PHARMACOEPIDEMIOLOGY AND PRESCRIPTION

Ascertainment of acute liver injury in two European primary care databases

A. Ruigómez · R. Brauer · L. A. García Rodríguez · C. Huerta · G. Requena · M. Gil ·
Francisco de Abajo · G. Downey · A. Bate · M. Feudjo Tepie · M. de Groot · R. Schlienger ·
R. Reynolds · O. Klungel

Abstract

Purpose The purpose of this study was to ascertain acute liver injury (ALI) in primary care databases using different computer algorithms. The aim of this investigation was to study and compare the incidence of ALI in different primary care databases and using different definitions of ALI.

Methods The Clinical Practice Research Datalink (CPRD) in UK and the Spanish "Base de datos para la Investigación Farmacoepidemiológica en Atención Primaria" (BIFAP) were used. Both are primary care databases from which we selected individuals of all ages registered between January 2004 and December 2009. We developed two case definitions of idiopathic ALI using computer algorithms: (i) restrictive definition (definite cases) and (ii) broad definition (definite and probable cases). Patients presenting prior liver conditions were excluded. Manual review of potential cases was performed to confirm diagnosis, in a sample in CPRD (21 %) and all potential cases in BIFAP. Incidence rates of ALI by age, sex and calendar year were calculated.

Results In BIFAP, all cases considered definite after manual review had been detected with the computer algorithm as potential cases, and none came from the non-cases group. The restrictive definition of ALI had a low sensitivity but a very high specificity (95 % in BIFAP) and showed higher rates of agreement between computer search and manual review compared to the broad definition. Higher incidence rates of definite ALI in 2008 were observed in BIFAP (3.01 (95 % confidence interval (CI) 2.13–4.25) per 100,000 person-years than CPRD (1.35 (95 % CI 1.03–1.78)).

Conclusions This study shows that it is feasible to identify ALI cases if restrictive selection criteria are used and the possibility to review additional information to rule out

A. Ruigómez (✉) · L. A. G. Rodríguez
Centro Español de Investigación Farmacoepidemiologica (CEIFE),
Almirante 28, 2°, 28004 Madrid, Spain
e-mail: aruigomez@ceife.es

R. Brauer
Faculty of Epidemiology and Population Health, London School of
Hygiene and Tropical Medicine, Keppel Street,
London WC1E 7HT, UK

C. Huerta · G. Requena · M. Gil · F. de Abajo
Division of Pharmacoepidemiology and Pharmacovigilance,
Medicines for Human Use Department, Agencia Española de
Medicamentos y Productos Sanitarios (AEMPS), Madrid, Spain

F. de Abajo
Pharmacology Section, Department of Biomedical Sciences II,
University of Alcalá, Madrid, Spain

R. Brauer · G. Downey · M. F. Tepie
Amgen NV, London, UK

A. Bate
Epidemiology, Pfizer Ltd, Walton Oaks, Dorking Road,
Tadworth, UK

M. de Groot · O. Klungel
Division of Pharmacoepidemiology & Clinical Pharmacology,
Faculty of Science, Utrecht University, Utrecht, The Netherlands

R. Schlienger
Novartis Pharma AG, Basel, Switzerland

R. Reynolds
Epidemiology, Pfizer Research & Development,
New York, NY, USA

differential diagnoses. Our results confirm that idiopathic ALI is a very rare disease in the general population. Finally, the construction of a standard definition with predefined criteria facilitates the timely comparison across databases.

Keywords Primary care databases · Incidence · Acute liver injury · Computer algorithms · Diagnosis definition · BIFAP · CPRD

Introduction

Acute liver injury is an important cause of adverse drug reactions and a common reason for regulatory actions [1, 2]. Idiopathic acute liver injury (ALI), unrelated to a specific cause, is rare [3]. Few studies have provided frequency estimates for ALI in the general population [4]. There is a great variation in the reported incidence rates of ALI, depending on the type of data source, characteristics of the study population and, especially, the case definition [5–9]. There is some research on the capability of observational databases to accurately capture ALI. The FDA mini-Sentinel project [10, 11] and FNIH's Observational Medical Outcomes Partnership [12] have specifically analyzed the predictive value of the outcome across multiple US databases, primarily focussed on insurance claims data [13, 14].

Primary health care databases (PCDB) are available in several countries and have been proven to be an excellent source to perform epidemiological and post-marketing drug safety research [15, 16]. The recording of diseases with a non-specific diagnostic test or those that can easily be mixed up with other diseases is difficult to ascertain in PCDBs, and researchers may need to construct more complex definitions to identify all potential cases. ALI is one of these diseases, with multiple aetiologies to rule out, and one that can only be identified after detailed reviews of medical history [17].

