

Validation and incidence of community-acquired pneumonia in patients with type 2 diabetes in the BIFAP database

L. C. SAIZ¹, J. GARJÓN¹, J. GORRICO², J. ERVITI¹, M. J. GIL-GARCÍA³ AND E. MARTÍN-MERINO^{3*}

¹ *Unit of Drug Assessment, Advice and Research, Navarre Regional Health Service, Plaza de la Paz s/n, Pamplona, Navarre 31002, Spain*

² *Planning, Evaluation and Management Service, General Directorate of Health, Government of Navarre, Calle Irunlarrea 3 (Pabellón Docencia del Hospital de Navarra), Pamplona, Navarre 31008, Spain*

³ *BIFAP. Division of Pharmacoepidemiology and Pharmacovigilance. Spanish Agency of Medicines and Medical Devices (AEMPS). Calle Campezo no 1, Edificio 8, Madrid 28022, Spain*

Received 10 November 2016; Final revision 7 July 2017; Accepted 31 July 2017; first published online 31 August 2017

SUMMARY

Oral anti-diabetic drugs (OADs) have been associated with community-acquired pneumonia (CAP). We aimed to validate the recording of CAP in the Spanish Database for Pharmacoepidemiological Research in Primary Care (BIFAP) for the future evaluation of OAD–CAP association. The incidence rate (IR/1000 person-years) of CAP in type 2 diabetes mellitus (T2DM) was also determined. In total, 2966 pneumonia records (2040 listed as diagnosis and 926 as identified from comments added by physicians) were identified from 76 009 patients with T2DM after the first OAD in 2002–2013. Data around the CAP date were reviewed: 1803 (60.9%) were classified as ‘probable CAP’ (confirmed by X-ray/laboratory, referral letters or CAP lung site); 589 (19.8%) as ‘no-case’ (486 had other illness, 78 previous CAP, 25 cancer); and 574 (19.4%) as ‘possible CAP’ (441 without confirmatory information, 133 with uncertain diagnosis or uncertain diagnosis date). In total, 74.2% and 31.4% of pneumonia records in the diagnosis and comments, respectively, were ‘probable cases’ (IR: 6.04), which increased to 90.5% and 42.9%, respectively, when the 441 ‘possible cases’ without confirmatory information were included (IR: 7.52). In summary, diagnosis had a high positive predictive value, and adding cases automatically detected from comments decreased that value significantly.

Key words: Electronic health records, incidence, pneumonia, type 2 diabetes mellitus, validation studies.

INTRODUCTION

Community-acquired pneumonia (CAP) is a common infectious disease with an incidence of 2.7–11 per 1000 adult population-year and is a major cause of

hospitalisation and mortality (1–4 admissions per 1000 adult population-year) [1, 2].

The incidence of pneumonia is higher in diabetic patients than in the general population [3]. In addition to the increased risk of infection from hyperglycaemia, several groups of oral anti-diabetic drugs (OADs) have been associated with alterations in the immune system, which may increase the risk for infections and, more specifically, pneumonia [4, 5].

* Author for correspondence: E. Martín-Merino, BIFAP, Spanish Agency of Medicines and Medical Devices (AEMPS), C/Campezo no 1, Edificio 8, Madrid 28022, Spain.
(Email: emartin_fcsai@bifap.aemps.es)

Primary care electronic health records are a useful source of data for investigating drug-related adverse effects. Electronic health records convey detailed clinical and treatment data on a large number of patients recorded by physicians in a real setting [6]. However, the validity and completeness of records need to be verified before being used for research. In relation to pneumonia, confirmed cases must be distinguished from suspected cases through validation. Once the episodes registered in an electronic database have been detected automatically, the episode and date are verified by reviewing data in the patient's medical record and/or by contacting the physician who entered the data [7].

The purpose of this study was to identify episodes of CAP among cases of pneumonia recorded on the Spanish Database for Pharmacoepidemiological Research in Primary Care (BIFAP) database to derive the validity criteria and design a strategy for the automatic detection of CAP in BIFAP. The secondary objective of this study was to estimate the incidence rate (IR) of CAP in patients with type 2 diabetes mellitus (T2DM) after the onset of oral anti-diabetic therapy (OAD). This study has been performed within the framework of a larger study on the association between OADs and pneumonia [8].

METHODS

Data source

Data were retrieved from the BIFAP of the Spanish Ministry of Health [9]. BIFAP is a project of the Spanish Agency of Medicines and Medical Devices (AEMPS) where nine regions are involved. This database includes data from primary care medical records and it is intended to support pharmacoepidemiological research studies aimed at assessing the safety and effectiveness of medications.

