The value of a health insurance database to conduct pharmacoepidemiological studies in oncology

C. Conte, C. Vaysse, Pauline Bosco-Levy, Pernelle Noize, Annie Fourrier-Reglat, F. Despas, M. Lapeyre-Mestre

▶ **To cite this version:**

C. Conte, C. Vaysse, Pauline Bosco-Levy, Pernelle Noize, Annie Fourrier-Reglat, et al.. The value of a health insurance database to conduct pharmacoepidemiological studies in oncology. Thérapie, 2019, 74 (2), pp.279-288. 10.1016/j.therap.2018.09.076 . hal-03212201

HAL Id: hal-03212201
https://hal.science/hal-03212201v1

Submitted on 25 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Version of Record: https://www.sciencedirect.com/science/article/pii/S0040595719300241
Manuscript_91ba5163d89a1073651a7c748aadf337

THERAPIES

HEADING: PHARMACOEPIDEMIOLOGY

EPUB AHEAD OF PRINT THEN NUMBER 2 (March April) 2019

The value of a health insurance database to conduct pharmacoepidemiological studies in oncology

Pharmacoepidemiological study in oncology using claims database

Cécile Conte[a,b,c,*], Charlotte Vaysse[d], Pauline Bosco[e,f,g], Pernelle Noize[e,f,g], Annie Fourrier-Reglat[e,f,g], Maryse Lapeyre-Mestre[a,b,c], Fabien Despas[a,b,c]

[a] *Laboratoire de pharmacologie médicale et clinique, faculté de médecine, université de Toulouse III, 31000 Toulouse, France*

[b] *Service de pharmacologie médicale et clinique, CIC 1436, CHU de Toulouse, 31000 Toulouse, France*

[c] *UMR 1027 Inserm-université de Toulouse, 31000 Toulouse, France*

[d] *Département de chirurgie oncologique et gynécologique, CHU de Toulouse, 31000 Toulouse, France*

[e] *Université de Bordeaux, INSERM, Bordeaux population health research center, team pharmacoepidemiology, UMR 1219, 33000 Bordeaux, France*

[f] *Bordeaux PharmacoEpi, INSERM CIC1401, 33000 Bordeaux, France*

© 2019 published by Elsevier. This manuscript is made available under the CC BY NC user license
https://creativecommons.org/licenses/by-nc/4.0/

[g] *CHU de Bordeaux, pôle de santé publique, service de pharmacologie médicale, 33000 Bordeaux, France*

Received 25 September 2018; accepted 29 September 2018

*__Corresponding author__. Service de pharmacologie médicale et clinique, 37 allées Jules Guesde, 31000 Toulouse ;

E-mail adress: cecile.conte@univ-tlse3.fr (C. Conte)

Abbreviations

ALD: affection de longue durée

CepiDC: *centre d'épidémiologie sur les causes médicales de décès*

CMS: centers for Medicare & Medicaid services

CNODES: Canadian Network for Obervational Drug Effect Studies

DDD: defined daily dose

EMA: European Medicne Agency

ICD-10: international statistical classification of diseases and related health problems 10th revision

LTD: long term disease

M-CAHPS: Medicare consumer assessment of healthcare providers and systems

MPR: medication possession ratio

NCI: National Cancer Institute

NorPEN: Nordic PharmacoEpidemiological Network

OMOP: observational medical outcomes partnership

OTC: over-the-counter

PIC: personal identity code

PMSI: French hospital discharge database (*programme de medicalisation des systèmes d'information*)

PROTECT: pharmacoepidemiological research on outcomes of therapeutics

SEER: surveillance, epidemiology and end results program

SNDS: French national health insurance database (*système national des données de santé*)

Summary

Some concerns have emerged about the evidence of benefits on survival outcomes or quality of life of new anticancer drugs. In parallel, the decreased cancer mortality leads to an increased number of patients exposed to cancer-treatment-related consequences. In this context, pharmacoepidemiology is crucial to assess anti-cancer drug use, effectiveness and safety in real life conditions. We aimed to describe strengths, limitations and considerations associated with the use of the French national health insurance database (*système national des données de santé* [SNDS]) to conduct pharmacoepidemiological studies in oncology. The SNDS represents a powerful tool in pharmacoepidemiology owing to its extensive coverage, accurate description and quantification of drug exposure and individual data on patients. The main limitations of this database ensue from the administrative nature resulting in technical difficulties in its management and gaps in availability of data. Another limitation is the lack of accurate identification of diseases, comorbidities or outcomes and potential confounding with notably the lack of data regarding cancer stage, prognosis or risk factors. Finally, the accurate identification of the nature of chemotherapy received by patients is sometimes complex. To minimize these limitations, several approaches and statistical methods could be used as highlighted by national or international initiatives. First, the SNDS may be linked with cancer registry or clinical data. Then, several data sources could be combined using meta-analytical methods. The development of methodological tools and the use of standardized methods are crucial to enhance the quality of studies that can impact clinical practice and guide public decision. Pharmacoepidemiological approaches and pharmacovigilance represent an important cornerstone in oncology for signal detection or long-term follow up of cancer patients. In this context, validated methods to identify cancer patients and to describe chemotherapy regimens within these data should be promoted and remain too scarce despite international guidelines. Moreover, limits and strength of each data sources should be systematically discussed according to the research question. Optimized and framed use of claims database represents a future challenge in onco-pharmacoepidemiology.

