis. Although routinely collected data can only represent the visible part of the clinical iceberg, there is a growing body of literature that has used well-curated electronic healthcare records (EHR) databases to study disease characteristics and epidemiology in large numbers of people [20-22].

In many European countries where healthcare is largely state-funded and there are low or absent primary care co-payments, the population has unrestricted access to healthcare with primary care physicians acting as gatekeeper (including referral to secondary care) [23]. Healthy people register with primary care centres when they move to an area in order to access healthcare at a later date when it will be needed and so primary care EHR represent data that are as close to a ‘general’ population as possible, with near universal coverage of the population in the region the data is collected. Recording of diagnosis in European primary care databases is not driven by reimbursement and the patient population is relatively stable compared to other types of EHRs such as US claims databases. Primary care databases hold comprehensive medical records that include diagnoses, prescriptions, laboratory values, lifestyle and health measures, and demographic information for a large and representative sample of patients. Concerns around the degree of data completeness are now largely historic as the vast majority of practices are paper-free and therefore these data represent the only clinical record for care, administration and re-imbursement purposes. Thus, within the areas that utilise these databases, coverage is near-universal. If a practice joins the database, all the patients at that practice are registered in the database. Although there is an option for individual patients to opt out, this is minimal (<1%).

In this study, we harmonized healthcare records for 17.7 million adults from 4 large European primary healthcare databases to estimate the prevalence and incidence of recorded diagnoses of NAFLD in patients in primary care, and, where available, NASH, and compare with estimates from cohort studies. We sought to ascertain the changes from 2007 – 2015 in prevalence and incidence of recorded diagnoses of NAFLD, and the effect of age and sex. We compared the characteristics of patients with NAFLD diagnosis in the different databases and reported, where possible, the proportion of patients with markers of advanced disease in the diagnosed population.

Methods

Databases

Ethical approval was obtained by data custodians of each primary care database according to local institutional review board requirements. Anonymised data were extracted from the Health Search Database (HSD) in Italy [24], the Integrated Primary Care Information (IPCI) in the Netherlands[25], The Health Improvement Network (THIN) in the UK [26]; and the Information System for Research in Primary Care (SIDIAP) in the Catalonia region of Spain [27] (**Supplementary Table 1**).

THIN, HSD and IPCI all had reached high levels of patient registration from January 2004 onwards. SIDIAP started data collection in 2005 and had high quality data from 2006. Data entered between 01/01/2004 (SIDIAP from 01/01/2007 only) and up to 31/12/2015 were included in incidence estimates. Individuals were excluded if they had less than one year of follow-up post registration into the database. Individuals with a diagnosis of NAFLD were not included in the analyses if they also had a recorded history of alcohol abuse. To maximize data completeness, we only included patients whose NAFLD diagnosis occurred within ± 6 months of a GP visit when describing patients' characteristics (**Table 1 & Supplementary Table 3**).

Patient involvement

All eligible patients were included in the study. Routine healthcare records were collected from patients at each encounter with a healthcare practitioner. Following local regulations, patients who did not wish to share their data were able to withdraw from the databases.

Semantic harmonisation and case ascertainment

The four databases each use different coding systems (**Supplementary Table 1**). As a result, capture of NAFLD and NASH diagnoses differed across databases. In HSD and IPCI, NAFLD and NASH were captured in a single code as "NAFLD or NASH". In SIDIAP and THIN, NAFLD and NASH were coded separately, branching out of a "NAFLD or NASH" code. In this study, we extracted all "NAFLD or NASH" diagnoses as well as "NASH only diagnoses" where available; and for simplicity we labelled "NAFLD or NASH" as "NAFLD", and "NASH only" as "NASH". Code lists were generated in the four terminologies that all mapped to the same Unified Medical Language System concepts [28] (**Supplementary Table 2**).

Clinical diagnoses were defined using code lists which mapped to UMLS concepts using the same process of harmonization (**code lists available on request**). In SIDIAP, we used a combination of clinical codes and answers to questionnaires on alcohol consumption to identify alcohol abuse.

Given the absence of a code for NAFLD in the IPCI terminology, we additionally used text-mining in this database. The algorithm for the identification of NAFLD in IPCI database is detailed in **Supplementary Figure 1**. Patients with the following search terms were extracted: “NASH”, “NAFLD”, “steatohepatitis” or “fatty liver disease” as distinct words preceded by a space and followed by a space, or at the beginning or end of a sentence. Patients with relevant search terms preceded by a negation term (e.g. “no” or “not”) were excluded. To validate the text-mining, 100 individuals identified using free-text were randomly sampled. Their complete medical charts were manually reviewed to confirm that the clinical data support the text-mining derived diagnosis.

Use of historical data

Governance rules differed between the different databases. In HSD and SIDIAP, there were no records available prior to the primary care practice joining the database. In THIN, data from patients who had already left the practice were available; and so NAFLD/NASH diagnoses made prior to the patient’s primary care practice joining THIN were counted in both incidence and prevalence estimates. However, in IPCI records that pre-dated their primary care practice joining the database were available only for patients who remained in the practice (leavers not having the opportunity to refuse to participate). Therefore, historic diagnoses could be included in point prevalence. However, given that both the number of new diagnoses made as well as the total number of patients at risk in a given time-period were unknown we could not include diagnoses made before the patient joined a practice in incidence estimates in IPCI.

Other data extraction

Demographic information, lifestyle and medical history on relevant morbidities were also extracted for all NAFLD or NASH patients identified in the four databases. Medical records for type 2 diabetes and hypertension at any time prior to NAFLD or NASH diagnosis were extracted. Code lists for those diagnoses were harmonised across databases using the process of semantic harmonization described in the Method section which aligns all terminologies on the same list of UMLS concepts (code lists available on request).

Laboratory values for aspartate transaminase (AST), alanine transaminase (ALT) and platelet count were extracted, taking values closest to the time of NAFLD diagnosis (up to 2 years prior to diagnosis or less than 6 months after). Body Mass Index (BMI) was calculated for all NAFLD patients with weight recorded within the time frame of 2 years prior to 6 months after diagnosis; and with height recorded anytime in adulthood. We excluded values that were likely to be implausible: BMI below 15kg/m^2; laboratory values greater than the mean in the database plus 3 times the standard deviation; AST and ALT less than 5iU/L; and platelet counts below 5x10^9/L.

The Fib-4 index was calculated to provide an estimate of severity of fibrosis in patients at the time of their NAFLD diagnosis. The formula for the Fib-4 is: Age[years] x AST[U/L]/(platelet[10^9]X√ALT[U/L]) [29]. The cut-offs for Fib-4 scoring for NAFLD are <1.30: low risk of advanced fibrosis or cirrhosis; between 1.30 and 2.67: indeterminate score; 2.67: high risk of advanced fibrosis or cirrhosis [30].

Statistical methods

Quantitative variables were reported as mean and 95% confidence interval (CI) of the mean assuming a normal distribution, and qualitative variables as percentages. Differences in patients' characteristics between the 4 databases was tested by an ANOVA test for quantitative characteristics and a Chi-square test for categorical characteristics.

Incidence in the adult population aged ≥18 years old was estimated by dividing the number of individuals with a diagnosis of NAFLD (or NASH where relevant) by the total number of person-years at risk. Incidence was reported by predefined age categories, gender and calendar year.

