Therapie xxx (xxxx) xxx—xxx

Disponible en ligne sur ScienceDirect www.sciencedirect.com

Elsevier Masson France EM|consulte www.em-consulte.com

PHARMACOEPIDEMIOLOGY

Use of the French National Health Data System (SNDS) in pharmacoepidemiology: A systematic review in its maturation phase

Olivier Maillard[a,b,c,1,*], René Bun[a,b,c], Moussa Laanani[a,d], Amandine Verga-Gérard[a,e], Taylor Leroy[a,e], Nathalie Gault[a,f], Candice Estellat[a,g], Pernelle Noize[a,h], Florentia Kaguelidou[a,i,j], Agnès Sommet[a,k], Maryse Lapeyre-Mestre[a,k], Annie Fourrier-Réglat[a,h], Alain Weill[a,l,2], Catherine Quantin[a,m,n,2], Florence Tubach[a,g,2]

[a] *Réseau de recherche en épidémiologie clinique et en santé publique/French Clinical Research Infrastructure Network (RECaP F-CRIN) Inserm network, 54500 Vandoeuvre-lès-Nancy, France*
[b] *Department of Public Health and Research, CHU de La Réunion, 97400 Saint-Pierre, Ile de La Reunion, France*
[c] *Clinical Investigation Center, INSERM CIC 1410, CHU de La Réunion, 97400 Saint-Pierre, Ile de La Reunion, France*
[d] *French National Health Insurance, 75000 Paris, France*
[e] *INSERM, CIC-EC 1433, 54100 Nancy, France*
[f] *INSERM, CIC-EC 1425, hôpital Bichat, 75018 Paris, France*
[g] *Sorbonne Université, INSERM, institut Pierre-Louis d'épidémiologie et de Santé publique, AP—HP, hôpital Pitié-Salpêtrière, département de Santé publique, centre de pharmacoépidémiologie (Cephepi), CIC-1901, 75000 Paris, France*
[h] *Université de Bordeaux, INSERM, BPH, U1219, Team AHeaD, CHU de Bordeaux, pôle de santé publique, service de pharmacologie médicale, 33000 Bordeaux, France*
[i] *INSERM, CIC-EC 1426, Department of Pediatric Pharmacology and Pharmacogenetics, Clinical Investigations Center, hôpital Robert-Debré, 75019 Paris, France*
[j] *UMR-1123, ECEVE, université Paris Diderot, Sorbonne Paris Cité, 75013 Paris, France*
[k] *Service de pharmacologie médicale et clinique, faculté de médecine, CIC 1436, CHU, université de Toulouse, 31000 Toulouse, France*

* Corresponding author. INSERM CIC 1410, centre hospitalier universitaire de La Réunion, BP 350, 97448 Saint-Pierre cedex, Ile de La Reunion, France
Adresse e-mail : olmaillard@yahoo.fr (O. Maillard).

[1] French research network supported by Inserm and French Clinical Research Infrastructure Network (F-CRIN) — https://recap-inserm.fr/home.html.

[2] These authors contributed equally.

https://doi.org/10.1016/j.therap.2024.05.003

Pour citer cet article : O. Maillard, R. Bun, M. Laanani et al., Use of the French National Health Data System (SNDS) in pharmacoepidemiology: A systematic review in its maturation phase, Therapies, https://doi.org/10.1016/j.therap.2024.05.003

[l] *Epiphare (French National Medicines Agency ANSM and French National Health Insurance CNAM), 93200 Saint-Denis, France*
[m] *Service de biostatistiques et d'information médicale (DIM), CHU Dijon Bourgogne, INSERM, université de Bourgogne, CIC 1432, module épidémiologie clinique, 21000 Dijon, France*
[n] *Université Paris-Saclay, UVSQ, Inserm, CESP, 94807 Villejuif, France*

Reçu le 30 janvier 2024 ; accepté le 16 mai 2024

KEYWORDS
Pharmacoepidemiology ;
Healthcare databases ;
Real world data ;
Systematic review

Summary
Aim of the study. — The French National Health Data System (SNDS) comprises healthcare data that cover 99% of the population (over 67 million individuals) in France. The aim of this study was to present an overview of published pharmacoepidemiological studies using the SNDS in its maturation phase.
Methods. — We conducted a systematic literature review of original research articles in the Pubmed and EMBASE databases from January 2012 until August 2018.
Results. — A total of 316 full-text articles were included, with an annual increase over the study period. Only 16 records were excluded after screening because they did not involve the SNDS but other French healthcare databases. The study design was clearly reported in only 66% of studies of which 57% were retrospective cohorts and 22% cross-sectional studies. The reported study objectives were drug utilization (65%), safety (22%) and effectiveness (9%). Almost all ATC groups were studied but the most frequent ones concerned the nervous system in 149 studies (49%), cardiovascular system drugs in 104 studies (34%) and anti-infectives for systemic use in 50 studies (16%).
Conclusion. — The SNDS is of growing interest for studies on drug use and safety, which could be conducted more in specific populations, including children, pregnant women and the elderly, as these populations are often not included in clinical trials.