KEYWORDS

Pharmacoepidemiology; Oncology; Electronic claims databases; Electronic health care databases; Anticancer drugs; Safety; Benefit; Utilization

Introduction

Nearly 40 % of men and women will be diagnosed with cancer of any site at some point during their lifetime with a median age at diagnosis of 66 years. Breast, bronchopulmonary, prostate and colorectal cancers account for approximately half of all new cancer cases. Cancer represents the leading cause of death for patients less than 65 years of age. The five year relative survival rate varies between 4 and 98% according to cancer types [1]. Cancer can be treated with systemic cytotoxic chemotherapy and/or with a new heterogeneous group of anticancer drugs presented as "targeted" (monoclonal antibodies, protein-kinase inhibitors). This group represented 60% of anticancer drugs' marketing authorizations between 2012 and 2015 [2]. These drugs' marketing authorizations are based on results of randomized controlled trials with strict selection criteria, low number of exposed patients and limited duration of exposure [3–7]. In addition, concerns have emerged about evidence of benefits on survival outcomes or quality of life of anticancer drugs. From 2009 to 2013, the European Medicines Agency (EMA) approved the use of 48 anticancer drugs in 68 indications while only 51% of them showed a significant improvement on survival or quality of life [8]. In parallel, there is an average decreased of death rates of 1.5 % each year over 2006-2015 with an increased number of patients exposed to treatment cancer-related consequences [1]. Adverse drug reactions, potential misuse, second cancer, polypharmacy and drug-drug interactions with cancer specific drugs or non-cancer specific drug are examples of these adverse outcomes. Therapeutic progresses have led to the development of new paradigms in the management of some cancer types considered today as chronic diseases. In this context, most of patients are now essentially treated as outpatients with the development of oral anticancer drugs. Taking into account the above considerations, pharmacoepidemiological studies became crucial in oncology to describe patterns of use and to assess benefits and harms of anticancer drugs in real life conditions during the active treatment phase of cancer but also during survivorship. Owing to numerous advantages, health insurance administrative databases are frequently used to conduct pharmacoepidemiological studies and to guide public decision in many countries [9–17]. The French "*système national des données de santé*" (SNDS) is one of the largest health insurance databases in the World. This database contains anonymous claims from outpatients and inpatients care linked with the French hospital discharge database (*programme de médicalisation des systèmes d'information* [PMSI]). The SNDS has been also recently implemented with cause of death by the linkage with death registry (*centre d'épidémiologie sur les causes médicales de décès* [CepiDC]). It

may be used as a pertinent and complementary tool for onco-pharmacoepidemiology studies because of several strengths that can minimize classic bias related to other sources. However, some biases and technical difficulties need to be acknowledged before widely using these complex and administrative data developed for management purpose. In this way, reflections in future challenge and perspectives are needed to compensate these limits and to improve reliability of future studies. Moreover, validated and standardized methods remain scarce and need to be developed to improve validity of future onco-pharmacoepidemiology studies.

In this context, we aimed:

1) To describe methodological tools, strengths, limits and considerations related to the use of the SNDS to conduct pharmacoepidemiological studies in oncology;

2) To present international initiatives proposed to optimize the use of a health insurance database and minimize potential biases.

The *système national des données de santé* (SNDS)

Exhaustive coverage

The SNDS has been described in details elsewhere [14,18–21]. This national database provides extensive data encompassing a population of more than 65 million inhabitants. It provides anonymous and individual data on patient characteristics with demographic data, long-term chronic diseases (long term disease [LTD] so called in French: *affections de longue durée* [ALD]) and vital status. The access to ambulatory health care consumption (reimbursed drugs and medical acts) and the linkage with data from the national hospital database *(*medical information systems programme, so called in French: *programme de médicalisation des systèmes d'information* [PMSI]) gives a complete overview of patients care pathway for several years all over France [14,18–21]. The large

number of patient recorded in this database allows increasing statistical power of analyses especially when studying rare cancer [22–24]. Moreover, completeness of the data could minimize selection bias related to the constitution of specialized cancer center's cohorts and attrition bias related to long-term follow-up. Selection bias is an important issue with results not always transposable to the target population. However, it seems crucial to understand that SNDS provides exhaustive data on all reimbursement of care but not exhaustive data for the identification of prevalent and incident cancer cases as highlighted in the literature.

Identification of incident cases of cancer

Accurate and exhaustive identification of incident cases of cancer has been studied in oncology in order to estimate cancer incidence even in areas were no cancer registry exists. In this context, validation studies have been performed using cancer registries or medical records as gold standard. Performance of algorithms detection varies between 30% and 94% for sensitivity and between 20% and 90% for predictive positive value according to type of cancer, type of sources used and type of codes used to detect cases. It is important to note that some algorithms gave incidence close to those from registries, but this is due to the fact that false-negatives and false-positives have similar frequencies and counterbalanced each other [22,25–35]. Table 1 and table 2 summarize validated algorithms of incident cancer patient selection in the SNDS available in the literature. We identified only five validation studies for solid tumors conducted before 2006. Yet, PMSI data reliability improved after 2007 and we can assume that the results of these studies would be different today. Heterogeneity in performances of detection can be explained by the heterogeneity of cancer care pathway and questioned on the relevance of the use of PMSI data to select new cases in certain cancer subtypes. For example, some hematologic malignancies such as chronic lymphocytic leukemia are non-progressive at diagnosis and do not require active treatment and related hospitalization. When exploratory analyses were conducted, they showed that cancer patients not detected as incident or identified in the PMSI database are more likely to be old, with localized stage of cancer [22,28]. These results suggest that sensitivity analyses including only treated cancer patients would have been of interest but this information is often lacking in cancer registry database. Hence, false-negatives may concern patients never hospitalized for their cancer because of different

disease management and/or a gap between diagnosis and treatment. Moreover, the use of ICD-10 to classify cancer by subtypes often lacks of precision and is sometimes ambiguous because of the multiplicity of codes. Finally, coding errors could be another reason of misclassification. Therefore, sources of detection are of primary importance. The use of LTD data alone for identifying cancer patients in claims database resulted in poor performances with sensitivity lower than 50% for many cancers [27]. Despite these limits, the SNDS remains a source of interest in onco-pharmacoepidemiology for many cancer types notably with regard to precise quantification of drug exposure. In fact, exhaustive data is not needed in pharmacoepidemiology but only representative sample of patients requiring drug therapy. Moreover, exhaustive data are not available in other data sources and some rare cancers such as Hodgkin lymphoma are well detected using these databases.