Point-prevalence was estimated on the 1st of January of each calendar year available in the data, by gender and by predefined age categories. Point prevalence was defined as total number of individuals with a recorded NAFLD diagnosis at or prior to the 1st of January of a calendar year and who were still active in the database, divided by the total number of active patients in the database on that date.

In addition, the 1-year period prevalence was estimated in a sensitivity analysis to account for potential differences in length of follow-up across databases, and over time within databases. The 1-year period-prevalence was defined for each calendar year available as the number of new individuals with a recorded diagnosis of NAFLD in a calendar year divided by the average number of active patients in that year (defined as the number on the 1st of January plus the number on the 31st of December divided by 2).

Age was computed as age at the end of the time-period for period prevalence, i.e. on the 31st of December of that year; and on the 1st of January of the year of interest for point prevalence. Within each database, incidence estimates were compared by calendar year (assuming a linear relationship), sex (Males are reference group) and age group (age 60-69 as reference group) fitting Poisson distributions; and prevalence were compared fitting logistic regressions and performing chi-square tests. P-value <0.001 were considered as significant, although it is important to bear in mind that in such large datasets high level of significance can be achieved even for minimal absolute differences in prevalence and incidence levels.

Incidence and prevalence estimates were pooled for each calendar year across the four databases using random effects meta-analysis after natural log-transformation (weighting by the inverse of the variance). We reported the I^2 statistic which gives the percentage of variation among databases attributable to heterogeneity, and the p-values of heterogeneity (p-het) tested using Q-statistics. To investigate sources of heterogeneity, we tested for a linear association between incidence and point-prevalence with calendar year by fitting a meta-regression.

Data were extracted and analysed using the European Medical Information Framework (EMIF) Platform of a distributed network approach that allows data custodians to maintain control over their protected data [31]. Each data custodian extracted data from their database into four common files: a prescriptions, measurements, events, and patients file. These files were transformed locally by the data transformation tool called Jerboa Reloaded which produces analytical datasets that can be shared with the data analysts for further post-processing in a central remote research environment. The analytical datasets contained characteristics for each patient with a NAFLD diagnosis, as well as aggregated results on incidence and prevalence by age, gender and calendar year. Quality controls were run on each database and the research team communicated with data custodians to confirm results.

Statistics and graphics were generated in the remote research environment using the statistical software Stata/SE 14.1.

Results

Semantic harmonisation to identify European NAFLD cohort

In total, the four European databases held data on 21,981,019 patients, of whom 17,699,973 adults had been registered for at least one year in adulthood (**Table 1**). Using semantic harmonisation, we identified 176,114 patients who had a recorded diagnosis of NAFLD (including NASH). This represents 1.0% of the total population, ranging from 0.3% in the UK (THIN) to 2.7% in The Netherlands (IPCI). The largest number of NAFLD patients was in the Spanish cohort (SIDIAP, n=77,547, **Table 1**). Recording of NASH diagnoses was only possible in Spain (SIDIAP, n=1,887) and the UK (THIN, n=1,133) as the other two databases did not contain specific codes distinguishing NAFLD from NASH. Given the small numbers overall, we did not pursue analysis of NASH incidence and prevalence further and we included these within the total number of patients with a recorded diagnosis of NAFLD.

In the Dutch database (IPCI), the majority of patients were identified via free-text mining with seed words “NAFLD”, “NASH”, “fatty liver” or “steatosis”, and a minority from diagnostic codes only (see **Supplementary Figure 1**). The code for “Liver steatosis” (D97.05), identified 1,282 patients; and the code for “Cirrhosis/Other Liver Disease” (D97.00) identified 4,228 patients when combined with a free-text search on the code label; and 1,214 additional patients when combined with a free-text anywhere in the medical records. Searching for free-text in the absence of a relevant code identified 44,442 additional patients. Of these, 19,048 patients had an incident NAFLD diagnosis (recorded at a time when the patient’s GP practice was contributing to IPCI). In the sample of 100 cases that were manually reviewed, the positive predictive value for a text-mined diagnosis of NAFLD was 98%.

We only identified a small proportion of patients with a recorded diagnosis of NAFLD who also drank alcohol in excess of recommended limits; 3,130 (7.0%) NAFLD patients in IPCI, 921 in HSD (3.3%), 12,461 in SIDIAP (14.1%), and 925 in THIN (3.8%). These patients were excluded from the statistical analysis.

The characteristics of the populations of patients with an incident diagnosis of NAFLD, after exclusions, made during the study period in individual databases are shown in **Table 2**. There were minor differences in the mean age, proportion of patients with impaired fasting glucose or diabetes and platelet count for patients in each of the four databases. However, we observed that patients in HSD had statistically significantly higher proportions of males and patients with hypertension than other databases. There was considerable variation in recorded BMI (29.7kg/m^2 in HSD to 32.4kg/m^2 in THIN), in alanine transaminase (ALT) levels (median 28iU/l in HSD to 39iU/l in THIN) and aspartate transaminase (AST) levels (median 24iU/l in HSD to 32iU/l in THIN). Moreover, we observed variation in clinical practice with higher rates of BMI recording and ALT requesting in THIN and SIDIAP compared to IPCI and HSD (**Table 2** and **Supplementary Table 3**).

Non-invasive scores that estimate the degree of liver fibrosis can be calculated from clinical parameters and are used to risk-stratify patients with NAFLD. Although both ALT and AST are required to calculate the majority of such non-invasive scores, ALT was more frequently available than AST in all four databases (**Supplementary Table 3**). An AST result was available in 21% (THIN) to 68% (HSD) and an ALT result in 67% (IPCI) to 86% (SIDIAP). This is reflected in the proportion of patients in whom a Fib-4 non-invasive assessment of liver fibrosis could be calculated, ranging from 11% in THIN to 54% in SIDIAP. Despite having the smallest number (and percentage) of patients in whom we could calculate Fib-4, the THIN database had the highest proportion of patients with high-risk scores indicative of advanced fibrosis or even cirrhosis (10.0% vs 2.9–4.3%, p<0.001). Practically, patients with indeterminate or high risk scores are often managed with further assessment leading to a liver biopsy. The proportion of patients with intermediate/high risk scores was lower in IPCI (29.8%) compared to the other databases (35.0–35.7%); albeit the number of people in whom we could calculate Fib-4 was variable.

The rising prevalence of NAFLD diagnosis

The overall (pooled) prevalence of NAFLD diagnosis was low at 1.85% (95% CI: 0.91–2.79) (I^2=99.99%, p-het<0.001) on 1st of January 2015, but it had trebled from 0.60% (0.41-0.79) (I^2=99.97%, p-het<0.001) on 1st of January 2007 (**Figure 1** and **Supplementary Table 4**).

The prevalence of recorded NAFLD diagnosis rose over time in all databases albeit levels and rates of rise differed between databases; highest in the Netherlands (IPCI) and lowest in the UK (THIN). To confirm that those trends were not due to more complete medical records being available in more recent years, we also estimated 1-year period prevalence and observed rising trends for the four databases (**Supplementary Table 5**).