Abbreviations

ATC	anatomic therapeutic chemical
COVID-19	coronavirus disease 2019
DCIR	French inter-scheme consumption datamart (*datamart des consommations interrégimes*)
EGB	permanent sample of beneficiaries (*échantillon généraliste des bénéficiaires*)
GDPR	General Data Protection Regulation
ICD-10	International Classification of Diseases-10th revision
LTD	long-term disease
MeSH	Medical Subject Headings
NHIS	National Health Insurance System
PMSI	French national hospital discharge database (*programme médicalisé des systèmes d'information*)
RECORD-PE	reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology
SI-DEP	population-based screening information system (*système d'information sur le dépistage populationnel*)
SNDS	French National Health Data System (*système national des données de santé*)

Introduction

Healthcare databases refer to all systems gathering information during patient care, and in particular electronic health records and claims databases. Electronic health records include data on demographics, clinical data, results of exams (biology, imaging, functional tests…), reports of surgery and procedures during hospital (hospital data warehouses) or ambulatory care. Claims databases arise from health insurance systems for reimbursement purpose, and include data on reimbursable care (hospitalizations, drug deliveries, procedures, exams, medical visits, physiotherapy…) and provide data on primary care, secondary care or mixed. These healthcare databases can be used to generate real world evidence that could support regulatory decisions but also guidelines on drug use [1]. The use of these complex databases, whose primary purpose was not research, in epidemiological (and in particular pharmacoepidemiological) studies has grown with the increase

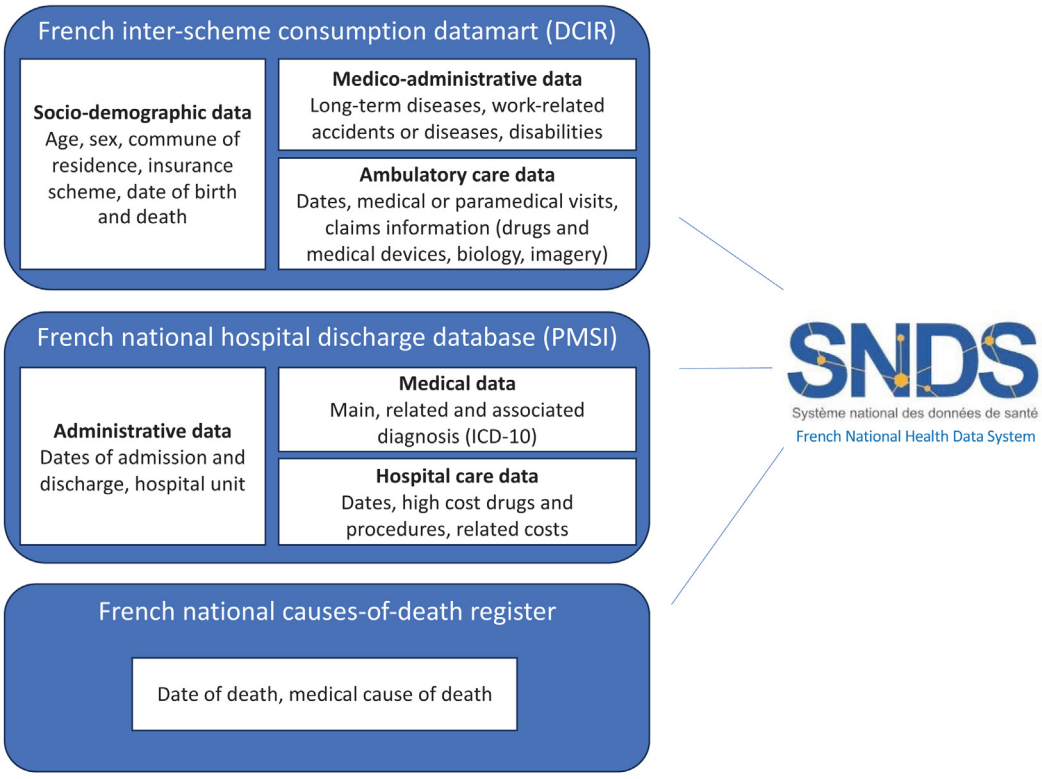

Figure 1. Presentation of the French National Health Data System (SNDS). ICD-10: International Classification of Diseases-10th revision.

in computer processing power [2] and the development of methods for causal inference in an observational context [3,4].

In France, the ''*système national des données de santé*'' (the French National Health Data System — SNDS) represents the most important national healthcare dataset. A detailed description of the SNDS database and its implications for pharmacoepidemiological studies is provided elsewhere [5,6]. It consists of individual data from three databases (Fig. 1): the inter-scheme consumption datamart, i.e. the national claims database (DCIR), the national hospital discharge database (PMSI), and the national causes-of-death register. The three databases are linked through a unique personal identifier allocated to each individual, from birth to death, covered by the National Health Insurance System (NHIS). The SNDS covers continuously around 99% of the French population, i.e. more than 67 million people [7]. The DCIR database gathers pseudonymized demographic data (including gender, dates of birth and death) and claims information, including dates of medical or paramedical visits, drugs and medical devices dispensed, the realization (but not the results) of laboratory tests, imaging procedures and other complementary exams, from all insurance schemes. The DCIR database also provides information on serious and costly long-term diseases (LTDs) status, encoded in the International Classification of Diseases — 10th revision (ICD-10), work-related accidents or diseases, disabilities and multiple pathologies eligible for full health insurance coverage. The PMSI database includes information on hospitalizations with ICD-10 codes for diagnoses, dates of admission and discharge, high cost drugs and procedures, and related costs. Recently, the national causes-of-death register has been linked to DCIR and PMSI, providing medical causes of death [7—11]. The SNDS database thus offers great opportunities to conduct pharmacoepidemiological studies on almost the entire French population. In recent years, due to the complexity of raw data analysis (for researchers), the ''*échantillon généraliste des bénéficiaires*'' (EGB), which is the 1/97th random permanent representative sample of the SNDS [9] has been developed for research purposes, particularly in pharmacoepidemiology.

