



HAL
open science

Development and validation of algorithms to predict left ventricular ejection fraction class from healthcare claims data

Damien Logeart, Maxime Doublet, Margaux Gouysse, Thibaud Damy,
Richard Isnard, François Roubille

► To cite this version:

Damien Logeart, Maxime Doublet, Margaux Gouysse, Thibaud Damy, Richard Isnard, et al.. Development and validation of algorithms to predict left ventricular ejection fraction class from healthcare claims data. ESC Heart Failure, In press, 10.1002/ehf2.14725 . hal-04489541

HAL Id: hal-04489541

<https://hal.science/hal-04489541>

Submitted on 8 Mar 2024

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.

Development and validation of algorithms to predict left ventricular ejection fraction class from healthcare claims data

Damien Logeart ¹, Maxime Doublet ², Margaux Gouysse ², Thibaud Damy ³, Richard Isnard ⁴, François Roubille ⁵

¹ Department of Cardiology, Paris Cité University, AP-HP Hôpital Lariboisière, Inserm U942, 2 rue Ambroise Paré, Paris, France.

² Clinityx, Boulogne-Billancourt, France.

³ Department of Cardiology and French National Reference Centre for Cardiac Amyloidosis, Hôpitaux Universitaires Henri-Mondor AP-HP, IMRB, Inserm, Université Paris-Est Créteil, Créteil, France.

⁴ Hôpital Pitié-Salpêtrière, AP-HP, Paris, France.

⁵ Department of Cardiology, INI-CRT PhyMedExp Inserm CNRS, CHU de Montpellier, Université de Montpellier, Montpellier, France.

Keywords:

Claims database; Heart failure; Left ventricular ejection fraction; Machine learning; Registry records.

Abstract

Aims:

The use of large medical or healthcare claims databases is very useful for population-based studies on the burden of heart failure (HF). Clinical characteristics and management of HF patients differ according to categories of left ventricular ejection fraction (LVEF), but this information is often missing in such databases. We aimed to develop and validate algorithms to identify LVEF in healthcare databases where the information is lacking.

Methods and results:

Algorithms were built by machine learning with a random forest approach. Algorithms were trained and reinforced using the French national claims database [Système National des Données de Santé (SNDS)] and a French HF registry. Variables were age, gender, and comorbidities, which could be identified by medico-administrative code-based proxies, Anatomical Therapeutic Chemical codes for drug delivery, International Classification of Diseases (Tenth Revision) coding for hospitalizations, and administrative codes for any other type of reimbursed care. The algorithms were validated by cross-validation and against a subset of the SNDS that includes LVEF information. The areas under the receiver operating characteristic curve were 0.84 for the algorithm identifying $LVEF \leq 40\%$ and 0.79 for the algorithms identifying $LVEF < 50\%$ and $\geq 50\%$. For $LVEF \leq 40\%$, the reinforced algorithm identified 50% of patients in the validation dataset with a positive predictive value of 0.88 and a specificity of 0.96. The most important predictive variables were delivery of HF medication, sex, age, hospitalization, and testing for natriuretic peptides with different orders of positive or negative importance according to the LVEF category.

Conclusions:

The algorithms identify reduced or preserved LVEF in HF patients within a nationwide healthcare claims database with high positive predictive value and low rates of false positives.

Introduction

Data from real-life patient registries are highly valuable for the assessment of efficiency, safety, and outcomes of treatments, as well as the quality and performance of healthcare provided in everyday practice. Claims databases include data on very large numbers of individuals and are increasingly used in comparative effectiveness and safety studies of drug treatments.¹ However, claims databases do not always capture all information relevant to such analyses and may need to be enriched by other sources such as registries, the quality of which often suffers from data gaps and losses to follow-up.² Hence, there is a major need for improved data quality overall.

Heart failure (HF) is an important public health issue with high prevalence—up to 1–2% of the population—and a severe impact on morbidity and mortality despite considerable progress in therapies and management over the last decades.³ Most of the available treatments are effective in HF with reduced ($\leq 40\%$) left ventricular (LV) ejection fraction (LVEF).⁴ However, there is a relative lack of pharmacological agents with proven efficacy in patients with preserved LVEF ($\geq 50\%$) and little data on patients with mildly reduced LVEF between 41% and 49%. Thus, information on LV class is important for an informative analysis of the care pathways and outcomes of HF patients. Yet such data are often not captured in claims databases.

The French health system offers universal health protection to anyone who works or resides in France on a stable and regular basis. The single-payer nationwide healthcare database [Système National des Données de Santé (SNDS)] includes health-related reimbursement information on close to 99% of the French population or more than 68 million persons, making it one of the world's largest continuous claims databases.⁵ The SNDS also includes data on mortality, hospitalizations, and chronic diseases. By law, all SNDS data are anonymized. The International Classification of Diseases (Tenth Revision, ICD-10) codes are used to identify reasons for hospitalizations, including HF, but up to 2019, the SNDS contained no information on LVEF class. We have previously reported on the feasibility of anonymous matching of the population in the SNDS and a registry cohort of patients, either hospitalized for HF or identified as outpatients [FRENch Survey on HeartFailure (FRESH); NCT01956539].⁶ As the FRESH registry captures LVEF dimensions, the linked population can be viewed as an enriched SNDS cohort, which enables outcomes analyses according to LVEF class. Because of the small size of this cohort as well as its imperfect representativeness of real life, we need to study HF in a much larger sample of patients such as the SNDS database. However, LVEF class remains poorly informed in the nationwide healthcare database.

The availability of two subsets of SNDS data, the linked cohort from before 2019 and the later database, which includes LVEF information, provides data that may be used to develop proxy-based algorithms to identify LVEF categories in administrative healthcare databases where this information is missing. We here report the development and validation of such an algorithm.

Methods

Data sources

Two data sources were used: the SNDS database from the year 2019 ('2019-SNDS') and the cohort from the linked FRESH/SNDS databases ('FRESH-SNDS').