There were no significant differences in prevalence between sexes in any database, but prevalence did vary by age. Peak prevalence was in patients aged 60-79 in whom it was >20 times higher than in 18-29 years old in IPCI (4.89% versus 0.24%) and 10-14 times higher in the other databases (**Figure 2 and Supplementary Table 6**).

Incidence of NAFLD has doubled since 2007

The overall (pooled) incidence of recorded NAFLD diagnoses was 2.35 (1.29 - 3.40; I^2=99.92%, p-het<0.001) per 1,000 person years in 2015, having approximately doubled since 2007 (1.32; 0.83-1.82)) (see **Figure 3 and Supplementary Table 7**).

We observed heterogeneity between databases. In IPCI and SIDIAP, there was a clear and consistent rise in incidence with a 2.7-fold increase from 2004-2015 to 4.09 per 1,000 person-years in IPCI and 3.2-fold increase from 2007-2015 to 2.61 per 1,000 person-years in SIDIAP. In HSD, there was no statistically significant change in incidence between the years 2005 and 2015 (**Supplementary Table 6**). Although the rate of rise in THIN was comparable to IPCI and SIDIAP, the very low starting rate meant that despite a 5-fold increase, the absolute increase was still modest and the incidence in 2014 was 1.08 per 1,000 person-years.

There was a significant difference between sexes in HSD and SIDIAP (p<0.05) but not in IPCI and THIN. In HSD, IPCI and SIDIAP, peak incidence was in 60-69 year olds; and in 50-59 year olds in THIN (but estimate not significantly different from that in 60-69 year olds) and then decreased in older age groups (**Figure 4**, **Supplementary Table 8**).

Discussion

In the largest real-world study of its kind to date, we report the incidence and prevalence of recorded NAFLD diagnoses among 17.7 million adults in 4 different European countries.

The databases used have been validated, are broadly representative of the population of the country; and have been extensively used for pharmaco-epidemiology research [17, 20] (**Supplementary Table 1**). Despite a rise in incidence, our study found a large shortfall in Europe between the expected number of patients with NAFLD and NASH and the number with recorded diagnoses. Although other have suggested that this might be the case at a local level or in small questionnaire-based exercises [32], this study has identified the scale of that diagnostic gap across four European territories. Under-recording of NAFLD in primary care may reflect missed opportunities to make the diagnosis by investigating abnormal liver enzyme values or imaging findings, confidence to make the diagnosis even if liver enzymes are in the reference range and under-recognition of the diagnosis when made in secondary care. Furthermore, many patients who do have the diagnosis have not had the investigations required to make appropriate risk-stratification and therefore offer specialist care to those at greatest need. The current study represents a departure from existing population-level study design in NAFLD. Notwithstanding the limitations discussed below, by using real world data, we have gained insight into current practice and attitudes to NAFLD and into the changing face of NAFLD in primary care.

We used UMLS semantic harmonisation to extract primary care EHR data and identify 176,114 patients with a recorded diagnosis of NAFLD. Despite variation in coding systems, in the characteristics of the populations and in the health care systems in each country, the results from all four territories are broadly consistent. They show rising incidence and prevalence of NAFLD, but the levels of recorded NAFLD in EHR primary care databases is many-fold lower than those anticipated based on prior observation studies, which estimated the prevalence of NAFLD in the general European population to be 20-30%[33]. The characteristics of patients in that study were comparable with those with NAFLD in a recent systematic review of the literature and meta-analysis that included 101 studies [13]. That study reported the European prevalence of NAFLD diagnosed by imaging to be 24% (95% CI: 16%-34%) and diagnosed by blood tests to be 13% (95% CI: 4%-33%). Thus, our pooled prevalence in European EHR databases of 1.9% is at best ~1/6 and more likely only ~1/12 of the estimates based on cohort data. Our estimates of incidence in 2015 ranged from 1.1 to 4.1

per 1,000 and are approximately 10 times lower than expected based on cohort studies: 28 (95% CI: 19 - 41) per 1,000 person-years in Israel and 52 (95% CI: 28-97) per 1,000 in Asia [13].

Prevalence of NAFLD diagnosis has trebled and incidence doubled over the time-period of this study. The rising rates of co-morbid conditions such as diabetes and obesity may be responsible for this. Other probable factors that play a part include increased awareness among primary care and non-liver physicians, improved communication of the diagnosis from secondary to primary care and the increased use of blood tests and imaging to investigate common complaints such as abdominal pain or monitoring long-term conditions. Our data do not allow us to test these hypotheses further, however, studies from other groups also suggest that the total number of people developing NAFLD is rising as is the number of people with NAFLD who develop life-threatening complications [13].

Despite the consistency in overall findings, the differences between the databases are indicative of differing practices. SIDIAP had a relatively large proportion of patients with a history of alcohol abuse (14.1%), although all databases included at least some NAFLD patients with recorded alcohol abuse. This reflects uncertainty in the community as to whether an individual can have fatty liver disease associated with metabolic syndrome even if they drink alcohol in excess of recommended limits, or indeed have any other cause of chronic liver injury such as viral hepatitis. While clinical trials make a very precise distinctions between alcoholic and non-alcoholic fatty liver disease, the reality is that an obese, diabetic and hypertensive patient can consume alcohol in excess of recommended limits and have liver injury. There is no way to distinguish which aetiology is the dominant cause and so clinicians are quite comfortable with co-existing diagnoses. Indeed, some authors now refer to BAFLD – both alcoholic and fatty liver disease. An alternative explanation may be that specialists making the diagnosis of fatty liver are unaware of the high alcohol use – either because of under-reporting by patients or poor communication from the GP records.

In HSD, prevalence increased over time whereas incidence decreased in recent years. This can be explained by a relatively stable population in which nearly all patients were enrolled in 2000, see Supplementary Figure 3, and remained in the data until December 2015.

Text-mining in IPCI increased the number of NAFLD diagnoses by over 8-fold. This suggests that while the diagnosis of NAFLD is being made, GPs are not recording it despite a code for liver steatosis existing in IPCI. IPCI had the lowest level of ALT recording. A recent survey of Dutch GPs explored attitudes to the 'importance' of NAFLD[34]. Only 47% of doctors used liver tests in patients with NAFLD and non-invasive scores were 'never used' by 73% of respondents (we were able to calculate Fib-4 scores in only 27% in IPCI).

The UK THIN database appears to outlie from the others in several ways. The prevalence of recorded NAFLD in THIN (0.2%) is much lower than the other databases and markedly lower than that found in a study of almost 700,000 adults in a primary care EHR study in London, UK (0.9%) [35]. Higher rates of alcohol recording in the UK alone are unlikely to account for all this difference. The median ALT was highest in THIN. This may suggest that the diagnosis of NAFLD is more likely to be arrived at in the UK by investigation of abnormal liver enzymes than in other territories. However, the data required to calculate Fib-4 were available in only 11% of patients in THIN (Supplementary Table 3). NAFLD patients in THIN had the highest mean BMI, proportion of patients with diabetes or impaired fasting glucose and the highest proportion of patients with high-risk Fib-4 scores. Large scale liver biopsy-based cross sectional studies or replication of the current study in cohorts with systematic ascertainment of the component of Fib-4 would be needed to confirm that patients are diagnosed with NAFLD at more advanced stages in the UK compared to other European countries.