The aim of this investigation was to study and compare the incidence of ALI using different routine health care databases. With this purpose, we developed a specific computer algorithm to ascertain ALI, considering different scenarios to test the validity of this algorithm for the future use on drug safety issues [18] and to assess the impact of different definitions on the incidence rates. This study was performed within the framework of the IMI-Pharmacoepidemiological Research on Outcomes of Therapeutics by a European ConsorTium (PROTECT) project (http://www.imiprotect.eu/).

Methods

Data source

This study was performed using two European databases: the Clinical Practice Research Datalink (CPRD; http://www.cprd.

com) in the UK and the Spanish "Base de datos para la Investigación Farmacoepidemiológica en Atención Primaria" (BIFAP; http://www.bifap.org). Both are nationwide primary care databases. The CPRD contains data from more than 5 million active patients (8.3 % of the population) provided by primary care general practitioners (GPs) based throughout the UK. A very comprehensive dictionary of clinical terms, READ codes, enables GPs to effectively record medical conditions. The list of READ codes used to identify ALI is included Online Resource 1.

The BIFAP database was developed in recent years and includes anonymized clinical and prescription data collected by primary care physicians during their consultations, covering around 3.9 million patients and representing close to 8.6 % of Spanish population. Validation studies have been performed on other diagnoses [19, 20]. In the BIFAP database medical events are recorded using the *International Classification of Primary Care* (ICPC) codes, less granulated than the READ codes [21]. In addition to ICPC codes, the GP software (OMI-AP) includes a list of semantic terms to better describe the disease of interest where GPs can enter additional information as free text notes. BIFAP does not contain personal data; thus, patients cannot be identified nor their records be linked with other databases. See Online Resource 1 for codes and additional text mining search performed to identify ALI

Study population

The study period started in January 2004 and ended in December 2009. The study population encompassed patients of all ages registered during the study period. Within this period, we defined the start date for an individual, once the patient had at least 1 year of enrolment with the GP and 1 year of computerized prescription history. We excluded individuals with prior history of cancer, alcoholism, alcohol-related problems, gallbladder disease, pancreatic disease and other chronic liver diseases with clear aetiology such as viral, alcoholic or autoimmune (see computer codes in Online Resource 2). Individuals were followed from start date until the earliest of the following: the date a patient had a liver injury-related code (Tables 1 and 2), the patient died, the patient had a record of an exclusion criteria, the patient was transferred out of the practice, the end of the study period or the end of the practice data collection. Flow chart of study is available in Online Resource 3.

Outcome definition

We adapted as ALI definition the classification issued from a consensus meeting on drug-induced liver disorders [22] and widely used in drug safety studies with databases [4, 6, 8, 23]. We developed computer algorithms that could be applied to both databases, relying on three data elements: (1) recorded

Table 1 Computer search algorithms to ascertain acute liver injury. Operational case definition

Case status	Ia. Diagnosis of liver injury or symptoms recorded by specific codes or text[a] for liver injury	Ib. Diagnosis of liver injury or symptoms recorded by unspecific codes or text[a] indicating only positive results for liver tests	II. Complete laboratory criteria: an increase of more than two times ULN in ALT or a combined increase in AST, AP and total bilirubin provided one of them was two times ULN within 2 months of the event	III. A referral to a specialist orhospital within 2 weeks of a recorded diagnosis of liver injury
Definite	Yes	No	Yes	Yes
Probable A	Yes	No	Yes	No
Probable B	No	Yes	Yes	Yes
Possible	No	Yes	Yes	No
No case	Yes	No	No (normal LFTs or just increased values not with complete criteria)	Yes
	No	Yes	No (normal LFTs or just increased values not with complete criteria	No

ULN upper limit of normal, ALT alanine aminotransferase, AP alkaline phosphatase, AST aspartate aminotransferase

[a] In BIFAP, database ICPC codes were used along with computer search of keywords in text

medical information: a predefined code in CPRD and a predefined code from the BIFAP thesaurus list or in free text (Online Resource 1). Codes selected include either specific codes identifying ALI and unspecific codes that for themselves do not represent ALI but instead only implied a suggestion of a sign to further study; (2) biochemical parameters in laboratory tests recorded: including an increase of more than two times the upper limit of the normal range (ULN) in alanine aminotransferase (ALT) or a combined increase in aspartate aminotransferase (AST), alkaline phosphatase (AP) and total bilirubin provided one of them is two times ULN within 2 months; and (3) a referral or hospitalization entry within 2 weeks of recorded diagnostic code. Based on these three criteria, we classified computer-detected patients as potential cases including the following: definite, probable, possible ALI and non-cases (Table 1)

Two distinct case definitions of ALI for all ages were then used:

The restrictive definition of ALI, including only definite cases of ALI, defined as having the three components mentioned above: specific diagnostic code (or specific keyword in BIFAP) as listed in Online Resource 2, laboratory test criteria and hospitalization or referral to a specialist.