The SNDS contains data on care pathways (medical consultations, nursing care, biology, imaging, etc.), hospitalizations (types and causes described by ICD-10 classification), drug deliveries, and deaths.⁵ All reimbursed provision of care, devices, and drugs are encoded with specific administrative codes. Age and sex are included in the database, but there are no clinical data or results from tests. Following an ICD-10 update, the LVEF category is taken into account in coding of HF hospitalization since 2019. All data are anonymized, and patients are registered by a unique personal identification number that allows linking of hospitalization registries, the national healthcare reimbursement database, and the national registry of deaths.

We obtained authorization to identify and analyse HF patients within the SNDS. Patients were identified by at least one of the following within SNDS: (i) hospitalization with a main, associated, or related diagnosis corresponding to HF (ICD-10 Codes I50, I500, and I501); (ii) at least one dispensation of sacubitril/valsartan; (iii) pacemaker with biventricular resynchronization with or without an implantable automatic defibrillator; and (iv) long-term disease status for HF (full reimbursement of care related to chronic HF). From this dataset, we extracted the subset of HF patients who were hospitalized during the first quarter of 2019 with a main, associated, or related diagnosis coding corresponding to HF and with an HF coding that included LVEF categories: I50-1, I500-1, or I501-1 for LVEF \leq 40%; I50-2, I500-2, or I501-2 for LVEF 41–49%; and I50-3, I500-3, or I501-3 for LVEF \geq 50%. We assumed that the LVEF class for a patient was the same in 2019 as in the 1 or 2 years before. Only patients who were alive at least 3 months after hospitalization were included. Patients with no dispensation of at least one cardiac treatment during the period of interest (6 months before and 12 months after the index date) were excluded.

The construction of the FRESH-SNDS database has been described previously.⁶ The FRESH multicentre cohort includes 2719 HF patients enrolled between 2014 and 2018, either during a hospitalization for decompensation or during an outpatient clinic visit. The matched population consists of 1656 patients with claims data on care pathways and outcomes for 2 years before inclusion and up to 3 years after inclusion.

Included variables

Variables of interest were identified according to literature reviews and validated by the FRESH Scientific Committee. Four types of variables were extracted from the SNDS: demographics (age, sex, and chronic diseases), treatments [Anatomical Therapeutic Chemical (ATC) codes for drugs], comorbidities (proxies based on ICD-10, chronic diseases, ATC, and others related to specific medical or surgical procedures), and care pathway (consultations, biology, and hospitalization). All variables were initially binary (0 or 1) according to the presence or absence of the event and then censored (2) according to death occurring before the period or in the first half of the observed period. Age was stratified into five categories: \leq 60, 60–70, 71–80, 81–90, and $>$ 90 years. The index date was set to the first day of the

quarter associated with the LVEF diagnosis. Treatments and care pathway variables were captured at 3 and 6 months before and 12 months after the index date. Comorbidities were included at 6 months and for the maximum time available from the index date. Among 547 variables identified as relevant for the HF algorithm, the final selection of variables used in algorithms was obtained during the training process of the respective algorithms.

Three LVEF categories are currently used in the guidelines. However, the category HF with midrange LVEF (41–49%) comprises an exceptionally heterogeneous population, and many patients with midrange LVEF can change class over time.⁷ Thus, we did not develop a specific algorithm for this class of patients. We tested the LVEF thresholds at 40% and 50%.

LVEF categories for the FRESH-SNDS cohort were obtained directly from the FRESH data files.

For the 2019-SNDS cohort, LVEF categories were obtained by using the updated ICD-10 codes (Supporting Information, *Table S1*): I50-1, I500-1, or I501-1 for LVEF \leq 40%; I50-2, I500-2, or I501-2 for LVEF 41–49%; and I50-3, I500-3, or I501-3 for LVEF \geq 50% (and 9 when LVEF was unknown). All patients with ambiguous diagnoses were excluded from this analysis.

Generation of algorithms

Algorithms were built using random forest machine learning⁸ with a three-step approach (*Figure 1*). Algorithms were generated with scikit-learn (Python ML library).

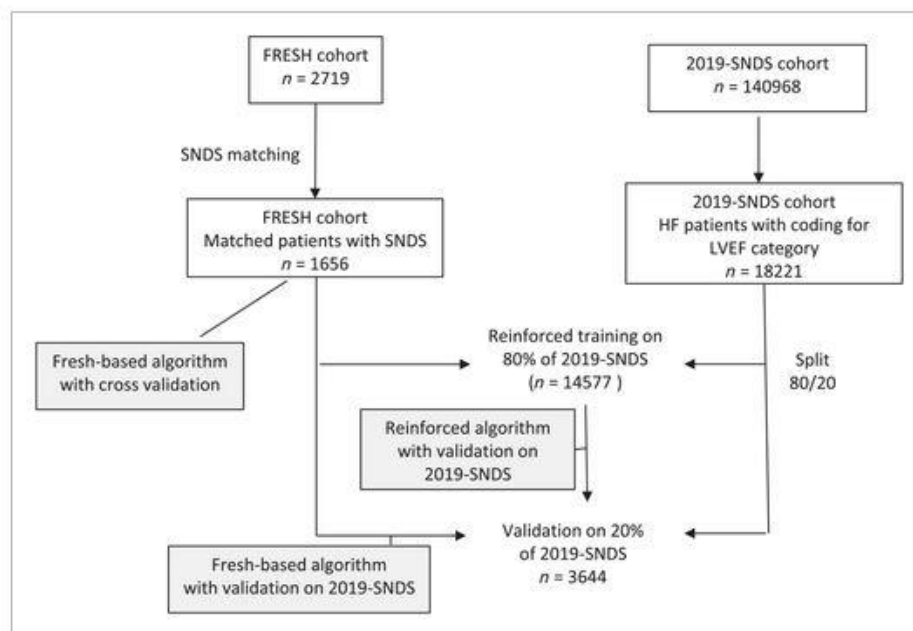


Figure 1: Study design with flow charts of datasets that were used for algorithms. FRESH, FRENch Survey on HeartFailure; LVEF, left ventricular ejection fraction; SNDS, Système National des Données de Santé.