Limitations of the study: When interpreting the data, it is important to consider the following issues. In IPCI, a diagnostic code for NAFLD was not available, therefore we devised an algorithm based on the diagnostic code 'liver steatosis' and excluding excess alcohol consumption. We did not do this for all databases because IPCI terminology only contains 1073 clinical terms and therefore general practitioners often utilize free-text to record information with greater precision, whereas the other coding systems contain many more such concepts: ICD9CM contains 40,855 terms, ICD10 contains 13,505 terms and Read Codes 347,568 terms [36].

The number of cases of recorded NASH is too small to make meaningful estimates of incidence and prevalence: 2-4% of patients with NAFLD in THIN and SIDIAP in which NASH was coded. This is far short of the 12.2% estimated from a US biopsy-based study[37]. This shortfall between coded NASH and the true burden of disease is probably due to the same factors that result in under-recording of NAFLD diagnosis: recognition, referral and coding in primary care, under-diagnosis or poor communication in secondary care.

It is not possible to verify the accuracy or origin of recorded diagnoses, although the characteristics of the patients derived from the four databases are in keeping with the population one would expect with a NAFLD diagnosis. Some individuals may have undiagnosed NAFLD who do not appear in this study. Therefore, our results do not represent the 'true' disease burden in the epidemiological sense, rather they tell us what is actually happening with people who currently have a diagnosis of NAFLD and can inform the arguments for or against greater action in this area. While we cannot exclude the possibility (however unlikely) that all the other millions of expected NAFLD patients exist in other databases, we are not making any conclusions about people outside this dataset. Although primary care data contain a large body of information, this does not diminish the value of well-phenotyped cohort studies in which NAFLD can be ascertained systematically using standardized screening methods (e.g. measuring liver enzymes or performing ultrasound in all patients). That said, the databases included in this study have been extensively used for research, have been validated for diagnoses other than NAFLD [24, 27, 38].

Conclusions

Clinical practice is evolving in this emerging field and as yet there are no recommendations to formally screen for NAFLD, even in high risk groups [39, 40]. One school of thought is that if the only available intervention for NAFLD or NASH is lifestyle change, then doctors are already giving such advice to their patients, although the extent to which patients take up such advice varies. However, hepatic steatosis is an independent predictor of diabetes [41, 42] and could therefore identify patients who stand to benefit from lifestyle changes to prevent diabetes and hepatic complications. Furthermore, the emerging data suggesting hepatic steatosis is an independent cardiovascular risk factor may provide additional incentive for physicians to increase their awareness of NAFLD at the early stages. At the more severe end of the scale, novel therapies targeted at NASH and fibrosis are already in phase III clinical trials and are expected to be available in the next few years, changing the treatment paradigm. Therefore, the scale of the healthcare challenge posed by NAFLD and its sequelae cannot simply be side-stepped by dismissing NAFLD as pre-disease. Further research is required to quantify the associations of NAFLD with outcomes and to determine whether Wilson’s criteria for effective screening can be fulfilled [43], thereby informing the screening debate.

Abbreviations

ALT	Alanine transaminase
ANOVA	Analysis of variance
AST	Aspartate transaminase
BMI	Body Mass Index
CI	Confidence interval
EHR	Electronic Health Record
EMIF	European Medical Information Framework
GP	General Practitioner
HSD	Health Search Database
IPCI	Information System for Research in Primary Care
LFT	Liver function tests
NAFLD	Non-alcoholic fatty liver disease
NASH	Non-alcoholic steatohepatitis
SIDIAP	Information System for Research in Primary Care
THIN	The Health Improvement Network
UK	United Kingdom
US	United States

References

1. Sattar, N., E. Forrest, and D. Preiss, *Non-alcoholic fatty liver disease.* BMJ, 2014. **349**: p. g4596.
2. Tai, F.W., W.K. Syn, and W. Alazawi, *Practical approach to non-alcoholic fatty liver disease in patients with diabetes.* Diabet Med, 2015. **32**(9): p. 1121-33.
3. Mantovani, A., et al., *Nonalcoholic Fatty Liver Disease and Risk of Incident Type 2 Diabetes: A Meta-analysis.* Diabetes Care, 2018. **41**(2): p. 372-382.
4. Mantovani, A., et al., *Nonalcoholic fatty liver disease increases risk of incident chronic kidney disease: A systematic review and meta-analysis.* Metabolism, 2018. **79**: p. 64-76.
5. Targher, G., et al., *Non-alcoholic fatty liver disease and risk of incident cardiovascular disease: A meta-analysis.* J Hepatol, 2016. **65**(3): p. 589-600.
6. Ekstedt, M., et al., *Fibrosis stage is the strongest predictor for disease-specific mortality in NAFLD after up to 33 years of follow-up.* Hepatology, 2015. **61**(5): p. 1547-54.
7. Söderberg, C., et al., *Decreased survival of subjects with elevated liver function tests during a 28-year follow-up.* Hepatology, 2010. **51**(2): p. 595-602.
8. Sanna, C., et al., *Non-Alcoholic Fatty Liver Disease and Extra-Hepatic Cancers.* Int J Mol Sci, 2016. **17**(5).
9. Lonardo, A., et al., *Epidemiological modifiers of non-alcoholic fatty liver disease: Focus on high-risk groups.* Dig Liver Dis, 2015. **47**(12): p. 997-1006.
10. Targher, G., C.P. Day, and E. Bonora, *Risk of cardiovascular disease in patients with nonalcoholic fatty liver disease.* N Engl J Med, 2010. **363**(14): p. 1341-50.
11. Targher, G., A. Lonardo, and C.D. Byrne, *Nonalcoholic fatty liver disease and chronic vascular complications of diabetes mellitus.* Nat Rev Endocrinol, 2018. **14**(2): p. 99-114.
12. Allen, A.M., et al., *Nonalcoholic Fatty Liver Disease Incidence and Impact on Metabolic Burden and Death: a 20 Year-Community Study.* Hepatology, 2017.
13. Younossi, Z.M., et al., *Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes.* Hepatology, 2016. **64**(1): p. 73-84.
14. Williamson, R.M., et al., *Prevalence of and risk factors for hepatic steatosis and nonalcoholic Fatty liver disease in people with type 2 diabetes: the Edinburgh Type 2 Diabetes Study.* Diabetes Care, 2011. **34**(5): p. 1139-1144.
15. Vernon, G., A. Baranova, and Z.M. Younossi, *Systematic review: the epidemiology and natural history of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis in adults.* Aliment. Pharmacol. Ther., 2011. **34**(3): p. 274-285.
16. Bedossa, P., et al., *Systematic review of bariatric surgery liver biopsies clarifies the natural history of liver disease in patients with severe obesity.* Gut, 2017. **66**(9): p. 1688-1696.
17. Zezos, P. and E.L. Renner, *Liver transplantation and non-alcoholic fatty liver disease.* World J Gastroenterol, 2014. **20**(42): p. 15532-8.
18. Rinella, M.E., *Screening for nonalcoholic fatty liver disease in patients with atherosclerotic coronary disease?--In principle yes, in practice not yet.* Hepatology, 2016. **63**(3): p. 688-90.

19. Wong, V.W. and N. Chalasani, *Not routine screening, but vigilance for chronic liver disease in patients with type 2 diabetes.* J Hepatol, 2016. **64**(6): p. 1211-3.

