



Feasibility and accuracy of linking a heart failure registry to the national claims database using indirect identifiers

Damien Logeart, Thibaud Damy, Maxime Doublet, Muriel Salvat, Christophe Tribouilloy, Fabrice Bauer, Jean-Christophe Eicher, François Picard, Gérald Roul, Jean-Noël Trochu, et al.

► To cite this version:

Damien Logeart, Thibaud Damy, Maxime Doublet, Muriel Salvat, Christophe Tribouilloy, et al.. Feasibility and accuracy of linking a heart failure registry to the national claims database using indirect identifiers. Archives of cardiovascular diseases, 2022, <10.1016/j.acvd.2022.11.002>. <hal-03911741>

HAL Id: hal-03911741

<https://hal.science/hal-03911741v1>

Submitted on 23 Dec 2022

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.



HAL Authorization

Research Feasibility and accuracy of linking a heart failure registry to the national claims database using indirect identifiers

Damien Logeart^{a,*}, Thibaud Damy^b, Maxime Doublet^c, Muriel Salvat^d, Christophe Tribouilloy^e, Fabrice Bauer^f, Jean-Christophe Eicherg^g, François Picard^h, Gérard Rouliⁱ, Jean-Noël Trochuj^j, Pascal De Groot^k, Nicolas Bihry^l, Emmanuelle Berthelot^m, Guillaume Jondeaunⁿ, Marie-France Seronde^o, François Roubille^p, Richard Isnard^q, on behalf of the FRESH investigators

^aParis Cité University, Hôpital Lariboisière, AP-HP, 75010 Paris, France

^bHôpital Henri-Mondor, AP-HP, 94000 Créteil, France

^cClinityx, 75008 Paris, France

^dHôpital Michallon, 38700 La Tronche, France

^eUniversity Hospital Amiens, 80000 Amiens, France

^fUniversity Hospital Rouen, 76000 Rouen, France

^gUniversity Hospital Dijon, 21000 Dijon, France

^hUniversity Hospital Bordeaux, 33000 Bordeaux, France

ⁱUniversity Hospital Strasbourg, 67000 Strasbourg, France

^jUniversity Hospital Nantes, 44000 Nantes, France

^kUniversity Hospital Lille, 59000 Lille, France

^lSaint-Joseph and Saint-Luc Hospital, 69007 Lyon, France

^mHôpital Kremlin Bicêtre, AP-HP, 94270 Le Kremlin-Bicêtre, France

ⁿHôpital Bichat, AP-HP, 75018 Paris, France

^oUniversity Hospital Besançon, 25000 Besançon, France

^pUniversity Hospital Montpellier, 34295 Montpellier, France

^qHôpital Pitié-Salpêtrière, AP-HP, 75013 Paris, France

•Follow-up of patients with HF in registries is a critical issue.

•Medicoadministrative databases contain long-term exhaustive data

•Linking accurately anonymized medicoadministrative database and HF registry is feasible

•The linking process increases the relevance of registries.

Keywords:

Heart failure Outcome Registries Claims databases

Abbreviations:

CKD, chronic kidney disease; COPD, chronic pulmonary obstructive disease; FRESH, FRENch Survey on Heart Failure; HF, heart failure; ICD-10, International Classification of Diseases, Tenth Revision; LVEF, left ventricular ejection fraction; NIR, Numéro d'Inscription au Répertoire; SNDS, Système National des Données de Santé (French National Healthcare System database); TAVI, transcatheter aortic valve implantation; TAVR, transcatheter aortic valve replacement.

***Corresponding author at:**

Hôpital Lariboisière, 2, rue Ambroise Paré, 75010 Paris, France.

E-mail address: damien.logeart@aphp.fr (D. Logeart).

SUMMARY

Background:

Heart failure (HF) registries include rich data on patient inclusion characteristics, but follow-up information is often incomplete. Medico administrative databases may provide less clinical information than registries, e.g. on left ventricular ejection fraction (LVEF), but long-term data are exhaustive and reliable. The combination of the two types of database is therefore appealing, but the feasibility and accuracy of such linking are largely unexplored.

Aims:

To assess the feasibility and accuracy of linking an HF registry (FRESH; FRENch Survey on Heart Failure) with the French National Healthcare System database (SNDS).

Methods: A probabilistic algorithm was developed to link and match patient data included in the FRESHHF registry with anonymized records from the SNDS, which include: hospitalizations and diagnostic codes; all care-related reimbursements by national health system; and deaths. Consistency was assessed between deaths recorded in the registry and in the SNDS. A comparison between the two databases was carried out on several identifiable clinical characteristics (history of HF hospitalization, diabetes, atrial fibrillation, chronic broncho-pneumopathy, severe renal failure and stroke) and on events during 1-year follow-up after inclusion.

Results: Of 2719 patients included in the FRESH registry (1049 during decompensation; 1670 during outpatient follow-up), 1885 could be matched with a high accuracy of 94.3% for deaths. Mortality curves were superimposable, including curves according to type of HF and LVEF. The rates of missing data in the FRESH registry were 2.3–8.4% for clinical characteristics and 17.5% for hospitalizations during follow-up. The discrepancy rate for clinical characteristics was 3–13%. Hospitalization rates were significantly higher in the SNDS than in the registry cohort.

