scientific reports

Check for updates

OPEN

Development and indirect validation of a model predicting frailty in the French healthcare claims database

Hana Lahbib[1], Laurence Mandereau-Bruno[1], Sarah Goria[1], Vérène Wagner[1], Marion J Torres[1], Catherine Féart[2], Catherine Helmer[2], Karine Pérès[2] & Laure Carcaillon-Bentata[1]✉

This study aimed to build a predictive model to identify frailty in the French national health data system (SNDS) so as to create a new tool to monitor and anticipate the disability burden associated with population ageing. We developed the model using the 2012 wave of the French Health, Healthcare, and Insurance Survey (ESPS) linked to the SNDS (n = 2,829). This survey used Fried's frailty phenotype as the gold standard. We compared two statistical approaches – logistic regressions (stepwise and LASSO selection) and random forest – to predict frailty probability based on different SNDS healthcare claims. We indirectly validated the model by comparing (1) the predicted frailty prevalence in the overall French population in the SNDS with the expected prevalence and (2) the predictive ability of the model for 6-year mortality with that of Fried's frailty phenotype. Logistic regression with LASSO selection was retained as the best method to predict frailty. After stratification for age, we obtained three models for individuals aged 55–64, 65–74, and ≥ 75 years (AUC = 0.61, 0.76, and 0.80 respectively). Applying these models to the SNDS, frailty prevalence was comparable to expected prevalence in all sex and age groups: overall prevalence = 12.9% (95%CI: 12.9–12.9) in the SNDS versus 12.0% (95%CI: 10.8–13.2) in the ESPS. Predicted frailty probabilities in the SNDS showed a similar strong association with 6-year mortality ($HR_{frail\_probability}$=2.6, 95%CI: 1.9–3.5) compared with Fried's phenotype ($HR_{frail\_phenotype}$=2.9, 95%CI: 2.1–3.8). Our predictive models are thus useful for estimating frailty probability in the SNDS.

Keywords French national health data system, Algorithm, Predictive model, Frailty

According to the World Health Organization, one in five people will be aged over 60 years by 2050, totaling 2.1 billion people worldwide[1]. Quality of life is highly dependent on functional ability, and age-related disability is a burden for the affected individuals, their families, and society as a whole.

Frailty is a highly prevalent geriatric condition[2], defined as a "clinically recognizable state of increased vulnerability resulting from age-associated decline in reserve and function across multiple physiologic systems such that the ability to cope with every day or acute stressors is compromised"[3]. In the literature, frailty is assessed using multiple scales, mainly based on two widespread approaches: Fried's phenotype[4] and deficit-accumulation frailty indices first introduced by Rockwood et al.[5]. One of the most common approaches is the phenotype of Fried et al.[4], which defines frailty as a clinical syndrome with at least three of the following five criteria: unintentional weight loss, self-reported exhaustion, weakness, slow walking speed, and low physical activity. Frailty is a known risk factor for functional decline and adverse outcomes such as falls, hospitalization, and death[3–5]. Recent research further suggests a possible continuum between frailty and severe functional disability[6,7]. To anticipate the future burden of age-related disability, improve knowledge about its determinants, and identify preventive actions to reduce its risk, frailty has become the subject of growing interest due to its central place in the ageing process. Unlike functional dependency, frailty is a reversible state that constitutes a useful target for preventive programs to diminish the risk of dependency[8].

[1]Santé publique France, French National Public Health Agency, Saint-Maurice 94410, France. [2]Bordeaux Population Health Research Center, Bordeaux University, INSERM, UMR U1219, Bordeaux 33000, France. ✉email: laure.carcaillon-bentata@u-bordeaux.fr

Over the past decade, large claims databases and electronic medical records have been increasingly used to describe and assess population health (e.g., chronic disease surveillance), as they provide historic national data for the general population. Several recent international studies have developed frailty instruments (including indices or scores) using claims databases or electronic medical records[9–19]. These instruments have generally proven to be valid prognostic tools for predicting adverse outcomes, including mortality, hospitalization, and disability. Some of them have further been validated using a clinical assessment of frailty and found to have an acceptable-to-good diagnostic accuracy, thus supporting their use in primary care or inpatient databases[13,17,20–23]. Some frailty instruments based on medical and administrative data are now used by government departments to make healthcare decisions. For example, the electronic frailty index is currently used as part of the frailty management guidelines of the British Geriatrics Society and National Health Service (NHS)[19,24].

Some algorithms may sometimes be adapted from one database to another[25]. However, unless there is a high similarity between available variables in databases, it is most of the time challenging to apply or adapt a country-specific algorithm to another country. Our study aimed to develop a new predictive model to identify frailty in the French claims-based database (*Système National de Données de Santé*, SNDS). The SNDS allows the monitoring of conditions such as diabetes[26,27], coronary syndrome[28], and Parkinson's disease[29]. However, contrary to specific diseases, frailty is difficult to identify, since the SNDS does not include the medical evaluations necessary to determine frailty (e.g., walking speed, physical activity level, laboratory test results). A recent study from England[17] developed the Hospital Frailty Risk Score based on the diagnosis of chronic diseases during hospital stays. This score was applied to the SNDS and proved its ability to predict 30-day inpatient mortality[30].

In our study, we aim to go beyond the Hospital Frailty Risk Score by developing and validating a model to estimate the probability of frailty for individuals aged ≥ 55 years in the general population. For this purpose, we will use data from a national population-based health survey linked to the SNDS in order to develop a predictive model of frailty, which will be assessed using Fried's phenotype reference definition.

Methods

Data sources

We used data from the French Health, Healthcare, and Insurance Survey (*Enquête Santé Protection Sociale*, ESPS)[31], a national survey based on interviews conducted every 2 years since 1988 by the French Institute for Research and Information on Health Economics. This survey is sampled from the national health insurance database (SNDS). Randomly selected individuals as well as their households are then interviewed. Data are weighted to be representative of the general population (excluding people living in institutions like nursing homes). Data collection combined a general questionnaire conducted by telephone (for sociodemographic and social protection data) and a self-administered questionnaire completed by email (for the health, lifestyle, and functional evaluations). To develop the predictive model, this study used the 2012 wave of the ESPS, which is the only ESPS wave with complete data on Fried's frailty criteria. Among the sample taken from the SNDS, individuals and their household members were asked to participate; they could refuse to participate or they may not have been reachable at the time of the survey. In 2012, the overall participation rate was 48%, resulting in 23,047 participants aged ≥ 15 years. The linkage of the ESPS with vital statistics allowed for the assessment of mortality up to December 31, 2017. The ESPS received institutional review board approval from the National Council for Statistical Information (*Conseil National de l'Information Statistique*, CNIS) and the national data protection agency (*Commission Nationale de l'Informatique et des Libertés*, CNIL) and followed the declaration of Helsinki principles. All participants and/or their parent and/or legal guardian provided informed consent in compliance with the General Data Protection Regulation.

