Development and validation of a code-based algorithm using in-hospital medical records to identify patients with pulmonary arterial hypertension in a French healthcare database

Clément Jambon-Barbara [1,2], Alex Hlavaty[1,2], Claire Bernardeau[1], Hélène Bouvaist[3], Marie-Camille Chaumais [4,5,6], Marc Humbert [4,7,8], David Montani [4,7,8], Jean-Luc Cracowski[1,2] and Charles Khouri [1,2,9]

[1]Pharmacovigilance Unit, Grenoble Alpes University Hospital, Grenoble, France. [2]Univ. Grenoble Alpes, HP2 Laboratory, Inserm U1300, Grenoble, France. [3]Cardiology Unit, Grenoble Alpes University Hospital, Grenoble, France. [4]INSERM UMR_S 999, Hôpital Marie Lannelongue, Le Plessis Robinson, France. [5]Assistance Publique-Hôpitaux de Paris (AP-HP), Department of Pharmacy, Hôpital Bicêtre, Le Kremlin-Bicêtre, France. [6]Faculty of Pharmacy, Université Paris-Saclay, Saclay, France. [7]Faculty of Medicine, Université Paris-Saclay, Le Kremlin-Bicêtre, France. [8]AP-HP, Department of Respiratory and Intensive Care Medicine, Pulmonary Hypertension National Referral Centre, Hôpital Bicêtre, DMU 5 Thorinno, Le Kremlin-Bicêtre, France. [9]Grenoble Alpes University Hospital, Clinical Pharmacology Department INSERM CIC1406, Grenoble, France.

Corresponding author: Charles Khouri (ckhouri@chu-grenoble.fr)

Abstract

Introduction Pulmonary arterial hypertension (PAH) is a rare and severe disease for which most of the evidence about prognostic factors, evolution and treatment efficacy comes from cohorts, registries and clinical trials. We therefore aimed to develop and validate a new PAH identification algorithm that can be used in the French healthcare database "Système National des Données de Santé (SNDS)".

Methods We developed and validated the algorithm using the Grenoble Alpes University Hospital medical charts. We first identified PAH patients following a previously validated algorithm, using in-hospital ICD-10 (10th revision of the International Statistical Classification of Diseases) codes, right heart catheterisation procedure and PAH-specific treatment dispensing. Then, we refined the latter with the exclusion of chronic thromboembolic pulmonary hypertension procedures and treatment, the main misclassification factor. Second, we validated this algorithm using a gold standard review of in-hospital medical charts and calculated sensitivity, specificity, positive and negative predictive value (PPV and NPV) and accuracy. Finally, we applied this algorithm in the French healthcare database and described the characteristics of the identified patients.

Results In the Grenoble University Hospital, we identified 252 unique patients meeting all the algorithm's criteria between 1 January 2010 and 30 June 2022, and reviewed all medical records. The sensitivity, specificity, PPV, NPV and accuracy were 91.0%, 74.3%, 67.9%, 93.3% and 80.6%, respectively. Application of this algorithm to the SNDS yielded the identification of 9931 patients with consistent characteristics compared to PAH registries.

Conclusion Overall, we propose a new PAH identification algorithm developed and adapted to the French specificities that can be used in future studies using the French healthcare database.

Introduction

Pulmonary arterial hypertension (PAH) corresponding to group 1 of the pulmonary hypertension (PH) classification is a rare and severe disease with an estimated prevalence of 48 to 55 cases per million [1–3]. Given its scarcity, most of the evidence about risk and prognostic factors, evolution of the disease and treatment efficacy comes from cohorts and registries, on the one hand, and clinical trials on the other.