Conclusions: The anonymous matching of an HF research cohort with a national health database is feasible, with a significant proportion of patients being accurately matched, and facilitates combination of clinical data and a reduced rate of losses to follow-up.

1. Background

The prevalence of heart failure (HF) in France is similar to that in other European countries, affecting around 2.3% of the adult population, with rates rising to 15.1% in those aged > 85 years [1]. To provide the best possible care to the most inclusive patient population, clinicians, caregivers and healthcare authorities need to be able to assess the efficiency, safety and outcomes of treatments, as well as gain insight into the quality and performance of the care provided in everyday practice.

For such assessments, data from real-life registries are valuable. Such registries can include clinical datasets with very high levels of details and granularity. A common problem with registries is that they depend on the availability of the information; however, the manual entry of information adds to the ordinary workload of healthcare personnel, and frequently leads to gaps in the records and major losses to follow-up [2]. Independent verification of collected data increases the data reliability, but with a dramatic impact on costs, and thus is rarely done.

The explosive growth in electronic health databases in recent years has made it possible to enrich the completeness of registries and enhance the value of analyses by linking multiple data to claims databases [3]. The French single-payer National Healthcare System database (SNDS; *Système National des Données de Santé*) includes reimbursement information on close to 99% of the French population (or > 68 million people), making it one of the world's largest continuous claims databases [4]. The SNDS also includes data on mortality, hospitalizations and long-term diseases. The French version of the International Classification of Diseases, Tenth Revision (ICD-10) [5] is used to identify reasons for hospitalizations and long-term diseases; this includes HF (ICD-10 codes I50.x and R570). However, although the accuracy of ICD-10 is good, it does not capture all patients with HF. Only a proportion of patients with HF are hospitalized during the analysed time, and some are not coded as having HF-related long-term disease because of previous coding for co-morbidities, such as ischaemic heart disease or diabetes. The sensitivity of coding for HF has been estimated at between 29% and 86% in several studies [6–10]. Furthermore, only few clinical data are captured when using ICD-10 codes. As a relevant example for HF, information on left ventricular ejection fraction (LVEF) was not available in the database until the 2019 update of ICD-10 codes. On the other hand, records of events, hospitalization, discharge and mortality tend to be much more complete and accurate than in registries.

We have previously demonstrated the feasibility of linking a registry and SNDS data for patients undergoing transcatheter aortic valve implantation (TAVI) to minimise follow-up losses and close data gaps [11]. However, patients undergoing TAVI represent a comparatively homogeneous (older) population undergoing a specific and easily identified treatment, whereas HF is a syndrome that manifests in a multitude of ways, and its treatment, management and outcomes vary widely between individuals.

Hence, a linkage attempt will face greater challenges than for populations undergoing specific procedures. The objective of the current study was to assess the accuracy of linking a research HF registry (FRESH; *FRENch Survey on Heart Failure*) to the SNDS, as well as to evaluate differences and discrepancies between the two data sources.

2. Methods

2.1. Data sources

The FRESH registry was set up in 2014 with the aim of providing detailed information on several aspects of HF and its treatment nationally. The registry gathers data on the characteristics of the population with HF, including socioeconomic status, aetiology of the disease, medical history and co-morbidities, information about the care provided to manage HF at baseline, and follow-up data on rehospitalization and death. A total of 2719 patients were included at 32 centres across France between 18 February 2014 and 31 December 2017. The registry is registered on ClinicalTrials.gov (identifier: NCT01956539), and it is supported by the French Cardiology Society

The FRESH registry enrolled patients aged > 18 years with a diagnosis of HF during both outpatient consultation and unplanned hospitalization for acute HF. Recruitment was 1 day per week. To enable analysis of the course of disease and changes in treatment patterns over time in individual patients, including the impact of new guideline recommendations, patients were followed up at 12 and 24 months. Patients included at the time of an episode of acute HF were also followed up 3 months after the event. There was no independent verification of collected data. The SNDS combines health insurance data (SNIIRAM database), hospital data (PMSI database), medical causes of death (Cépi DC database), disability-related data and data from complementary health insurance organizations, but no clinical, biological or echocardiographic data. The organization of the SNDS has been described previously [11–13]. The database contains more than 1.2 billion yearly claims records and 11 million hospitalizations [14]. Entries are anonymized as a unique number “Numéro d’Inscription au Répertoire” (NIR) for each patient, and records contain no information that allows direct identification of individuals [12]. SNDS data were analysed for the period 2012 to 2018, starting 2 years before, and covering up to 5 years after enrolment in the FRESH registry. Data collected up to 31 December 2017 in the FRESH registry were included in the FRESH-SNDS linkage.

2.2. Linkage process

The development of the linking algorithm and its core features has been described previously [11]. Because current regulations prohibit identification of individuals in the SNDS to match them directly with registries, the linkage process employs a probabilistic approach, based on matching SNDS data as closely as possible to the profiles in the FRESH registry. An initial filtering was performed, utilising a selection of primary anonymous identifiers, including sex, year of birth, department of residence, date of consultation or hospitalization, date of discharge for hospitalized patients and hospital identification number. For patients with missing data, the match was refined by the inclusion of secondary identifiers, which were not utilised in the primary linking process, i.e. month of birth and area of residence.