When this study was carried out (2013), BIFAP included the medical records of 4.8 million patients (25.9 million person-years (p-y) of follow-up) completed by 2600 primary care physicians (PCP), accounting for 28% of the population of the regions involved in the project.

BIFAP conveys data including age, sex, medical diagnoses, prescriptions, laboratory results, lifestyle factors (such as smoking, body mass index and alcohol use, etc.) and referrals to specialists. In particular, the Diagnosis File in BIFAP data structure provides diagnostic data, which include the medical diagnoses,

date of diagnoses and a field for the PCP to add free-text comments. Thus, each diagnosis may have associated comments. Diagnoses are coded based on the International Classification of Primary Care (ICPC) [10]. ICPC is a classification of the most frequent health problems in primary care with limited granularity (~700 codes). The software used by physicians provides a list of specific descriptors associated with an ICPC term for PCP to select. New descriptors can be added when the existing list of descriptors is incomplete.

The most common descriptors are indexed in BIFAP by adding a fourth digit to the ICPC code of reference (1,2,...*n*) (the list of indexed descriptors will be referred to hereinafter as the 'ICPC-BIFAP Dictionary'). ICPC code descriptors that were not indexed were registered by adding 0 to the ICPC code of reference.

This project was approved by the Ethics Committee for Clinical Research of Navarra, Spain, on 18/01/2012 (Project 86/11).

Study cohort

The study cohort included patients >18 years of age diagnosed with T2DM who had been registered in the database for at least 1 year and whose first-recorded OAD therapy had been initiated between 2002 and 2013. The start date of follow-up was the date of the first recorded prescription of OAD therapy. Patients with a history of malignant cancer or use of OAD or insulin before the start date were excluded as well as patients aged ≥ 70 years who had not visited their PCP at least twice during the study follow-up period (ghost patients).

A follow-up on the patients was performed from the start date to the first occurrence of one of the following events (stop date): a record of pneumonia (potential cases to be reviewed), death, loss to follow-up, exclusion (record of malignant cancer, aspiration or hospital-acquired pneumonia) or end of the study (31/12/2013).

Identification of the potential cases of CAP by automatic algorithms

Potential cases of CAP were identified by searching for the term 'pneumonia' on the medical diagnosis of Diagnoses File. The term was found in indexed ICPC-BIFAP Dictionary codes for pneumonia as a unique diagnosis (i.e. R81.0 to R81.13) (Table 1) as well as in other ICPC-BIFAP codes together with terms referring to respiratory system dysfunction or

Table 1. *Type of record searched for community-acquired pneumonia in the Diagnosis File in BIFAP data structure*

Type of record	
(A) Medical diagnoses	
Codes indexed in ICPC-BIFAP R81*	Descriptors
R81·1	Q FEVER, PNEUMONIA
R81·2	INFLUENZA, FLU (CONF.) WITH PNEUMONIA
R81·3	LEGIONNAIRES DISEASE
R81·4	ATYPICAL PNEUMONIA NC
R81·5	BACTERIAL PNEUMONIA (CONF./PRO.)
R81·6	INFECTIOUS PNEUMONIA (UNKNOWN ORIGIN) NC
R81·7	VIRAL PNEUMONIA (CONF./PRO.)
R81·8	BRONCHOPNEUMONIA
R81·9	PNEUMONIA
R81·10	PNEUMOCOCCAL PNEUMONIA NC
R81·11	VIRAL PNEUMONIA
R81·13	POSSIBLE PNEUMONIA
R81·0	PNEUMONIA (not otherwise indexed)
Other ICPC-BIFAP codes	Including string text: *NEUMONIA*;
(B) Record in free-text comments	Excluding string text for aspiration or hospital-acquired pneumonia: *ASPIRACIO*, *NOSOCOMIAL*
	Excluding string text for suspect, negation or familiarity were also added to algorithm

ICPC-BIFAP, International Classification of Primary Care BIFAP indexed descriptors.

* The ICPC Dictionary code for generic pneumonia is R81.

general and unspecified health problems. Aspiration (R99·6) and nosocomial (R81·12) pneumonia were not included in case definition. In addition, a semantic search for the term 'pneumonia' was performed in the free-text comments associated with any medical diagnosis. Once the potential episodes of pneumonia were detected automatically (Table 1), the medical records of the patients were reviewed manually using the validation method described hereunder.