There is therefore a need to conduct large-scale epidemiological studies at the populational level to complement evidence from selected patients in expert centres included in registries and clinical trials [4]. However, routinely collected data (*i.e.* electronic health records and claims databases) are still underused in PAH, although they have great potential to assess risk factors of PAH or to conduct comparative effectiveness research. However, identifying patients with PAH in these databases is challenging due to the complexity and heterogeneity of the diagnostic criteria, as well as the lack of specific diagnostic codes dedicated to PAH. In the International Classification of Diseases Tenth Revision (ICD-10) coding system, PH diagnosis codes do not mirror the current World Health Organization (WHO) clinical classification [3, 5]. A systematic review of algorithms used to identify PAH in administrative databases showed that algorithms solely using ICD-10 PH diagnosis codes perform poorly with a positive predictive value (PPV) as low as 3% [6]. To overcome the limitation of the ICD-10 coding system, an increasing number of algorithms have been developed, but most of them have not been validated against patient medical charts [6, 7]. In response, a panel of PH experts convened to create best practices for the development of PAH identification algorithms in administrative databases [7]. Following these recommendations, GILLMEYER *et al.* [8] created and validated a collection of algorithms that can be used to identify PAH in administrative databases. However, these algorithms have never been validated in France using French healthcare administrative database "Système National des Données de Santé (SNDS)". The French healthcare database covers almost the entire French population and includes demographic data, healthcare encounters visits, drug and medical devices dispensing, chronic medical conditions, hospitalisations diagnoses, date and duration, procedures, diagnostic-related groups, and cause of death, making it powerful to study rare diseases and perform observational studies [9–11]. We therefore selected the algorithm displaying the best performance developed by GILLMEYER *et al.* [8] as a starting point to develop and validate an algorithm from the medical records of a PAH reference centre, *i.e.* Grenoble Alpes University Hospital, that can be used in the French SNDS to identify patients with PAH.

Methods

Data source

This study used two distinct data sources. Firstly, the algorithm development and validation study was conducted using data extracted from the Grenoble Alpes University Hospital medical charts. The Grenoble Alpes University Hospital is one of the PAH reference centres in France involved in the French PH Network PulmoTension [12, 13]. Secondly, we applied the final algorithm in the French healthcare database SNDS and compared the characteristics of the identified patients with those of the registries.

Algorithm development

Firstly, we applied the best performing GILLMEYER *et al.* [8] algorithm in identifying all adult patients hospitalised in Grenoble Alpes University Hospital with a PH diagnostic code (I27.0) between 1 January 2010 and 31 December 2022 [8]. We then excluded patients without specific PAH treatment 6 months prior or any time after PH diagnosis. The Anatomical Therapeutic Chemical (ATC) codes and "Unités Communes de Dispensation (UCD)" codes used to identify PAH treatments are detailed in supplementary table S1. Finally, we excluded patients without right heart catheterisation (RHC) performed 1 year prior or after PAH diagnosis to compose our first group of PAH cases.

Secondly, given patients with a chronic thromboembolic pulmonary hypertension (CTEPH or group 4 PH) often undergo RHC and receive treatments with pulmonary vasodilators, they could easily be misclassified as PAH. This was in fact identified by GILLMEYER *et al.* [8] as one of the limitations of their algorithm [8]. Therefore, we further excluded patients with CTEPH-specific procedures (*i.e.* pulmonary endarterectomy and balloon pulmonary angioplasty) and patients with a riociguat prescription to compose our second group of PAH cases. Codes used to identify RHC and CTEPH-specific procedures are detailed in supplementary table S2.

Algorithm validation

We used all medical charts of patients identified as PAH cases by the algorithms. We randomly selected the same number of medical records to compose our non-cases from the excluded patients (*i.e.* adult patients hospitalised in Grenoble Alpes University Hospital with a diagnostic code I27.0 and no RHC nor PAH treatment or no RHC and treated with PAH treatment or RHC and no-PAH treatment or with a CTEPH-specific condition).

Three authors (A. Hlavaty, C. Bernardeau and C. Jambon-Barbara) read two-thirds of all medical charts in parallel. Therefore, at least two persons read all the medical records blindly; divergences were resolved by discussion and case review among the team. Agreement between reviewers regarding presence of PAH was high, with a Cohen's κ value of 0.83 [14].

Statistical analysis

We created a 2×2 contingency table using reviewer-determined PAH as the gold standard, and we calculated performance characteristics for the final algorithm, *i.e.* specificity, sensitivity, PPV, negative predictive value (NPV) and accuracy. In line with previous studies, we have defined the performance results as follows: high when values were ⩾70%, modest between 50 and 70% and poor when values were <50% [8, 15].

Identification of PAH patients in the French healthcare database

We applied our final algorithm to identify all patients with PAH from 2008 to 2022 in the French healthcare database SNDS, which covers almost the entire French population. We compared patients' characteristics between ones identified in the SNDS by the algorithm and recent registry data in the literature [5, 12, 16]. The codes used for characteristics identification have previously been validated by French institutions [17–21] and are presented in supplementary table S3.