The final quality of linking was assessed by the consistency rates for deaths between the two data sources. The objective was to generate a linked database with the highest consistency rate possible, which was estimated at $\geq 90\%$ for deaths between the two databases. A death recorded in the SNDS was considered identical to one in the FRESH registry if the registry date fell within 60 days of that recorded in the SNDS. Because of delay between the date of a death and its inclusion in the SNDS, deaths were considered in the SNDS up to 31 December 2018. In the case of insufficient consistency rate for deaths, we planned additional scoring methods with several identifiers, including some of previous administrative variables, but also vital status and highly identifying clinical variables, such as diabetes. The most reliable model had to result in a consistency rate of $> 90\%$ for deaths, and with the highest rate of matched patients.

2.3. Comparison of the FRESH and SNDS-FRESH linked databases

In the FRESH registry, clinical characteristics were entered on case report forms by local investigators at the time of a consultation or hospitalization. In the SNDS, only some of these variables could be approximated using a case-identifying algorithm, including ICD-10 code-

related hospitalizations or long-term diseases, administrative codes related to the reimbursement of procedures (Classification Commune des Actes Medicaux (CCAM) or Anatomical Therapeutic Chemical (ATC) codes.

The two data sources were compared on the following variables, which were present in both databases: age; sex; diabetes; ischaemic heart disease; chronic pulmonary obstructive disease(COPD); end-stage chronic kidney disease (CKD); atrial fibrillation; stroke; and HF hospitalization within the previous 12 months. Algorithms used in the SNDS for each variable are detailed in Appendix B. These characteristics were compared for exhaustiveness of the data (proportion of missing data) and consistency rates, i.e. the proportion of patients with matching information between the two data sources.

In addition, deaths and hospitalizations occurring in the two data sources during the 1-year follow-up after inclusion were compared. Only hospitalizations lasting at least 2 nights were taken into account. Hospitalization in follow-up care and rehabilitation were identified and excluded. In the SNDS, HF hospitalization was identified by ICD-10 codes I50 and R570 as the main or associated diagnosis.

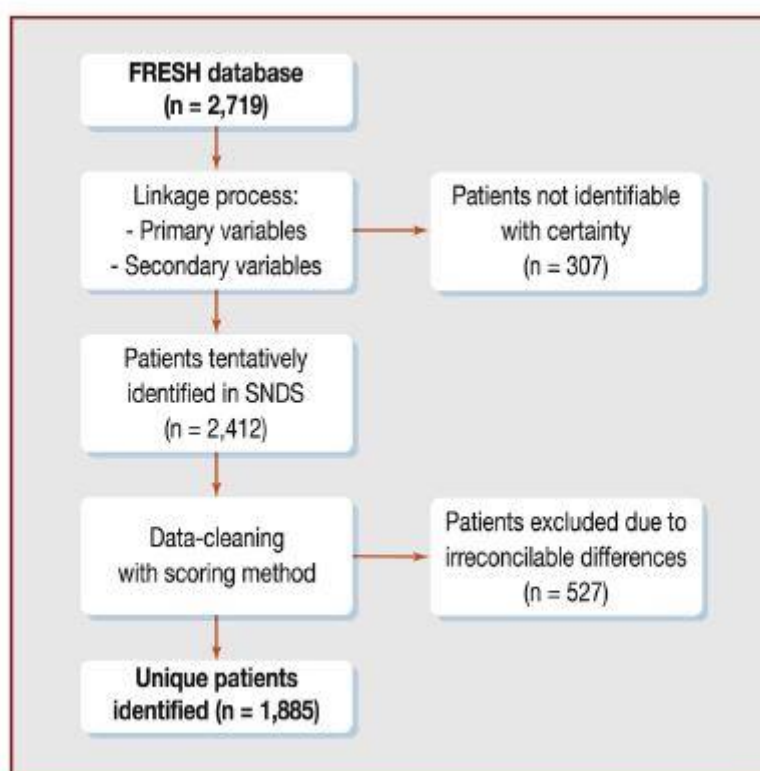


Fig. 1. Schematic representation of the linkage process between the FRENch Survey on Heart Failure (FRESH) registry and the French single-payer National Healthcare System database (Système National des Données de Santé [SNDS]).

2.4. Ethical considerations

All patients in the FRESH cohort provided written informed con-sent. Ethics approval to link the FRESH registry and the SNDS was obtained from the French Expert Committee on Research, Studies and Assessments in Health (CEREES) on 25 April 2019 (TPS347113bis), and from the Commission Nationale de l'Informatique et des Libertés (CNIL) on 19 November 2019 (DR-2019-34). All data were analysed anonymously. Each individual record in the SNDS was randomly assigned a numerical identity (different from the NIR), which included no information about the identity of the patient or centre. This number was used in the analysis, with no reverse identification possible. A different individual identity was assigned to each patient in the FRESH cohort. All transmissions of data were encrypted and stored on secure servers. At no point did the analysts have access to original names or NIRs. Because all patient data were from a registry that had already obtained informed consent and ethical clearance, no additional ethical approval was necessary.