Drug exposure

The SNDS is a powerful tool to describe and quantify drug exposure in patients as described elsewhere [19,20]. First, it contains claims from drugs dispensed by community pharmacies. Each drug packages are identified by a unique CIP (*club inter-pharmaceutique*) code providing for each drug, their anatomical therapeutic chemical (ATC) codes, the number of items and dosage for each item. For each drug, date of prescribing, date of dispensing and information on prescribers are available. Hence, the SNDS enables to estimate a large amount of drug exposure indicators such as duration of treatment, discontinuation and number of treatment cycles, number of defined daily doses (DDDs) or medication possession ratio (MPR). These data allow to assess compliance and associated factors to oral anticancer drugs which become of great importance as regards the increasing marketing of this route of administration for cancer treatment. Hence, this database provides accurate information to model drug exposure with minimized information bias (recall bias, non-response bias or reporting bias).

However, data availability in the SNDS is not continuous and gaps in availability must be accounted when designing a pharmacoepidemiological study in order to minimize bias. Drugs dispensed over the counter or during hospital stays are not available from the SNDS with potential classification bias. As described previously by Palmaro et al., other factors of potential gap in

availability of drug exposure should also be checked to minimize bias [19]. However, innovative and costly drugs (established by decree), drugs dispensed by hospital to ambulatory patients (so called: “retrocession”) and drugs not approved on the basis of the programme of temporary authorization of use are available. This is of great interest in oncology, because new anticancer drugs are often included in this category.

Another fundamental issue in onco-pharmacoepidemiology is the lack of method to identify multidrug chemotherapy regimens in these databases with only one publication in multiple myeloma patients in France [36]. In fact, sequence of treatments (chemotherapy, radiotherapy or stern cells transplant) can be identified using diagnosis procedure codes in the PMSI database but the type of chemotherapy protocol is not always fully available. In fact, old and conventional anticancer drugs are not recorded. Yet, it is essential to identify the type of chemotherapy regimen and number and rank of treatment lines to describe with accuracy care pathway and to assess survival, effectiveness or safety. One possible approach is using recorded drugs (costly or retrocession) as tracers of chemotherapy protocol when available or specific. Another approach could be the linkage of SNDS data with other hospital sources giving access to the nature of chemotherapy received by patients but also other potential confounding factors not available in the SNDS as done by several international initiatives.

Confusion bias

The major limitation to conduct onco-pharmacoepidemiology studies with the SNDS is confusion bias. The lack of clinical data such as cancer stage, prognosis factors, anatomo-pathology reports, result of diagnosis tests does not permit to consider data according to prognosis markers. Moreover, there is little social information and no information on lifestyle habit of patients (e.g.: smoking status, addictions etc.) [18,20]. Moreover, adverse drug reactions not requiring a specific treatment such as drug initiation, medical procedure or hospitalization are not detectable in the SNDS. Finally, the SNDS enables to identify the reimbursement of medical act but the results are not registered in the database. This is particularly the case for medical biology examinations including pharmacogenetic explorations and could be a limit as regards the request of genetic exploration for

the initiation of anticancer treatments [37,38]. The pharmacoepidemiology approach does not allow making a direct link between the status for a gene and the use of a drug at the population level. In addition, reimbursement of pharmacogenetics explorations is still carried out today only on non-nomenclature acts ("*biologie hors nomenclature*"), which makes their identification in the databases more complex. The attribution of dedicated codes could facilitate the identification of the realization of such acts and thus be able to study the proportion of exploration performed before the introduction of certain drug treatments. All these limitations highlight the complementarity of sources that can be used in onco-pharmacoepidemiology. Several cohort studies have been conducted to describe patterns of drug use and to assess effectiveness in real life conditions. These cohorts included data from university hospitals, cancer care centers, private clinics or community hospitals with patients identified by physicians or anticancer drugs by nominative hospital pharmacy dispensation [39–42]. Then, pharmacovigilance in oncology based on spontaneous report remains irreplaceable to detect and prevent adverse drug reaction in oncology [43,44]. To minimize confusion bias, linkage with observational cohort data is a perspective of interest as highlighted by international initiatives [45]. Then, statistical methods such as propensity scores or the use of self-controlled designs can be used to deal with confounding factors [20,46]. It should be noted that since 2016, owing respecting regulation processes, the link from field cohorts with reimbursement data is allowed which could be an enrichment data source.

International initiatives

Linkage of health insurance administrative databases with multiple sources

The Surveillance, Epidemiology and End Results (SEER) program in the United-States.