Case validation

Validation of the potential cases of pneumonia automatically detected was based on a manual review of the medical records dated up to 3 months before and after the episode of pneumonia. Anonymised records included PCPs comments, laboratory test results (radiological and/or microbiological data) and referral data. In order to establish the criteria for 'probable CAP case', 'possible CAP case' or 'no-CAP case', a pilot manual review of a randomised sample was previously performed. Next, eight investigators conducted a manual review of all the potential cases automatically detected following these established criteria:

- A case was considered 'probable' when diagnosis was supported by additional confirmatory data

including a report issued by a specialist, hospital or emergency unit, radiological or laboratory test confirmatory findings (blood test/culture), or when the specific site of pneumonia was detailed from physical examination findings (right upper lobe pneumonia, left lower lobe pneumonia, etc.).

- A case was considered 'possible' when there was a record of pneumonia, but wherein the patient did not meet the criteria for a 'probable case' or 'no case', had an uncertain diagnosis, the exact date of diagnosis was not clear or when there was no confirmatory or discarding information in any other section which enabled the confirmation or discard of pneumonia (aka possible cases without any additional information).
- A patient was classified as a 'no case' when confirmatory information discarding that the patient had pneumonia was provided (another diagnosis was confirmed or the diagnosis did not correspond to the patient of interest), the patient met an exclusion criterion (history of malignant cancer), CAP had occurred before the start date, or it was an episode of aspiration, nosocomial, interstitial, cryptogenic, eosinophilic, tuberculous or varicella pneumonia.

Each case was separately validated by two investigators. Discrepancies were solved by consensus. Given

that this study forms part of an OAD safety study project [8], the review was blind to the drug therapy.

Statistical analysis

The positive predictive value (PPV) of the potential cases of CAP was estimated regarding the detection of CAP overall and by the type of record (recorded in the medical diagnosis or in free-text comments). PPV was calculated by dividing the number of probable cases by the total number of potential cases.

Assuming that there were no cases other than the 'probable cases' in the study population, the sensitivity of the medical diagnosis was estimated by dividing the 'probable cases' detected in the medical diagnosis by the total of 'probable cases' (i.e. 'probable cases' were detected in the medical diagnosis and free-text comments) according to the true positive/(true positive + false negative) formula. Sensitivity was also estimated by using probable and possible cases without any additional information as a broader definition of the gold standard.

The IR of CAP per 1000 p-y was estimated by dividing the number of probable cases by the p-y to follow-up (main analysis), overall and by age at the start date. The IR was also estimated by using the probable and possible cases without any additional information as a broader definition of the gold standard. The main analysis represents the most conservative scenario.

RESULTS

Validation

The study cohort included 76 009 patients with T2DM. A total of 2966 patients had at least a record of pneumonia during the follow-up (2040 were detected in the medical diagnoses (1909 with ICPC R81 and 131 with a recorded diagnosis associated with other ICPC), and 926 were detected in free-text comments).

Of the 2966 records of pneumonia revised manually, 1803 (60.8%) were classified as 'probable cases' of CAP, 574 (19.4%) as 'possible cases', and 589 (19.9%) as 'no cases' (Fig. 1).

The site of pneumonia was recorded in 1392 (77.2%) of the 1803 'probable cases' of CAP, diagnosis was confirmed by a specialist in 1178 cases (65.3%) based on the radiological findings in 663 (36.8%) and based on positive laboratory test results in 119 cases (6.6%).

Of the 574 'possible cases' of CAP, no additional confirmatory or discarding information was provided in 441 patients (76.8%); 119 reports (20.7%) were

unclear regarding the diagnosis of CAP (with reference to the clinical signs of a suspected, unconfirmed CAP, unclear X-ray findings, suspicion of pneumonia other than CAP or disagreement between physicians on diagnosis); and the date of the episode of CAP was uncertain in 14 (2.4%).

Of the 589 'no cases' of CAP, 392 (66.6%) were not episodes of pneumonia, 94 (16.0%) were episodes of pneumonia other than CAP, 78 (13.2%) were episodes of CAP previous to the start date and 25 (4.2%) met an exclusion criterion. In total, 86.2% ($N = 338$) of 'no cases' were detected in free-text comments.

According to the type of record used for automatic detection of episodes of pneumonia, the PPV was 74.2% for episodes recorded in the medical diagnoses (74.5% for diagnoses recorded with ICPC R81, and 68.7% for other ICPCs) and 31.4% for diagnoses recorded in free-text comments. Percentages rose to 90.6% and 42.9% when possible cases without any additional confirmatory/discarding information were included in the analysis (Table 2).

Validation revealed that the codes *R81.5-Bacterial pneumonia* and *R81.9-Pneumonia* accounted for 70.7% of the probable cases detected in the medical diagnoses, with PPVs of 73% and 78%, respectively. The codes with the lowest PPV for probable CAP were *R81.13-Possible pneumonia* and *R81.8-Bronchopneumonia* (0% and 26%, respectively) (Table 3).