Algorithms were initially trained on the FRESH-SNDS dataset rather than the 2019-SNDS dataset because the FRESH cohort included HF patients with or without hospitalization over a wide period contrary to the 2019-SNDS cohort. The performance was calculated either using cross-validation with five-fold and bootstrap method ('FRESH-based with cross-validation') or against the 2019-SNDS dataset ('FRESH-based with SNDS validation'). Because of the limited number of patients and the inherent bias of the FRESH cohort, reinforcement learning was performed by enriching with new decision trees trained on 80% of patients in the 2019-SNDS cohort. The resulting 'reinforced algorithms' included 50% trees from FRESH-SNDS and 50% trees from 2019-SNDS. The performances of 'FRESH-based with SNDS validation' as well as 'reinforced algorithm' were calculated on the remaining 20% of patients from the 2019-SNDS dataset (validation dataset).

Finally, two sets of algorithms were constructed, one for discriminating patients with LVEF either $\leq 40\%$ or $>40\%$ and the second for discriminating patients with LVEF $\leq 50\%$ or $>50\%$.

Isotonic calibration was applied to adjust the model's predicted probabilities.⁹ The effectiveness of the calibration was evaluated by the Brier score (average of the squared differences between the predicted probabilities and the observed outcomes) and by calibration curves.

Statistical methods

A machine learning approach was carried out, and we tested the following models: logistic regression, random forest, and extreme gradient boosting (XGB). In this work, random forest resulted in better performances than other models. Random forest is an assembling model that trains and combines multiple tree-based models to predict the outcome of interest. It combines bagging (bootstrap aggregation) with random feature selection to build uncorrelated decision trees. Each tree yields predictions, which are then aggregated into one final result by a voting method to give a final, more accurate global prediction. The main benefits of these models are versatility, reduced variance, and bias, as well as the ability to determine the importance of each variable in the final output. In our model, the output was the LVEF category selected by most trees. The tuning of hyperparameters was obtained with the GridSearch method.

French legislation bars the merging of the two datasets. Hence, the reinforcement of the algorithm using the 2019-SNDS dataset was performed using the warm-start method,¹⁰ which allows the addition of decision trees trained on the 2019-SNDS dataset to the random forest results while preserving variables and parameters previously optimized using the FRESH-SNDS dataset. The algorithm parameters were optimized according to the metrics of interest [positive predictive value (PPV) and sensitivity]. Different probability thresholds were also tested to obtain different populations.

To evaluate the performance of the classification model, analyses of the confusion matrix of predictions as well as the receiver operating characteristic (ROC) curves were performed. Additional performance evaluation metrics included PPV or precision, sensitivity or recall, specificity, negative predictive value, accuracy, and Matthew's correlation coefficient. As the primary objective of the algorithm was to allow subsequent studies in the whole HF population in the SNDS according to LVEF categories, the main quality criterion was the maximal PPV while maintaining an acceptable sensitivity to minimize bias. The importance of included variables and their hierarchy in the algorithms were calculated and shown by their SHapley Additive exPlanations (SHAP) values.¹¹

Ethics

All patients in the FRESH cohort provided written, informed consent. Ethics approval to link the FRESH and SNDS databases was obtained from the French Expert Committee on Research, Studies and Assessments in Health (CEREES) on 25 April 2019 (TPS 347113bis) 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, 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 the 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.

Results

Study populations

In the 2019-SNDS dataset, 18 221 HF patients were hospitalized during the first quarter of 2019 with a main, associated, or related diagnosis of HF (I50x) and LVEF coding (Supporting Information, *Figure S1*). In this cohort, 37.0% and 58.4% of the population had LVEF \leq 40% and $<$ 50%, respectively. As compared with the 2019-SNDS dataset of HF patients with LVEF coding, HF patients were older and had a lower rate of coding for ischaemic heart disease, atrial fibrillation, diabetes, and HF drugs delivery than in the whole 2019-SNDS dataset (Supporting Information, *Table S1*). In the FRESH-SNDS cohort, 59.8% and 71.7% of HF patients had LVEF \leq 40% and $<$ 50%, respectively. *Table 1* shows the demographics and clinical characteristics of patients in the two datasets. Patients from the 2019-SNDS cohort were significantly older, more frequently female, and had more comorbidities than patients in the FRESH-SNDS cohort. In both cohorts, most characteristics differed according to LVEF categories: HF patients with reduced LVEF were younger, had less comorbidities and more ischaemic heart disease, and received more HF drugs. However, outcomes did not differ between LVEF classes.