3. Results

3.1. Linkage process

The result of the linkage process is shown schematically in Fig. 1. The FRESH cohort included 2719 patients with available data:1049 (38.6%) subjects were included during a hospitalization for acute HF; and 1670 (61.4%) were included during an outpatient consultation. The initial linkage process identified 2412 (88.7%) individuals with matching identifiers in the SNDS and the FRESH registry. After this first linking process, 25% of deaths were irreconcilable between the two databases. The rate of inconsistencies was similar for patients who were included during HF-related hospitalization and outpatients. The linking process was refined further using scoring methods. The most reliable model required that at least three of the following four identifiers – birth month, consultation/hospitalization date, diabetes, vital status – were identical in the two databases. This resulted in a cohort of 1885 (69.3%) linked patients, with a 94.3% consistency rate for deaths.

Table 1

Demographic and clinical characteristics of the linked population ($n = 1885$), based on data from the FRENCH Survey on Heart Failure registry and the French National Healthcare System database.

	FRESH registry Missing	Median or %	SNDS Median or %	Consistency
Birth year	0	1947	1947	100%
Male sex	0	69.7%	69.7%	100%
Diabetes	0	28.2%	25.7%	91.1%
COPD	2.5%	15%	14.2 %	85.1%
End-stage CKD	4.8%	1.4%	3.4%	97.0%
Atrial fibrillation	8.4%	43.7%	42.5%	76.9%
Stroke	2.3%	12.8%	7.3%	89.0%
HF hospitalization ^a within the latest year	3.7%	55%	56.3%	91.3%

COPD: chronic obstructive pulmonary disease; end-stage CKD: chronic kidney disease with glomerular filtration rate < 15 mL/min; FRESH: FRENch Survey on Heart Failure; HF: heart failure; SNDS: Système National des Données de Santé (French National Healthcare System database).a Hospitalization for HF was defined by using International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) code I50 or R570 as the main diagnosis.

3.2. Comparison of clinical characteristics between the two databases

A comparison of demographic and clinical characteristics for the linked population, based on data in the two source databases, is shown in Table 1. In the FRESH registry, demographic variables were available for all patients, but some data were missing for co-morbidities (COPD, CKD, atrial fibrillation, stroke; 2.3–8.4%), and data on history of HF hospitalization before inclusion were unavailable in 3.7% of patients. Recorded rates of diabetes, stroke and atrial fibrillation were higher in the FRESH registry than in the SNDS, whereas rates of COPD, end-stage CKD and previous HF hospitalization were higher in the SNDS.

There was 100% consistency between the FRESH registry and the SNDS for age and sex, which were used for the linkage process. For co-morbidities, the consistency varied between 76.9% and 97%. In an analysis of subpopulations, the concordance rate for diabetes and atrial fibrillation was higher among patients included during hospitalization (96.3 and 82.5%, respectively) than among those included during outpatient consultation or day hospitalization (87.6% and 73.0%, respectively).

3.3. Outcomes and comparison between the two databases

Rates of deaths at 1 year were 12.0% in the SNDS and 11.5% in the FRESH registry. Fig. 2 compares survival curves from the FRESH registry and the SNDS. Fig. 3A and Fig. 3B show survival curves from the FRESH registry and the SNDS, respectively, according to the following subgroups, based on type of HF at inclusion: patients hospitalized for acute HF; outpatients (chronic HF); patients with a reduced LVEF; and patients with a preserved LVEF.

Because of different definitions and causes of hospitalization in the two databases, no direct comparison between the FRESH registry and the SNDS was carried out for this variable. In the FRESH registry, data on all-cause hospitalization were missing in 17.5% of patients, and the cause of hospitalization was unknown in 29.3%. For patients with collected data, the rate of patients with at least one all-cause hospitalization was 46.2% at 1 year. This rate fell to 29.6% if only unplanned hospitalizations were considered. In the SNDS, the rate of patients with at least one all-cause hospitalization at 1 year was 69.7%; the rate was 55.9% if only hospitalizations with at least 2 nights were considered. The rate of HF-related hospitalizations at 1 year in the SNDS (ICD-10 codes I50 and R570 as main diagnosis and hospitalizations ≥ 2 nights) was 20%; the rate was 30.2% if HF hospitalizations included ICD-10 codes I50 and R570 as the main or associated diagnosis. Fig. 4 shows all-cause and HF-related hospitalizations in the SNDS. Fig. 5 shows the occurrence of all-cause hospitalizations, according to the type of HF at inclusion.