The sensitivity of the medical diagnosis to detect all 'probable cases' was 83.9% (1513/1803), and to detect all 'probable cases' and 'possible cases without any confirmatory/discarding information' was 82.3% (1848/2244) (Table 3).

Incidence of CAP in patients with T2DM

The estimated IR of CAP in patients with T2DM was 6.04 probable cases per 1000 p-y after the onset of OAD therapy.

The IR increased with age and was higher in men (7.12 per 1000 p-y) compared with women (4.80 per 1000 p-y), especially in patients over 65 years (Fig. 2).

When possible cases without confirmatory/discarding information were also included ($N = 441$), the IR for CAP was 7.52 cases per 1000 p-y.

DISCUSSION

Case validation

The contribution of this study is twofold: (i) it identifies cases of CAP recorded in a cohort of patients with

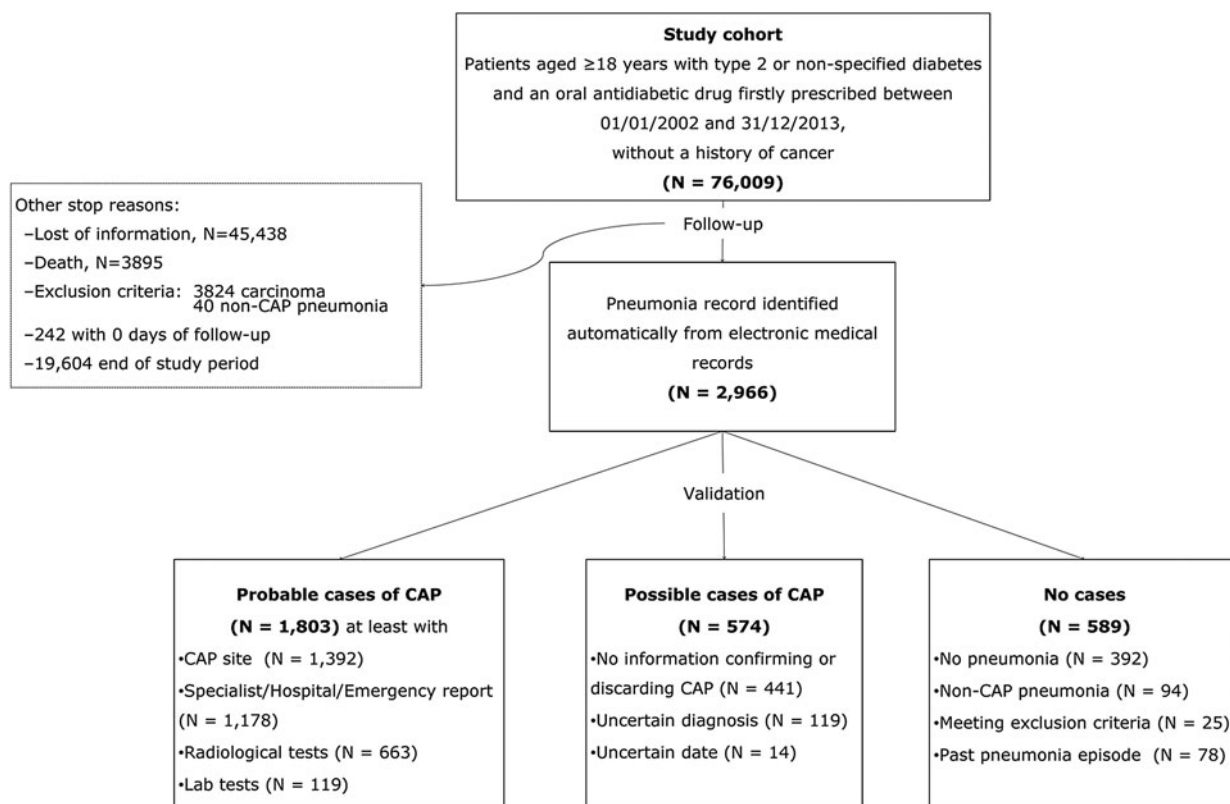


Fig. 1. Inclusion criteria, follow-up from the onset of oral antidiabetic drug therapy until pneumonia record in patients with type 2 diabetes mellitus and results of validation of CAP cases.