Ethics approval

For the Grenoble University Hospital medical records analysis, we received the authorisation from the Grenoble Institutional Medical Committee. For the algorithm use on the SNDS, neither committee approval nor informed consent was required because only anonymous data were used under the French Data Protection Supervisory Authority (CNIL) agreement.

Results

Study cohort

In the Grenoble Alpes University Hospital between 1 January 2010 and 30 June 2022, 2899 unique patients were hospitalised at least once for a PH (ICD-10 diagnosis code I27.0). We subsequently excluded 183 patients aged <18 years. Finally, we excluded 2219 patients with no PAH-specific treatment 6 months pre-diagnosis or at any time post-diagnosis and 120 patients with no RHC performed 1 year pre- or post-PH diagnosis. Overall, 377 adult patients were identified with PAH by the first algorithm and composed our first PAH group and 2339 patients composed our non-PAH group (figure 1). Then, we excluded 116 patients with a CTEPH-specific procedure and nine patients with a prescription of riociguat, thus our second PAH group was composed of 252 patients and 2464 patients were identified as non-PAH from which we randomly selected 252 patients to make up our non-case PAH group (figure 1).

Performance of algorithms

Among the 377 adult patients identified with PAH by the first algorithm, which composed our first PAH group, we retrieved only 177 genuine cases of PAH on reviewing medical charts (46.9%). Characterisation of the other 200 patients identified by the algorithm revealed considerable confusion with group 4 PH, *i.e.* CTEPH diagnosed in 122 patients (supplementary table S4).

Among the 252 adult PAH patients identified by the final algorithm (*i.e.* in further excluding 125 patients with a CTEPH-specific procedure or riociguat dispensing), we retrieved 171 PAH cases on reviewing medical charts (67.9%). Characterisation of the 81 patients wrongly classified as PAH retrieved only seven CTEPH cases, the majority being group 3 PH, which are associated with chronic respiratory diseases (table 1). Characterisation of the 252 randomly selected patients without PAH (*i.e.* non-cases) is detailed in table 1, and the 2×2 contingency table for the calculation of the different performance parameters is presented in supplementary table S5. Overall, the algorithm performed better than the first one with 67.9% PPV, a sensitivity of 91.0%, a specificity of 74.3%, an NPV of 93.3% and 80.6% accuracy.

Identification of PAH patients in the French healthcare database

Application of the final algorithm in the French administrative claims database SNDS retrieved 9931 patients with a PAH diagnosis between 1 January 2008 and 31 December 2022; that is, patients hospitalised with a diagnostic code I27.0 and with a PAH-specific treatment between 6 months pre-diagnosis or any time after and with an RHC procedure 1 year before or after diagnosis and no CTEPH-specific procedures nor riociguat dispensing.

Characteristics of patients identified from the SNDS are presented in table 2. As sickle cell disease is associated with group 5 PH, this characterisation informs us that 43 (0.43%) patients were misclassified as PAH.

Discussion

In this study we validated an algorithm to identify patients with PAH in the French healthcare database from medical charts of a French university hospital. We developed this algorithm based on previous validation studies conducted in the US administrative database Veterans Health Administration [8]. When

| 2899 | Patients hospitalised in Grenoble University Hospital with ICD-10 diagnostic code I27.0 between 1 January 2010 and 30 June 2022 |

| 2464 | Excluded patients |

| 183 | Excluded patients: aged <18 years |

| 252 | Patients randomly selected (1:1) PAH "non case" group |

| 2219 | Excluded patients: no treatment with pulmonary vasodilators 6 months prior to or any time after PH diagnosis |

| 120 | Excluded patients: no right heart catheterisation performed 1 year prior to or after the PH diagnosis |

| 377 | Adult patients hospitalised in Grenoble University Hospital with diagnostic code I27.0 and right heart catheterisation, and treated with pulmonary vasodilators
1st PAH group |