The broad definition of ALI, including definite and probable cases of ALI, required the occurrence of a diagnostic code (specific or non-specific) together with laboratory test criteria with or without a referral or hospitalization.

Patients with the following conditions were not considered as valid cases: (a) with liver function tests (LFTs) in the normal range or elevated but not in the magnitude proposed in the above definition, (b) with increased LFT values as results of routine investigations but without specific symptomatology recorded (e.g., fever, malaise, jaundice and abdominal pain), (c) with any exclusion criteria mentioned above within 6 months of the diagnosis date and (d) patients with clinical or laboratory alterations after 6 months of the initial onset date were considered chronic liver injury cases and excluded automatically in CPRD and during the manual review process in BIFAP.

Table 2 Computer case ascertainment and manual review process in BIFAP database

Pre-review computer case status[a]	Patients with a ICPD code of ALI (N=19,074)	Status after manual review of free text					
		Definite-confirmed	%	Probable-confirmed	%	No-case confirmed[b]	%
1. Definite	179	43	24.02	19	10.61	117	65.36
2. A-Probable A	119	14	11.76	22	18.49	83	69.75
2. B-Probable B	1,038	51	4.91	122	11.75	865	83.33
3. Possible	1,537	16	1.04	149	9.69	1372	89.26
4. No case	16,201	Manually reviewed a sample n=120, 100 % no case					

[a] As in Table 1

[b] Reason for exclusion during manual review: other liver disease (691), cancer (23), alcohol-related problems (186), gallbladder and pancreatic disease (120), routine testing (1,322) or not confirmed cases (95)

Case ascertainment and validation of ALI diagnosis

In BIFAP database, the diagnosis of liver injury was validated by reviewing the clinical profile of all potential cases including free text comments annotated by the GP. The computerized patient profiles and free text fields of all potential cases, initially identified by computer, were reviewed individually by two researchers (AR, GR), blinded to computer assignment. In case of discrepancy, consensus was reached by joint review of the case files (Table 2).

In CPRD, a random sample of definite and probable cases was selected (20 %), and extra information recorded in free text was requested to validate the ALI diagnosis. For practical reasons, only part of the available text could be provided, including 20 words either side around the keywords. An external specialist, blinded to the classification based on our search algorithm, reviewed and classified the cases (Table 3).

Statistical analyses

Incidence rates (IRs) of first-time ALI by age (in 10-year categories) and sex were calculated in both databases per calendar year (2004–2009) and according to both case definitions (narrow and broad definitions). Incidence rates were estimated using as numerator the cases of ALI (using the confirmed cases after manual review in BIFAP and the computer detected cases in CPRD) and as denominator the number of person-years in each year, overall and in each age- and sex-specific categories. For the comparison of the IRs in the general population across databases and over time, we carried out a direct sex and age standardization using the European Union population in 2008 (EUROSTAT) as standard (http://epp.eurostat.ec.europa.eu/portal/page/portal/population/data/database).

Results

Case ascertainment and validation in BIFAP

In BIFAP, by an automated computer search, we initially identified 19,074 patients with a first-ever record of a liver-related code or related keyword during the study period (2004–2009). Of these, only 2,873 (15.1 %) were retained as potential cases applying the computer algorithms shown in Table 1 and categorized in mutually exclusive categories as definite ($n=179$), probable ($n=1157$) or possible ($n=1537$) (Table 2). After manual review, 2,437 (85 %) were not considered as valid cases, more than half of them were patients with just incidental LFT findings. Among the 179 computer-detected definite cases, 65 % were not considered definite cases (Table 2). Among the computer-detected probable cases that were identified with specific codes related to liver

diseases (labelled as probable group A, in Table 1) ($N=119$), the proportion of no-confirmed cases was slightly higher (69.7 %) than for definite. Finally, among computer-detected probable cases identified with unspecific codes only including abnormal LFTs (labelled as probable B in Table 1) ($N=1,038$), the majority were considered non-cases (83.3 %). We also reviewed a sample of computer-detected non-cases, and none was considered ALI. All cases considered definite after manual review had been detected with the computer algorithm as potential cases, and none came from the random sample of 120 test negative cases. Most of the definite-confirmed cases, 87 % (108/124), were initially classified by the algorithm as either computer-detected definite or probable.

When we tested the computer algorithm for the restricted definition (computer-detected definite cases ($N=179$)), only 43 (24 %) were definite-confirmed ALI. This restricted definition had a low probability to detect all the real cases (sensitivity of 34.6 % (43/124) but was very specific, with less than 5 % of identified patients being false positive cases (specificity=95.2 % (2,613/2,746).

When broadening the inclusion criteria, for those patients categorized by computer algorithm as definite or probable ALI ($N=1336$), the manual reviewed confirmed only 271 (20.3 %) of them as valid ALI cases. This broad definition increased the sensitivity up to 62.1 %, but the specificity decreased to 56.3 % (1,372/2,437), meaning that close to half of the patients detected with this definition were false positive cases.