20. Booth, H., et al., *Incidence of type 2 diabetes after bariatric surgery: population-based matched cohort study.* Lancet Diabetes Endocrinol, 2014. **2**(12): p. 963-8.

21. Farmer, R.D., et al., *Population-based study of risk of venous thromboembolism associated with various oral contraceptives.* Lancet, 1997. **349**(9045): p. 83-8.

22. Hobbs, F.D.R., et al., *Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007-14.* Lancet, 2016. **387**(10035): p. 2323-2330.

23. Kringos, D., et al., *The strength of primary care in Europe: an international comparative study.* Br J Gen Pract, 2013. **63**(616): p. e742-50.

24. Gini, R., et al., *Chronic disease prevalence from Italian administrative databases in the VALORE project: a validation through comparison of population estimates with general practice databases and national survey.* BMC Public Health, 2013. **13**: p. 15.

25. Vlug, A.E., et al., *Postmarketing surveillance based on electronic patient records: the IPCI project.* Methods Inf Med, 1999. **38**(4-5): p. 339-44.

26. Blak, B.T., et al., *Generalisability of The Health Improvement Network (THIN) database: demographics, chronic disease prevalence and mortality rates.* Inform Prim Care, 2011. **19**(4): p. 251-5.

27. Garcia-Gil Mdel, M., et al., *Construction and validation of a scoring system for the selection of high-quality data in a Spanish population primary care database (SIDIAP).* Inform Prim Care, 2011. **19**(3): p. 135-45.

28. Avillach, P., et al., *Harmonization process for the identification of medical events in eight European healthcare databases: the experience from the EU-ADR project.* J Am Med Inform Assoc, 2013. **20**(1): p. 184-92.

29. Sterling, R.K., et al., *Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection.* Hepatology, 2006. **43**(6): p. 1317-1325.

30. Shah, A.G., et al., *Comparison of noninvasive markers of fibrosis in patients with nonalcoholic fatty liver disease.* Clin Gastroenterol Hepatol, 2009. **7**(10): p. 1104-12.

31. Coloma, P.M., et al., *Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR Project.* Pharmacoepidemiol Drug Saf, 2011. **20**(1): p. 1-11.

32. Nascimbeni, F., et al., *From NAFLD in clinical practice to answers from guidelines.* J Hepatol, 2013. **59**(4): p. 859-71.

33. Loomba, R. and A.J. Sanyal, *The global NAFLD epidemic.* Nat Rev Gastroenterol Hepatol, 2013. **10**(11): p. 686-90.

34. van Asten, M., et al., *The increasing burden of NAFLD fibrosis in the general population: Time to bridge the gap between hepatologists and primary care.* Hepatology, 2017. **65**(3): p. 1078.

35. Alazawi, W., et al., *Ethnicity and the diagnosis gap in liver disease: a population-based study.* Br J Gen Pract, 2014. **64**(628): p. e694-702.

36. Medicine, N.L.o. *Unified Medical Language System*. 2017; Available from: https://www.nlm.nih.gov/research/umls/sourcereleasedocs/mrsabfields.html.

37. Williams, C.D., et al., *Prevalence of nonalcoholic fatty liver disease and nonalcoholic steatohepatitis among a largely middle-aged population utilizing ultrasound and liver biopsy: a prospective study.* Gastroenterology, 2011. **140**(1): p. 124-131.
38. Coloma, P.M., et al., *Identification of acute myocardial infarction from electronic healthcare records using different disease coding systems: a validation study in three European countries.* BMJ Open, 2013. **3**(6).
39. Chalasani, N., et al., *The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases.* Hepatology, 2018. **67**(1): p. 328-357.
40. European Association for the Study of the, L., D. European Association for the Study of, and O. European Association for the Study of, *EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease.* J Hepatol, 2016. **64**(6): p. 1388-402.
41. Sung, K.C. and S.H. Kim, *Interrelationship between fatty liver and insulin resistance in the development of type 2 diabetes.* J Clin Endocrinol Metab, 2011. **96**(4): p. 1093-7.
42. Zelber-Sagi, S., et al., *Non-alcoholic fatty liver disease independently predicts prediabetes during a 7-year prospective follow-up.* Liver Int, 2013. **33**(9): p. 1406-12.
43. Wilson, J.M.G. and G.j.a. Jungner, *Principles and practice of screening for disease [by] J. M. G. Wilson [and] G. Jungner*. 1968: Geneva, World Health Organization.

Figure Legends

Figure 1: Point-Prevalence of NAFLD (per 100 persons) by calendar year and gender. Results are shown for each database and pooled across databases by meta-analysis. Pooled estimate is provided from 2007 only as data from SIDIAP was only available from that year onward. The pooled estimate confidence interval is shaded grey.

Figure 2: Point-Prevalence of NAFLD (per 100 persons) by age group on the 1st of January 2015 in a) Males and b) Females.

Figure 3: Incidence of NAFLD (per 1,000 person-years) by calendar year in four primary care databases, and pooled across databases by random effects meta-analysis. Pooled estimate is provided from 2007 only as data from SIDIAP was only available from that year onward. The pooled estimate confidence interval is shaded grey.

Figure 4: Incidence of NAFLD (per 1,000 person-years) by age group for the four primary care databases for the year 2015 in a) Males and b) Females.

Declarations

Ethics approval and consent to participate: We followed local data laws in all four territories from which data were obtained and in all countries, specific ethical approval was not required for this study that used anonymised data. However, approval was sought and obtained from the scientific research committee for THIN, the IPCI Governing Board (ref 2015/18) and the IDIAP Ethics Committee (Reference P15/167) and the scientific committee of the Italian College of General Practitioners and Primary Care.

Consent for publication: Not applicable

Availability of data and material: This work uses data provided by patients and collected by the different healthcare systems involved as port of their care and support. All data relevant to the study purpose are within the paper and its Supporting Information files. Original, individual-level data are in custody to local partners, and the possibility to access them may vary depending on local governance rules. Local restrictions on publicly sharing original study data may vary on a case-by-case basis and depend on institutional review board, ethics committee or law. Further information on data request and access should be sent individually to the authors of this paper responsible for the data provided by the relevant organizations: SIDIAP (tduarte@idiapjgol.org), HSD (lapi.francesco@simg.it), THIN (d.ansell@bham.ac.uk), IPCI (j.vanderlei@erasmusmc.nl).

Competing interests: MA was contracted to work at and J F-B, PE, SK, DMW are employees of Glaxosmithkline which has conducted clinical research including trials of therapeutic agents in NAFLD. AKL is an employee of Pfizer which is conducting clinical research including trials of therapeutic agents in NAFLD. DP-A: unrestricted research grants from UCB, Amgen, Servier, and consultancy fees (paid to his department/research group) from UCB Pharma. DA: consultancy and advice to many pharmaceutical companies on undertaking outcomes studies using real world evidence. FL: consultancy for AlfaSigma, Bayer and Abbvie. PE, SK: Employee and stock holder, GlaxoSmithKline. NS: Consulted for Boehringer Ingelheim, Eli Lilly, Novo Nordisk, Janssen, and grants from Astrazeneca and BI. WA: Consultant and sponsored lectures: UCB Pharma, Gilead, Intercept, Medimmune. TDS: none to declare.