| 116 | Excluded patients: with a CTEPH-specific procedure |

| 9 | Excluded patients: with a CTEPH-specific treatment: riociguat |

| 252 | Adult patients hospitalised in Grenoble University Hospital with diagnostic code I27.0 and right heart catheterisation, treated with pulmonary vasodilators, and without a CTEPH-specific condition
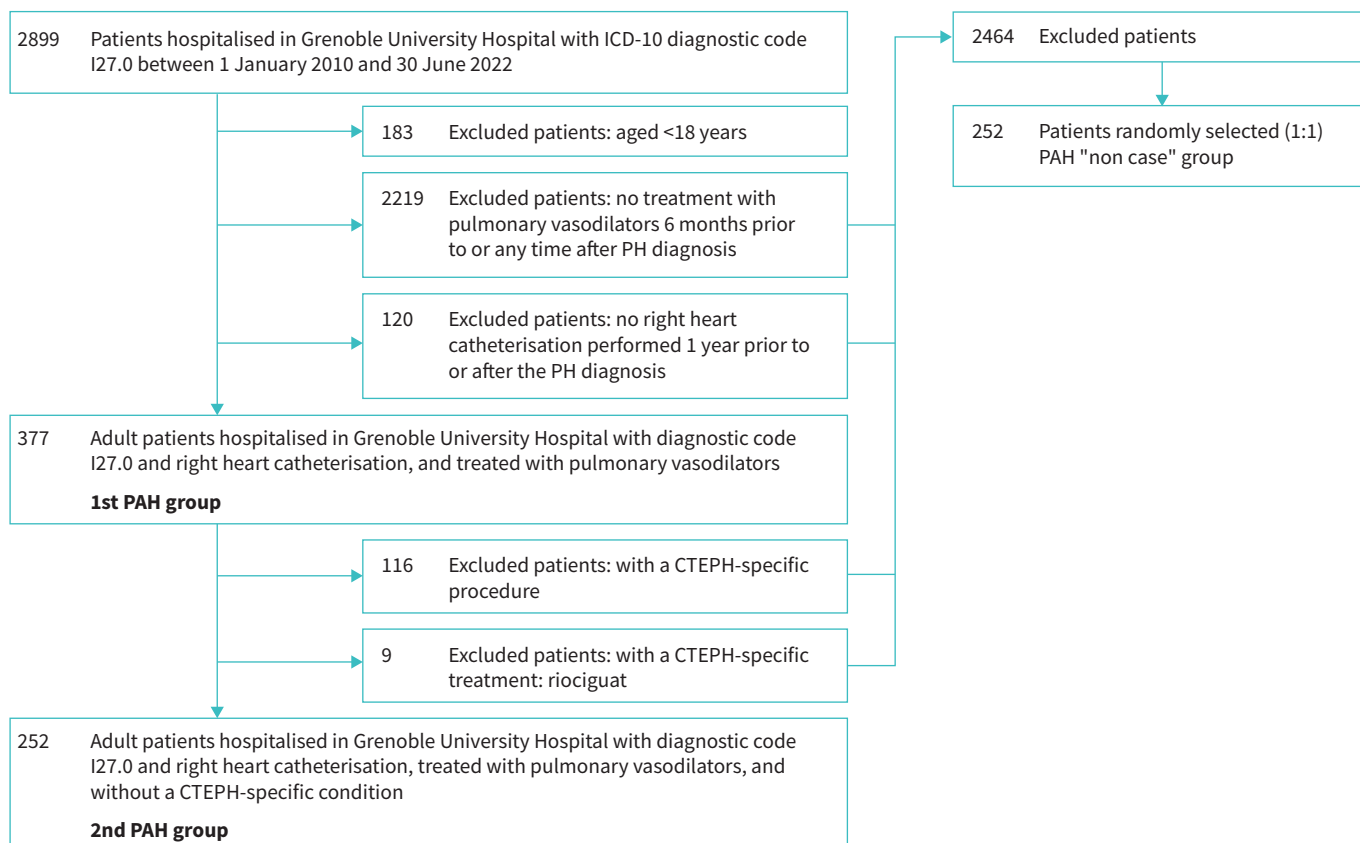**2nd PAH group** |

FIGURE 1 Flowchart of the algorithm identification of pulmonary arterial hypertension (PAH) cases and non-cases. ICD-10: International Classification of Diseases 10th revision; PH: pulmonary hypertension; CTEPH: chronic thromboembolic PH.

replicating the same algorithm in the Grenoble Alpes University Hospital, we found less convincing results with a PPV of about 50%. The addition of other decision rules, *i.e.* exclusion of patients with a CTEPH-specific procedure and treated by riociguat, has significantly improved the performance. Overall, the final algorithm achieved a rather good performance with a sensitivity higher than 90%, which means that the algorithm misses only a few PAH cases. Specificity was about 75%, and we can suppose that the patients who had undergone RHC and are under PAH-specific treatment have probably a mixed PH classified as disproportionate pulmonary hypertension in medical records but identified as group 3 PH in this study [22, 23].