Finally, 124 patients fulfilled all the criteria to be considered definite-confirmed cases, and 312 patients fulfilled the criteria for probable cases.

Case ascertainment and validation in CPRD

By computer search and applying the criteria proposed in the study, there were 112,157 patients with a READ code suggestive of liver injury. Of these, 11,363 (10 %) had more than one code spread over more than 180 days since the onset of the first diagnosis and were excluded as chronic cases. Furthermore 99,639 (89 %) patients were excluded as they presented with exclusion criteria that could be the cause of their liver disease. Finally, 269 patients were considered as definite ALI cases and 729 as probable.

Case validation was undertaken on a sample of 208 cases, 21 % of the total number of definite and probable cases identified by computer search in the CPRD. The review of the 208 ALI cases, either definite or probable, gave an overall agreement rate of 58.6 % (Table 3). When restricted to the 101 definite ALI cases, the reviewer confirmed 64 cases as definite-confirmed idiopathic ALI (agreement rate of 63.4 %). Additional free text was available for a limited number of patients 59 (28 %), and in the definite category, only 47 cases had this additional information for clinical review by the expert (Table 3).

Table 3 Computer case ascertainment and manual review process in CPRD database

Pre-review computer case definition[a]	Sample cases to review	Status after manual review	
		Confirmed	%
1. Definite	101 (47 with free text)	64	63.4
2. Definite+probable	208 (59 with free text)	122	58.6

[a] As in Table 1

ALI incidence in BIFAP and CPRD

We observed that for all age groups, there was a higher incidence of definite ALI among females than males in both datasets and that there was an increased rate with increasing age (Fig. 1a). The overall incidence rate of definite ALI in 2008 was 3.01 (95 % confidence interval (CI) 2.13–4.25) per 100,000 person-years in BIFAP and 1.35 (95 % CI 1.03–1.78) per 100,000 person-years in CPRD.

Figure 1b shows the comparison between BIFAP and CPRD on the incidence of ALI using the broad definition. In CPRD, there was a higher incidence in females, while in BIFAP, there was a higher rate of ALI in males for most age groups. For both countries, we observed an increase in the incidence of ALI with age, but there was a decrease in the very old age groups. The overall incidence rate in 2008 when using the broad definition of ALI was 27.5 (95 % CI 24.6–20.9) per 100,000 person-years in BIFAP and 4.9 (95 % CI 4.2–6.6) per 100,000 person-years in CPRD.

After standardization to the European population, definite ALI incidence rates were higher in BIFAP (ranging from 2.41 in 2004 to 4.44 in 2009) than in CPRD (range 0.55–1.31). There was a suggestion of a time trend over the study period with a slight increase in later years in BIFAP, while this pattern was not seen in CPRD. Using the broad definition of ALI, CPRD rates were markedly lower (ranging from 2.47 in 2004 to 5.03 in 2009) while in BIFAP, these estimates were much higher, ranging from 15.0 to 29.9 per 100,000 person-years.

Discussion

This study demonstrates that when taking into consideration the complexity of the ascertainment of this disease and the heterogeneity of recorded information, it is feasible to identify ALI cases in two European databases applying a common methodology. In order to accurately capture idiopathic ALI in these databases, operational criteria need to be very specific, and a broad definition has demonstrated to be unsuitable to detect valid cases of ALI.

Most epidemiologic studies have adapted the classification issued from a consensus meeting on drug-induced liver disorders [22], with no special distinction by age groups or sex. Recently, there have been some approaches to create definitions based on coded information from computerised databases, as well as on laboratory and clinical data. The OMOP project and the FDA mini-Sentinel project evaluated different definitions of ALI across a combination of insurance claim databases and estimated the predictive value of this outcome in different sources [10–14], but not a general agreement was reached in the algorithm or codes used for the definition of idiopathic ALI. In our study, we tested two distinct case definitions: a restricted ALI definition and a broader one, resulting in substantial variation in the incidence rate. As expected, we observed that the more restrictiveness of case definition, the lower the number of cases in both databases with greater validity and smaller number of false positives. Aithal et al. have proposed changing the criteria for drug-induced liver injury by raising the cut-off level of ALT elevation to five times ULN [24]. While this criterion would exclude clinically unimportant and self-limited drug-related events, we believe that its application will bias the results towards more serious cases of ALI, and therefore, it would be less appropriate for the objectives of our study.

Our results confirm that idiopathic ALI is a very rare disease in the general population with an incidence rate between 1 and 4 cases per 100,000 person-years when ALI is defined with restricted criteria. These results are in line with previous reported estimates. Several drug safety studies on liver injury have been done using primary care UK databases [8, 23]. The incidence of idiopathic ALI ranged from 2.4 (95 % CI 2.0–2.8) per 100,000 person-years up to 14 per 100,000 person-years although differences in methodological approaches between studies need to be keep in mind [4, 5, 25, 26].