The aim of our study was to conduct a systematic review to document the use of the SNDS for pharmacoepidemiological studies, in its maturation phase.

Methods

Search strategy

Eligibility criteria were articles reporting pharmacoepidemiological studies using French healthcare databases. A systematic literature search was performed in Medline via PubMed platform and Embase to identify articles published from January 1, 2012 corresponding to the start of the availability of the SNDS in its current form, to August 28, 2018 because of the change in access rules and the application of the General Data Protection Regulation (GDPR).

The search strategy combined Medical Subject Headings (MeSH) terms and free-text words as detailed in Appendix

1. A complementary search was performed using the names of well-known scientists and databases in the field of pharmacoepidemiology in France. No restrictions were made on the study design.

Review process

All papers from the literature search were screened on the basis of title and abstract. Articles were divided in 5 blocks, which were assigned to 5 pairs of reviewers. Within each pair, the two members reviewed each article separately to validate whether it met the eligibility criteria. In case of doubt or disagreement, the full-text article was reviewed by the whole working group, and inclusion or exclusion was decided by consensus. The reason for exclusion was recorded.

The systematic review was registered in Prospero as 42018096544 [12].

Data collection

Data were extracted using a standardized extraction form developed and validated by the working group. The extraction form was accessible via a LimeSurvey tool and is presented in Appendix 2. The form was tested on 8 articles extracted by all reviewers for validation.

The following items were captured: general characteristics of the publication (year of publication, journal, country of corresponding author, type of collaboration, institutions involved, funding sources), study features (study objectives and design, observation period, study population, exposure of interest and outcome), and databases involved. The study population was divided into specific groups: adults, children, pregnant women, and the elderly.

Type of study was categorized as drug utilization, effectiveness, safety, health economics or methodology. Drug utilization studies were defined as studies focusing on quantification of drugs, misuse or consistency with recommendations, patient compliance, or determinants of use. Studies were classified as cohort, cross-sectional, case-control or case-based studies including self-controlled case series, case crossover and case-time-control studies, or other less common designs. The exposure of interest was reported by pharmacological group (second-level anatomic therapeutic chemical [ATC] group) or medical device name, and outcomes were reported as disease or medical event according to ICD-10, use of care, death or drug/device use.

The use of an algorithm to identify the outcome of interest and whether this algorithm had been validated were also considered. By algorithm, we mean the use of combined data sources to obtain a composite outcome. Algorithms can combine ICD-10 codes, drug or device dispensations, use of care and/or causes of death [13—15]. All extracted characteristics were primarily as reported by the authors. Disagreements were solved by consensus within the working group. When no clear conclusion could be stated, the item was considered unclear.

Analysis

Data were described using numbers and proportions.

Secondary analyses were performed based on the geographic level (regional, national or international) of the studies and the time from the end of the study period to publication.

Analyses were conducted using Microsoft Excel® (Microsoft Corporation, Microsoft Excel, version 16.0, 2016).

Results

The initial search generated a list of 8657 articles over the study period (Fig. 2). Of these, 1028 were screened on the basis of title and abstract, of which 413 were assessed for eligibility on full text. A total of 316 studies were included in the analysis (Appendix 3), 86% of which were published in English-language journals.

The number of studies per year globally increased from 37 (12%) to 67 (21%) over the study period. The mean time from the end of the study period to publication was 3.7 years ranging from 1 to 13 years. Academic researchers were involved in 91% of the studies while 44% of the publications involved public health authorities of which 67% the French National Health Insurance and 20% regulatory agencies, and 12% private companies or associations. The corresponding author was from France in 98% of studies.

Twenty-six percent of studies involved at least one foreign-affiliated author (international collaboration). The studies were mainly published in specialized journals (82%) of which 30% in the field of pharmacology or pharmacoepidemiology, and 12% in the field of public health, epidemiology or health economics. The scope of specialized journals reported was broad: psychiatry, neurology, pain, cardiovascular, infectious diseases and diabetes were the most common.

Funding sources were not indicated in 40% of studies, while 25% reported public funding, 11% private funding, 6% mixed funding, 15% no funding and it was unclear in 3% of studies.

The objectives of the study were drug utilization in 65%, safety in 22% and effectiveness in 9%. Other objectives were health economics (9%) and methodological issues (7%) (Table 1).

Most studies (63%) were conducted on a national scale, 35% focused on one or more regions and 2% were limited to the hospital level.

Of the total, 56% were retrospective cohorts and 24% were cross-sectional studies (Table 2). The study design was clearly reported in the methods' section in only 66% of studies. Data were mainly collected before the study was conducted (88%) or secondary data were mixed with primary collections (12%). The vast majority of data were extracted from the claims databases (91%) and/or from the PMSI (17%). Details are provided in Table 3. SNDS databases were combined or linked with registers (4%), or clinical study databases (3%).