Funding: FP7 Ideas: European Research Council Award Number 115372. ERC had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript. DPA is funded by a National Institute for Health Research Clinician Scientist award (CS-2013-13-012). This article presents independent research funded by the National Institute for Health Research (NIHR). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health. This work was partially supported by the NIHR Biomedical Research Centre, Oxford. WA is in receipt of a Medical Research Council New Investigator Award.

Authors' contributions:

Study Design: MA, AKL, PE, SK, DW, NS, WA

Extracted data: TDS, PA (semantic harmonization), PR (data transformation and federated data analysis)

Analysed Data: MA, JFB,

Interpreted results: All authors

Wrote manuscript: MA, NS, WA

Edited manuscript: All authors

Approved for submission: All authors

Acknowledgements: The European Medical Information Framework (EMIF) is a collaboration between industry and academic partners that aims to develop common technical and governance solutions to facilitate access to diverse electronic medical and research data sources. These analyses were supported by the Innovative Medicines Initiative Joint Undertaking under EMIF grant agreement n° 115372, resources of which are composed of financial contributions from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies’ in kind contribution. The authors would like to acknowledge Nicholas Galwey for his advice on the statistical methods, Alba Jene for her administrative support and support during submission to ethical review boards, and Derek Nunez for support early on a protocol design stage.

Table 1: Flow-chart of identification of NAFLD patients

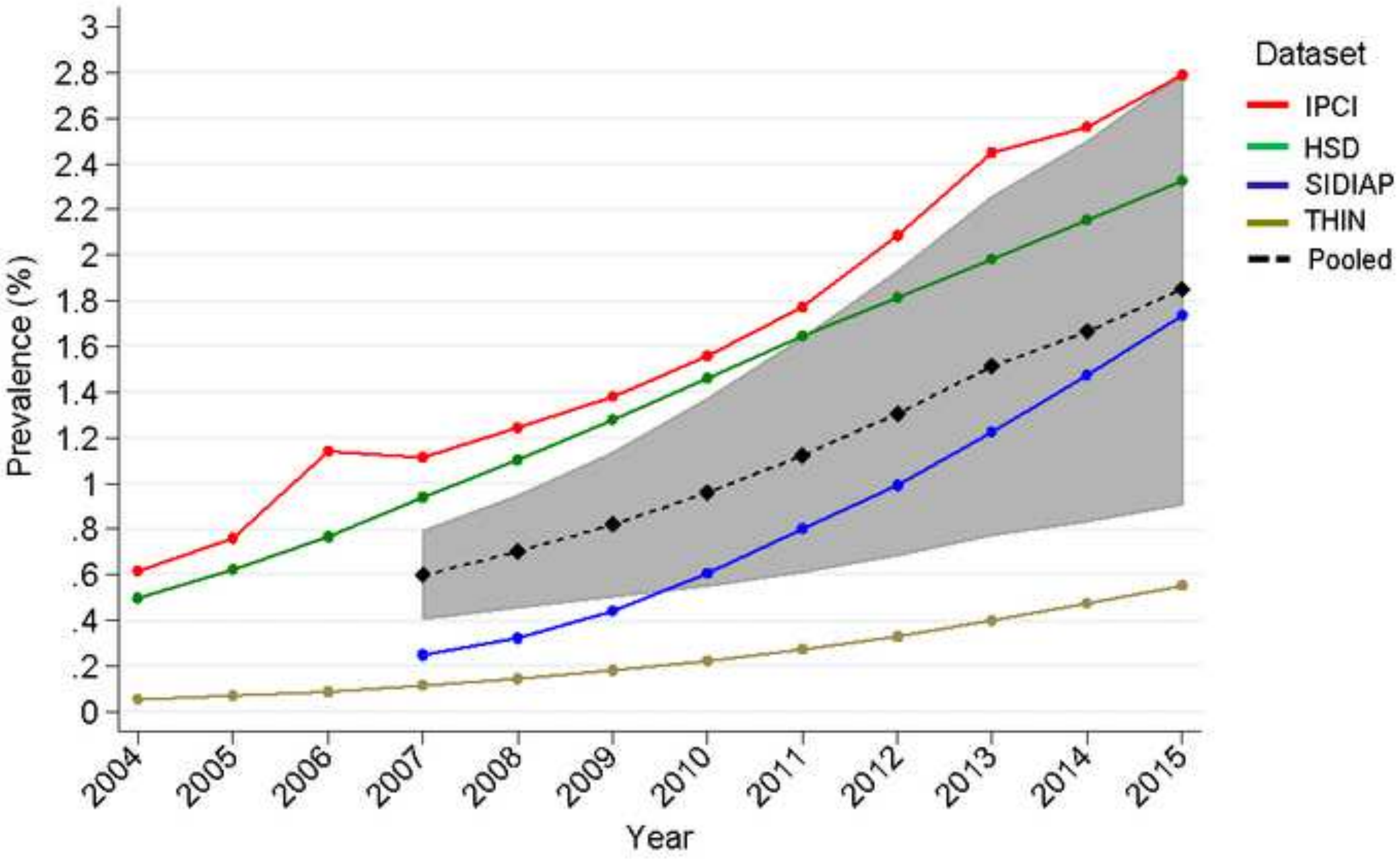
Flow-chart	**HSD (Italy)**	**IPCI (The Netherlands)**	**SIDIAP (Spain)**	**THIN (UK)**	**Total**
Total number of individuals ever enrolled by December 2015	1,571,651	2,225,925	5,488,397	12,695,046	21,981,019
Number of individuals with ≥1 year of registration in adulthood	1,544,573	1,780,500	5,259,575	9,085,325	17,669,973
Number of NAFLD patients*	NAFLD: 27,002	NAFLD: 48,036 (19,048 were incident post IPCI starting date)	NAFLD: 77,547 NASH only: 1,887	NAFLD: 23,529 NASH only: 1,133	NAFLD: 176,114