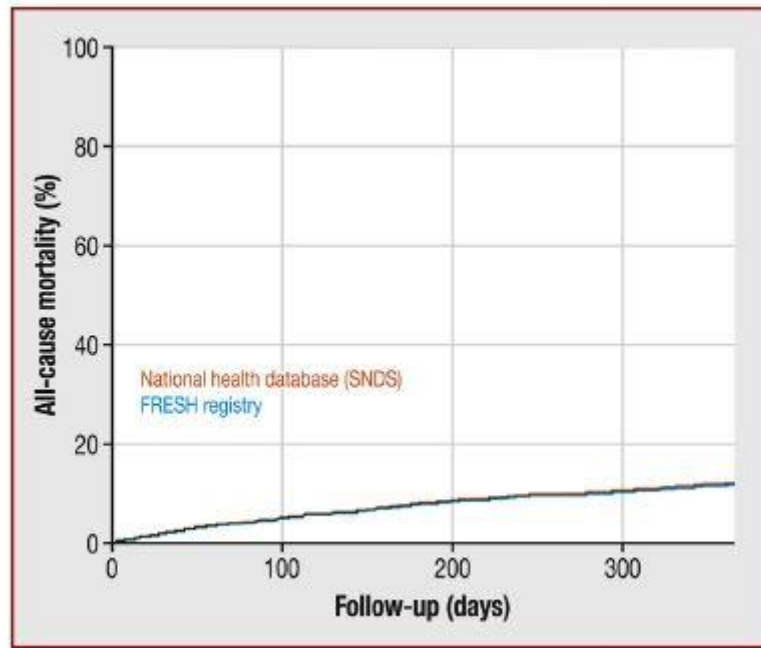


Fig. 2. Survival curves from the FRENch Survey on Heart Failure (FRESH) registry and the French National Healthcare System database (Système National des Données de Santé [SNDS]) for all-cause mortality.

3. Discussion

The results presented here show that it is possible to link and match records in the national healthcare database with patient data from the case report forms of a pre-existing HF registry. Such linkage markedly improves the quality of follow-up data.

The development of a probabilistic algorithm for HF is considerably more challenging than the earlier reported linkage of transcatheter aortic valve replacement (TAVR) registries–FRANCE-2 and FRANCE-TAVI. HF is a frequent disease, with various patient profiles and treatments; this creates at least two serious problems for linking processes. The first is the inclusion of outpatients who have no ICD-10 code-related hospitalization to help their identification in the SNDS during the study period. The second is the lack of a medical procedure common to all patients with HF, which contrasts with, for example, valve implantation. Considering these hurdles, the 69.3% completeness achieved with the multistep algorithm appears quite respectable, even if lower than the 89.9% completeness achieved for the TAVR registries and the SNDS [11].

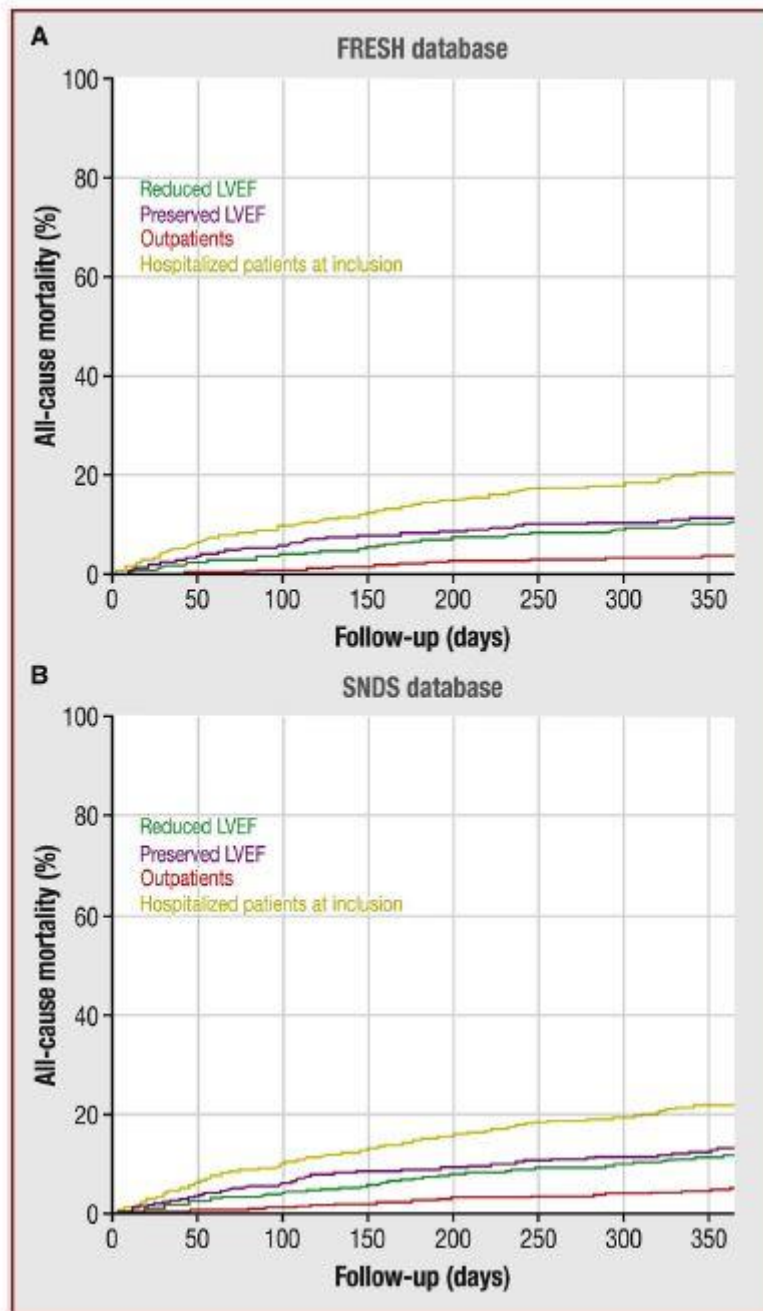


Fig. 3. A–B. Survival curves from the FRENch Survey on Heart Failure (FRESH) registry (A) and the French National Healthcare System database (Système National des Données de Santé [SNDS]) (B) for all-cause mortality, according to the following subgroups: patients hospitalized for acute heart failure; outpatients; patients with a left ventricular ejection fraction (LVEF) $\leq 45\%$; and patients with an LVEF $> 45\%$.