The study population was mainly identified by the condition of interest using specific drugs (62%), specific ICD-10 codes from PMSI data (27%) or LTD status (18%) and/or reimbursed procedure codes (8%). Age was an inclusion criterion in 54% of studies. Studies on specific age groups involved children in 19 studies (6%), elderly over 65 years of age in 20 (6%), or over 75 in 6 studies (2%). These studies were primarily focused on drug utilization or safety. Exposure was

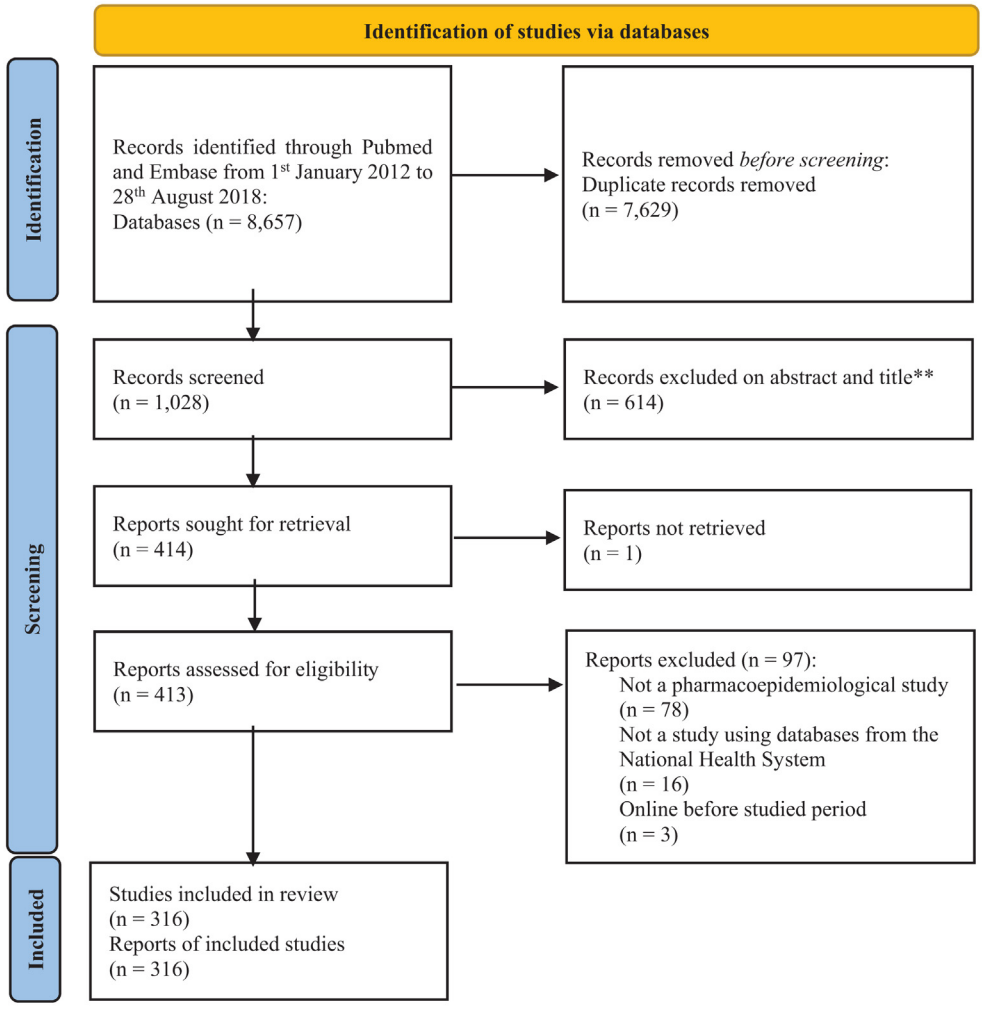

Figure 2. Flow chart of the data source selection.

Table 1 Distribution of included studies according to study objectives ($n = 316$).

Study objectives[a]	Total	%
Drug utilization	205	64.9
Quantification of drug utilization (incidence, prevalence, volume, care pathway-)	158	50.0
Consistency with recommendations/misuse	49	15.5
Patient compliance	27	8.5
Determinants of use	61	19.3
Effectiveness	28	8.9
Safety	68	21.5
Methodology	21	6.6
Health economics	28	8.9
Others	15	4.7

[a] Each study could have multiple objectives.

Table 2 Distribution of the included studies according to study design ($n = 316$).

Study design	Total	%
Number of studies in which study design is clearly reported in the methods' section	207	65.5
Retrospective cohort study	178	56.3
Cross-sectional study	75	23.7
Case-control study	15	4.7
Case-only designs (sub-question associated)	13	4.1
Self-controlled case series	3	0.9
Case cross-over	9	2.8
Ecological studies	7	2.2
Others	12	3.8

vaccine in 40% of studies targeting children (half in women only), and psycholeptics, blood or cardiovascular drugs in the elderly. Gender was used to define the study population in 43 studies (14%), mainly to select women (93%), of which one third involved pregnant women only for studying the use and safety of drugs when exposed to nervous system drugs or anti-infectives for systemic use in particular.

The exposure of interest was drugs in most studies (97%), including 6% of vaccines. The remaining 3% were medical

Table 3 Distribution of included studies according to databases from which data were extracted ($n = 316$).

Databases	Total	%
Claims databases	288	91.1
DCIR or subsystems[a]	131	41.5
EGB	86	27.2
Regional database	92	29.1
National hospital discharge database (PMSI)	54	17.1
French national causes-of-death register	3	0.9
Other registers[b]	10	3.2
Electronic hospital records	2	0.6
Database from a clinical study[c]	9	2.8
Others[d]	54	17.1

DCIR: Inter-scheme consumption datamart; EGB: permanent sample of beneficiaries (of the French national health insurance information system - SNIIRAM); PMSI: French national hospital discharge database.
[a] ERASME database (Extraction, Research, Analysis, Medico-economy).
[b] From national biomedicine agency, regional registers (on cancer, neuro-cardiovascular diseases, congenital malformations, rare diseases).
[c] Linked to the French National Health Data System (SNDS).
[d] Using the French National Health Data System (SNDS) and other databases, including databases of pharmacies, mother and child protection services, pharmacovigilance, EFEMERIS on risk assessment in pregnant women, police database of injurious crashes, MedQual-Ville on outpatient bacterial resistance, sentinel network of GPs for syndromic surveillance, ELIPPSE 40 cohort on breast cancer, regional addictovigilance databases, Medicam system for teleconsulting, spontaneous reporting system of the national public health agency.

devices (Table 4). Almost all ATC groups were studied but the most frequent ones concerned the nervous system in 149 studies (49%), of which psycholeptics and psychoanaleptics in 64% and analgesics in 14%, followed by cardiovascular system drugs in 104 studies (34%) and anti-infectives for systemic use in 50 studies (16%).