When using the broad definition, the age and sex standardized incidence of ALI in a year ranged from 2.5 to 5 per 100.000 person-years in CPRD and 14.9 to 29.9 1 in the BIFAP database. These results that include less specific liver injury were more in line to other population-based studies with different approaches that reported higher incidence rates [27, 28].

This study observed a higher incidence rate of ALI in BIFAP than in the CPRD database during 2004–2009 (Fig. 2); these discrepancies could be real or could reflect remaining non-identical methods of ascertainment, including differences in the electronic health records between both databases, the codification dictionaries used (READ versus ICPC) and the user interface with different structure for entry data [29], as well as the patterns of recording by the GPs. Moreover, in the UK-CPRD, the READ dictionary allows GPs to record a specific code for most health conditions only using additional comments to better describe the patient's

Fig. 1 Incidence rate of acute liver injury (ALI) in BIFAP and in CPRD (*dashed line*) by age and sex. Using the ALI narrow definition (**a**) and broad definition (**b**)
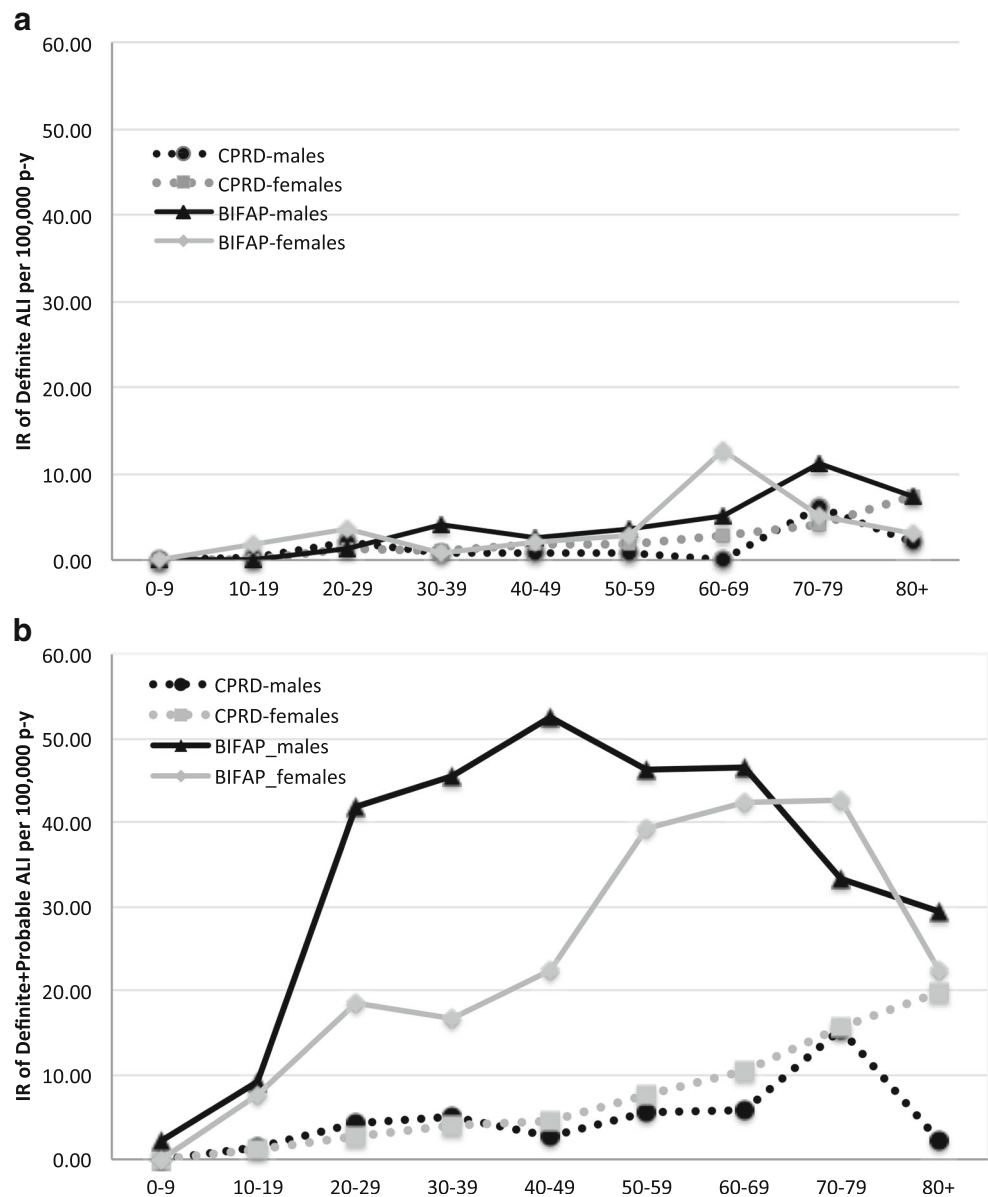