Individual data collected during the survey were linked to the SNDS data (i.e., healthcare consumption data). This linkage was performed for selected beneficiaries drawn at random (the personal identifier used for linkage purposes was available given that these individuals were initially selected from the SNDS) as well as some of their household members (when a personal identifier was likewise available). Overall, just over half of the sample was linked, representing 12,021 participants in the ESPS 2012 survey. SNDS data cover almost the entire French population and include individual information relating to the sociodemographic (e.g., sex, year of birth, health insurance affiliation scheme, area of residence) and medical characteristics of beneficiaries (i.e., individuals receiving the care). The SNDS also includes all hospital and non-hospital care reimbursements coded according to various systems (e.g., diagnosis, products and services, medical acts)[32]. The International Classification of Diseases, 10th version French adaptation (ICD-10)[33] diagnosis codes are available for acute care (medical, surgical, and obstetrical), aftercare and rehabilitation, psychiatry, and home hospitalization (along with the date, duration of hospitalization, and medical procedures performed). They are also available for outpatient reimbursed healthcare expenditures such as general practitioner and specialist visits, medical procedures, nursing procedures, physiotherapy, medical imaging, laboratory tests, drugs, medical devices, and medical transportation. The SNDS also includes patients with long-term diseases (e.g., chronic conditions) for whom the associated medical expenses are fully reimbursed. Both the ESPS and SNDS have been described in detail elsewhere[31,32].

Access by Santé Publique France to ESPS data linked to the SNDS received ethical approval from the relevant institutional review board (French national data protection agency, DE-2017-361). Database access took place in a secure environment provided by the Secure Data Access Centre (Ref. 10.34724/CASD).

Fried's frailty phenotype as the gold standard

The ESPS 2012 assessed frailty using Fried's phenotype adapted to the declarative data[34,35]. All five frailty criteria (involuntary weight loss, self-reported exhaustion, low physical activity, muscular weakness, and slow walking speed) were collected from the self-administered questionnaire of the ESPS 2012 study (Supplementary Table S1).

Individuals were considered frail if they met at least three criteria and otherwise robust. In addition, individuals not initially identified as frail, who nevertheless expressed significant difficulty in at least one of Katz's activities of daily living (ADL) in the self-administered questionnaire[36] such as self-feeding, dressing, and toileting, were redefined as frail ($n = 22$) based on the previously published hierarchy between frailty and ADL disability[6].

Candidate variable selection in the SNDS

To identify the relevant SNDS variables, we performed a literature review using the following search equation in Medline: "frailty"[MeSH Major Topic] AND ("claims data"[All Fields] OR "health administrative database"[All Fields] OR "electronic health records"[All Fields]). We also scrutinized references cited in the retrieved articles to identify other relevant publications. Supplementary Table S2 presents the list of references explored in order to select the candidate variables. A multidisciplinary group (including epidemiologists specialized in ageing, nutrition, cognition, frailty, and disability) evaluated each variable selected as a potential parameter of the predictive model to determine its relevance to the biological concept of frailty, its consequences, and its availability or operationalization in the SNDS (Supplementary Table S3). We selected candidate variables for diseases (e.g., diabetes, depression, dementia, cardiovascular diseases), geriatric syndromes (e.g., undernutrition, falls/traumas, bladder dysfunction, vertigo), healthcare utilization (e.g., hospitalizations, general or specialist consultations), and specific healthcare consumption (e.g., walker, wheelchair, therapeutic oxygen, hearing aids). To ensure the robustness of the statistical analyses, in addition to age and sex, the analyses only included variables ($n = 49$) observed in at least five subjects in both the robust and frail categories. Further details regarding these variables are provided in Supplementary Tables S4 and S5.

Statistical analysis

For the descriptive analysis, categorical variables are presented in numbers and percentages. We performed Chi-square and exact Fisher tests.

Development and internal validation of a predictive model for Fried's frailty phenotype using ESPS 2012 data

To develop a parsimonious and reproducible predictive model, we considered three models using two statistical approaches: logistic regression with two types of variable selection methods, namely stepwise and least absolute shrinkage and selection operator (LASSO) selection[37], and the machine learning random forest[38] using recursive feature elimination for variable selection[39].

The models included all selected candidate variables in addition to age and sex (Supplementary Table S4). For the logistic models, we tested the interactions of the variables with age and sex, because healthcare consumption as well as disease symptoms and impact differ depending on the age and sex of individuals[40–43]. We performed internal validation using a 100-fold cross-validation technique on a dataset randomly split into training and test sets (80/20).

We comprehensively assessed the models' performance using multiple evaluation metrics, including area under the curve (AUC), sensitivity (Se), specificity (Sp), and Nagelkerke's R^2[44], to provide a complete picture of their effectiveness. More specifically, we compared the performances of the models in terms of Se, Sp, and positive and negative predictive values (PPV and NPV, respectively), estimated using the threshold determined by the Youden index[45] (i.e., cut-off maximizing the number of correctly classified individuals). We chose to use the Youden index, because our objective was to accurately identify the different classes. This required balancing between sensitivity and specificity, which is effectively captured by the Youden index. We chose the best model based on the balance between performance (assessed using the various metrics detailed above) and parsimony.

Predictive validation of our developed model using SNDS 2013 data

Our development sample (ESPS 2012) did not include subjects living in nursing homes (or long-term care). In the SNDS, information about nursing homes is only available from 2013 onwards[32], so we applied our models to 2013 instead of 2012. We selected all individuals aged ≥ 55 years who were not living in a nursing home (or long-term care) in 2013. We compared estimated frailty prevalence with that of the ESPS 2012, which is the study of reference to estimate frailty prevalence in France.