The most common exposures of interest in drug utilization studies were psycholeptics and psychoanaleptics (20% each), analgesics and antithrombotic agents (10% each), drugs used for diabetes (9%), vaccines and lipid modifying agents (7% each), antineoplastic agents, antibacterials for systemic use, antihypertensives and obstructive airway diseases (6% each). In safety studies, the most common exposures were psycholeptics (15%), drugs used in diabetes (12%), antithrombotic agents (10%), psychoanaleptics (9%), antiepileptics (7%), vaccines, antibacterials and corticosteroids for systemic use (6% each). Antithrombotic agents were the most frequently reported drugs in effectiveness studies (29%), followed by beta blocking agents, lipid modifying and agents acting on the renin—angiotensin system (25% each).

Outcomes concerned use of drug or medical device in 66% of studies, disease or medical event in 31%, use of care in 19% and death in 9% of studies (Table 5). In 27 studies (9%), outcome was identified using an algorithm, which had been previously validated in 5 studies.

Discussion

A total of 316 pharmacoepidemiologic studies involving the French National Health Data system (SNDS) published between January 1, 2012 and August 28, 2018 were identified. Only 16 records were excluded after screening because they did not involve the SNDS but other French healthcare databases.

To our knowledge, other reviews of pharmacoepidemiological studies involving healthcare databases are rare in the literature and most often outside the study period.

Furthermore, they are mainly aimed at describing the structure or availability of secondary healthcare data and assessing their usefulness, rather than describing their use in pharmacoepidemiology, or raise methodological issues, that do not allow us to compare our results [6,16—23]. In a literature review on the Nordic prescription databases which covered 26 million individuals, i.e. the entire population of 5 countries (Sweden, Denmark, Finland, Norway and Iceland), 515 publications in the field of pharmacoepidemiology were retrieved between 2005 and 2010, of which 262 (51%) were from Denmark [24]. Indeed, Danish data were available since 1989 and 1995 for the whole population (2005 for Sweden), which could explain the greater number of publications per year. Furthermore, 7% concerned children, 8% elderly over 65 yo and 5% pregnant women which is in agreement with our results.

In most of the articles, funding sources were not mentioned. This may be the effect of inconsistent submission guidelines for authors across journals. It could also be assumed that commercial funding sources were mentioned in most of cases, and that if no source of funding was mentioned, the study probably received public or nonprofit funding, or no funding at all. In any case, observational studies in pharmacoepidemiology were rarely funded.

Most articles reported on drug utilization studies (65%). But only 24% of them aimed at evaluating the consistency with guidelines of prescription although this represents one-third of drug utilization studies in the Nordic prescription databases over 2005—2010.

Indeed, healthcare databases are a very useful tool in this kind of studies, particularly because they allow the description of drug use in real life. The SNDS is of particular interest since it covers continuously around 99% of the French population, i.e. more than 67 million people, with a very low attrition rate as all social insurance schemes are combined in the SNDS [7]. Some of these drug utilization studies also intended to describe the penetration of newly marketed drugs such as the direct oral anticoagulants that were launched in France during the review period [25].

In our analysis, 30% of studies focused on the effectiveness and/or safety of drugs; it was 51% in the Nordic prescription databases over 2005—2010. When focusing on specific groups, they represented 46% of studies conducted in pregnant women, 32% of studies in children but only 2 studies (10%) in elderly over 65 years of age, compared with 74%, 42% and 28% of the studies, respectively, in the Nordic prescription databases.

Table 4 Distribution of included studies according to exposure and study objectives ($n = 306$[a]).

Exposure/study objectives[b]	Utilization	Effectiveness	Safety	Health economics	All	%[c]
Medical device	2		2	4	8	2.6
All drugs	8		2	1	11	3.6
Specific to a pharmacological group (ATC second level)					287	93.8
A00: Alimentary tract and metabolism	3		1		4	
A02: Drugs for acid related disorders	4				4	
A03: Drugs for functional gastrointestinal disorders	1		1		2	
A04: Antiemetics and antinauseants	1		1		2	
A07: Antidiarrheals, intestinal anti-inflammatory/anti-infective agents	1		1		2	
A10: Drugs used in diabetes	19		8	2	29	9.5
A11: Vitamins	4	1	1		6	
A12: Mineral supplements	3				3	
A16: Other alimentary tract and metabolism products	1				1	
B00: Blood and blood-forming organs	1				1	
B01: Antithrombotic agents	21	8	7	3	33	10.8
B02: Antihemorrhagics	1				1	
B03: Antianemic preparations	3				3	
B05: Blood substitutes and perfusion solutions	2				2	
C00: Cardiovascular system	5		1		6	
C01: Cardiac therapy	5	3	1		8	
C02: Antihypertensives	12	2	3		18	5.9
C03: Diuretics	8	4	2		12	3.9
C04: Peripheral vasodilators	1				1	
C05: Vasoprotectives			1		1	
C07: Beta blocking agents	10	7	2		17	5.6
C08: Calcium channel blockers	6	1	2		9	
C09: Agents acting on the renin—angiotensin system	10	7	3		17	5.6
C10: Lipid modifying agents	14	7	3		24	7.8
D00: Dermatologicals	2				2	
D05: Antipsoriatics	1				1	
D07: Corticosteroids, dermatological preparations	1				1	
D10: Anti-acne preparations			1		1	
G00: Genitourinary system and sex hormones	2		1		3	
G02: Other gynecologicals	1				1	
G03: Sex hormones and modulators of the genital system	7		3		9	
G04: Urologicals	2				2	
H00: Systemic hormonal preparations, excluding sex hormones and insulins	3		1		4	
H01: Pituitary and hypothalamic hormones and analogues			2		2	
H02: Corticosteroids for systemic use	6	2	4		11	3.6
H03: Thyroid therapy	2				2	
H05: Calcium homeostasis	1		1		2	
J00: Anti-infectives for systemic use	2		1		3	
J01: Antibacterials for systemic use	12	2	4		15	4.9
J02: Antimycotics for systemic use			1	1	2	
J04: Antimycobacterials	2				2	