T2DM treated with OAD registered in the BIFAP database [9], and (ii) it establishes an algorithm with high predictive value for the automatic detection of CAP in patients with T2DM, which reduces the need for manual review. The algorithm for detecting recorded episodes of CAP (mostly with ICPC code R81.9) had a PPV of 74.2%, as confirmed by recorded radiological or laboratory findings, specialists and/or details on the site of pneumonia. The sensitivity to detect CAP was 83.9% (remaining cases were only identifiable through manual review of free-text comments). With a broader gold standard accepting CAP records without any confirmatory information, PPV was 90.6%, and sensitivity was 82.3%.

Other studies have been conducted to validate diagnoses recorded in BIFAP [11, 12]. The rate of confirmed diagnoses of CAP obtained in our study is consistent with that reported in a previous study based on a former version of BIFAP [2] (53.7% for pneumonia recorded in free-text comments and medical diagnoses; and 75.4% when only medical diagnoses were used). In both studies, the predictive value of free-text comments was substantially lower (20.9% and 31.4%, respectively). Whereas the case criteria established by Chacón-García

et al. only included confirmation by radiological findings or a specialist, ours incorporated laboratory test results or records of the lung site of pneumonia. The site of pneumonia was detailed in free-text clinical notes and might partially explain the higher percentage of confirmed cases of CAP obtained in our study. Given that similar PPVs were obtained in patients with T2DM, we recommend adopting these parameters in further studies on pneumonia based on BIFAP provided that this database maintains homogeneity in its data structure and data origin.

A number of previous studies have validated records of pneumonia in a variety of databases [13–18], mostly using ICD codes (International Classification of Diseases 9 and 10) for hospitalised patients, where PPV ranged from 73% to 96%. In the current study, PPV also ranged from 74% to 91%. Thus, PPV was 74% when only recorded confirmed cases of CAP ('probable cases') were used and 91% when both confirmed episodes of CAP and just recorded episodes without additional information (most 'possible cases') were included. The PPV obtained indicates a moderate-to-high accuracy in detecting new episodes of CAP.

Most of the discarded episodes of pneumonia detected in the medical diagnosis were episodes of influenza

Table 2. PPV for detection of episodes of CAP by type of record of pneumonia in the Diagnosis File

Type of record	Possible cases											
	Potential cases detected automatically		Probable cases		Without additional information		Uncertain diagnosis/date		No cases		Probable cases + possible cases without additional information	
	N	%	N	% (PPV)	N	%	N	%	N	% False positives		
Total	2966	100	1803	60.8	441	14.9	133	4.5	589	19.9	2244	75.7
Medical diagnoses:	2040	68.8	1513	74.2	335	16.4	70	3.4	122	6.0	1848	90.6
Codes indexed in ICPC R81	1909	64.4	1423	74.5	323	16.9	64	3.4	99	5.2	1746	91.5
Diagnosis of pneumonia recorded in other ICPC	131	4.4	90	68.7	12	9.2	6	4.6	23	17.6	102	77.9
Record in free-text Comments	926	31.2	291	31.4	106	11.4	63	6.8	466	50.3	397	42.9

PPV, positive predictive value; ICPC, International Classification of Primary Care; CAP, community-acquired pneumonia.

without complications, bronchitis, suspected CAP by the PCP not finally mentioned to be confirmed and cases of pneumonia other than CAP. Therefore, some codes had a low predictive value, specifically *R81.2-Influenza, flu with pneumonia*, *R81.8-Bronchopneumonia* or *R81.13-Possible pneumonia*. Code R81.13 showed no PPV, its descriptor shows PCP's uncertainty and it should not be used for the automatic detection of CAP. In addition, codes R81.2 and R81.8 should be used carefully taking into account their low PPVs.

Most potential episodes of pneumonia recorded in free-text comments were discarded, as they actually were episodes of a disease other than pneumonia. Half of these cases were episodes of *Klebsiella pneumoniae* urinary tract infections detected through excessively broad semantic searches, i.e. *pneumoniae*. Thanks to this validation exercise, future studies interested in high sensitivity may be benefited from using a corrected text-mining algorithm to detect pneumonia recorded in free-text comments. This corrected algorithm would exclude records mentioning *Klebsiella* or urinary infection close to *pneumonia*, increasing the PPV of probable CAP detected through comments.

Incidence of CAP

The IR of CAP obtained in this study ranged from 6.04 to 7.52 cases per 1000 p-y following the start of the OAD treatment. The first IR was obtained when only probable cases were considered, whereas the second corresponds to the incorporation of probable and possible cases without any confirmatory/discarding information. Incidence increased with age and – as expected – was higher in men [19, 20]. The reason for this association between CAP and gender has not been clearly determined yet [21]. Some authors suggest that, in men, a greater lifetime exposure to smoke, dust and to higher rates of chronic lung disease could be behind these findings. In fact, however, the role of smoking as a confounding variable has also not been adequately tested in an epidemiological study [22].