Longitudinal assessment of our predictive model for 2-, 4-, and 6-year mortality

Using the ESPS 2012 sample, we used Cox proportional hazard models to compare the ability of our predictive model to assess 2-, 4-, and 6-year mortality with that of Fried's phenotype. In addition to crude estimates, we successively applied two models to evaluate the influence of sex and age (model 1) and co-morbidity (model 2) on the association between frailty and mortality. Given that frailty and multimorbidity often coexist[46], we adjusted for the locally adapted Charlson co-morbidity index[47] in order to evaluate the added value of our frailty predictive model with regard to co-morbidity. We then divided the estimated frailty probabilities into percentiles for the 55–64 age group (which includes individuals from 55 to 64 years) and into deciles for the two other age groups (65–74 age group, which includes individuals from 65 to 74 years, and the ≥ 75 group, which includes subjects aged 75 years and older). We further categorized them into two groups to mimic Fried's expected phenotype prevalence based on a frailty prevalence of around 5% in the 55–64 age group, 10% in the 65–74 age group, and 30% in the ≥ 75 age group. We categorized frailty probabilities into two groups using the 95th percentile as a cut-off in the younger group. Deciles in the 65–74 age group were grouped into the 1st to 9th and 10th deciles. Among the ≥ 75 age group, deciles were grouped into the 1st to 6th and 7th to 10th deciles. We tested the proportional hazard assumption using graphical methods while assessing interactions between time to death and frailty measures. We estimated AUC to compare the predictive values of the models.

Analyses were performed using R, *caret*[48], *glmnet*[49], and *rocr* packages, and SAS 9.4.

Results

Study sample characteristics

The study development population (ESPS 2012) included 3,659 participants aged ≥ 55 years (Fig. 1). We excluded 830 individuals due to missing frailty data or ADL measures, leading to a final dataset of 2,829 individuals. Excluded subjects were younger than included subjects (64.5 years vs. 68.1 years, respectively) but had the same sex ratio (51.6% women vs. 52.8%).

Frailty prevalence was 12.0% (95%CI: 10.8–13.2) ($n = 339$). It increased with age (60.5% of frail individuals were aged ≥ 75 years) and was higher in women (63.5% of frail individuals were women) (Table 1).

Development of predictive models

We retained the LASSO logistic regression approach as the best method to develop our model. Although the stepwise logistic regression showed better parsimony (13 vs. 31 variables), its performances were slightly inferior to the LASSO model regarding the R^2 Nagelkerke. Compared with the random forest, the LASSO model showed similar AUC performances, better sensitivity, better R^2, and better parsimony (Supplementary Table S6).

The full LASSO model revealed multiple interactions between age and covariates. We therefore stratified our analysis and ran three models according to the three age groups 55–64, 65–74, and ≥ 75 years. To achieve sufficient statistical power, we performed these analyses on the whole sample without using random training test sets. Table 2 shows the performances of the three age-group models. For the youngest group, the model included 13 variables and presented moderate performances (AUC = 0.61, 95%CI: 0.54–0.67; Se = 50.8%, PPV = 16.0%). For the second group (65–74 years), the model also included 13 variables and presented good performances but poor PPV (AUC = 0.79, 95%CI: 0.72–0.85; Se = 70.4%; PPV = 22.1%). The model for the oldest age group (≥ 75 years) included 22 variables and had a good level of discrimination (AUC = 0.80, 95%CI: 0.76–0.84; Se = 62.0%; PPV = 59.9%). Table 3 provides details on the variables selected in the three predictive models based on the

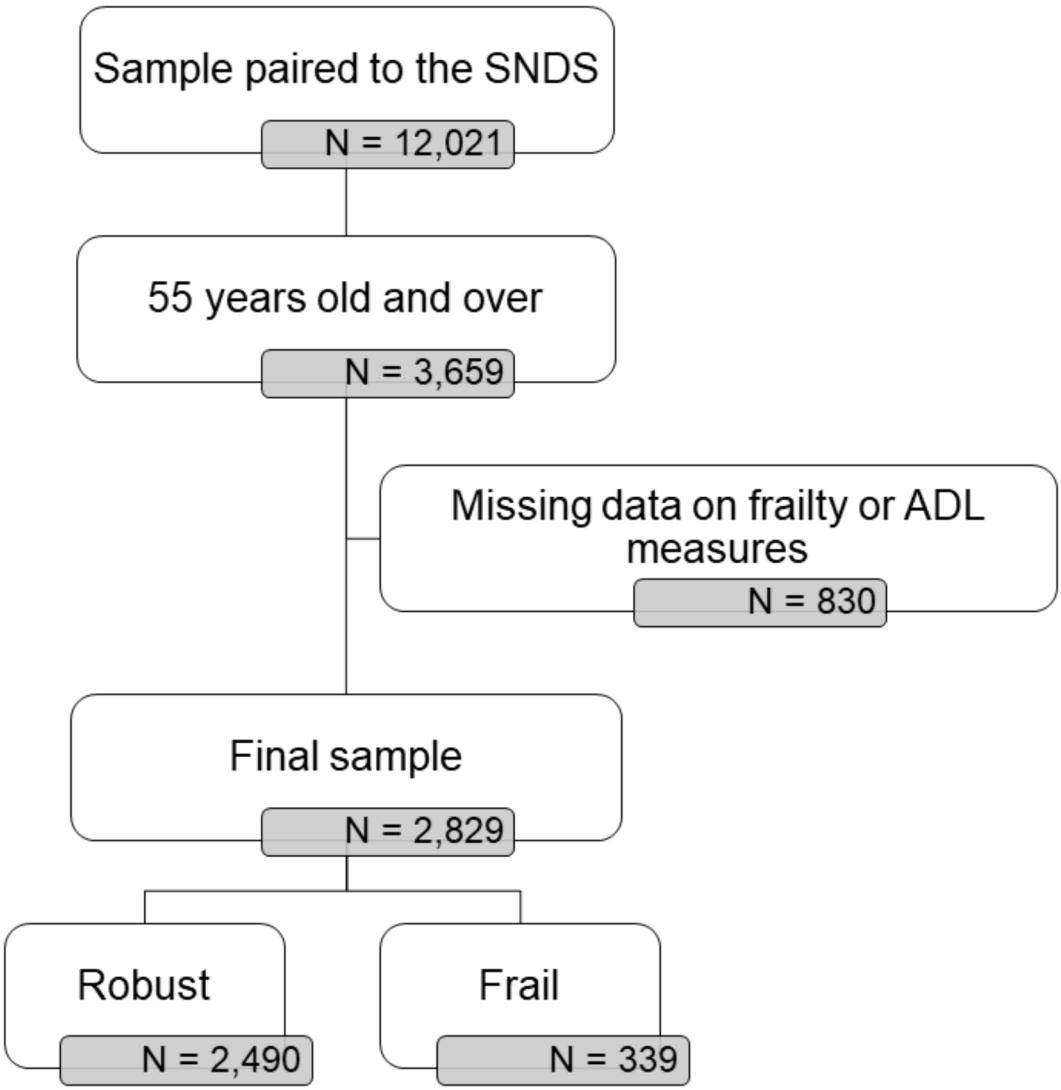

Fig. 1. Flowchart of the study sample (ESPS 2012).