medical condition, allowing the confirmation of GP recordings in validation studies [17, 30]. In the BIFAP database, free text section is often used by GPs to record specific diagnosis. Hence, the information included in this section needs to be retrieved by semiautomatic search of keywords (data mining process) in addition to ICPC codes. Furthermore, routine laboratory testing, under health promotion programs with in some instances, lack of additional information from hospital or consultant visits, is recorded in BIFAP, leading to a higher number of potential computer-based probable cases (broad definition).

An important challenge of this study was the development of accurate definitions that enable comparisons of incidence of ALI and risk estimates in different datasets. Contrary to specific outcomes such as cancer where simple code-based algorithms are valid [31], the present study reinforces the importance of validation when pursuing more challenging outcomes, as relying only on automatic computer search may overestimate the true incidence of the disease. The consequence for drug safety studies using a restricted definition would be an increase in the specificity of the outcome at expense of potential underestimation of true cases and reduced statistical power to detect associations.

The main strength of this study is that it is population-based, using databases with large numbers of patients registered over long periods of time [16]. Furthermore, these patients are a good representation of the total population of their countries [15, 32]. Both Spain and the UK have a national health care system with universal coverage in which general practitioners (GPs) are the gatekeeper of the system for most

Fig. 2 Incidence rate of acute liver injury in BIFAP and in CPRD by year (standardised by age and sex EURO weights 2008). Using the ALI narrow definition (**a**) and broad definition (**b**)
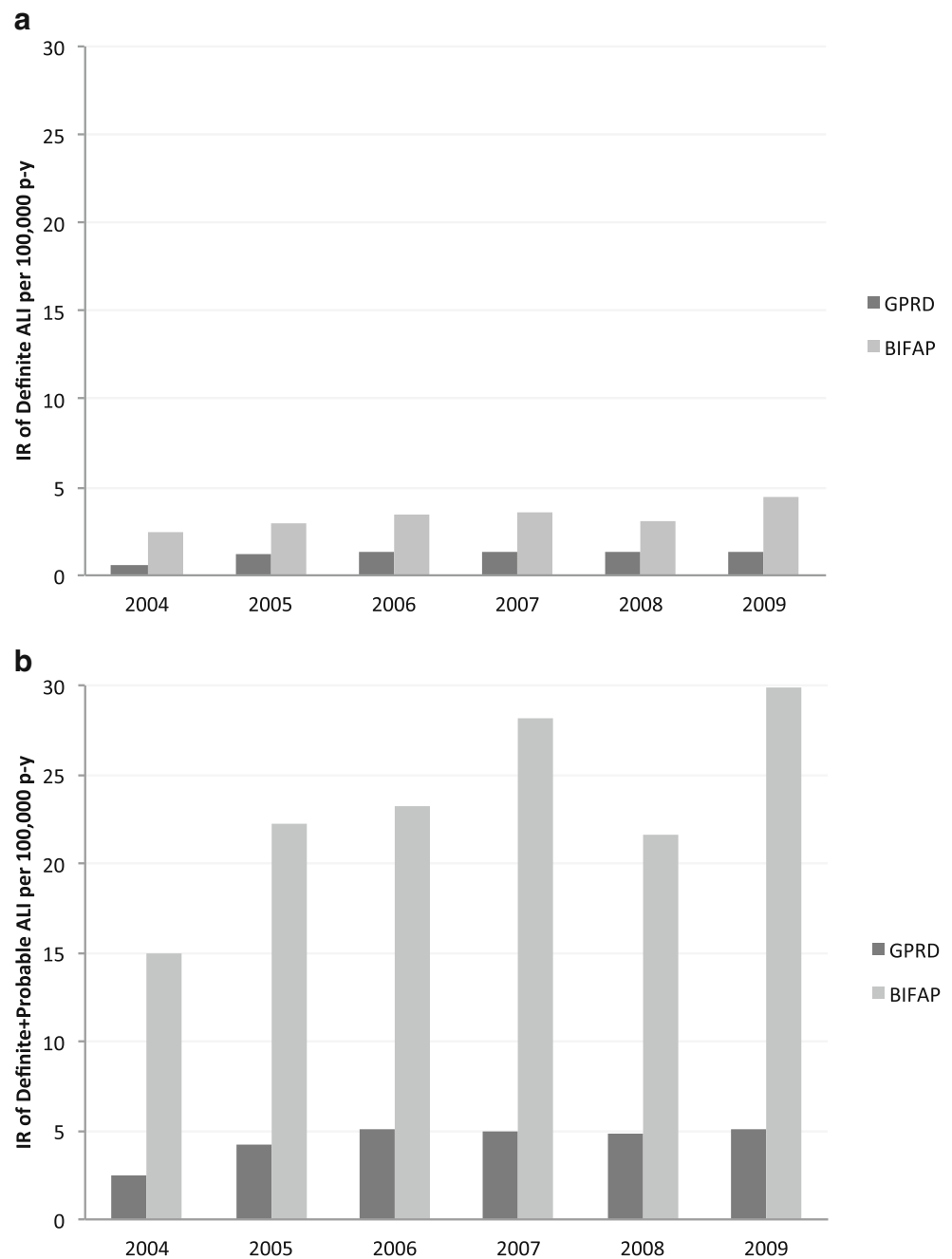
health problems, and they prescribed most of the medications issued, around 80 % in BIFAP [32].