Overall and age-stratified IRs doubled the ones reported in previous studies (overall [2]: 2.69 cases/1000 p-y and in patients aged >60 years [2, 23]: near 4–5/1000 p-y). Such a difference might be explained by the higher risk of pneumonia in patients with diabetes [3] or the higher probability of detection due to the continuous monitoring of this population. In another study in the Spanish population [20], the reported IR for patients with diabetes aged ≥65 years reached 15.0 per 1000 p-y, which doubles the

Table 3. PPV for detection of episodes of CAP by type of record in the medical diagnoses of patient's medical record

Type of record in the medical diagnoses Codes indexed in ICPC-BIFAP R81* (N = 1909)	Descriptors	Cases detected		Probable cases	PPV
		automatically			
		N	%	N	%
R81-1	Q FEVER, PNEUMONIA	8	0.4	6	75.0
R81-2	INFLUENZA, FLU (CONF.) WITH PNEUMONIA	22	1.1	10	45.5
R81-3	LEGIONNAIRES DISEASE	2	0.1	2	100.0
R81-4	ATYPICAL PNEUMONIA NC	65	3.2	51	78.5
R81-5	BACTERIAL PNEUMONIA (CONF./PRO.)	724	35.5	530	73.2
R81-6	INFECTIOUS PNEUMONIA (UNKNOWN ORIGIN) NC	144	7.1	99	68.8
R81-7	VIRAL PNEUMONIA (CONF./PRO.)	34	1.7	19	55.9
R81-8	BRONCHOPNEUMONIA	19	0.9	5	26.3
R81-9	PNEUMONIA	719	35.2	563	78.3
R81-10	PNEUMOCOCCAL PNEUMONIA NC	3	0.1	3	100.0
R81-11	VIRAL PNEUMONIA	0	0.0	0	–
R81-13	POSSIBLE PNEUMONIA	2	0.1	0	0.0
R81-0	PNEUMONIA (not otherwise indexed)	167	8.2	134	80.2
Diagnosis of pneumonia recorded in other ICPC [†]		131	6.4	90	68.7
Total		2040	100	1512	74.2

PPV, predictive positive value; ICPC, International Classification of Primary Care; BIFAP, Spanish Database for Pharmacoepidemiological Research in Primary Care; CAP, community-acquired pneumonia.

* The ICPC Dictionary code for generic pneumonia is R81.

† In these records, pneumonia was mostly recorded with an ICPC of respiratory system dysfunction, general and unspecified health problem, although it could be associated with any ICPC suggesting that pneumonia was related to other concomitant disease.

incidence obtained in our study for that age group. That study included CAP cases from both primary care and hospital records, which may yield higher IR compared with studies based exclusively on primary care cases. In any case, there is wide variability in the IRs reported in studies in the Spanish population [2, 20, 23, 24] probably due to the differences in the characteristics of the studied patients (IR would increase with the age, in men, winter seasons or specific geographical areas) and the methods used to collect data (field studies *vs.* studies based on electronic health records), case definition (a broader definition would result in a higher incidence) or sampled patients (studies using small samples sizes will have less power to include real incidence than studies using the overall population).

Strengths and limitations

This study provides an estimation of the IR of CAP in Spanish patients with T2DM, who were under-represented in the literature [2, 13–18].

Another relevant strength of this study is that all the records detected were pair reviewed and discussed that contribute rigour to the final classification.

In addition, the large number of patients and volume of data recorded in BIFAP ensures its robustness and reliability. Another relevant aspect facilitating the detection of cases of pneumonia – no matter whether it resulted in hospitalisation or not – was that, like CPRD [25] and THIN [26], BIFAP is based on primary care records. This fact provides a comprehensive view of the global clinical care provided to the patient.

A limitation of this study is that validation was based only on a review of PCP records. Although most CAP diagnostic data were rigorously recorded by PCPs, interpretation errors may occur by the reviewers, especially in records without confirmatory information. However, for logistic reasons, PCPs could not be asked to access other sources of information (specialist reports, discharge letters, etc.) to validate the diagnoses of pneumonia or to provide the radiologic criteria used to diagnose pneumonia. To prevent an underestimation of the IR, a broader