	Robust	Frail		Total
	N = 2,490	*N* = 339	p	*N* = 2,829
Age	n (%)	n (%)		n (%)
55–64	1,148 (46.1)	63 (18.6)		1,211 (42.8)
65–69	504 (20.2)	37 (10.9)	< 0.01	541 (19.1)
70–74	299 (12.0)	34 (10.0)		333 (11.8)
75+	539 (21.7)	205 (60.5)		744 (26.3)
Sex				
Men	1,210 (48.6)	125 (36.9)	< 0.01	1,335 (47.2)
Women	1,280 (51.4)	214 (63.1)		1,494 (52.8)
Diploma level (MD *n* = 95)				
No diploma	358 (14.7)	106 (35.8)		464 (17.0)
< High school diploma	1,340 (55.0)	156 (52.7)		1,496 (54.7)
≥ High school diploma, 2-year university degree	416 (17.1)	21 (7.1)	< 0.01	437 (16.0)
≥ 3-year university degree	277 (11.4)	10 (3.4)		287 (10.5)
Other/unknown	47 (1.9)	3 (1.0)		50 (1.8)
Socioprofessional category (MD *n* = 96)				
Farmer	142 (5.8)	32 (10.9)		174 (6.3)
Craftsperson, salesperson, company head	185 (7.6)	18 (6.1)		203 (7.4)
Executive, intellectual profession	416 (17.1)	16 (5.4)	< 0.05	432 (15.8)
Middle manager	420 (17.2)	19 (6.4)		439 (16.1)
Employee	630 (25.8)	81 (27.5)		711 (26.0)
Worker	576 (23.6)	103 (34.9)		679 (24.8)
Pensioner	11 (0.5)	5 (1.7)		16 (0.6)
No professional activity	58 (2.4)	21 (7.1)		79 (2.9)
Frailty criteria				
Involuntary weight loss (MD *n* = 33)	150 (6.1)	104 (31.4)		254 (9.1)
Exhaustion (MD *n* = 189)	966 (41.4)	271 (87.7)		1,237 (46.9)
Low physical activity (MD *n* = 100)	267 (11.1)	209 (64.3)		476 (17.4)
Muscular weakness (MD *n* = 20)	206 (8.3)	308 (90.9)		514 (18.3)
Slow walking speed (MD *n* = 16)	33 (1.3)	254 (75.2)		287 (10.2)
ADL criteria: difficulty with…				
Self-feeding (MD *n* = 14)	0 (0)	25 (7.4)		25 (0.9)
Transferring (MD *n* = 18)	0 (0)	54 (16.2)		54 (1.9)
Dressing (MD *n* = 15)	0 (0)	56 (16.9)		56 (2)
Toileting (MD *n* = 16)	0 (0)	35 (10.5)		35 (1.2)
Bathing (MD *n* = 11)	0 (0)	86 (25.7)		86 (3.1)

Table 1. Characteristics of the study sample according to Fried's frailty phenotype (ESPS 2012, *n* = 2,829). *ADL* activities of daily living, *MD* missing data

	n	Se	Sp	PPV	NPV	AUC	95%CI	*N* True positive	*N* False positive	*N* True negative	*N* False negative
Age-group models											
55–64 years	1,211	50.8%	85.4%	16.0%	96.9%	0.61	0.54–0.67	32	168	980	31
65–74 years	874	70.4%	78.1%	22.1%	96.8%	0.76	0.72–0.85	50	176	627	21
≥ 75 years	744	62.0%	84.2%	59.9%	85.3%	0.80	0.76–0.84	127	85	454	78

Table 2. Performances of the models according to the three age groups (ESPS 2012, *n* = 2,829). *Se* sensitivity, *Sp* specificity, *PPV* positive predictive value, *NPV* negative predictive value, *AUC* area under the operative curve, *CI* confidence interval.

age group and their coefficients. The regression coefficients and their weight varied across the three age-group predictive models. In the youngest groups (55–64 and 65–74 years), the variables with higher weight included respiratory diseases, medical devices for respiratory diseases and for sensory and physical impairments, depression, and cardiovascular diseases. In the oldest group (≥ 75 years), the same variables remain important,

Model5564 for individuals aged between 55 and 64 years	β[a]	SD[c]	Model6574 for individuals aged between 65 and 74 years	β[a]	SD[c]	Model75 for individuals aged 75 years and older	β[a]	SD[c]
Intercept	-3.4438		Intercept	-7.0713		Intercept	-10.4395	
Ophthalmology consultation	-0.3311	0.139	Age	0.0529	0.001	Age	0.1035	0.003
Medical devices (respiratory diseases)	2.1793	0.553	ENT[b] consultation	-0.0581	0.008	Sex	0.1303	0.019
Medical devices (sensory and physical impairments)	0.3712	0.115	Medical devices (sensory and physical impairments)	0.2904	0.115	Cardiology consultation	-0.2367	0.032
Depression	0.6013	0.180	Diabetes	0.9682	0.029	Ophthalmology consultation	-0.0103	0.019
Respiratory diseases	0.7387	0.237	Depression	0.1789	0.184	Rheumatology consultation	-0.0597	0.036
Cancer	0.0569	0.055	General practitioner consultation (number)	0.0490	0.050	Medical devices (respiratory diseases)	0.2458	0.087
Cardiac diseases	0.0557	0.047	Medical transport (number)	0.0817	0.010	Medical devices (sensory and physical impairments)	0.1404	0.030
General practitioner consultation (number)	0.0203	0.006	Nursing procedure (number)	0.0015	0.0001	Diabetes	0.3287	0.041
Medical transport (number)	0.0399	0.011	Physiotherapy act (number)	0.0124	0.002	Dementia	0.7468	0.111
Physiotherapy act (number)	0.0084	0.003	Sex*Endocrinology consultation	0.3696	0.033	Depression	0.7026	0.051
Sex*Diabetes	0.5188	0.180	Sex*Diabetes	0.7084	0.161	Lipid-lowering drug	-0.0032	0.020
Sex*Coronary diseases	-1.2816	0.764	Sex*Coronary diseases	0.5439	0.829	Cardiac diseases	0.7955	0.072
Sex*Cardiac diseases	0.6241	0.232	Sex*Hospitalization acute care field (number)	0.0601	0.005	Hospitalization acute care field (number)	0.1153	0.017
Sex*Medical transport (number)	0.2231	0.075	Sex* Medical specialist consultation (number)	0.0003	0.002	General practitioner consultation (number)	0.0273	0.003
Sex*Nursing procedure (number)	0.0036	0.001				Medical transport (number)	0.0540	0.009
						Nursing procedure (number)	0.0008	0.0001
						Physiotherapy act (number)	0.0138	0.0007
						Sex*ENT[b] consultation	-0.9857	0.098
						Sex*Rheumatology consultation	-0.0742	0.060
						Sex*Medical devices (respiratory diseases)	1.5579	0.115
						Sex*Medical devices (sensory and physical impairments)	0.2603	0.040
						Sex*Hypertension	0.0220	0.016
						Sex*Respiratory diseases	0.6136	0.075
						Sex*Cancer	0.2366	0.060
						Sex*Coronary diseases	-0.2794	0.095
						Sex*Cardiac diseases	0.0822	0.061
						Sex*Hospitalization acute care field (number)	0.0281	0.019