	FRESH-SNDS				2019-SNDS		
	All	LVEF ≤ 40	LVEF < 50	LVEF ≥ 50	All	LVEF ≤ 40	LVEF < 50
Chronic disease	40.0%	37.9%	37.8%	45.7%	78.3%	77.4%	77.3%
Comorbidities							
Hypertension	52.2%	47.7%	48.4%	62.0%	72.4%	66.1%	68.0%
Diabetes	24.6%	23.3%	23.1%	28.4%	33.4%	33.4%	33.1%
Atrial fibrillation	32.2%	31.9%	31.6%	34.0%	60.2%	58.3%	57.6%
Infarction	21.6%	23.4%	23.6%	16.4%	17.2%	21.2%	21.7%
Cardiomyopathy	24.3%	28.6%	27.4%	16.4%	31.3%	48.1%	40.4%
PAD	12.4%	11.5%	11.5%	14.7%	15.6%	15.6%	15.5%
Stroke	7.5%	7.5%	7.2%	8.1%	9.1%	9.1%	9.2%
COPD	13.0%	11.9%	11.9%	15.8%	22.1%	18.7%	19.5%
Sleep apnoea	11.1%	11.0%	11.2%	10.9%	12.6%	12.0%	12.3%
Valvulopathy	17.4%	17.56%	17.4%	17.5%	34.4%	35.6%	34.3%
Treatment							
Loop diuretics	65.3%	64.7%	63.0%	71.1%	85.6%	88.5%	84.4%
MRA	38.8%	47.1%	44.5%	24.1%	30.5%	46.5%	37.9%
ACE inhibitors	50.3%	55.0%	52.9%	43.6%	51.0%	59.8%	58.3%
Beta-blockers	66.2%	67.7%	67.1%	64.1%	78.8%	88.9%	85.8%
Calcium blockers	19.7%	13.8%	15.1%	31.6%	30.5%	18.6%	23.5%
1 year outcome							
GP consultation	94.0%	93.6%	93.8%	94.7%	94.7%	94.9%	94.9%
All-cause hospitalization	71%	73.4%	72%	72%	67.5%	70.5%	68.3%
HF hospitalization	21%	20.0%	20%	19%	19.6%	23.6%	19.5%

Table 1. Baseline characteristics of the registry linked to the claims database (FRESH-SNDS) and 2019 claims database (2019-SNDS) cohorts according to the LVEF category. ACE, angiotensin-converting enzyme; COPD, chronic obstructive pulmonary disease; GP, general practitioner; LVEF, left ventricular ejection fraction; MRA, mineralocorticoid antagonist; PAD, peripheral arterial disease. Treatments were cardiac drugs delivered during the 1 year follow-up. HF hospitalization was defined as the main diagnosis.

Performance of algorithms

Reinforcing algorithms with new decision trees trained on 80% of patients in the 2019-SNDS cohort resulted in the best global performance. [Table 2](#) shows the performance of reinforced algorithms. Supporting Information, [Figure S2](#) shows calibration curves for reinforced algorithms before and after applying isotonic calibration. Brier scores before and after calibration were 0.16 and 0.14 for $LVEF \leq 40\%$ and 0.19 and 0.18 for $LVEF < 50\%$, respectively.

LVEF class	AUC	PPV	SE	NPV	SPE	ACC	MCC
$\leq 40\%$	0.83	0.64	0.74	0.84	0.76	0.75	0.49
$> 40\%$	0.83	0.84	0.76	0.75	0.64	0.75	0.49
$< 50\%$	0.79	0.73	0.81	0.70	0.60	0.72	0.42
$\geq 50\%$	0.79	0.70	0.60	0.72	0.80	0.72	0.41

Table 2. Performances of the reinforced algorithms for predicting $LVEF \leq 40\%$ and $> 40\%$ and $LVEF < 50\%$ and $\geq 50\%$. ACC, accuracy; AUC, area under the receiver operating characteristic curve; LVEF, left ventricular ejection fraction; MCC, Matthew's correlation coefficient; NPV, negative predictive value; PPV, positive predictive value; SE, sensitivity; SPE, specificity. These metrics were calculated for a probability threshold of 0.5 and after isotonic calibration.

[Figure 2](#) shows ROC curves obtained with the reinforced algorithms. For $LVEF \leq 40\%$, the reinforced algorithm identified 50% of patients (true positives) in the validation dataset with a PPV of 0.88 and a specificity of 0.96. The clinical characteristics of true and false positives predicted by the reinforced algorithm for $LVEF \leq 40\%$ in the validation dataset and using a probability threshold of 0.5 were highly similar (Supporting Information, [Table S4](#)) as were the characteristics of true and false negatives.

For HF and $LVEF < 50\%$, the reinforced algorithm identified 50% of patients (true positives) with a PPV of 0.85. With the false-positive rate reduced to 5%, the algorithm identified 40% of true positives. For HF and preserved LVEF ($\geq 50\%$), the area under the ROC curve (AUC) was similar to the $LVEF < 50\%$ group, but the cut-off for identifying 50% of patients (true positives) had a 17% rate of false positives (PPV 0.71). With a false-positive rate of $\leq 5\%$, 22% of true positives were identified ([Figure 2](#)).

Supporting Information, [Table S3](#) shows the performances of algorithms based only on the FRESH-SNDS dataset: Model 1 with cross-validation and Model 2 with external validation (SNDS database).