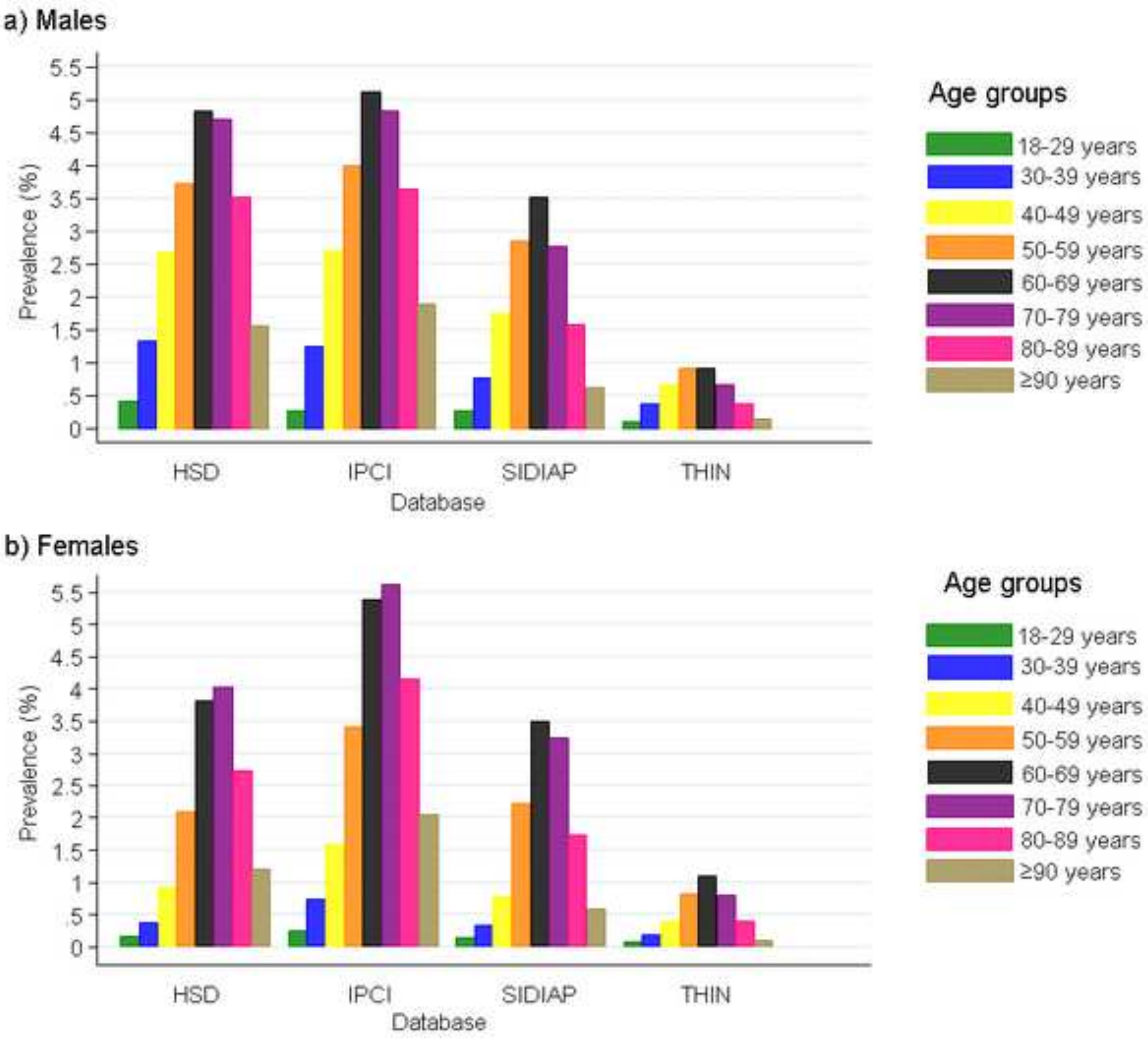
- In the descriptive tables (Table 2 & Supplementary Table 3), only patients with an incident diagnosis made within the study period and with a record of a GP visit within +/- 6 months of diagnosis. Numbers for NAFLD/NASH were as follows: HSD 24,027; IPCI 18,865; SIDIAP 77,107; and THIN: 12,385 individuals. Note in HSD and IPCI, 'NAFLD' is likely to include patients with NASH as no separate term for NASH exists in these databases. The number in the 'Total' column includes patients with in SIDIAP and THIN who have NASH.

Table 2: Descriptive characteristics of patients with an incident diagnosis of NAFLD in four European primary care databases

Characteristics	HSD (N: 24,027)	IPCI (N: 18,865)	SIDIAP (N: 77,107)	THIN (N: 23,385)	Test of difference p value
	% / Mean (SD) / Median (IQR)	% / Mean (SD) / Median (IQR)	% / Mean (SD) / Median (IQR)	% / Mean (SD) / Median (IQR)	
Age in years	56.2 (14.3)	56.8 (13.9)	56.0 (13.4)	53.7 (13.4)	<0.0001
Gender (Males)	57.3%	49.3%	52.7%	51.5%	<0.0001
Body Mass Index in kg/m2	29.7 (5.0)	30.8 (5.4)	31.3 (5.1)	32.4 (5.9)	<0.0001
History of diabetes or impaired fasting glucose	18.0%	20.5%	20.0%	21.0%	<0.0001
History of hypertension	47.5%	36.0%	42.8%	40.5%	<0.0001
Aspartate transaminase (IU/L)	24 (19 – 33)	29 (22 – 40)	29 (22 – 40)	32 (24 – 47)	<0.0001
Alanine transaminase (IU/L)	30 (20 – 48)	37 (25 – 56)	34 (22 – 53)	45 (28 – 68)	<0.0001
Platelet counts (10^9/L)	238 (65)	262 (68.6)	244 (61.2)	250 (75.3)	<0.0001
AST to ALT ratio	0.87 (0.34)	0.80 (0.36)	0.83 (0.38)	0.82 (0.38)	<0.0001
FIB4 score					<0.0001
Low risk (FIB4 <1.30)	64.3%	70.4%	65.5%	65.0%	
Indeterminate risk (FIB4: 1.30-2.67)	31.4%	26.7%	30.3%	25.0%	
High Risk (FIB4>2.67)	4.3%	2.9%	4.2%	10.0%	

N: Number of individuals; %: percentage; n: number; IQR: Interquartile range. Arithmetic means were reported for age, BMI, platelet counts and AST to ALT ratio; median (IQR) were reported for albumin, AST and ALT. P-values are from ANOVA test of difference between means for continuous variables (for log-transformed AST and ALT), and Chi-2 test of difference for categorical variables. Number of patients with data available on each of these variables is provided in Supplementary Table 3.


Prevalence (%)
3
2.8
2.6
2.4
2.2
2
1.8
1.6
1.4
1.2
1
.8
.6
.4
.2
0
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
Year
Dataset
IPCI
HSD
SIDIAP
THIN
Pooled


a) Males
Prevalence (%)
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
.5
0
HSD
IPCI
SIDIAP
THIN
Database
Age groups
18-29 years
30-39 years
40-49 years
50-59 years
60-69 years
70-79 years
80-89 years
≥90 years
b) Females
Prevalence (%)
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
.5
0
HSD
IPCI
SIDIAP
THIN
Database
Age groups
18-29 years
30-39 years
40-49 years
50-59 years
60-69 years
70-79 years
80-89 years
≥90 years