In recent years, numerous code-based algorithms have been developed to identify PAH patients in administrative databases with highly variable performance [6, 24]. Overall, our study is in line with others showing that sensitivity decreases and specificity increases when further components are added to algorithms [6, 24, 25]. While very simple algorithms based solely on ICD-10 codes have shown insufficient performance to conduct studies in administrative databases, the choice of adding complexity to PAH identification must be tailored to the specific research question and the acceptable balance between sensitivity and specificity. Yet, in any case, sensitivity analyses by modifying algorithm definition and assessing the impact of misclassification error seem important to appraise the robustness of studies using PAH code-based algorithms.

After an in-depth review of all medical records of patients with genuine PAH but wrongly classified by the final algorithm, it appeared that for the vast majority they had undergone RHC earlier than 1 year pre-diagnosis. Further study may therefore explore the impact of the lag time between RHC and PH diagnosis on the performance of the algorithm. Moreover, recent studies using machine-learning approaches to identify and classify PH patients on electronic health records have proven their ability to identify patients who are likely to have pulmonary hypertension [26]. These methods could be used in combination with decision rules or may have the ability to identify new features to be included in algorithms to improve their performance. Lastly, linkage of in-hospital data repository or PAH registries

TABLE 1 Pulmonary hypertension (PH) characterisation of patients identified by the second algorithm

	Group 1 PH PAH	Group 2 PH	Group 3 PH	Group 4 PH CTEPH	Group 5 PH	No PH
PAH group	171 (67.9)	10 (3.97)	53 (21.0)	7 (2.78)	9 (3.57)	2 (0.79)
No-PAH group	17 (6.75)	105 (41.7)	41 (16.3)	38 (15.1)	3 (1.19)	48 (19.0)

Data are presented as n (%). PAH: pulmonary arterial hypertension; CTEPH: chronic thromboembolic PH.

with electronic health records may allow the development of new algorithms or machine-learning approaches against a high-quality gold standard group of patients with a harmonised procedure for diagnosis [27, 28].

The application of this algorithm in the French healthcare database retrieved 9931 patients with PAH between 1 January 2008 and 31 December 2022. These numbers are in line with the incidence of PAH in developed countries, which is estimated to be ~6 cases/million adults, corresponding to about 6300 newly diagnosed patients with PAH during the study period [2, 5]. The characteristics of the patients identified by our algorithm in the French SNDS were consistent with recent registry data highlighting the changing epidemiology of PAH, with older patients, a lower predominance of women and a higher burden of comorbidities at diagnosis [5, 16].

Several limitations of these algorithms should be noted. First, although the claims database can confirm that a patient underwent RHC, the result of this exam is not available. Thus, the possibility of patient

TABLE 2 Patients' characteristics in the Système National des Données de santé (SNDS) (n=9931)

Female	5222 (52.6)
Age years, mean±SD	61.4±16.9
PAH treatment	9931 (100)
Phosphodiesterase 5 inhibitors	4007 (40.4)
Sildenafil	2279 (23.0)
Tadalafil	1728 (17.4)
Endothelin receptor antagonists	5505 (55.4)
Ambrisentan	1483 (14.9)
Bosentan	3994 (40.2)
Sitaxentan	28 (0.28)
Guanylate cyclase stimulators	0 (0.00)
Riociguat	0 (0.00)
Prostacyclin analogues	392 (3.93)
Epoprostenol	279 (2.80)
Iloprost	46 (0.46)
Treprostinil	67 (0.67)
Prostacyclin receptor agonist	27 (0.27)
Selexipag	27 (0.27)
PAH-associated conditions	
Systemic scleroderma	1200 (12.1)
Portal hypertension	542 (5.46)
HIV positive	276 (2.78)
Congenital heart disease	673 (6.78)
Comorbidities	
COPD	995 (10.0)
Heart failure	4537 (45.7)
Diabetes	3122 (31.4)
Hypertension	8473 (85.3)
VTE	1449 (14.6)
Chronic dialysis	163 (1.64)
Sickle cell disease	43 (0.43)