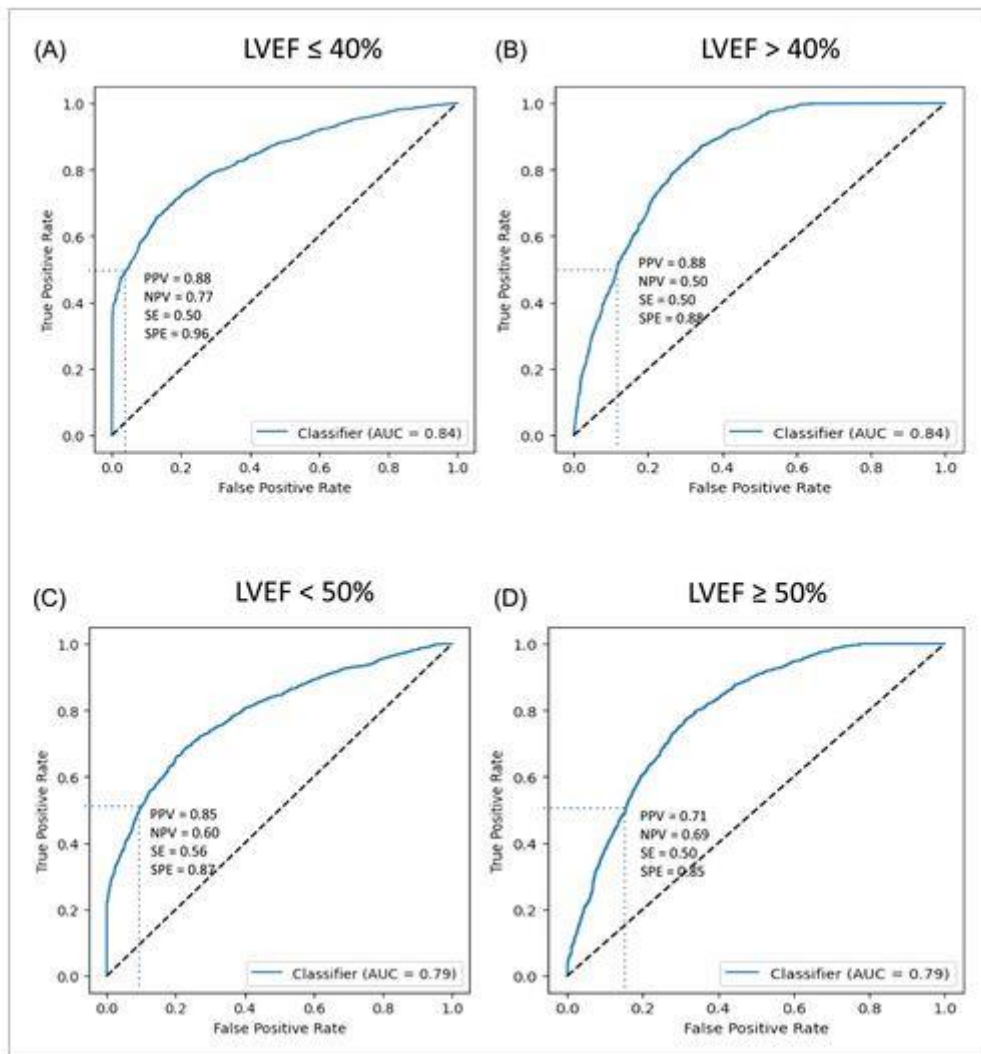


Figure N°2 Receiver operating characteristic curves of reinforced algorithms to predict (A) left ventricular ejection fraction (LVEF) $\leq 40\%$, (B) LVEF $> 40\%$, (C) LVEF $< 50\%$, and (D) LVEF $\geq 50\%$. The area under the receiver operating characteristic curve (AUC) is indicated, as well as metrics for different probability thresholds. NPV, negative predictive value; PPV, positive predictive value; SE, sensitivity; SPE, specificity.

The importance of the 14 most significant variables used by the reinforced algorithms to predict LVEF categories is shown in *Figure 3*. Many variables, for example, administration of drugs, were included for different time periods (12–6 months before hospitalization, 6 months before hospitalization, or 6 months after hospitalization). As there were no differences between the periods for such variables, the values were aggregated in the final analysis. The most important variables that predicted LVEF $\leq 40\%$ were delivery of mineralocorticoid receptor antagonists (MRAs), beta-blockers, loop diuretics, and angiotensin-converting enzyme (ACE) inhibitors; male sex; age ≤ 75 years; high rate of natriuretic peptide assays; and cardiomyopathy. The same variables predicted LVEF $< 50\%$, differing only in ranking.

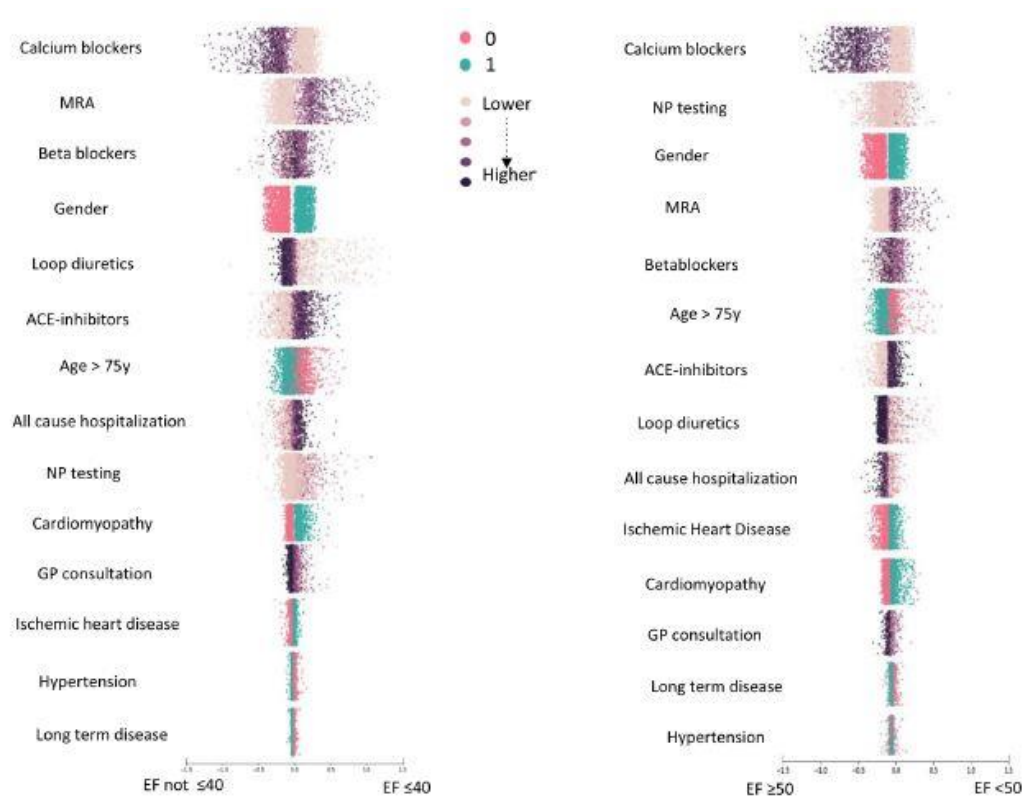


Figure N°3 : Weighting of variables in reinforced algorithms based on SHapley Additive exPlanations values to predict left ventricular ejection fraction (LVEF) $\leq 40\%$ and LVEF $< 50\%$ and $\geq 50\%$. ACE, angiotensin-converting enzyme; EF, ejection fraction; GP, general practitioner; MRA, mineralocorticoid antagonist; NP, natriuretic peptide.