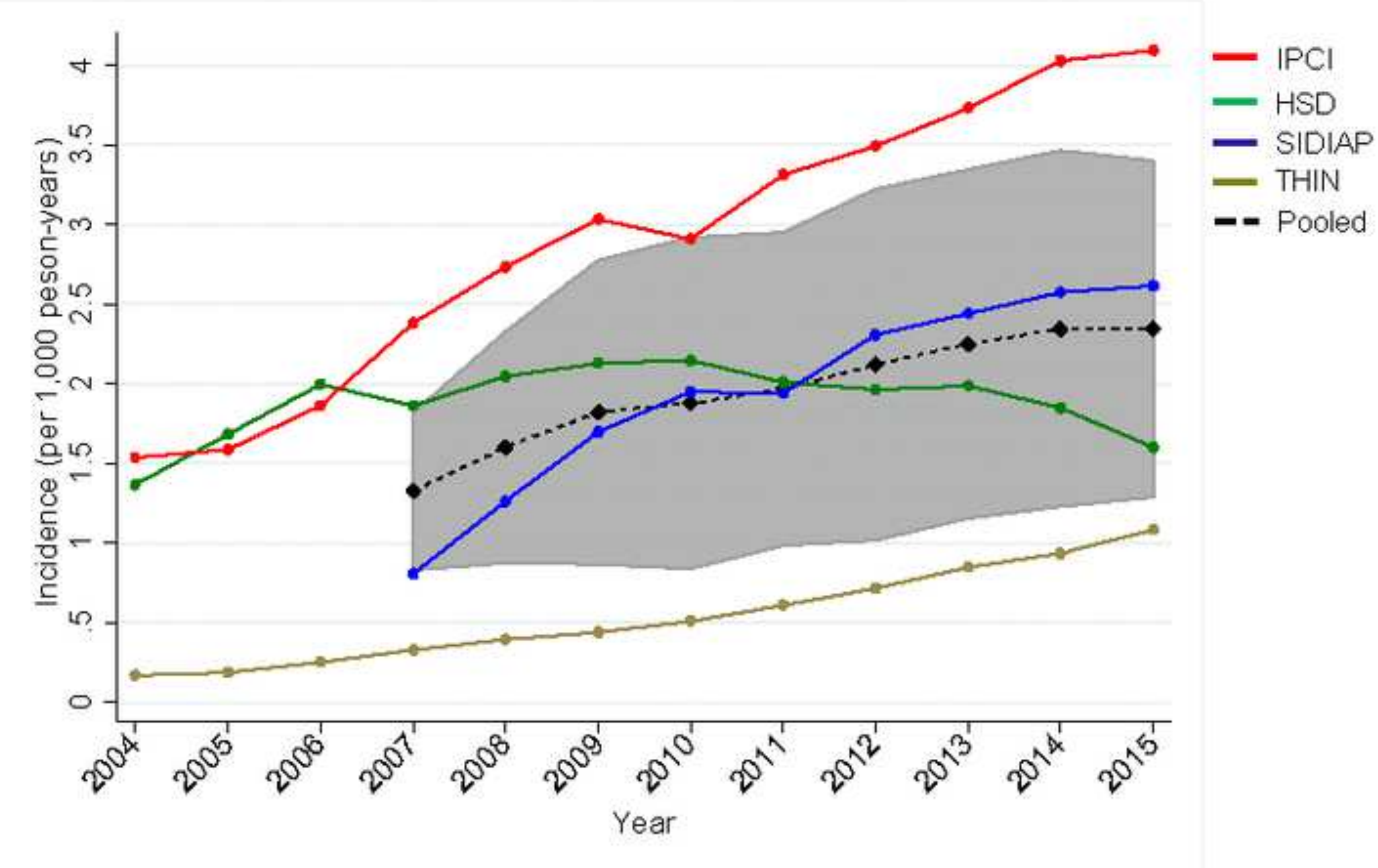
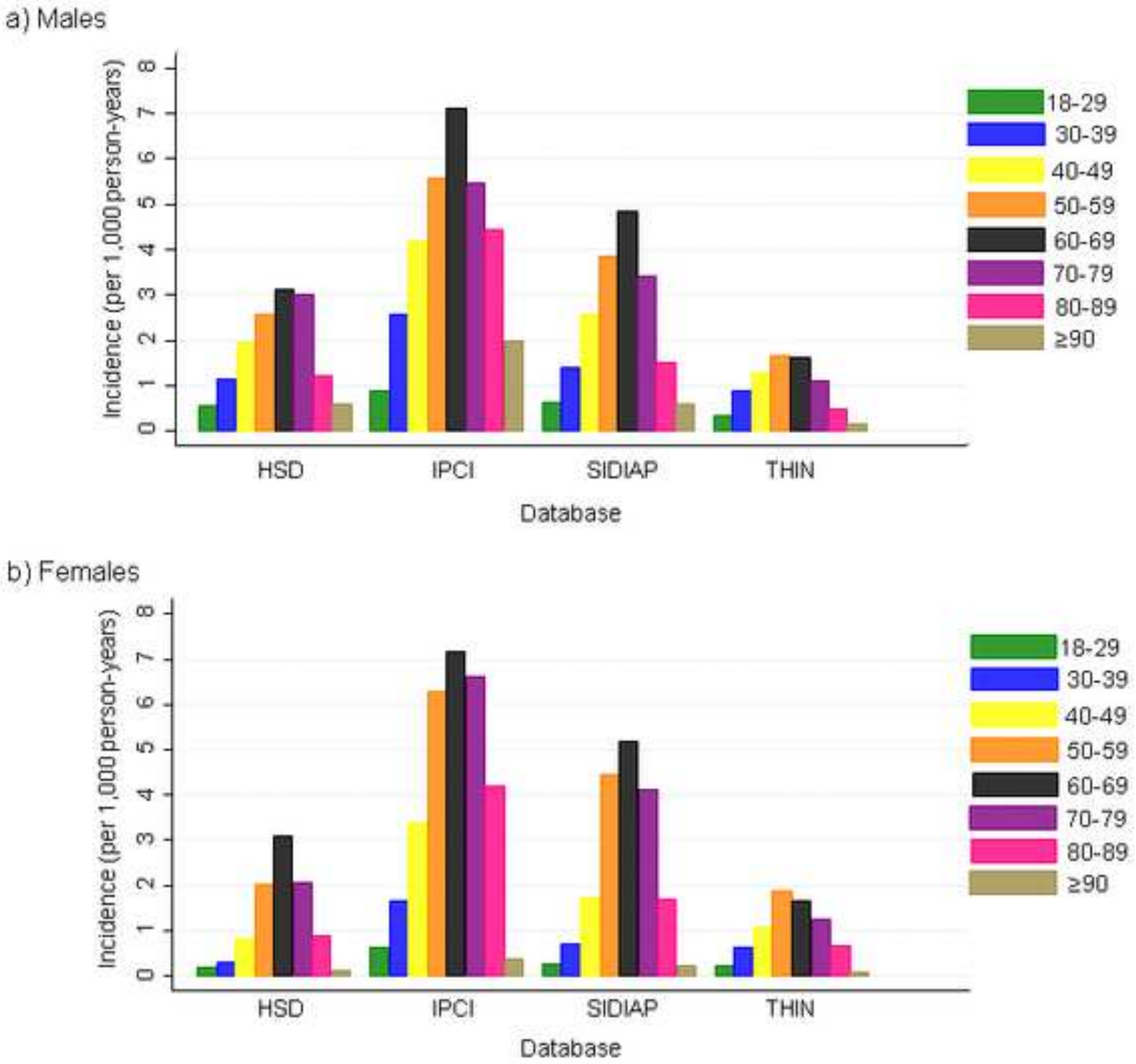

Incidence (per 1,000 peson-years)
0
.5
1
1.5
2
2.5
3
3.5
4
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
Year
IPCI
HSD
SIDIAP
THIN
Pooled

Figure 4


a) Males
Incidence (per 1,000 person-years)
0
1
2
3
4
5
6
7
8
HSD
IPCI
SIDIAP
THIN
Database
18-29
30-39
40-49
50-59
60-69
70-79
80-89
≥90
b) Females
Incidence (per 1,000 person-years)
0
1
2
3
4
5
6
7
8
HSD
IPCI
SIDIAP
THIN
Database
18-29
30-39
40-49
50-59
60-69
70-79
80-89
≥90

Supplementary material