The National Cancer Institute's (NCI) Surveillance, Epidemiology and End Results (SEER) programme is a population-based registry. It collects informations from some geographic areas which represent 28% of the US population. It aims to produce epidemiological data on incidence prevalence and mortality. There are additional supplementary data that SEER collects and makes available through databases that are not part of the standard SEER Research. In Pharmacoepidemiology, two specialized databases are of interest: the SEER-MEDICARE database and the SEER-CAHPS data set. These databases are based on a linkage between SEER data and the Centers for Medicare & Medicaid Services' (CMS) Medicare Consumer Assessment of Healthcare Providers and Systems (M-CAHPS®) patient surveys. The SEER-CAHPS data set links 3 types of data: clinical/registry (SEER), patient-reported information from the Medicare Consumer Assessment of Healthcare Providers and Systems (M-CAHPS), and administrative/billing (Medicare claims and enrollment files). Beneficiaries included with cancer are Medicare beneficiaries residing in SEER areas who responded to a CAHPS survey between 1997 and 2015. The SEER-MEDICARE links 2 types of data: clinical/registry (SEER) and administrative/billing (Medicare claims and enrollment files). Beneficiaries with cancer included are all Medicare-eligible persons appearing in the SEER data who were diagnosed with cancer between 1973 and 2013 and their Medicare claims through 2014. This resource permits examination of sociodemographic, clinical characteristics and medical care of patients diagnosed with cancer, including cancer survivors. However, Medicare is not representative of the whole US population including only beneficiaries older than 65 years old, beneficiaries aged <65 years with certain disability and beneficiaries of any age with renal end-stage disease [47–49]. Yet, it is crucial to assess long-term consequences of cancer chemotherapy in young adults in terms of somatic and psychologic complications, quality of life or social impact such as difficulties in returning to work. Moreover, this database has geographical limits contrary to the national coverage of the SNDS.

The Nordic prescription databases

The five Nordic countries (Denmark, Finland, Iceland, Norway and Sweden) have nationwide prescription databases on prescriptions dispensed in pharmacies. In these countries, all residents are

identified through a unique personal identity code (PIC) at birth or time of permanent residency since 1960s. PIC permits the linkage between various population-based data sources, such as Nordic cancer registries or biobanks. Available data are categorized in 4 categories patient-specific data, prescriber data, drug data and pharmacy data [50,51]. Between 2005 and 2010, 515 studies including drug utilization studies, effect studies and validation studies were published using these databases. A linkage with other registries was done in 63% (N = 356) of these studies with only 5% (N = 25) concerning antineoplastic and immunomodulation agents [52]. However, some limitations, similar to that observed with the SNDS, must be highlighted. First, OTC drugs and drugs dispensed in hospital and other institutions are not available while anticancer drugs are mainly administered in hospital. Another limit is possible confounding resulting from a lack of data regarding lifestyle habits or detailed clinical data in the database and in other registries. Finally, these databases are limited by small population's coverage ranging from 0.3 to nearly 9 million inhabitants and consequently for studying rare diseases or events [51–53]. One approach to obtain large population study and get sufficient statistical power is to combine information from multiple health care databases as describe in the review by Trifiro et al. [54]. In this way, the Nordic PharmacoEpidemiological Network (NorPEN) has been created in 2008 in order to facilitate research within pharmacoepidemiology in the Nordic countries.

Combination of multiple health care databases

Several international initiatives have been conducted combining multiple health care sources in order to increase statistical power. In Europe, several projects combining multiple national health care databases are conducted to evaluate specific safety issues but none projects were conducted so far in oncology [46,55–58]. In Canada, the Canadian Network for Observational Drug Effect Studies (CNODES) was created in order to link health care databases and to collect data on nearly 40 million of subjects with meta-analytical approach of individual provinces data [59,60]. The difficulty is to combine heterogeneous medical data with different type of systems, sources, medical event coding and languages. However, this approach is of great interest in oncology drug safety and in the detection of signal but harmonization of methods is needed to detect diseases or events of interest.

Discussion

Harmonization process

These considerations highlight that the relevance of the use of health care database for research purpose must be based on a case by case reflection process. In this way, several aspects should be considered in order to improve results' validity of future studies conducted on these databases, such as intrinsic features of diseases and management, type, design and aims of study conducted according to data sources. The development of validated tools and standardized methods is crucial for the validity of future active surveillance studies in oncology. In fact, different published studies used distinct definitions for the same event or disease or to quantify drug exposure. Moreover, the concept of "expert users" should be promoted given the complexity of these data sources. In this way, several initiatives and projects are conducted with the aims to harmonize detection of medical event in claims database in the USA and in Europe (mini sentinel program, observational medical outcomes partnership, pharmacoepidemiological research on outcomes of therapeutics by a European consortium) [61–63]. The development of such initiatives contrasts with the difficulty to publish validation studies, which are needed for national research or international collaboration. Moreover, the use of health care databases is increasing to conduct pharmacoepidemiological studies in oncology with a growing interest in the "big data" context but study validation remains scarce despite guidelines published since 2013 by the ISPOR oncology special interest group. A systematic review conducted to identify secondary data use to conduct health research in oncology between 2006 and 2010, revealed that 72.8% of studies were using claims data with very little studies reporting the use of published algorithms and only 4.8% discussing the impact of their selection algorithm on the results [64]. The development of methodological tools is crucial to enhance quality of studies that can impact clinical practice and to guide public decision. As described above, it is essential to accurately identify stage of cancers, biomarkers, type and line of

treatments, relevant clinical confounders such as comorbidities but also patients reported outcomes. Such information could be obtained with the combination of multiple data sources which constitute a future challenge in onco-pharmacoepidemiology. The use of health care database without linkage with other sources is of great interest to study health care pathway as a whole with significant statistical power but often only in a macroscopic way. However, this increase in statistical power can lead to statistical significance due to very minor effects without clinical relevance [65,66].

Value of pharmacoepidemiology in oncology

In oncology, systemic older anti-cancer drugs present a well-known safety profile for patients contrary to "innovative" ones. In March 2018, a French press release states the request of patient organizations suffering from hematological malignancies to obtain early access to "innovative drugs". This request takes place in a general context suggesting that a rapid access to innovative drugs is necessarily valuable for cancer patients. However, some concerns have emerged in the literature about this new heterogeneous group of anticancer drugs which are presented as "innovative, targeted and safer".