Table 4 (*Continued*)

Exposure/study objectives[b]	Utilization	Effectiveness	Safety	Health economics	All	%[c]
J05: Antivirals for systemic use	6		2		8	
J06: Immune sera and immunoglobulins	2	1	1		2	
J07: Vaccines	15	2	4	1	18	5.9
L00: Antineoplastic and immunomodulating agents	1		1		2	
L01: Antineoplastic agents	13	4	5	4	19	6.2
L02: Endocrine therapy	7	2	3		8	
L03: Immunostimulants				1	1	
L04: Immunosuppressants	8		3	4	13	4.2
M00: Musculoskeletal system	1		1		2	
M01: Anti-inflammatory and antirheumatic products	8		2		8	
M02: Topical products for joint and muscular pain	1				1	
M03: Muscle relaxants	5			1	6	
M05: Drugs for treatment of bone diseases	5				5	
M09: Other drugs for disorders of the musculoskeletal system		1		1	1	
N00: Nervous system	2		1		3	
N01: Anesthetics	1		1		1	
N02: Analgesics	20		3		21	6.9
N03: Antiepileptics	11	1	5	1	16	5.2
N04: Anti-parkinson drugs			2		2	
N05: Psycholeptics	42		10	1	50	16.3
N06: Psychoanaleptics	42		6	2	45	14.7
N07: Other nervous system drugs	8	1	3		11	3.6
P01: Antiprotozoals	1				1	
R00: Respiratory system	1		1		2	
R01: Nasal preparations	1		1		2	
R02: Throat preparations	1				1	
R03: Drugs for obstructive airway diseases	11	2		1	12	3.9
R05: Cough and cold preparations	1				1	
R06: Antihistamines for systemic use	2		2		4	
S00: Sensory organs	1				1	
S01: Ophthalmologicals	5	1			6	
S02: Otologicals	1				1	
V00: Various	4		1		5	
V03: All other therapeutic products	2				1	

[a] After exclusion of 10 purely methodological studies.
[b] Each study could have multiple objectives.
[c] Percentage is reported above 10 studies.

Comparative designs to assess effectiveness or safety were less frequent, as they could challenge the definition of the study population or disease-specific outcomes [13]. Indeed outpatient diagnoses are missing in the SNDS and populations selected on the basis of hospital diagnoses are often unsuitable for real-life studies. And drug use may be difficult to evaluate when there are multiple indications for the same drug. However, innovative designs have emerged in recent years, which could increase the proportion of studies aimed at assessing drug efficacy or safety [26—30].

Studies based on the SNDS linked to other databases were rare. Such a linkage, whether direct on the unique personal identification number, or indirect when linked on several patient characteristics such as age, gender, or commune of residence, remained infrequent during the study period. Indeed linkages were rarely allowed in accordance with French privacy policy and, since 2018, the GDPR because it challenged the security of the linkage. Today, new procedures aim at making easier to obtain the authorization for linking field data to SNDS data in France.

The most frequently reported drug exposures (nervous system, cardiovascular, anti-infectives, metabolism), are consistent with the growing amount of drugs marketed in these areas in relation to the quantity of prescriptions, pharmacovigilance alerts, and the increasing incidence of certain conditions such as diabetes. These results are also consistent

Table 5 Distribution of included studies according to outcome ($n = 316$).

Outcomes	Total	%
Disease or medical event	99	31.3
A00-B99: Certain infections and parasitic diseases	5	
C00-D48: Neoplasms	7	
D50-D89: Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	2	
E00-E90: Endocrine, nutritional and metabolic diseases	4	
F00-F99: Mental and behavioural disorders	4	
G00-G99: Diseases of the nervous system	5	
H00-H59: Diseases of the eye and adnexa	4	
H60-H95: Diseases of the ear and mastoid process	0	
I00-I99: Diseases of the circulatory system	16	
J00-J99: Diseases of the respiratory system	3	
K00-K93: Diseases of the digestive system	4	
L00-L99: Diseases of the skin and subcutaneous tissue	0	
M00-M99: Diseases of the musculoskeletal system and connective tissue	6	
N00-N99: Diseases of the genitourinary system	1	
O00-O99: Pregnancy, childbirth and the puerperium	3	
P00-P96: Certain conditions originating in the perinatal period	0	
Q00-Q99: Congenital malformations, deformations and chromosomal abnormalities	2	
R00-R99: Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified	2	
S00-T98: Injury, poisoning and certain other consequences of external causes	4	
V01-Y98: External causes of morbidity and mortality	5	
Z00-Z99: Factors influencing health status and contact with health services	1	
U00-U99: Codes for special purposes	3	
Use of care	60	19.0
Hospitalization for any reason	11	3.5
Doctor's consultation	9	2.8
Other	38	12.0
Death	29	9.2
Drug/medical device use	207	65.5
Algorithm	27	8.5
Of which validated	5	

with studies conducted in other countries in Europe or elsewhere [14,24]. Most reported pharmacological classes are also of high worldwide relevance in terms of mortality and morbidity [15,31]. However, certain drugs mainly used in hospitals, such as injectable antineoplastics, are not studied in detail, as their cost is not shown separately in the database, but is included in the cost of the hospital stay.