The weight of the variables differed between the ‘FRESH-based’ and ‘reinforced’ algorithms with the reinforced algorithm weighting attributing a lower weight to age and sex (Supporting Information, [Figure S2](#)). In addition to these two variables, for LVEF $< 50\%$, the reinforced algorithm attributed a lower weight to the type of heart disease (ischaemic heart disease and cardiomyopathy).

Neither the use of specific algorithms for outpatients and inpatients nor the attribution of different weights to these groups improved the results compared with the final algorithm (data not shown).

Discussion

Claims databases are important sources of information for the evaluation of health outcomes, healthcare utilization, and cost of care among HF patients in routine care. However, LVEF measurements are often not available, which limits the possible range of analyses. The present communication describes the development of machine learning algorithms to identify patients with reduced or preserved LVEF in the French SNDS national claims database based on the international diagnostic codes of hospitalizations (ICD-10), pharmacological treatments (ATC), and codes related to the billing of other treatments reimbursed by the national health system, but without access to demographic or clinical data except age and gender. As the

database covers ~99% of the population, these algorithms have the potential to enable scientific and health-economic population-based studies on HF patients with either reduced or preserved LVEF nationwide. The significant disparities between these types of HF in clinical characteristics, treatments, care pathways, and outcomes⁴ mean that the groups need to be analysed separately to provide useful information.

Several claims-based models to predict LVEF class have been developed internationally in recent years. Our study is the first such exercise in a French HF population and the first to utilize a random forest-based algorithm using only administrative codes from a national health insurance system. As each health system is unique, capturing differing variables, it can be difficult to compare results between algorithms. In a previous study, Desai *et al.*¹² built an administrative claims data-based model to identify patient subgroups with specific LVEF class by linking Medicare data to electronic medical records containing LVEF measurements from two academic medical centres and using logistic regression models. In their study, the best model was obtained using the LVEF threshold of 0.45 and the algorithm required up to 35 predictor variables defined in Medicare claims. Recently, the authors published an external validation with a sample of commercial insurance enrollees.¹³ Other studies have utilized clinical data from electronic medical records.^{14, 15} Such algorithms could be useful for research purposes to identify HF and LVEF categories from databases where the specific information is not present.

Other algorithms have identified similar key predictors to ours. HF medication is repeatedly confirmed as one of the strongest differentiators between reduced and preserved LVEF.¹⁶ In our model, the use of calcium channel blockers was strongly negatively related to reduced LVEF, which is in accordance with the contraindication of this drug class in HF with reduced LVEF.⁴ Natriuretic peptide lab code was an important differentiator between preserved and reduced LVEF; similarly, natriuretic peptide lab code predicted preserved LVEF in a study of data from inpatient and outpatient visits.¹⁷

The key predicting variables did not differ to any relevant degree between the algorithms, but the relative importance of the variables differed between the predicted LVEF classes. That sex was strongly predictive is in line with the higher prevalence of reduced LVEF in men. Conversely, greater age, female sex, and hypertension are known risk factors for HF with preserved LVEF. The predictive value of the variable ‘long-term disease’ (full reimbursement for specific chronic disease) related to other diseases than HF reflects the weight of comorbidities. The highly similar set of important predictive variables in algorithms for LVEF $\leq 40\%$ and $<50\%$ are in agreement with the most recent studies and guidelines, which have suggested that HF patients with mildly reduced LVEF (41–49%) are close to HF patients with reduced LVEF ($\leq 40\%$) and should probably be treated in the same way.¹⁸ Interestingly, we also observed that the clinical characteristics of false-positive patients identified by the algorithm for LVEF $\leq 40\%$ (for the usual probability threshold of 0.50) were close to those in true-positive patients (Supporting Information, *Table S2*), which is reassuring for subsequent analyses on healthcare databases using these algorithms to identify LVEF.

The lower PPV and higher rate of false positives for predicting preserved LVEF compared with LVEF $\leq 40\%$ in our model show the continuing difficulties in identifying (and treating) this type of HF. Among previous studies that used only medico-administrative codes, Sundaram *et al.*¹⁶ and Cohen *et al.*¹⁷ reported even lower performance of their predicting algorithms for preserved LVEF with PPV $\leq 35\%$. However, higher performance such as PPV at 81% could be obtained to predict HF with preserved ejection fraction (HFpEF), but such

performance was obtained by adding numerous comorbidities.¹³ In our study, the final algorithm for preserved LVEF was the same as for LVEF < 50%, that is, patients predicted as 'not LVEF < 50%' were the same as those predicted as $\geq 50\%$. By the same token, the AUCs were similar for predicting the two LVEF categories.

To develop algorithms, several approaches were tested including standard logistic regression and machine learning models. The best results were obtained with the random forest approach, which resulted in algorithms with an AUC above 0.75 in most cases when logistic regression resulted in algorithms with AUCs that were never above 0.70 (data not shown). The random forest approach is particularly robust with the noisy data that typically characterize administrative claims and electronic health records datasets and has demonstrated high reliability to predict outcomes or identify specific populations within large databases.¹⁹⁻²¹ The variables selected by the random forest algorithm were consistent with regard to the literature describing the characteristics of HF patients according to their LVEF category.^{4, 22}

Even when data are supposed to be captured, databases are rarely complete. Changes to the SNDS in 2019 include ICD-10 codes for the LVEF category and specific diagnostic codes for HF hospitalization as the main, associated, or related diagnosis. Yet, in the 2019-SNDS dataset used for the current algorithm development, the LVEF category was coded in <20% of patients hospitalized with an HF diagnosis. What is more, coding inconsistencies were identified in more than half of HF patients with LVEF coding. With time, LVEF coding in the SNDS will probably improve, but there will likely be a need for identification algorithms in the foreseeable future. In addition, a number of HF patients are not hospitalized because of HF or may not be hospitalized during the time of analyses and will only be identifiable by models such as the one presented here.