First, as depicted in the introduction, analysis of EMA marketing authorization between 2009 and 2013 showed that only half of new anti-cancer drugs actually improves survival or quality of life in cancer patients. When survival gains were found, they were frequently marginal compared to existing drugs. These concerns contrast with the substantial cost of these drugs. Moreover, 16% of drug approvals received orphan designation. In onco-hematology, this percentage increases to 47% [8]. As depicted in the literature, orphan drugs in oncology generated the largest amount of profit. Some authors comment this concept with headlines such as "*from blockbusters to niche-busters"* [67,68]. In fact, the development and authorization of orphan drugs are facilitated and centralized. After authorization, companies can ask extension of marketing authorization to other indications with a large number of treated patients. Finally, a recent study showed that the median total number of patients studied prior to marketing authorization was 1,708 (interquartile range [IQR] 968-3,195) for standard drugs and 438 (IQR 132-915) for orphan drugs. Then duration of

exposure for chronic medication was insufficient in 20% of studies conducted [7]. In this context, pre-marketing results may be controversial and assessment of these results' transposition in terms of effectiveness and safety in real life conditions is crucial at the population level. In addition, the long-term adverse events of these new drugs or combinations remain unknown.

Conclusion

Pharmacoepidemiology and pharmacovigilance constitute an important cornerstone in oncology for signal detection and long-term follow up of cancer patients. Optimized use of claims data represents a future challenge in Pharmacoepidemiology according to research issues but also with limits and strengths specific to every data sources. In the future, other complementary data sources such as social networks could be considered in a new field of cyberpharmacovigilance or cyberpharmacoepidemiology [69].

Disclosure of interest

The authors declare that they have no conflict of interest.

References

[1] Cancer of Any Site - Cancer Stat Facts. 2019. https://seer.cancer.gov/statfacts/html/all.html [Accessed 25 January 2019].

[2] INCA. Les thérapies ciblées dans le traitement du cancer en 2015 - État des lieux et enjeux - Ref : APDEMOMED16. 2016. ISBN: 978-2-37219-214-9. http://www.e-cancer.fr/Expertises-et-publications/Catalogue-des-publications/Les-therapies-ciblees-dans-le-traitement-du-cancer-en-2015-Etat-des-lieux-et-enjeux. [Accessed 25 January 2019].

[3] Al-Refaie WB, Vickers SM, Zhong W, Parsons H, Rothenberger D, Habermann EB. Cancer trials versus the real world in the United States. Ann Surg 2011;254:438–42; discussion 442-443. doi:10.1097/SLA.0b013e31822a7047.

[4] Penberthy LT, Dahman BA, Petkov VI, DeShazo JP. Effort required in eligibility screening for clinical trials. J Oncol Pract 2012;8:365–70. doi:10.1200/JOP.2012.000646.

[5] Kwiatkowski K, Coe K, Bailar JC, Swanson GM. Inclusion of minorities and women in cancer clinical trials, a decade later: Have we improved? Cancer 2013;119:2956–63. doi:10.1002/cncr.28168.

[6] Freemantle N, Marston L, Walters K, Wood J, Reynolds MR, Petersen I. Making inferences on treatment effects from real world data: propensity scores, confounding by indication, and other perils for the unwary in observational research. BMJ 2013;347:f6409. doi:10.1136/bmj.f6409.

[7] Duijnhoven RG, Straus SMJM, Raine JM, de Boer A, Hoes AW, De Bruin ML. Number of patients studied prior to approval of new medicines: a database analysis. PLoS Med 2013;10:e1001407. doi:10.1371/journal.pmed.1001407.

[8] Davis C, Naci H, Gurpinar E, Poplavska E, Pinto A, Aggarwal A. Availability of evidence of benefits on overall survival and quality of life of cancer drugs approved by European Medicines Agency: retrospective cohort study of drug approvals 2009-13. BMJ 2017;359:j4530.

[9] Kimura T, Matsushita Y, Yang YH, Choi NK, Park BJ. Pharmacovigilance systems and databases in Korea, Japan, and Taiwan. Pharmacoepidemiol Drug Saf 2011;20:1237–45. doi:10.1002/pds.2244.

[10] Kreis K, Neubauer S, Klora M, Lange A, Zeidler J. Status and perspectives of claims data analyses in Germany-A systematic review. Health Policy. 2016 Feb;120(2):213-26. doi: 10.1016/j.healthpol.2016.01.007.

[11] Platt R, Carnahan RM, Brown JS, Chrischilles E, Curtis LH, Hennessy S, et al. The U.S. Food and Drug Administration's Mini-Sentinel program: status and direction. Pharmacoepidemiol Drug Saf 2012;21 Suppl 1:1–8. doi:10.1002/pds.2343.

[12] Gavrielov-Yusim N, Friger M. Use of administrative medical databases in population-based research. J Epidemiol Community Health 2014;68:283–7. doi:10.1136/jech-2013-202744.

[13] Morrato EH, Elias M, Gericke CA. Using population-based routine data for evidence-based health policy decisions: lessons from three examples of setting and evaluating national health policy in Australia, the UK and the USA. J Public Health Oxf Engl 2007;29:463–71. doi:10.1093/pubmed/fdm065.

[14] 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. doi:10.1016/j.respe.2017.05.004.

[15] Bosco-Lévy P, de Boissieu P, Gouverneur A, Noize P, Molimard M, Fourrier-Réglat A, et al. National trends in use and costs of oral anticancer drugs in France: An 8-year population-based study. Pharmacoepidemiol Drug Saf 2017;26:1233–41. doi:10.1002/pds.4282.