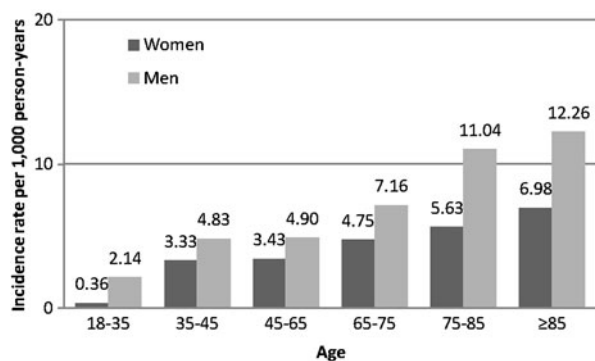


Fig. 2. Incidence rate of probable cases of CAP per 1000 person-years of follow-up after first oral antidiabetic drug prescription by age and sex recorded in BIFAP between 2002 and 2013.

gold standard was utilised providing a range that probably included the real IR. However, even the 'probable CAP' diagnosis can depend on the clinical impression of the clinician without X-ray or laboratory evidence. It is possible that a PCP or specialist may overdiagnose pneumonia.

Some cases of CAP may have not been detected, as they may have been confirmed by a specialist and not recorded by the PCP. On the other hand, pneumonia diagnosed shortly after discharge from hospital or following a nursing facility stay could not be excluded as potential nosocomial pneumonia.

The results of this study cannot be applied to pneumonias other than non-acquired ones (such as aspiration, nosocomial, interstitial, cryptogenic, eosinophilic, tuberculous or varicella pneumonia) or cancer patients who were excluded in order to prevent underlying confounders in future research studies. On the other hand, IRs might vary in regions not represented in BIFAP.

CONCLUSIONS

We concluded that the studies intended to have a high PPV for CAP in BIFAP – as is the case of case-control studies – should only include pneumonias recorded in the medical diagnosis (PPV will be 74% for the criteria of CAP requiring additional confirmatory information, and 91% for broader definitions). This study proves that PCPs frequently record additional data confirming the diagnosis of pneumonia (radiological or laboratory findings, and/or the site of pneumonia) in the free-text comment section. String text for detecting confirmatory data might be included in future automatic algorithms with a high predictive value.

When sensitivity is a relevant factor – as is the case of incidence or prevalence studies – researchers should

take into account that the algorithm detects 83.9% of episodes of CAP, whereas the remaining cases can only be detected from data recorded in free-text comments after careful corrections.

In Spain, the incidence of CAP seems to be higher in patients with T2DM than in the general population.

This study helps refine the search criteria and contributes valuable data for future studies on CAP in BIFAP, proving the potential of this useful database. The extent to which the reported PPV may depend on patients' characteristics, such as age, gender, co-morbidity or health care service use, deserves future research.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the excellent collaboration of the primary care general practitioners, paediatricians, nurses and the support of regional governments taking part in BIFAP and they thank Ana Azparren Andía, M^a Concepción Celaya Lecea, Antonio López Andrés and Lourdes Muruzábal Sitges for their participation in the case validation process. In addition, the authors acknowledge Nuria H. Buendía of Traducalia (Spain) (funded by the Navarre Health Service) and Michael Harlan Lyman of Rayma GLS (Global Linguistic Services) (funded by the AEMPS) who provided translation and edition of the original draft. This study was performed within the framework of a project on the study of the association between anti-diabetic drugs and pneumonia funded by the Carlos III Health Institute (EC11-356), and in kind contribution from AEMPS in the form of the BIFAP staff expertise and the provision of resources by enabling access to the database.

AUTHORSHIPS

EM-M conceived the validation and incidence study. Data extraction and a pilot review were performed by EM-M and MJG-G. JE, JGa, JGo and LCS designed the validation process and reviewed and categorised the data on clinical episodes. Data analysis was conducted by EM-M. All the authors contributed to the interpretation of the results (LCS, JGa, JGo, JE, MJG-G and EM-M). LCS, JGa and EM-M prepared the draft manuscript. All the authors (LCS, JGa, JGo, JE, MJG-G and EM-M) made significant contributions to the final version of the manuscript and approved it. The views expressed in this article are

those of the authors and do not necessarily represent the views of the co-authors' respective organisations.

DECLARATION OF INTEREST

The authors declare no conflict of interests.