Data are presented as n (%) unless otherwise specified. PAH: pulmonary arterial hypertension; VTE: venous thromboembolic events (*i.e.* pulmonary embolism, phlebitis and thrombophlebitis, other venous embolism and thrombosis).

misclassification remains. In addition, the unavailability of exam results in the French healthcare database prevented us from using ventilation/perfusion scan and computed tomography pulmonary angiography to further exclude CTEPH patients given that these procedures are also routinely performed in PAH patients to exclude a possible case of CTEPH. Second, with the multiplication of steps in an attempt to maximise specificity and minimise the risk of identifying CTEPH, the algorithm necessarily excludes some patients with genuine PAH, which likely reduces the sensitivity. Third, despite the fact that the non-PAH group was selected randomly, the possibility of sampling error remains, which can affect the precision of the estimates. Fourth, given that the algorithm validation was monocentric in the Grenoble University Hospital, one of the PAH reference centres in France, the generalisability elsewhere remains uncertain. In fact, as a reference centre the population is necessarily enriched for PAH as compared with the general population, and this could lead to an overestimation of the PPV. As prevalence of the disease is therefore lower in the SNDS, we could expect a lower PPV but a higher NPV than what we found in the algorithm validation. It would be interesting to replicate this study in other French hospitals. Moreover, the fact that the study was monocentric implied that we only had access to RHC and PAH-specific treatment performed and dispensed in the Grenoble Alpes University Hospital, meaning that we may have missed some patients who were diagnosed in Grenoble but performed RHC or took their treatment at another hospital. However, the number of patients who moved to another city or changed hospital during their medical care is supposed to be quite low, and this limitation should have only decreased PAH prevalence and therefore underestimated PPV.

Conclusion

Correctly identifying PAH is mandatory in order to conduct robust pharmaco-epidemiological studies. Starting from a US validated algorithm developed by GILLMEYER *et al.* [8] and adapted to the French specificities, we have developed and validated an algorithm with adequate performance for use in future studies using the French healthcare database.

References

1   Humbert M, Kovacs G, Hoeper MM, *et al.* 2022 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension. *Eur Respir J* 2023; 61: 2200879.

2   Leber L, Beaudet A, Muller A. Epidemiology of pulmonary arterial hypertension and chronic thromboembolic pulmonary hypertension: identification of the most accurate estimates from a systematic literature review. *Pulm Circ* 2021; 11: 2045894020977300.

3   Simonneau G, Montani D, Celermajer DS, *et al.* Haemodynamic definitions and updated clinical classification of pulmonary hypertension. *Eur Respir J* 2019; 53: 1801913.

4   Talwar A, Garcia JGN, Tsai H, *et al.* Health disparities in patients with pulmonary arterial hypertension: a blueprint for action. An Official American Thoracic Society Statement. *Am J Respir Crit Care Med* 2017; 196: e32–e47.

5   Humbert M, Kovacs G, Hoeper MM, *et al.* 2022 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension. *Eur Heart J* 2022; 43: 3618–3731.

6   Gillmeyer KR, Lee MM, Link AP, *et al.* Accuracy of algorithms to identify pulmonary arterial hypertension in administrative data: a systematic review. *Chest* 2019; 155: 680–688.

7   Mathai SC, Hemnes AR, Manaker S, *et al.* Identifying patients with pulmonary arterial hypertension using administrative claims algorithms. *Ann Am Thorac Soc* 2019; 16: 797–806.

8   Gillmeyer KR, Nunez ER, Rinne ST, *et al.* Development and validation of algorithms to identify pulmonary arterial hypertension in administrative data. *Chest* 2021; 159: 1986–1994.

9   Bezin J, Duong M, Lassalle R, *et al.* The national healthcare system claims databases in France, SNIIRAM and EGB: powerful tools for pharmacoepidemiology. *Pharmacoepidemiol Drug Saf* 2017; 26: 954–962.