[16] Bosco-Lévy P, Jové J, Robinson P, Moore N, Fourrier-Réglat A, Bezin J. Persistence to 5-year hormonal breast cancer therapy: a French national population-based study. Br J Cancer 2016;115:912–9. doi:10.1038/bjc.2016.276.

[17] Létinier L, Mansiaux Y, Pariente A, Fourrier-Réglat A. Impact of cancer diagnosis on persistence of oral antidiabetic drugs. Diabetes Res Clin Pract 2018;139:323–30. doi:10.1016/j.diabres.2018.03.011.

[18] 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 Jun;36(6):411-7. doi: 10.1016/j.revmed.2014.11.009.

[19] 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 Dec;30(6):616-624. doi: 10.1111/fcp.12214.

[20] 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. doi:10.1002/pds.4233.

[21] Kaguelidou F, Sommet A, Lapeyre-Mestre M. Use of French healthcare insurance databases in pediatric pharmacoepidemiology. Therapie 2018;73:127–33. doi:10.1016/j.therap.2017.11.008.

[22] Conte C, Palmaro A, Grosclaude P, Daubisse-Marliac L, Despas F, Lapeyre-Mestre M. A novel approach for medical research on lymphomas: A study validation of claims-based algorithms to identify incident cases. Medicine (Baltimore) 2018;97:e9418. doi:10.1097/MD.0000000000009418.

[23] Moulis G, Sailler L, Adoue D, Lapeyre-Mestre M. Pharmacoepidemiology of Immune Thrombocytopenia: protocols of FAITH and CARMEN studies. Therapie 2014;69:437–48. doi:10.2515/therapie/2014056.

[24] Moulis G, Palmaro A, Montastruc JL, Godeau B, Lapeyre-Mestre M, Sailler L. Epidemiology of incident immune thrombocytopenia: a nationwide population-based study in France. Blood 2014;124:3308–15. doi:10.1182/blood-2014-05-578336.

[25] Mitton N, Colonna M, Trombert B, Olive F, Gomez F, Iwaz J, et al. A Suitable approach to estimate cancer incidence in area without cancer registry. J Cancer Epidemiol 2011;2011:418968. doi:10.1155/2011/418968.

[26] Desandes E, Lacour B, Belot A, Molinie F, Delafosse P, Tretarre B, et al. Cancer incidence and survival in adolescents and young adults in France, 2000-2008. Pediatr Hematol Oncol 2013;30:291–306. doi:10.3109/08880018.2012.762569.

[27] Uhry Z, Remontet L, Colonna M, Belot A, Grosclaude P, Mitton N, et al. Cancer incidence estimation at a district level without a national registry: a validation study for 24 cancer sites using French health insurance and registry data. Cancer Epidemiol 2013;37:99–114. doi:10.1016/j.canep.2012.10.010.

[28] Quantin C, Benzenine E, Hägi M, Auverlot B, Abrahamowicz M, Cottenet J, et al. Estimation of national colorectal-cancer incidence using claims databases. J Cancer Epidemiol 2012;2012:298369. doi:10.1155/2012/298369.

[29] Ganry O, Taleb A, Peng J, Raverdy N, Dubreuil A. Evaluation of an algorithm to identify incident breast cancer cases using DRGs data. Eur J Cancer Prev 2003;12:295–9. doi:10.1097/01.cej.0000082601.47188.01.

[30] Couris CM, Seigneurin A, Bouzbid S, Rabilloud M, Perrin P, Martin X, et al. French claims data as a source of information to describe cancer incidence: predictive values of two identification methods of incident prostate cancers. J Med Syst 2006;30:459–63.

[31] Palmaro A, Gauthier M, Conte C, Grosclaude P, Despas F, Lapeyre-Mestre M. Identifying multiple myeloma patients using data from the French health insurance databases: validation using a cancer registry. Medicine (Baltimore) 2017;96:e6189. doi:10.1097/MD.0000000000006189.

[32] Remontet L, Mitton N, Couris CM, Iwaz J, Gomez F, Olive F, et al. Is it possible to estimate the incidence of breast cancer from medico-administrative databases? Eur J Epidemiol 2008;23:681–8. doi:10.1007/s10654-008-9282-y.

[33] Coureau G, Baldi I, Saves M, Jaffre A, Barat C, Gruber A, et al. Performance evaluation of hospital claims database for the identification of incident central nervous system tumors compared with a cancer registry in Gironde, France, 2004. Rev Epidemiol Sante Publique 2012;60:295–304.

[34] Carré N, Uhry Z, Velten M, Trétarre B, Schvartz C, Molinié F, et al. Predictive value and sensibility of hospital discharge system (PMSI) compared to cancer registries for thyroïd cancer (1999-2000). Rev Epidemiol Sante Publique 2006;54:367–76.

[35] Hafdi-Nejjari Z, Couris CM, Schott AM, Schot AM, Perrot L, Bourgoin F, et al. Role of hospital claims databases from care units for estimating thyroid cancer incidence in the Rhône-Alpes region of France. Rev Epidemiol Sante Publique 2006;54:391–8.

[36] Palmaro A, Gauthier M, Despas F, Lapeyre-Mestre M. Identifying cancer drug regimens in French health insurance database: An application in multiple myeloma patients. Pharmacoepidemiol Drug Saf 2017;26:1492–9. doi:10.1002/pds.4266.

[37] Quaranta S, Thomas F. Pharmacogenetics of anti-cancer drugs: State of the art and implementation - recommendations of the French National Network of Pharmacogenetics. Therapie 2017;72:205–15. doi:10.1016/j.therap.2017.01.005.