REFERENCES

1. **Lim WS, et al.** BTS guidelines for the management of community acquired pneumonia in adults: update 2009. *Thorax* 2009; **64**: iii1–iii55. doi: 10.1136/thx.2009.121434.
2. **Chacón-García A, Ruigómez A, García-Rodríguez LA.** Incidence rate of community acquired pneumonia in a population cohort registered in BIFAP. *Atención Primaria* 2010; **42**: 543–551.
3. **Torres A, et al.** Which individuals are at increased risk of pneumococcal disease and why? Impact of COPD, asthma, smoking, diabetes, and/or chronic heart disease on community-acquired pneumonia and invasive pneumococcal disease. *Thorax* 2015; **70**: 984–989.
4. **Willemens MJ, et al.** Use of dipeptidyl peptidase-4 inhibitors and the reporting of infections: a disproportionality analysis in the World Health Organization Vigibase. *Diabetes Care* 2011; **34**: 369–374.
5. **Sing S, Loke YK, Furberg CD.** Long-term use of thiazolidinediones and the associated risk of pneumonia or lower respiratory tract infection: a systematic review and meta-analysis. *Thorax* 2011; **66**: 383–388.
6. **Salvador A, et al.** The BIFAP project: database for pharmaco-epidemiological research in primary care. *Atención Primaria* 2002; **30**: 655–661.
7. **García Rodríguez LA, Ruigómez A.** Case validation in research using large databases. *British Journal of General Practice* 2010; **60**: 160–161.
8. **Gorricho J, et al.** Use of oral antidiabetic agents and risk of community-acquired pneumonia: a nested case-control study. *British Journal of Clinical Pharmacology* 2017; **83**: 2034–2044. doi: 10.1111/bcp.13288.
9. **BIFAP.** Base de Datos para la Investigación Farmacoepidemiológica en Atención Primaria (<http://bifap.aemps.es/>). Accessed 01-04-2016.
10. **Lamberts H, Wood M.** *International Classification of Primary Care (ICPC)*. Oxford: Oxford University Press, 1987.
11. **Ruigómez A, et al.** Ascertainment of acute liver injury in two European primary care databases. *European Journal of Clinical Pharmacology* 2014; **70**: 1227–1235.
12. **De Abajo FJ, et al.** Upper gastrointestinal bleeding associated with NSAIDs, other drugs and interactions: a nested case-control study in a new general practice database. *European Journal of Clinical Pharmacology* 2013; **69**: 691–701.
13. **Drahos J, et al.** Accuracy of ICD-9CM codes in identifying infections of pneumonia and herpes simplex virus in administrative data. *Annals of Epidemiology* 2013; **23**: 291–293.
14. **Aronsky D, et al.** Accuracy of administrative data for identifying patients with pneumonia. *American Journal of Medical Quality* 2005; **20**: 319–328.
15. **Yu O, et al.** Classification algorithms to improve the accuracy of identifying patients hospitalized with community-acquired pneumonia using administrative data. *Epidemiology and Infection* 2011; **139**: 1296–1306.
16. **Meropol SB, Metlay JP.** Accuracy of pneumonia hospital admissions in a primary care electronic medical record database. *Pharmacoepidemiology and Drug Safety* 2012; **21**: 659–665.
17. **Skull SA, et al.** ICD-10 codes are a valid tool for identification of pneumonia in hospitalized patients aged > or = 65 years. *Epidemiology and Infection* 2008; **136**: 232–240.
18. **Van de Garde EM, et al.** International classification of diseases codes showed modest sensitivity for detecting community-acquired pneumonia. *Journal of Clinical Epidemiology* 2007; **60**: 834–838.
19. **Gutiérrez F, et al.** Epidemiology of community-acquired pneumonia in adult patients at the Dawn of the 21st century: a prospective study on the Mediterranean coast of Spain. *Clinical Microbiology and Infection* 2005; **11**: 788–800.
20. **Vila-Corcoles A, et al.** Epidemiology of community-acquired pneumonia in older adults: a population-based study. *Respiratory Medicine* 2009; **103**: 309–316.
21. **Farr BM, et al.** Risk factors for community-acquired pneumonia diagnosed by general practitioners in the community. *Respiratory Medicine* 2000; **94**: 422–427.
22. **Farr BM, et al.** Risk factors for community-acquired pneumonia diagnosed upon hospital admission. *Respiratory Medicine* 2000; **94**: 954–963.
23. **Gutiérrez F, et al.** The influence of age and gender on the population-based incidence of community-acquired pneumonia caused by different microbial pathogens. *Journal of Infection* 2006; **53**: 166–174.
24. **Almirall J, et al.** New evidence of risk factors for community-acquired pneumonia: a population-based study. *European Respiratory Journal* 2008; **31**: 1274–1284.
25. **Williams T, et al.** Recent advances in the utility and use of the general practice research database as an example of a UK primary care data resource. *Therapeutic Advances in Drug Safety* 2012; **3**: 89–99.
26. **Lewis JD, et al.** Validation studies of the health improvement network (THIN) database for pharmacoepidemiology research. *Pharmacoepidemiology and Drug Safety* 2007; **16**: 393–401.