Table 3. Final models* including three age-specific predictive models for individuals aged 55–64, 65–74, and ≥ 75 years (variables selected by LASSO logistic regression) (ESPS 2012, $n = 2{,}829$). [a] Regression coefficients estimated using logistic regression with LASSO selection. [b] Ear, nose, and throat. [c] Standard deviation. *To apply our model, it is necessary for each individual to multiply the frailty model coefficients by all the variables included in the model, which gives the log odds of frailty (Y). This can in turn be converted into a predicted probability (p) of being frail using the following formula: $= \frac{\exp(Y)}{1+\exp(Y)}$ Use "model5564" for individuals aged between 55 and 64 years, "model6574" for those between 65 and 74 years, or "model75" for those older than 75. "Sex*XXX" is the interaction between sex and the given variable.

but other variables enter the model such as dementia, Ear Nose Throat consultation, and cancer. In all three groups, we identified interactions with sex, with some showing high beta regression coefficients.

Application of predictive models to the SNDS

In the SNDS, 17,477,879 individuals were aged ≥ 55 years in 2013. After excluding subjects living in nursing homes (or long-term care), the final sample included 17,050,508 individuals. Supplementary Table S7 provides descriptive details of the SNDS sample.

Applying our age-stratified models to the SNDS, we estimated the overall frailty prevalence as well as by age and sex subgroups (Fig. 2). Predicted frailty prevalences were consistent with the expected prevalences based on Fried's phenotype in the ESPS (from around 5% in the youngest group to 30% in the oldest). Prevalence was slightly overestimated among older men (20.4% in the ESPS 2012 vs. 23.0% in the SNDS 2013).

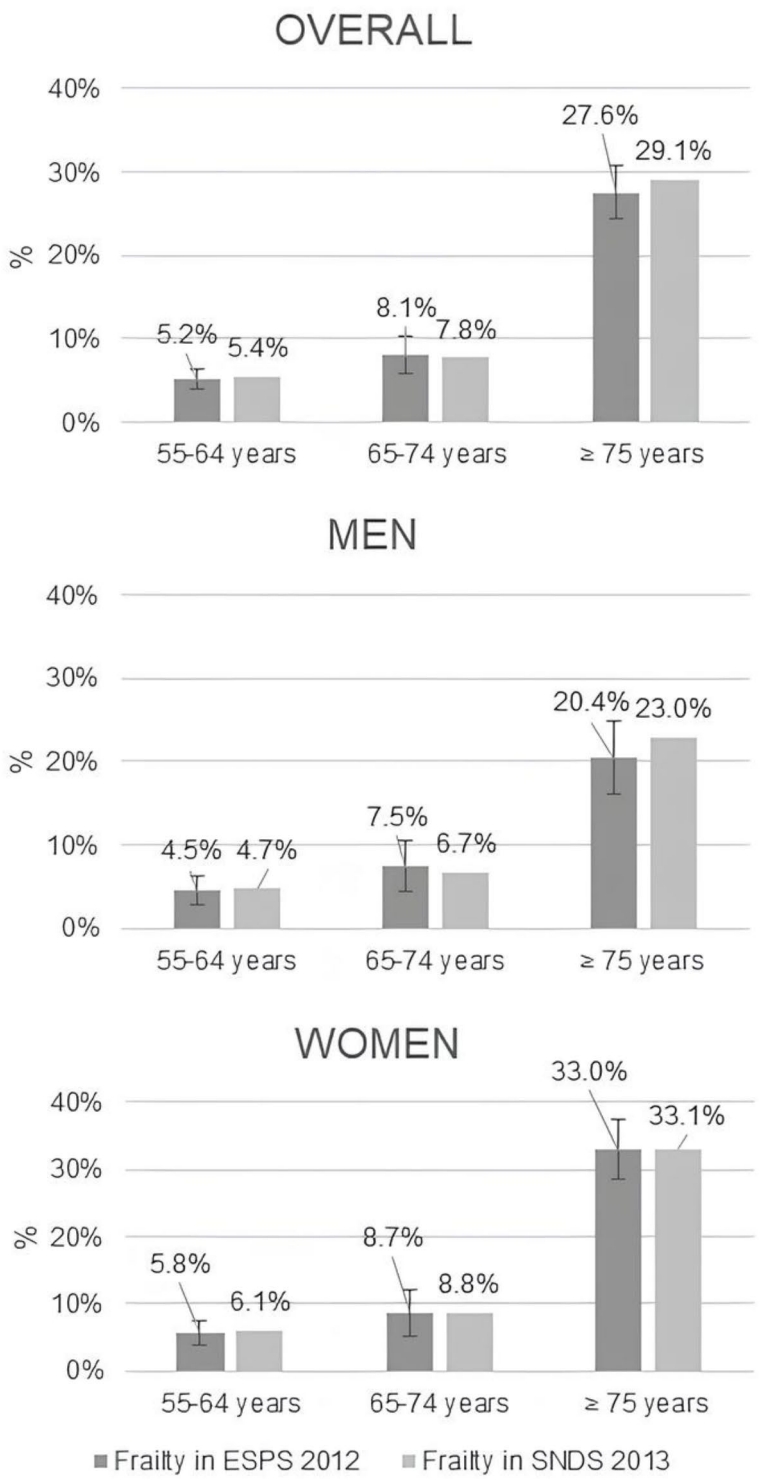

Fig. 2. Prevalence of frailty estimated in the SNDS sample according to age and sex (SNDS 2013, n = 17,050,508).

Longitudinal assessment of 2-, 4-, and 6-year mortality prediction

We compared the ability of our models to assess 6-year mortality (2012–2017) with that of Fried's phenotype. Compared with Fried's phenotype, frailty estimated by the age-stratified predictive models showed similar associations with 6-year mortality in the crude and adjusted analyses. Both approaches showed similar discrimination levels ($AUC_{predictive\_model}$=0.78 vs. $AUC_{Frailty\_phenotype}$=0.80) (Table 4). We drew the same conclusions with the prediction of 2- and 4- year mortality (Supplementary Table S8).