[38] Bondon-Guitton E, Despas F, Becquemont L. The contribution of pharmacogenetics to pharmacovigilance. Therapie 2016;71:223–8. doi:10.1016/j.therap.2016.02.005.

[39] Compaci G, Rueter M, Lamy S, Oberic L, Recher C, Lapeyre-Mestre M, et al. Ambulatory Medical Assistance--After Cancer (AMA-AC): a model for an early trajectory survivorship survey of lymphoma patients treated with anthracycline-based chemotherapy. BMC Cancer 2015;15:781. doi:10.1186/s12885-015-1815-7.

[40] Fourrier-Réglat A, Ravaud A, Jayles D, Balestra A, Rouyer M, Moore N. ETNA : étude de terrain sur les traitements innovants en cancérologie : un antiangiogénique l'Avastin® : données sur l'utilisation du bévacizumab en vie réelle. Rev Epidemiol Sante Publique 2009;57:S83–4. doi:10.1016/j.respe.2009.07.046.

[41] Fourrier-Réglat A, Noize P, Facon T, Fermand JP, Fitoussi O, Marit G, et al. Real-life patterns of use and effectiveness of bortezomib: the VESUVE cohort study. Leuk Lymphoma 2014;55:848–54. doi:10.3109/10428194.2013.806801.

[42] Fourrier-Réglat A, Smith D, Rouyer M, Bénichou J, Guimbaud R, Bécouarn Y, et al. Survival outcomes of bevacizumab in first-line metastatic colorectal cancer in a real-life setting: results of the ETNA cohort. Target Oncol 2014;9:311–9. doi:10.1007/s11523-013-0296-3.

[43] Miremont-Salamé G, Théophile H, Haramburu F, Bégaud B. Causality assessment in pharmacovigilance: The French method and its successive updates. Therapie 2016;71:179–86. doi:10.1016/j.therap.2016.02.010.

[44] Baldo P, De Paoli P. Pharmacovigilance in oncology: evaluation of current practice and future perspectives. J Eval Clin Pract 2014;20:559–69. doi:10.1111/jep.12184.

[45] Herk-Sukel MPP van, Lemmens VEPP, Poll-Franse LV van de, Herings RMC, Coebergh JWW. Record linkage for pharmacoepidemiological studies in cancer patients. Pharmacoepidemiol Drug Saf 2012;21:94–103. doi:10.1002/pds.2205.

[46] Andersen M. Research on drug safety and effectiveness using pharmacoepidemiological databases. J Intern Med 2014;275:548–50. doi:10.1111/joim.12235.

[47] Warren JL, Klabunde CN, Schrag D, Bach PB, Riley GF. Overview of the SEER-Medicare data: content, research applications, and generalizability to the United States elderly population. Med Care 2002;40:IV-3–18. doi:10.1097/01.MLR.0000020942.47004.03.

[48] Bach PB, Guadagnoli E, Schrag D, Schussler N, Warren JL. Patient demographic and socioeconomic characteristics in the SEER-Medicare database applications and limitations. Med Care 2002;40:IV-19–25. doi:10.1097/01.MLR.0000020934.40692.C0.

[49] Chawla N, Urato M, Ambs A, Schussler N, Hays RD, Clauser SB, et al. Unveiling SEER-CAHPS®: a new data resource for quality of care research. J Gen Intern Med 2015;30:641–50. doi:10.1007/s11606-014-3162-9.

[50] Langseth H1, Luostarinen T, Bray F, Dillner J. Ensuring quality in studies linking cancer registries and biobanks. Acta Oncol 2010 Apr;49(3):368-77. doi: 10.3109/02841860903447069..

[51] Furu K, Wettermark B, Andersen M, Martikainen JE, Almarsdottir AB, Sørensen HT. The Nordic countries as a cohort for pharmacoepidemiological research. Basic Clin Pharmacol Toxicol 2010;106:86–94. doi:10.1111/j.1742-7843.2009.00494.x.

[52] 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. doi:10.1002/pds.3457.

[53] Pukkala E, Engholm G, Højsgaard Schmidt LK, Storm H, Khan S, Lambe M, et al. Nordic Cancer Registries - an overview of their procedures and data comparability. Acta Oncol Stockh Swed 2018;57:440–55. doi:10.1080/0284186X.2017.1407039.

[54] Trifirò G, Coloma PM, Rijnbeek PR, Romio S, Mosseveld B, Weibel D, et al. Combining multiple healthcare databases for postmarketing drug and vaccine safety surveillance: why and how? J Intern Med 2014;275:551–61. doi:10.1111/joim.12159.

[55] Murray ML, Insuk S, Banaschewski T, Neubert AC, McCarthy S, Buitelaar JK, et al. An inventory of European data sources for the long-term safety evaluation of methylphenidate. Eur Child Adolesc Psychiatry 2013;22:605–18. doi:10.1007/s00787-013-0386-x.

[56] Coloma PM, Schuemie MJ, Trifirò G, Gini R, Herings R, Hippisley-Cox J, et al. Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR Project. Pharmacoepidemiol Drug Saf 2011;20:1–11. doi:10.1002/pds.2053.

[57] Boyd PA, Haeusler M, Barisic I, Loane M, Garne E, Dolk H. Paper 1: The EUROCAT network--organization and processes. Birt Defects Res A Clin Mol Teratol 2011;91 Suppl 1:S2-15. doi:10.1002/bdra.20780.

[58] de Jonge L, Garne E, Gini R, Jordan SE, Klungsoyr K, Loane M, et al. Improving information on maternal medication use by linking prescription data to congenital anomaly registers: a EUROmediCAT study. Drug Saf 2015;38:1083–93. doi:10.1007/s40264-015-0321-9.