Our literature review has some limitations. First, it may not have identified all the studies involving the French healthcare databases over the study period. Various translations of the names of healthcare database were used, as there is no official translation, so some studies may have been missed. In addition, for some studies, the use of databases was not clearly mentioned in the abstract. Second, reviewers were divided into pairs and assigned a list of articles, so not all the articles were evaluated by the same reviewers. However, an input guide was developed and validated by the group to avoid interpretation bias and regular meetings were held to exchange and find a collegial consensus. Third, the review period was limited to August 2018. This choice was made to document the use of healthcare databases in France before the change in regulatory requirements to access the SNDS for researchers as well as performing linkages of cohort databases to the SNDS. Another review could therefore be conducted in the next years to compare our results, in order to measure whether the number and type of pharmacoepidemiological studies have evolved. In addition, a specific review of pharmacoepidemiological studies published during the coronavirus disease 2019 (COVID-19) pandemic would be necessary to explore the use of the SNDS during this period and to present specific databases that have emerged, such as population-based screening information system (*système d'information sur le dépistage populationnel* — SI-DEP), a screening information system, which has recently been linked to the SNDS.

This systematic review highlights the strengths of the SNDS [10]. This is also the first one focusing on pharmacoepidemiological studies using the SNDS.

Overall, very few articles referred to reporting guidelines for such studies in the field of pharmacoepidemiology in healthcare databases. Therefore, the format of the articles is very heterogeneous, especially with regard to the methods section. This should be improved in the future by referring to the reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE) statement published in 2018 [32].

The expert group also suggests using either fixed translations of database names in foreign language articles or the French name to avoid misunderstanding (Appendix 4).

Conclusion

Studies based on healthcare databases are of increasing interest to researchers. The SNDS, the French National Health Data system, is the main healthcare database in France. It is of growing interest for studies on the effects and safety of drugs in the whole population, but also in specific populations, including children, pregnant women and elderly as these populations are often not included in clinical trials. As researchers gain greater access to the SNDS, and with the recent development of innovative designs, the number of studies would increase rapidly in the future, requiring further review and assessment of their potential for the implementation of public health policies and evidence-based guidelines.

Systematic review registration

Systematic review registration number in PROSPERO: CRD42018096544.

Support

Recherche en Épidémiologie Clinique et en Santé Publique (RECaP network), French Clinical Research Infrastructure Network (F-CRIN), Institut National de la Santé et de la Recherche Médicale (Inserm), France.

Funding statement

No funding.

Data availability statement

All relevant data are within the manuscript and its supporting information files.

Ethics approval statement

Not applicable.

Patient consent statement

Not applicable.

Authors' contributions

Conceptualization, supervision and methodology: AS, AW, CQ and FT.

Investigation: CE, AFR, FK, MLM, PN, CQ, AS, FT, AW, CYN, NG, ML, RB, OM.

Data curation and formal analysis: TL, AVG, ML, RB, OM.

Writing — original draft: OM, RB.

Writing — review and editing: ML, AVG, TL, CE, AFR, FK, MLM, PN, CQ, AS, FT, AW, CYN, NG.

Acknowledgments

This work was conducted in a working group of the national network RECaP of Research in Clinical Epidemiology and in Public Health supported by French National Institute for Health and Medical Research (Inserm) and French Clinical Research Infrastructure Network (F-CRIN), coordinated by Francis Guillemin.

Disclosure of interest

The authors declare that they have no competing interest.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.therap.2024.05.003.

Références

[1] Pacurariu A, Plueschke K, McGettigan P, Morales DR, Slattery J, Vogl D, et al. Electronic healthcare databases in Europe: descriptive analysis of characteristics and potential for use in medicines regulation. BMJ Open 2018;8:e023090.
[2] Dicuonzo G, Galeone G, Shini M, Massari A. Towards the use of big data in healthcare: a literature review. Healthcare (Basel) 2022;10:1232.
[3] Schneeweiss S. Automated data-adaptive analytics for electronic healthcare data to study causal treatment effects. Clin Epidemiol 2018;10:771—88.
[4] Wyss R, Yanover C, El-Hay T, Bennett D, Platt RW, Zullo AR, et al. Machine learning for improving high-dimensional proxy confounder adjustment in healthcare database studies: an overview of the current literature. Pharmacoepidemiol Drug Saf 2022;31:932—43.
[5] Martin-Latry K, Bégaud B. Pharmacoepidemiological research using French reimbursement databases: yes we can ! Pharmacoepidemiol Drug Saf 2010;19:256—65.
[6] Palmaro A, Moulis G, Despas F, Dupouy J, Lapeyre-Mestre M. Overview of drug data within French health insurance databases and implications for pharmacoepidemiological studies. Fundam Clin Pharmacol 2016;30:616—24.
[7] Tuppin P, Rudant J, Constantinou P, Gastaldi-Ménager C, Rachas A, de Roquefeuil L, 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 Epidemiol Sante Publique 2017;65(Suppl. 4):S149—67.
[8] Moulis G, Lapeyre-Mestre M, Palmaro A, Pugnet G, Montastruc JL, Sailler L. French health insurance databases: what interest for medical research? Rev Med Interne 2015;36:411—7.
[9] Bezin J, Duong M, Lassalle R, Droz C, Pariente A, Blin P, et al. The national healthcare system claims databases in France, SNIIRAM and EGB: powerful tools for pharmacoepidemiology. Pharmacoepidemiol Drug Saf 2017;26:954—62.