The study has a number of limitations. The reliability of any machine learning system depends on the quality and size of the training and validation datasets. The FRESH cohort comprises patients with confirmed HF diagnosis and LVEF dimensions. These patients may not be fully representative of the real-life HF population, and the sample size is relatively small. The quality of the 2019-SNDS dataset with updated ICD-10 codes for the LVEF category was not independently verified. We also assumed that patients did not change LVEF class in the analysis period between 2017 and 2019. In addition, there were differences in baseline characteristics between the subgroup of the 2019-SNDS dataset that we used for reinforced algorithms (HF patients with coding for LVEF class) and the whole 2019-SNDS dataset (with or without coding for LVEF class); these differences could be related to differences in the global accuracy of diagnostic coding. As an example, I50 was the main diagnostic code in 70% of the patients in the subgroup vs. only 50% of the whole dataset. The 'reinforced algorithms' that obtained the best predictive performances were trained on two independent cohorts and validated on the remaining part of one cohort. An additional and external validation (on a third independent cohort) would have been interesting. Because a number of HF drugs were included in the development, the populations identified by the trained algorithms may not be suitable for the analysis of optimal HF therapy. However, an optimization analysis of the most recent evidence-based HF drugs, angiotensin receptor/neprilysin inhibitor and sodium–glucose cotransporter-2 inhibitor, would be possible, as these agents were not included in the algorithms. The involvement of HF drugs in algorithms makes them dependent on geographical and socioeconomic factors as well as changes in national or international guidelines over time. Consequently, any algorithm would need to be reworked for a specific claim.

To get reliable predictive models, it is crucial to use at least two different datasets for training, reinforcement, and external validation. In our work, the FRESH cohort was matched to SNDS to collect medico-administrative data on treatments, care pathways, and clinical events before and after inclusion. This cohort was different from the 2019-SNDS cohort that we used for reinforcement and validation. Indeed, this 2019 SNDS cohort included only HF patients with hospitalization during the first quarter of 2019, while the FRESH cohort included patients without hospitalization and patients with hospitalization over a longer period. It is why we firstly trained predictive models on FRESH-SNDS. It should be noted that the performances of algorithms were lower when models were firstly trained on 2019-SNDS cohort and then reinforced on FRESH-SNDS cohort (data not shown).

In conclusion, these validated claims-based algorithms can identify reduced or preserved LVEF in HF patients within a nationwide healthcare database with high PPV and low rates of false positives. Using these tools, specific analyses of different classes of HF would be possible. They could also help identify, on a nationwide scale, patients who would benefit from targeted HF management information or preventive interventions.

Acknowledgements

The authors express sincere thanks to the French Society of Cardiology (SFC), promotor of the FRESH registry, and to all FRESH investigators at participating sites across France (Appendix 1). Under the guidance of the authors, editorial support was provided by Pelle Stolt.

Conflict of interest

D.L. received honoraria for consulting, lectures, or educational events from AstraZeneca, Bayer, Boehringer Ingelheim, Novartis, Pfizer, and Vifor. T.D. received honoraria for consulting, lectures, or educational events from Akcea Therapeutics, Alnylam Pharmaceuticals, AstraZeneca, Bayer, Boehringer Ingelheim, Novartis, Pfizer, Prothena, and Vifor and grants from Alnylam Pharmaceuticals, GSK, Neurimmune, Novartis, and Pfizer. F.R. received honoraria for consulting, lectures, or educational events from Abbott, Air Liquide, AstraZeneca, Bayer, Boehringer Ingelheim, Novartis, Novo Nordisk, Pfizer, Servier, and Vifor and grants from Air Liquide and Abbott. R.I. received honoraria for consulting, lectures, or educational events from AstraZeneca, Bayer, Boehringer Ingelheim, Novartis, Servier, and Vifor. M.D. and M.G. have no conflicts of interest.

Funding

This study was partially funded by grants from the Amgen Foundation, AstraZeneca, Bayer, Boehringer Ingelheim, and Novartis.

Appendix: List of centres and investigators

C. Augier (Grenoble), M. Bailly (Kremlin Bicetre), F. Bauer (Rouen), F. Beauvais (Paris), E. Berthelot (Kremlin Bicetre), N. Bihry (Lyon), K. Bourdariat (Lyon), N. Braik (Besançon), Y. Bohbot (Amiens), C. Casset (Grenoble), J. Cautela (Marseille), S. Champion (Le Chesnay), G. Clerici (Saint Pierre de la Réunion), R. Codjia (Senlis), P. DeGroote (Lille), I. DeSauniere (Roanne), B.D. Here (Rouen), J.C. Eicher (Dijon), M. Fayard (Chalon-sur-Saône), M. Fertin (Lille), A. Galat (Créteil), M. Galinier (Toulouse), M. Goralski (Orléans), E. Inorowicz-

Guillaume (Pontivy), A. Jobbe Duval (Saint-Nazaire), G. Jondeau (Paris), F. Massin (Montpellier), Z. Moutacalli (Villeneuve-Saint-Georges), N. Mewton (Lyon), J. Philip (Chalon-sur-Saône), F. Picard (Bordeaux), M. Poilane (Nantes), J. Roncalli (Toulouse), F. Roubille (Montpellier), S. Rouffiac (Créteil), R. Sabatier (Caen), M. Salvat (Grenoble), L. Sebbag (Lyon), M.F. Seronde (Besançon), J.M. Tartiere (Toulon), J.N. Trochu (Nantes), C. Tribouilloy (Amiens), J. Vonhunolstein (Strasbourg), and W. Yafi (Orléans).