10   Tuppin P, Rudant J, Constantinou P, *et al.* Value of a national administrative database to guide public decisions: from the système national d'information interrégimes de l'Assurance Maladie (SNIIRAM) to the système national des données de santé (SNDS) in France. *Rev Dépidémiologie Santé Publique* 2017; 65: S149–S167.

11   de Germay S, Conte C, Micallef J, *et al.* Performing pharmacoepidemiological studies using the French health insurance data warehouse (SNDS): how to translate guidelines into practice. *Therapie* 2023; 78: S0040-5957 (23)00026-4.

12   Humbert M, Sitbon O, Chaouat A, *et al.* Pulmonary arterial hypertension in France: results from a national registry. *Am J Respir Crit Care Med* 2006; 173: 1023–1030.

13   Montani D, Certain MC, Weatherald J, *et al.* COVID-19 in patients with pulmonary hypertension: a national prospective cohort study. *Am J Respir Crit Care Med* 2022; 206: 573–583.

14   Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159–174.

15   Carnahan RM, Moores KG. Mini-Sentinel's systematic reviews of validated methods for identifying health outcomes using administrative and claims data: methods and lessons learned. *Pharmacoepidemiol Drug Saf* 2012; 21: Suppl. 1, 82–89.

16   Lau EMT, Giannoulatou E, Celermajer DS, *et al.* Epidemiology and treatment of pulmonary arterial hypertension. *Nat Rev Cardiol* 2017; 14: 603–614.

17   Quantin C. Etude des algorithmes de définition de pathologies dans le Système Nationale d'Information Inter-Régimes de l'Assurance Maladie (SNIIRAM). Caisse National d'Assurance Maladie, 2015; p. 272.

18   Semenzato L, Botton J, Drouin J, *et al.* Chronic diseases, health conditions and risk of COVID-19-related hospitalization and in-hospital mortality during the first wave of the epidemic in France: a cohort study of 66 million people. *Lancet Reg Health Eur* 2021; 8: 100158.

19   Géodes – Santé publique France. Indicateurs: cartes, données et graphiques. Date last accessed: 25 October 2022. https://geodes.santepubliquefrance.fr/#view=map2&c=indicator

20   Hoisnard L, Pina Vegas L, Dray-Spira R, *et al.* Risk of major adverse cardiovascular and venous thromboembolism events in patients with rheumatoid arthritis exposed to JAK inhibitors *versus* adalimumab: a nationwide cohort study. *Ann Rheum Dis* 2023; 82: 182–188.

21   Marcombes C, Penso L, Weill A, *et al.* Persistence of second-line biologics in psoriasis after first-line biologic failure: a nationwide cohort study from the French health insurance database (SNDS). *Br J Dermatol* 2023; 189: 561–568.

22   Hoeper MM, Andreas S, Bastian A, *et al.* Pulmonary hypertension due to chronic lung disease: updated Recommendations of the Cologne Consensus Conference 2011. *Int J Cardiol* 2011; 154: Suppl. 1, S45–S53.

23   Seeger W, Adir Y, Barberà JA, *et al.* Pulmonary hypertension in chronic lung diseases. *J Am Coll Cardiol* 2013; 62: Suppl. 25, D109–D116.

24   Didden EM, Lu D, Hsi A, *et al.* Clinical evaluation of code-based algorithms to identify patients with pulmonary arterial hypertension in healthcare databases. *Pulm Circ* 2024; 14: e12333.

25   Sprecher VP, Didden E, Swerdel JN, *et al.* Evaluation of code-based algorithms to identify pulmonary arterial hypertension and chronic thromboembolic pulmonary hypertension patients in large administrative databases. *Pulm Circ* 2020; 10: 1–10.

26   Kogan E, Didden EM, Lee E, *et al.* A machine learning approach to identifying patients with pulmonary hypertension using real-world electronic health records. *Int J Cardiol* 2023; 374: 95–99.

27   Swinnen K, Quarck R, Godinas L, *et al.* Learning from registries in pulmonary arterial hypertension: pitfalls and recommendations. *Eur Respir Rev* 2019; 28: 190050.

28   Ong M, Klann JG, Lin KJ, *et al.* Claims-based algorithms for identifying patients with pulmonary hypertension: a comparison of decision rules and machine-learning approaches. *J Am Heart Assoc* 2020; 9: e016648.