[59] Suissa S, Henry D, Caetano P, Dormuth CR, Ernst P, Hemmelgarn B, et al. CNODES: the Canadian Network for Observational Drug Effect Studies. Open Med 2012;6:e134-140.

[60] Platt RW, Dormuth CR, Chateau D, Filion K. Observational studies of drug safety in multi-database studies: methodological challenges and opportunities. EGEMS (Wash DC). 2016 May 18;4(1):1221. doi: 10.13063/2327-9214.1221.

[61] Avillach P, Coloma PM, Gini R, Schuemie M, Mougin F, Dufour JC, et al. Harmonization process for the identification of medical events in eight European healthcare databases: the experience from the EU-ADR project. J Am Med Inform Assoc 2013;20:184–92. doi:10.1136/amiajnl-2012-000933.

[62] 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 1:82–9. doi:10.1002/pds.2321.

[63] Ehrenstein V, Petersen I, Smeeth L, Jick SS, Benchimol EI, Ludvigsson JF, et al. Helping everyone do better: a call for validation studies of routinely recorded health data. Clin Epidemiol 2016;8:49–51. doi:10.2147/CLEP.S104448.

[64] Schulman KL, Berenson K, Tina Shih YC, Foley KA, Ganguli A, de Souza J, et al. A checklist for ascertaining study cohorts in oncology health services research using secondary data: report of the ISPOR oncology good outcomes research practices working group. Value Health. 2013 Jun;16(4):655-69. doi: 10.1016/j.jval.2013.02.006.

[65] Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol 2016;31:337–50. doi:10.1007/s10654-016-0149-3.

[66] Bezin J, Bosco-Levy P, Pariente A. False-positive results in pharmacoepidemiology and pharmacovigilance. Therapie 2017;72:415–20. doi:10.1016/j.therap.2016.09.020.

[67] Sharma A, Jacob A, Tandon M, Kumar D. Orphan drug: development trends and strategies. J Pharm Bioallied Sci 2010;2:290–9. doi:10.4103/0975-7406.72128.

[68] Kumar Kakkar A, Dahiya N. The evolving drug development landscape: from blockbusters to niche busters in the orphan drug space. Drug Dev Res 2014;75:231–4. doi:10.1002/ddr.21176.

[69] Bagheri H, Lacroix I, Guitton E, Damase-Michel C, Montastruc JL. Cyberpharmacovigilance: What is the usefulness of the social networks in pharmacovigilance? Therapie 2016;71:235–9. doi:10.1016/j.therap.2015.09.002.

Table 1. Description of validated algorithms to identify incident cancer patients in the *système national des données de santé* (SNDS) by cancer type for solid tumors

Solid tumor						
Cancer type	ICD-10 Codes	Medical Acts or procedures	Other variable	Gold standard	Se*	PPV+
Colorectal *Quantin et al* *J Cancer Epidemiol 2012;2012:298369.*	C18; C19; C20	"endoscopic examination of the colon or rectum"; "partial or total exeresis of the colon or rectum"; "excision, exeresis or destruction of polyps or tumours in the colon or rectum"; "colostomy repair or closure"; "secondary restoration of continuity"; "implantation of a colon endoprosthesis"	-	Cancer registry	≈75%	≈75%
Breast *Couris et al.* *J Clin Epidemiol 2009 Jun;62(6):660-6*	C50	-	-	Cancer registry	64%	-
Prostate *Couris et al.* *J Med Syst 2006;30:459-63*	C61; D07.5	Surgical procedures and ultrasound procedures	-	Medical record	-	≈33%
Central nervous system *Coureau et al* *Rev Epidemiol Sante Publique 2012;60:295-304*	C70; C71; C72; D32;D33;D42; D43; C75.2;D35.3; D44.4; G93.0:	Surgical procedures Diagnosis act Chemotherapy and radiotherapy	No metastasis antecedent Age>70 years old	Cancer Registry	67%	69%
Thyroïd *Carré et al* *Rev Epidemiol Sante Publique 2006;54:367-76*	C73; D09.3;D44.0	Surgical acts	-	Cancer Registry	73%	89%

*Sensitivity; +Predictive positive value

ICD-10 codes: international statistical classification of diseases and related health problems 10th revision

Table 2. Description of validated algorithms to identify incident cancer patients in *the système national des données de santé* (SNDS) by cancer type for hematologic malignancies

Hematologic maligancies						
Cancer type	ICD-10 Codes	Medical Acts/procedures	Other variable	Gold standart	Se*	PPV+
Hodgkin Lymhoma[1]	C81	Chemotherapy	No lymphoma's antecedent	Cancer registry	94.2%	74.2%
B-Non Hodgkin Lymphoma[1]	C82 ; C83.3; C83.0; C83.1 ; C83.7 ; C83.8 ; C83.9 ; C85 ; C88.4	Chemotherapy	No lymphoma's antecedent	Cancer registry	74.6%	64.6%
T- Non Hodgkin Lymphoma[1]	C84;C86	Chemotherapy	No lymphoma's antecedent	Cancer registry	48.6%	78.3%
Chronic Lymphocytic Leukemia/ Small lymphocytic lymphoma[1]	C91.1	Chemotherapy	No lymphoma's antecedent	Cancer registry	19.0%	22.6%
Myeloma[2]	C90	-	No Myeloma's antecedent	Cancer registry	90%	66%

*Sensitivity; +Predictive positive value

ICD-10 codes: international statistical classification of diseases and related health problems 10th revision

[1]Conte et al. Medicine (Baltimore) 2018;97:e9418.

[2]Palmaro et al. Medicine (Baltimore) 2017;96:e6189.