	6-year mortality		Crude model				Model 1[a]				Model 2[b]			
	n	%	HR	95%CI	*p*-value	AUC	HR	95%CI	*p*-value	AUC	HR	95%CI	*p*-value	AUC
Fried's phenotype					< 0.0001	0.68			< 0.0001	0.77			< 0.0001	0.78
0: Robust	125	57.6%	1.0				1.0				1.0			
1: Frail	92	42.4%	6.6	5.1–8.7			3.5	2.6–4.8			2.9	2.1–3.8		
Frailty probabilities from the model					< 0.0001	0.67			< 0.0001	0.78			< 0.0001	0.80
0: Robust	127	58.4%	1.0				1.0				1.0			
1: Frail	90	41.5%	6.6	5.1–8.7			3.8	2.7–5.3			2.6	1.9–3.5		

Table 4. Hazard risk of 6-year mortality associated with the frailty phenotype and frailty probability estimated by the predictive models (ESPS 2012, $n = 2{,}829$). [a] Model 1 is adjusted for age and sex. [b] Model 2 is adjusted for age, sex, and Charlson Index. *AUC* area under the curve, *HR* hazard ratio (estimated using Cox regression models)

Frailty identification at the individual level

The low specificity and PPV of our models based on the Youden index prevented us from assigning a frailty state (yes/no) at the individual level. To provide a cut-off that could effectively identify frail individuals with a high level of confidence, we conducted supplementary analysis in the oldest age group (≥ 75 years), where the model showed good performances. In this population, we modulated the threshold of the frailty probability at the individual level and found that a probability of ≥ 0.4 (instead of > 0.32 with the Youden index) increased the PPV from 59.9 to 74.4% and specificity from 84.2 to 94.2% (Supplementary Table S9).

Discussion

We developed three age-stratified predictive models based on chronic diseases and healthcare consumption records to estimate frailty prevalence in individuals aged ≥ 55 years based on the French national health data system (SNDS). The models, which included between 13 and 22 variables depending on the age group, had moderate performances in the youngest group (55–64 years) and good performances in the two older age groups over 65 years. The indirect validation results (i.e., validation using secondary outcomes) were also good, highlighting the ability of the predictive models to estimate frailty prevalence in the SNDS and to predict 6-year mortality. Our models produced reliable estimates of frailty prevalence in all sex and age groups. However, the low specificity of the models at the threshold determined by the Youden index prevented us from using this threshold to assign a frailty status (yes/no) at the individual level, except for individuals over 75 years, where a threshold of ≥ 0.4 was found to be reliable for this purpose.

Our models estimated population-based frailty prevalence in individuals aged ≥ 55 years. Prevalence was slightly overestimated in men aged ≥ 75 years, probably due to the variables included in the predictive models, which mainly related to healthcare consumption and chronic diseases. Although our models considered several interactions between sex and covariates, sex differences in healthcare consumption behaviors may partly explain the better fit of our models in women, who seek care more frequently than men[42]. The lower number of cases of frailty in men than in women may also explain the poorer fit of the model in men. Unfortunately, we could not stratify our analyses by sex due to the lack of statistical power. While the prevalence of frailty in men aged ≥ 75 years may be slightly overestimated, it nevertheless remains within the estimated 95% confidence interval of the survey-based prevalence. This overestimation will not, however, influence health policy decisions drawn based on our frailty models.

Although our models produced reliable estimates of prevalence, their low specificity at the threshold determined by the Youden index argues against using this threshold to establish a frailty status at the individual level. We thus propose using a different threshold for individuals aged ≥ 75; at this threshold, 45% of frail individuals are identified with a PPV of nearly 75%. Even though this method may lead to a selection bias toward more severe cases, we recommend using this threshold to build a dichotomous variable, as it significantly reduces potential biases due to misclassification if the threshold determined by the Youden index is applied[50].

We chose logistic regression with LASSO selection, which reduced the number of variables with small effects and thus provided a limited number of predictors in the final model. Random forest showed similar performances, although we excluded this method due to parsimony and feasibility. Indeed, the final random forest model included all 49 variables, while age-stratified LASSO models required a maximum of only 22 variables. Other more sophisticated approaches such as neural networks are available to build predictive models. We chose not to use these methods, because their results are often difficult to interpret[51], even though they have proven to be promising for prediction purposes, including in the field of frailty[52,53]. Using a subsample from the Canadian Primary Care Sentinel Surveillance Network ($n = 5{,}466$) of individuals aged from 69 to 80 years, Aponte et al. ran seven commonly used binary supervised machine learning methods to predict the clinical frailty scale using primary care data from 2015 to 2019. They found that the best model had moderate performances (AUC = 0.83; Se = 40.5%; Sp = 93%). Another recent study among hospitalized patients aged ≥ 65[52] used the neural networks method to develop a predictive model of frailty within 1 year of admission. Good performances were found for most of the tested models (best model: AUC = 0.89; Se = 0.77; Sp = 0.85).

Our results are in line with studies using similar model-based predictive approaches[12,15,18,54,55]. Segal et al. developed[18] and validated[55] a claims-based frailty index (CFI) using data from the Cardiovascular Health

Study and the National Health and Aging trend study, which are both linked to American Medicare data. These studies used frailty phenotype as the gold standard. Applied in a sample of subjects aged ≥ 65 years, the final predictive model relying on LASSO logistic regression showed moderate performances (AUC = 0.75; PPV = 35%; NPV = 92%). Another study from Kim et al. focused on the development of a CFI using American Medicare data while relying on the deficit-accumulation frailty approach. Using two different waves of the Medicare Current Beneficiary Survey (MCBS) linked to claims data from 2006 to 2011, the authors found that the new CFI had a good correlation with the gold standard survey frailty index from 2006 and that compared with Charlson's comorbidity index, it had a similar predictive capacity for mortality but a better capacity for disability outcomes in 2011. Another validation study conducted with MCBS 2008 found that this CFI was predictive of low gait speed, weak grip strength, as well as ADL and instrumental ADL dependency, mortality, hospitalization, and prolonged stay in skilled nursing facilities[12]. Faurot et al. developed and validated a frailty proxy in the American electronic health records using ADL dependency instead of frailty as the reference[14]. Their approach identified the most severe stage of frailty[6]. Applying their algorithm to an independent population-based cohort, the Atherosclerosis Risk in Community study[10], they found moderate discrimination performances with Fried's frailty phenotype (AUC = 0.71) and good predictive abilities with adverse outcomes. Overall, external validation studies of indices initially developed without a model-based approach have provided inconsistent performances. Using data from a primary health center in Barcelona, Orfila et al.[13] reported the good predictive ability of their electronic frailty index (eFRAGICAP) with regard to two standardized clinical measures: the clinical frailty scale (AUC = 0.82) and the Risk Instrument for Screening in the Community (AUC = 0.85). However, their measure had low sensitivity and PPV for the clinical frailty scale (34% and 43%, respectively) but a better score for the Risk Instrument for Screening in the Community (55% and 76%, respectively).