Another strength of this study was the parsimonious procedure and strict criteria applied for case ascertainment by constructing computer-based algorithms that which can be replicated in future studies using multiple databases, keeping in consideration the inherited inter-database heterogeneity between resources.

Among limitations, it is possible that some ALI cases are not recorded by GP, for instance, those occurring in-hospital, and therefore, the incidence rates of ALI found in this study should be taken with caution. Also, we have to mention that

the validation process was different in the two databases, as in CPRD, only a sample of cases was available for this purpose. Finally, a limitation inherent in retrospective studies is that there may be missing information.

In conclusion, the construction of a standard definition with predefined criteria enables accurately identification of ALI cases and facilitates the timely comparison of incidence rates between different primary care databases. When the outcome to be studied is a rare and complex clinical condition such as ALI, the more restrictive definition and the possibility to review additional information to rule out differential diagnoses permit to capture more valid cases. The use of multiple

databases for epidemiological research as tested in the PROTECT project will be important to ensure power for valid comparisons and risk estimations when addressing rare outcomes and exposures. Hepatotoxicity of medical products has led to many marketing withdrawals and post-authorization safety studies. Often, such studies take many years to be performed, and the development of outcome definition is an important and time-consuming part of the study design. The ALI algorithms developed and tested in this study for two distinct primary care databases hold promise for the conduct of more rapid multiple database studies of ALI in the future.

Authorship statement Guarantor of article: Ana Ruigómez

Specific author contributions Ruigomez A, Garcia Rodriguez LA designed the research study and wrote the protocol.

Ruigomez A, Brauer R and Requena G performed the computer research and analyzed the data.

Ruigomez A, Garcia Rodriguez LA, Brauer R and Bate A drafted the manuscript.

Huerta C, Francisco de Abajo, Requena G and Gil M collected and managed the data in BIFAP and reviewed the manuscript.

Brauer R, Downey G and Feudjo Tepie M collected and managed the data in CPRD.

de Groot M, Schlienger R, Reynolds R and Klungel O reviewed the protocol and manuscript.

All authors contributed to the study design discussions, revising the manuscript, and given the approval of final versions.