Linkage generates an enriched population, which can be characterised in greater depth than the isolated underlying datasets. Registries are key to evaluating quality and effectiveness of care under everyday conditions; they also typically include variables that are not available because they do not correspond to ICD-10 codes in healthcare databases, e.g. renal function,

body mass index and functional class. However, the quality of a registry is only as good as that of the data, which are entered by investigators on a voluntary basis. For outcomes, a registry will rely on calling sub-jects or family or referring doctors. Such follow-up requires great and time-consuming efforts, and losses to follow-up can be significant, especially when the duration of follow-up is long, as in HF. The regulatory requirements and needs underlying a national administrative claims database mean that it will have a high cover-age of clinical events, which incur costs, as well as consumption of care and medication. Events such as death or hospitalizations are covered, with few gaps. This is reflected in the high rate of hospitalizations in the linked population compared with, for example,(completed) case reports.

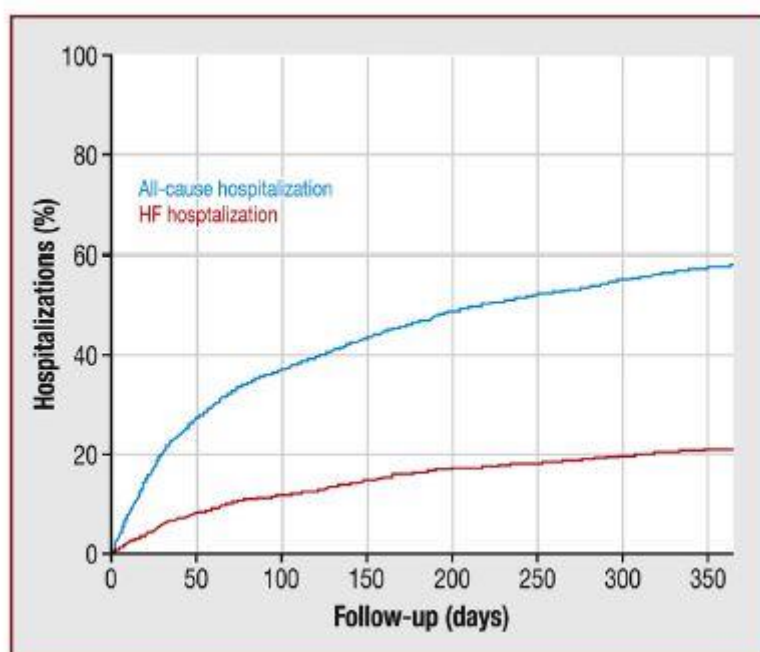


Fig. 4. All-cause hospitalizations and heart failure (HF)-related hospitalizations from the French National Healthcare System database (Système National des Données de Santé [SNDS]).

It is interesting to note that different co-morbidities were covered to different degrees of completeness in the two source databases. In the healthcare database, co-morbidities are determined by administrative codes and the use of case-identifying algorithms. In the FRESH registry, as in most registries, reliability of co-morbidity data depends on how patients are asked and the patient's knowledge of their conditions, which can be critical for some co-morbidities. We note that there was greater consistency for diabetes and atrial fibrillation in acute HF than in the outpatient setting. This may be partially explained by repeated and/or extended examination during hospitalization. The fact that atrial fibrillation and stroke were more frequently reported in the FRESH registry can be explained by the limited medical history captured in the SNDS. Atrial fibrillation was recorded only for the 2 latest years before inclusion, and on the basis of interventions or administered antiarrhythmic medications, which may not be considered necessary in many milder or paroxysmal cases. For

stroke, the identification in SNDS was based on stroke-related hospitalization, also only during the 2 latest years before inclusion, which will have underestimated the rate of previous strokes.

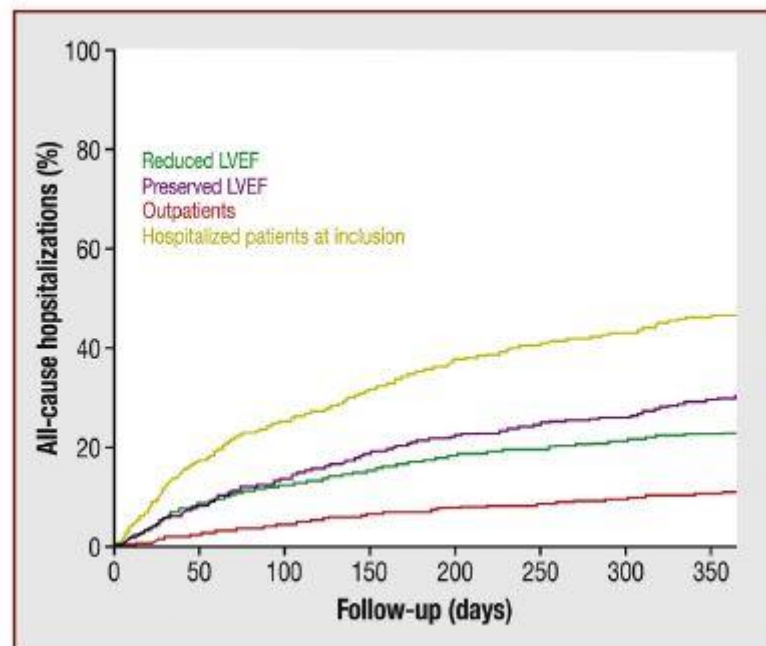


Fig. 5. All-cause hospitalizations, according to the following subgroups: patients hospitalized for acute heart failure; outpatients; patients with a left ventricular ejection fraction (LVEF) $\leq 45\%$; and patients with an LVEF $> 45\%$.