Compared with these previously described frailty models, our approach has several advantages. We used a population-based health survey linked to the SNDS to develop a predictive model of frailty probability using a reference measure (i.e., Fried's phenotype). In addition, our model included diverse types of data from the SNDS such as outpatient care and geriatric syndromes. The use of multiple data sources, not only disease-related diagnosis codes, allowed us to take into account the complexity of frailty and overcome the difficulty of identifying it in the absence of clinical examinations. Most importantly, using inpatient and outpatient claims data to estimate frailty proved to be just as challenging as approximating the underlying characteristics of biological ageing. In addition, although frailty and multimorbidity often coexist, they are two different entities and two separate aspects of the disablement process[46,56]. In our study, by adjusting for an adapted Charlson index[47], we found that frailty remained significantly associated with mortality. Using multiple data sources from the SNDS such as inpatient and outpatient healthcare claims data, we succeeded in providing age-specific predictive models of the probability of frailty, which contributes additional information to that provided by multimorbidity. Another strength of our study is that the model estimated frailty from the age of 55 years, a key age for preventive actions, whereas other studies usually begin at 65 years. Providing data on frailty prevalence from 55 years will allow us to study the early impact of ageing and chronic diseases in younger populations and the most vulnerable groups, where worrying health trends have recently been highlighted[57–59]. Stratifying the predictive models into three age groups further resulted in a model with good performances in the oldest group (≥ 75 years), which was better than other studies that usually start at 65 years. Finally, the comparison of several statistical methods allowed us to identify the best method.

Regarding the study limitations, we used an adapted version of Fried's phenotype for self-reported data. Although it is a very common way to measure frailty in epidemiological surveys, there are some key differences with the objective measures for grip strength and walking speed. Previous studies showed that the use of multiple questions to assess grip strength and walking speed, as done in our study, increased the concordance with a clinical assessment[60]. In addition, another french team compared the frailty prevalence estimated from the ESPS 2012 with that obtained from the 2011 French data of the large SHARE study (Survey of Health Ageing and Retirement in Europe[35]), which included objective measures for strength and walking speed. This comparison led to the conclusion that both approaches were very similar, particularly in subjects aged ≥ 65 years. The development of a standard operational tool remains a major topic in the field of geriatrics[61].

Regarding another study limitation, our study data were taken from the ESPS 2012 wave, which might appear quite outdated. This may have an impact on the relevance of the algorithm in more recent years given the changes to healthcare consumption patterns and heterogeneity in diagnostic coding over time[62]. We used the ESPS 2012 wave because Fried's frailty (our gold standard) was only available for this wave of the study. Indeed, we are not aware of more recent data available to conduct such a validation study at the national level in France. In addition, our dataset only included claims data for 1 year around the time of the survey interview, whereas it has been shown that using a longer look-back period could help improve the predictive performance of claims algorithms[63]. A new large nationwide study that is currently being conducted will provide data on Fried's phenotype linked to the historical SNDS data[64]. These data will represent a good opportunity to update the algorithm. Another limitation of our study is that due to the lack of statistical power, it was not possible to perform an internal validation of our predictive models, which may have affected the reliability and accuracy of our results. However, we applied our models directly to the target population (SNDS), which is a valuable approach for assessing real-world applicability when internal validation is not possible. Finally, nursing home residents, the majority of whom are ADL disabled, were excluded from the ESPS. Consequently, our predictive models should only be applied to individuals aged ≥ 55 living in the community. Although nursing home residents are not targeted by our predictive models, they are identifiable in the SNDS. Recently, Hucteau et al.[65] developed an algorithm to identify ADL dependency in individuals aged ≥ 65 years in the SNDS with very good sensitivity (75.4%) and specificity (95.9%). Combining their algorithm with our own could enable us to work on four distinct states along the continuum of dependency in the SNDS: robustness, frailty, ADL dependency, and dependency requiring nursing home care[6,7].

Conclusion

To estimate the individual probabilities of frailty using national claims-based records, we propose a new set of three predictive age-based models. We recommend applying each model to the corresponding age class. These models can be used for different purposes. In the context of population ageing, they can estimate frailty prevalence in the general population from the age of 55 years onwards. The probability of frailty can be used as a continuous variable to serve as an outcome variable or for adjustment purposes in all age groups. If frailty needs to be dichotomized, we recommend using the algorithm in subjects aged ≥ 75 years with a probability threshold of ≥ 0.4 to identify frailty, as this may represent a useful tool when the study population or the studied outcome needs a specific definition. These models can thus be used for surveillance purposes (i.e., time trends, regional prevalence) or to respond to broader research questions relating to prognostic, medico-economic, or pharmacological and epidemiological issues.

Data availability

The data that support the findings of this study are available from The National Health Data Hub www.health-data-hub.fr but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the corresponding authors upon reasonable request and with permission of The French national data protection agency (Commission Nationale de l'Informatique et des Libertés, CNIL).