References

1. Pastorino R, De Vito C, Migliara G, Glocker K, Binenbaum I, Ricciardi W, et al. Benefits and challenges of Big Data in healthcare: An overview of the European initiatives. *Eur J Public Health* 2019;29:23-27. doi:10.1093/eurpub/ckz168
2. Bufalino VJ, Masoudi FA, Stranne SK, Horton K, Albert NM, Beam C, et al. The American Heart Association's recommendations for expanding the applications of existing and future clinical registries: A policy statement from the American Heart Association. *Circulation* 2011;123:2167-2179. doi:10.1161/CIR.0b013e3182181529
3. Roger VL. Epidemiology of heart failure: A contemporary perspective. *Circ Res* 2021;128:1421-1434. doi:10.1161/CIRCRESAHA.121.318172
4. McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M, et al. 2021 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure: Developed by the Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure of the European Society of Cardiology (ESC). With the special contribution of the Heart Failure Association (HFA) of the ESC. *Eur Heart J* 2021;42:ehab368. doi:10.1093/eurheartj/ehab368
5. 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:S149-S167. doi:10.1016/j.respe.2017.05.004
6. Logeart D, Damy T, Doublet M, Salvat M, Tribouilloy C, Bauer F, et al. Feasibility and accuracy of linking a heart failure registry to the national claims database using indirect identifiers. *Arch Cardiovasc Dis* 2023;116:18-24. doi:10.1016/j.acvd.2022.11.002
7. Schiattarella GG, Tong D, Hill JA. Can HFpEF and HFrEF coexist? *Circulation* 2020;141:709-711. doi:10.1161/CIRCULATIONAHA.119.045171
8. Breiman L. Random forests. *Mach Learn* 2001;45:5-32. doi:10.1023/A:1010933404324
9. Berta E, Bach F, Jorda M. Classifier calibration with ROC-regularized isotonic regression. <https://arxiv.org/pdf/2311.12436.pdf>. Accessed 22 November 2023
10. Ash JT, Adams RP. On warm-starting neural network training. *Adv Neural Inf Process Syst* [Internet]. 2020. Available from: <http://www.scopus.com/inward/record.url?scp=85107944105&partnerID=8YFLogxK>. Accessed 31 December 2020
11. Lundberg SM, Lee S-I, Guyon I, Luxburg U. Advances in neural information processing systems. In: *Advances in Neural Information Processing Systems 30: 31st Annual Conference on Neural Information Processing Systems (NIPS 2017)*: Long Beach, California, USA; 2017.
12. Desai RJ, Lin KJ, Patorno E, Barberio J, Lee M, Levin R, et al. Development and preliminary validation of a Medicare claims-based model to predict left ventricular

- ejection fraction class in patients with heart failure. *Circ Cardiovasc Qual Outcomes* 2018;11:e004700. doi:10.1161/CIRCOUTCOMES.118.004700
13. Mahesri M, Chin K, Kumar A, Barve A, Studer R, Lahoz R, et al. External validation of a claims-based model to predict left ventricular ejection fraction class in patients with heart failure. *PLoS ONE* 2021;16:e0252903. doi:10.1371/journal.pone.0252903
 14. Patel YR, Robbins JM, Kurgansky KE, Imran T, Orkaby AR, McLean RR, et al. Development and validation of a heart failure with preserved ejection fraction cohort using electronic medical records. *BMC Cardiovasc Disord* 2018;18:128.
 15. Uijl A, Lund LH, Vaartjes I, Brugts JJ, Linssen GC, Asselbergs FW, et al. A registry-based algorithm to predict ejection fraction in patients with heart failure. *ESC Heart Fail* 2020;7:2388-2397. doi:10.1002/ehf2.12779
 16. Sundaram V, Zakeri R, Witte KK, Quint JK. Development of algorithms for determining heart failure with reduced and preserved ejection fraction using nationwide electronic healthcare records in the UK. *Open Heart* 2022;9:e002142. doi:10.1136/openhrt-2022-002142
 17. Cohen SS, Roger VL, Weston SA, Jiang R, Movva N, Yusuf AA, et al. Evaluation of claims-based computable phenotypes to identify heart failure patients with preserved ejection fraction. *Pharmacol Res Perspect* 2020;8:e00676. doi:10.1002/prp2.676
 18. Savarese G, Stolfo D, Sinagra G, Lund LH. Heart failure with mid-range or mildly reduced ejection fraction. *Nat Rev Cardiol* 2022;19:100-116. doi:10.1038/s41569-021-00605-5
 19. Trayanova NA, Popescu DM, Shade JK. Machine learning in arrhythmia and electrophysiology. *Circ Res* 2021;128:544-566. doi:10.1161/CIRCRESAHA.120.317872
 20. Benedetto U, Dimagli A, Sinha S, Cocomello L, Gibbison B, Caputo M, et al. Machine learning improves mortality risk prediction after cardiac surgery: Systematic review and meta-analysis. *J Thorac Cardiovasc Surg* 2022;163:2075-2087.e9. doi:10.1016/j.jtcvs.2020.07.105
 21. Ong M-S, Klann JG, Lin KJ, Maron BA, Murphy SN, Natter MD, et al. Claims-based algorithms for identifying patients with pulmonary hypertension: A comparison of decision rules and machine-learning approaches. *J Am Heart Assoc* 2020;9:e016648. doi:10.1161/JAHA.120.016648
 22. Kapłon-Cieślicka A, Benson L, Chioncel O, Crespo-Leiro MG, Coats AJS, Anker SD, et al. A comprehensive characterization of acute heart failure with preserved versus mildly reduced versus reduced ejection fraction-Insights from the ESC-HFA EORP Heart Failure Long-Term Registry. *Eur J Heart Fail* 2022;24:335-350. doi:10.1002/ejhf.2408