[10] Scailteux LM, Droitcourt C, Balusson F, Nowak E, Kerbrat S, Dupuy A, et al. French administrative health care database (SNDS): the value of its enrichment. Therapie 2019;74:215—23.
[11] Rey G. Death certificate data in France: production process and main types of analyses. Rev Med Interne 2016;37:685—93.
[12] Verga-Gerard A, Estellat C, Fourrier-Réglat A, Kaguelidou F, Lapeyre-Mestre M, Noize P, et al. The use of French medico-administrative databases for pharmacoepidemiology studies. PROSPERO; 2018. CRD42018096544 https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42018096544. [(Accessed 16 May 2024)].
[13] Weinstein EJ, Ritchey ME, Lo Re V. Core concepts in pharmacoepidemiology: validation of health outcomes of interest within real-world healthcare databases. Pharmacoepidemiol Drug Saf 2023;32:1—8.
[14] Fuentes AV, Pineda MD, Venkata KCN. Comprehension of top 200 prescribed drugs in the US as a resource for pharmacy teaching, training and practice. Pharmacy (Basel) 2018;6:43.
[15] Drug-related deaths and mortality in Europe: update from the EMCDDA expert network Luxembourg: Publications Office of the European Union; 2019. https://www.drugsandalcohol.ie/30861/. [Accessed 16 May 2024].
[16] Yoshida S, Takeuchi M, Tanaka-Mizuno S, Mizuno K, Nakashima M, Fukasawa T, et al. Clinical epidemiology and pharmacoepidemiology studies with real-world databases. Proc Jpn Acad Ser B Phys Biol Sci 2022;98:517—28.
[17] García Rodríguez LA, Pérez Gutthann S. Use of the UK general practice research database for pharmacoepidemiology. Br J Clin Pharmacol 1998;45:419—25.
[18] Franco JS, Vizcaya D. Availability of secondary healthcare data for conducting pharmacoepidemiology studies in Colombia: a systematic review. Pharmacol Res Perspect 2020;8:e00661.
[19] Durán CE, Christiaens T, Acosta Á, Vander Stichele R. Systematic review of cross-national drug utilization studies in Latin America: methods and comparability. Pharmacoepidemiol Drug Saf 2016;25:16—25.
[20] Sturkenboom M, Schink T.Databases for pharmacoepidemiological research. Cham, Switzerland: Springer International Publishing; 2021 [doi:10.1007/978-3-030-51455-6].
[21] Williams T, van Staa T, Puri S, Eaton S. Recent advances in the utility and use of the general practice research database as an example of a UK primary care data resource. Ther Adv Drug Saf 2012;3:89—99.
[22] Takahashi Y, Nishida Y, Asai S. Utilization of health care databases for pharmacoepidemiology. Eur J Clin Pharmacol 2012;68:123—9.
[23] Trifirò G, Gini R, Barone-Adesi F, Beghi E, Cantarutti A, Capuano A, et al. The role of European healthcare databases for post-marketing drug effectiveness, safety and value evaluation: where does Italy stand? Drug Saf 2019;42:347—63.
[24] Wettermark B, Zoëga H, Furu K, Korhonen M, Hallas J, Nørgaard M, et al. The Nordic prescription databases as a resource for pharmacoepidemiological research — a literature review. Pharmacoepidemiol Drug Saf 2013;22:691—9.
[25] Huiart L, Ferdynus C, Renoux C, Beaugrand A, Lafarge S, Bruneau L, et al. Trends in initiation of direct oral anticoagulant therapies for atrial fibrillation in a national population-based cross-sectional study in the French health insurance databases. BMJ Open 2018;8:e018180.
[26] Schneeweiss S, Schneeweiss M. Concepts of designing and implementing pharmacoepidemiology studies on the safety of systemic treatments in dermatology practice. JID Innov 2023;3:100226.
[27] Rassen JA, Glynn RJ, Brookhart MA, Schneeweiss S. Covariate selection in high-dimensional propensity score analyses of treatment effects in small samples. Am J Epidemiol 2011;173:1404—13.
[28] Wang SV, Schneeweiss S, RCT-DUPLICATE Initiative, Franklin JM, Desai RJ, Feldman W, et al. Emulation of randomized clinical trials with nonrandomized database analyses: results of 32 clinical trials. JAMA 2023;329:1376—85.
[29] Chow SC, Huang Z. Innovative design and analysis for rare disease drug development. J Biopharm Stat 2020;30:537—49.
[30] Gault N, Castañeda-Sanabria J, De Rycke Y, Guillo S, Foulon S, Tubach F. Self-controlled designs in pharmacoepidemiology involving electronic healthcare databases: a systematic review. BMC Med Res Methodol 2017;17:25.
[31] Glei DA, Preston SH. Estimating the impact of drug use on US mortality, 1999—2016. PLoS One 2020;15:e0226732.
[32] Langan SM, Schmidt SA, Wing K, Ehrenstein V, Nicholls SG, Filion KB, et al. The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). BMJ 2018;363:k3532.