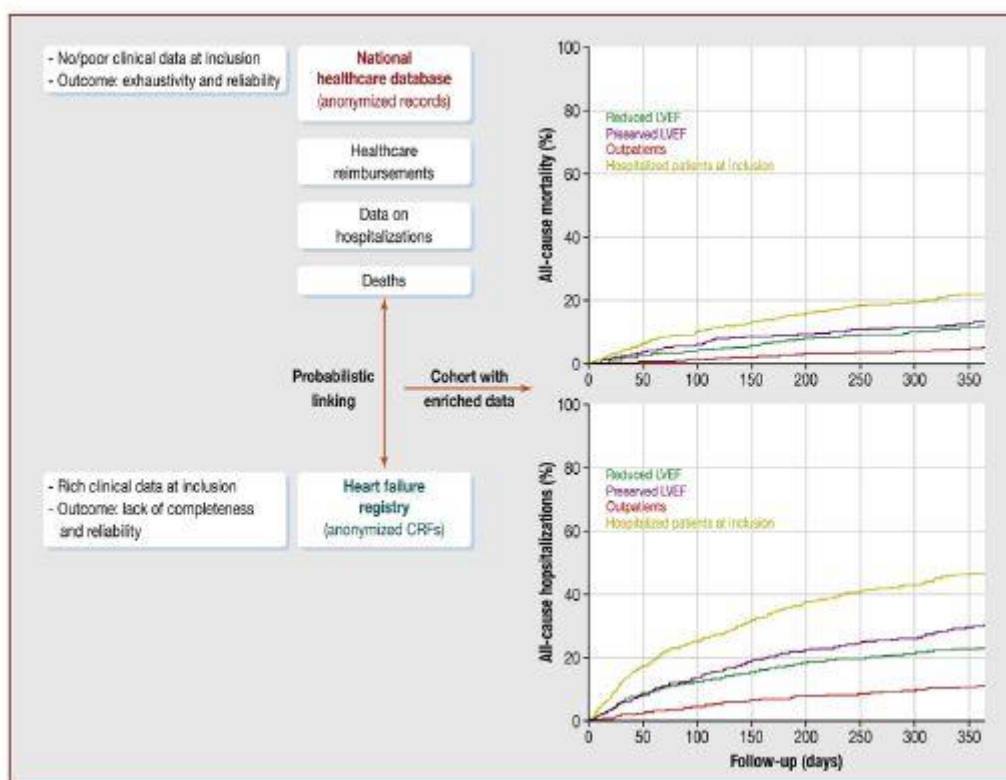
On the other hand, COPD and end-stage CKD were covered more completely in the SNDS. For COPD, the diagnosis was based on concomitant therapies in the SNDS. Although such medications are also prescribed for asthma, the age of the population makes this an unlikely confounder. The reasons for different degrees of completeness of co-morbidity entries would need further analysis, but it appears that both registries and medical claims databases have advantages and limitations, with no clear winner in this respect.

One of the main advantages of healthcare databases, such as the SNDS, over most registries is the completeness of events and their dates during follow-up. Such completeness is often difficult to obtain in academic registries because of the required capacity. In the FRESH registry, the rate of missing data on hospitalization was 17.5%, and the rate was even greater when the cause and/or the date of hospitalization were considered.

Consistency between the two databases can be altered by differences in dates. In the current analysis, we limited follow-up to 1 year; in studies with a longer follow-up, an even higher degree of completeness of the SNDS data compared with registries can be expected. For patients with data in the FRESH registry, the rate of unplanned hospitalizations was lower than in the SNDS, although it can be difficult to differentiate planned and unplanned hospitalization in the SNDS, which is a major limitation of medico administrative databases. To reduce the number of erroneously included planned hospitalizations in the analysis of unplanned hospitalizations in the SNDS, we removed day hospitalization and hospitalization

for < 2 nights, as these were unlikely to be hospitalizations for decompensation. Only hospitalizations ≥ 2 nights were considered in the SNDS, to avoid elective planned hospitalizations. Definitions of events in the SNDS are continually being revised with the arrival of new information and updates to guidelines as medical knowledge advances, which, although positive for the quality of the database, may weaken correlations with registry data.

Another issue with medico administrative databases is the presence of one main diagnosis and one or several associated and related diagnoses. For the same patient, HF code I50x can be used either as the main diagnosis or as an associated diagnosis, depending on the person in charge of administrative coding. For example, decompensated HF as a result of pneumopathy can lead to an “I50” coding as the main diagnosis or only as an associated diagnosis. Consequently, it is useful to show the two rates of HF-related hospitalizations according to either the single use of a main diagnosis or the use of main and associated diagnoses. Although it is imperfect, the reliability of HF coding in the SNDS was validated in a recent study [6]. A weakness of the current study is that events collected in the FRESH registry were not adjudicated by a blinded committee.