References

1. Trifiro G et al (2009) Data mining on electronic health record databases for signal detection in pharmacovigilance: which events to monitor? Pharmacoepidemiol Drug Saf 18:1176–1184
2. Lexchin J (2005) Drug withdrawals from the Canadian market for safety reasons, 1963-2004. CMAJ 172:765–767
3. Ibanez L, Perez E, Vidal X, Laporte JR (2002) Prospective surveillance of acute serious liver disease unrelated to infectious, obstructive, or metabolic diseases: epidemiological and clinical features, and exposure to drugs. J Hepatol 37:592–600
4. de Abajo FJ, Montero D, Madurga M, Garcia Rodriguez LA (2004) Acute and clinically relevant drug-induced liver injury: a population based case-control study. Br J Clin Pharmacol 58:71–80
5. El-Serag HB, Everhart JE (2002) Diabetes increases the risk of acute hepatic failure. Gastroenterology 122:1822–1828
6. Huerta C, Zhao SZ, Garcia Rodriguez LA (2002) Risk of acute liver injury in patients with diabetes. Pharmacotherapy 22:1091–1096
7. Garcia Rodriguez LA, Stricker BH, Zimmerman HJ (1996) Risk of acute liver injury associated with the combination of amoxicillin and clavulanic acid. Arch Intern Med 156:1327–1332
8. Garcia Rodriguez LA, Duque A, Castellsague J, Perez-Gutthann S, Stricker BH (1999) A cohort study on the risk of acute liver injury among users of ketoconazole and other antifungal drugs. Br J Clin Pharmacol 48:847–852
9. Bjornsson ES, Bergmann OM, Bjornsson HK, Kvaran RB, Olafsson S (2013) Incidence, presentation, and outcomes in patients with drug-induced liver injury in the general population of Iceland. Gastroenterology 144:1419–1425
10. Platt R et al (2012) The U.S. Food and Drug Administration's Mini-Sentinel program: status and direction. Pharmacoepidemiol Drug Saf 21(Suppl 1):1–8
11. Carnahan RM, Moores KG (2012) Mini-Sentinel's systematic reviews of validated methods for identifying health outcomes using administrative and claims data: methods and lessons learned. Pharmacoepidemiol Drug Saf 21(Suppl 1):82–89
12. Stang PE et al (2010) Advancing the science for active surveillance: rationale and design for the Observational Medical Outcomes Partnership. Ann Intern Med 153:600–606
13. Lo Re V 3rd (2013) Validity of diagnostic codes to identify cases of severe acute liver injury in the U.S. Food and Drug Administration's Mini-Sentinel Distributed Database. Pharmacoepidemiol Drug Saf 22:861–872
14. Katz AJ, Ryan PB, Racoosin JA, Stang PE (2013) Assessment of case definitions for identifying acute liver injury in large observational databases. Drug Saf 36:651
15. Garcia Rodriguez LA, Perez Gutthann S (1998) Use of the UK General Practice Research Database for pharmacoepidemiology. Br J Clin Pharmacol 45:419–425
16. Garcia Rodriguez LA, Perez-Gutthann S, Jick S (2000) In: Storm BL (ed) Pharmacoepidemiology, 3rd edn. Wiley, Ltd, Chichester, pp 375–385
17. Garcia Rodriguez LA, Ruigomez A (2010) Case validation in research using large databases. Br J Gen Pract 60:160–161
18. Abbing-Karahagopian V, Kurz X, de Vries F, van Staa TP, Alvarez Y, Hesse U, Hasford J, Dijk Lv, de Abajo FJ, Weil JG, Grimaldi-Bensouda L, Egberts AC, Reynolds RF, Klungel OH.(2014)Bridging differences in outcomes of pharmacoepidemiological studies: design and first results of the PROTECT project. Curr Clin Pharmacol. 9(2):130–8
19. de Abajo FJ et al (2013) Upper gastrointestinal bleeding associated with NSAIDs, other drugs and interactions: a nested case-control study in a new general practice database. Eur J Clin Pharmacol 69:691–701
20. Chacon Garcia A, Ruigomez A, Garcia Rodriguez LA (2010) Incidence rate of community acquired pneumonia in a population cohort registered in BIFAP. Aten Primaria 42:543–549
21. Connolly JP, McGavock H, Wilson-Davis K (1997) Research methodology: coding perceived morbidity in general practice—an evaluation of the Read Classification and the International Classification of Primary Care (ICPC). Pharmacoepidemiol Drug Saf 6:325–330
22. Benichou C (1990) Criteria of drug-induced liver disorders. Report of an international consensus meeting. J Hepatol 11:272–276
23. Garcia Rodriguez LA, Ruigomez A, Jick H (1997) A review of epidemiologic research on drug-induced acute liver injury using the general practice research data base in the United Kingdom. Pharmacotherapy 17:721–728

24. Aithal GP et al (2011) Case definition and phenotype standardization in drug-induced liver injury. Clin Pharmacol Ther 89:806–815

25. Garcia Rodriguez LA, Williams R, Derby LE, Dean AD, Jick H (1994) Acute liver injury associated with nonsteroidal anti-inflammatory drugs and the role of risk factors. Arch Intern Med 154:311–316

26. Garcia Rodriguez LA, Perez Gutthann S, Walker AM, Lueck L (1992) The role of non-steroidal anti-inflammatory drugs in acute liver injury. BMJ 305:865–868

27. Sgro C et al (2002) Incidence of drug-induced hepatic injuries: a French population-based study. Hepatology 36:451–455

28. Graham DJ, Drinkard CR, Shatin D (2003) Incidence of idiopathic acute liver failure and hospitalized liver injury in patients treated with troglitazone. Am J Gastroenterol 98:175–179

29. de Lusignan S et al (2001) A survey to identify the clinical coding and classification systems currently in use across Europe. Stud Health Technol Inform 84:86–89

30. Ruigomez A, Martin-Merino E, Rodriguez LA (2010) Validation of ischemic cerebrovascular diagnoses in the health improvement network (THIN). Pharmacoepidemiol Drug Saf 19:579–585

31. Goldberg DS, Lewis JD, Halpern SD, Weiner MG, Lo Re V 3rd (2013) Validation of a coding algorithm to identify patients with hepatocellular carcinoma in an administrative database. Pharmacoepidemiol Drug Saf 22:103–107

32. Salvador Rosa A, Moreno Perez JC, Sonego D, Garcia Rodriguez LA, de Abajo Iglesias FJ (2002) The BIFAP project: database for pharmaco-epidemiological research in primary care. Aten Primaria 30:655–661