Central illustration. Linking between Heart Failure registry and national healthcare database resulting in enriched data especially due to the exhaustive follow-up. CRF: case report form; LVEF: left ventricular ejection fraction; PMSI: Programme de Médicalisation des Systèmes d’Information (French hospital discharge database); SNDS: Système National des Données de Santé (French National Healthcare System database); SNIIR-AM: Système National d’Informations Inter-Régimes de l’Assurance Maladie (French outpatient claims database).

4. Conclusions

In summary, this study showed that a significant proportion of patients from an HF registry could be matched anonymously to the national healthcare system database (SNDS) with high accuracy. This generated a database that will permit analyses that are not possible in the separate registries – in our case, outcomes in sub-populations with preserved versus reduced LVEF or acute versus chronic HF. The success with a patient database typical of a heterogeneous syndrome, such as HF, indicates that the methodology maybe applied to the linking of other registries in less distinctly defined populations; this could improve quality of data (fewer losses to follow-up, better accuracy) and improve the analytical power. The potential of the FRESH-SNDS linked database will be explored in further studies (central illustration).

Sources of funding

Funded by grants from Amgen, Astra Zeneca, Bayer, Boehringer-Ingelheim and Novartis..

Disclosure of interest

The authors declare that they have no competing interest.

References

- [1] de Peretti C, Perel C, Tuppin P, Iliou MC, Juilliere Y, Gabet A, et al. Prévalence et statut fonctionnel des cardiopathies ischémiques et de l'insuffisance cardiaque dans la population adulte en France: apports des enquêtes déclaratives « Handicap-Santé ». *Bull Epidemiol Hebd* 2014;172–81.
- [2] Bufalino VJ, Masoudi FA, Stranne SK, Horton K, Albert NM, Beam C, et al. The American Heart Association's recommendations for expanding the application of existing and future clinical registries: a policy statement from the American Heart Association. *Circulation* 2011;123:2167–79.
- [3] Harron K, Dibben C, Boyd J, Hjertqvist A, Azimaee M, Barreto ML, et al. Challenges in administrative data linkage for research. *Big Data Soc* 2017;4.
- [4] Bezin J, Duong M, Lassalle R, Droz C, Pariente A, Blin P, et al. The national health-care system claims databases in France, SNIIRAM and EGB: Powerful tools for pharmacoepidemiology. *Pharmacoepidemiol Drug Saf* 2017;26:954–62.
- [5] Bousquet C, Trombert B, Souvignet J, Sadou E, Rodrigues JM. Evaluation of the CCAM Hierarchy and Semi Structured Code for Retrieving Relevant Procedures in a Hospital Case Mix Database. *AMIA Annu Symp Proc* 2010;2010:61–5.
- [6] Bosco-Levy P, Duret S, Picard F, Dos Santos P, Puymirat E, Gilleron V, et al. Diagnostic accuracy of the International Classification of Diseases, Tenth Revision, codes of heart failure in an administrative database. *Pharmacoepidemiol Drug Saf* 2019;28:194–200.
- [7] Ingelsson E, Arnlov J, Sundstrom J, Lind L. The validity of a diagnosis of heart failure in a hospital discharge register. *Eur J Heart Fail* 2005;7:787–91.
- [8] Khand AU, Shaw M, Gemmel I, Cleland JG. Do discharge codes underestimate hospitalisation due to heart failure? Validation study of hospital discharge coding for heart failure. *Eur J Heart Fail* 2005;7:792–7.
- [9] Kumler T, Gislason GH, Kirk V, Bay M, Nielsen OW, Kober L, et al. Accuracy of a heart failure diagnosis in administrative registers. *Eur J Heart Fail* 2008;10:658–60.
- [10] McCormick N, Lacaille D, Bhole V, Avina-Zubieta JA. Validity of heart failure diagnoses in administrative databases: a systematic review and meta-analysis. *PLoS One* 2014;9:e104519.
- [11] Didier R, Gouysse M, Eltchaninoff H, Le Breton H, Commeau P, Cayla G, et al. Successful linkage of French large-scale national registry populations to national reimbursement data: improved data completeness and minimized loss to follow-up. *Arch Cardiovasc Dis* 2020;113:534–41.
- [12] Boudemaghe T, Belhadj I. Data Resource Profile: The French National Uniform Hospital Discharge Data Set Database (PMSI). *Int J Epidemiol* 2017;46:392–400.

[13] Tuppin P, Rudant J, Constantinou P, Gastaldi-Menager C, Rachas A, deRoquefeuil L, et al. Value of a national administrative database to guide public decisions: From the systeme national d'information interregimes del'Assurance Maladie (SNIIRAM) to the systeme national des donnees de sante(SNDS) in France. *Rev Epidemiol Sante Publique* 2017;65(Suppl 4):S149–67.[

14] Système National des Données de Santé. Qu'est-ce que le SNDS ? 2020. Available at: <https://documentation-snds.health-data-hub.fr/introduction/01-snds.html#a-quoi-le-snds-peut-il-servir> [accessed date: 18th October 2021].