Received: 16 July 2024; Accepted: 24 March 2025
Published online: 02 April 2025

References

1. Decade of healthy ageing. (2021–2030) [Internet]. [cité 8 déc 2021]. Disponible sur: https://www.who.int/fr/initiatives/decade-of-healthy-ageing
2. O'Caoimh, R. et al. Prevalence of frailty at population level in European ADVANTAGE joint action member States: a systematic review and meta-analysis. *Ann. Ist Super Sanita*. **54** (3), 226–238 (2018).
3. Clegg, A., Young, J., Iliffe, S., Rikkert, M. O. & Rockwood, K. Frailty in elderly people. *Lancet* **381** (9868), 752–762 (2013).
4. Fried, L. P. et al. Frailty in older adults: evidence for a phenotype. *J. Gerontol. Biol. Sci. Med. Sci.* **56** (3), M146–156 (2001).
5. Rockwood, K. et al. A brief clinical instrument to classify frailty in elderly people. *Lancet* **353** (9148), 205–206 (1999).
6. Zamudio-Rodríguez, A. et al. The disability process: is there a place for frailty? *Age Ageing*. **49** (5), 764–770 (2020).
7. Zamudio-Rodríguez, A. et al. Towards an approach of disability along a continuum from robustness, pre-frailty, frailty to disability. *Age Ageing*. **51** (3), afac025 (2022).
8. Morley, J. E. et al. Frailty consensus: a call to action. *J. Am. Med. Dir. Assoc. Juin*. **14** (6), 392–397 (2013).
9. Kim, D. H. Measuring frailty in health care databases for clinical care and research. *Ann. Geriatr. Med. Res. Juin*. **24** (2), 62–74 (2020).
10. Cuthbertson, C. C. et al. Controlling for frailty in pharmacoepidemiologic studies of older adults: validation of an existing medicare Claims-based algorithm. *Epidemiology* **29** (4), 556–561 (2018).
11. Wen, Y. C., Chen, L. K. & Hsiao, F. Y. Predicting mortality and hospitalization of older adults by the multimorbidity frailty index. *PLoS One*. **12** (11), e0187825 (2017).
12. Kim, D. H. et al. Validation of a claims-based frailty index against physical performance and adverse health outcomes in the health and retirement study. *J Gerontol A Biol Sci Med Sci.* 12 juill. **74** (8) 1271–6. (2019).
13. Orfila, F. et al. Validation of an electronic frailty index with electronic health records: eFRAGICAP index. *BMC Geriatr. 7 Mai*. **22** (1), 404 (2022).
14. Faurot, K. R. et al. Using claims data to predict dependency in activities of daily living as a proxy for frailty. *Pharmacoepidemiol Drug Saf.* **24** (1), 59–66 (2015).
15. Kim, D. H. et al. Measuring frailty in medicare data: development and validation of a claims-based frailty index. *J. Gerontol. Biol. Sci. Med. Sci.* **73** (7), 980–987 (2018).
16. Kinosian, B. et al. Validation of the JEN frailty index in the National long-term care survey community population: identifying functionally impaired older adults from claims data. *BMC Health Serv. Res.* **18** (1), 908 (2018).
17. Gilbert, T. et al. Development and validation of a hospital frailty risk score focusing on older people in acute care settings using electronic hospital records: an observational study. *Lancet* **391** (10132), 1775–1782 (2018).
18. Segal, J. B. et al. Development of a claims-based frailty indicator anchored to a well-established frailty phenotype. *Med. Care*. **55** (7), 716–722 (2017).
19. Clegg, A. et al. Development and validation of an electronic frailty index using routine primary care electronic health record data. *Age Ageing 1 Mai*. **45** (3), 353–360 (2016).
20. Abbasi, M. et al. A cross-sectional study examining convergent validity of a frailty index based on electronic medical records in a Canadian primary care program. *BMC Geriatr.* 16 avr. 19(1) 109. (2019).
21. Ruiz, J. G. et al. Validation of an automatically generated screening score for frailty: the care assessment need (CAN) score. *BMC Geriatr.* 4 mai. **18** (1) 106. (2018).
22. Kim, D. H. et al. Measuring frailty in administrative claims data: comparative performance of four claims-based frailty measures in the U.S. medicare data. *J Gerontol A Biol Sci Med Sci.* 22 mai. **75** (6):1120–5. (2020).
23. Mak, J. K. L. et al. Development of an electronic frailty index for hospitalized older adults in Sweden. *J. Gerontol. Biol. Sci. Med. Sci.* **21** (11), 2311–2319 (nov 2022).
24. Turner, G. & Clegg, A. Best practice guidelines for the management of frailty: a British geriatrics society, age UK and Royal college of general practitioners report. *Age Ageing 1 Nov*. **43** (6), 744–747 (2014).
25. Kaier, K. et al. Reweighting and validation of the hospital frailty risk score using electronic health records in Germany: a retrospective observational study. *BMC Geriatr. 13 Juin*. **24**, 517 (2024).
26. Fuentes, S. et al. Identifying diabetes cases in health administrative databases: a validation study based on a large French cohort. *Int. J. Public. Health*. **64** (3), 441–450 (2019).
27. Haneef, R. et al. Use of artificial intelligence for public health surveillance: a case study to develop a machine Learning-algorithm to estimate the incidence of diabetes mellitus in France. *Arch. Public. Health*. **79**, 168 (2021).
28. Gabet, A. et al. Stroke incidence and death in atrial fibrillation patients newly treated with direct oral anticoagulants. *Clin. Epidemiol.* **13**, 131–140 (2021).
29. Moisan, F. et al. Prediction model of Parkinson's disease based on antiparkinsonian drug claims. *Am. J. Epidemiol.* **174** (3), 354–363 (2011).
30. Gilbert, T. et al. External validation of the hospital frailty risk score in France. *Age Ageing 26 Juin*. **51** (1), afab126 (2021).

31. Célant, N., Guillaume, S. & Rochereau, T. Enquête sur la santé et la protection sociale 2012. Rapport Irdes [Internet]. 1 juin 2014 [cité 8 déc 2021];(556). Disponible sur: https://www.irdes.fr/recherche/2014/rapport-556-enquete-sur-la-sante-et-la-protection-sociale-2012.html
32. Tuppin, P. et al. Value of a national administrative database to guide public decisions: From the système national d'information interrégimes de l'Assurance Maladie (SNIIRAM) to the système national des données de santé (SNDS) in France. Revue d'Épidémiologie et de Santé Publique. ;65:S149–67. (2017).
33. World Health Organization. International statistical classification of diseases and related health problems. Tenth revision. (1992).
34. Perrine, A. L. et al. and multimorbidity: Can French national surveys be used for the production of these indicators? Data analyses of ESPS 2012 and HSM 2008 surveys. Bull Epidémiol Hebd ;(16–17):301 – 10 [Internet]. 2017; Disponible sur: (2017). http://invs.santepubliquefrance.fr/beh/2017/16-17/2017_16-17_2. Html.
35. Sirven, N. Measuring frailty among the elderly in a general population setting a comparison between ESPS and SHARE [Internet]. Working Papers. IRDES institut for research and information in health economics; 2014 [cité 1 févr 2022]. (Working Papers). Report No.: DT60. Disponible sur: https://ideas.repec.org/p/irh/wpaper/dt60.html
36. Katz, S., Downs, T. D., Cash, H. R. & Grotz, R. C. Progress in development of the index of ADL. *Gerontologist* **10** (1), 20–30 (1970).
37. Tibshirani, R. Regression shrinkage and selection via the Lasso. *J. Royal Stat. Soc. Ser. B (Methodological)*. **58** (1), 267–288 (1996).
38. Breiman, L., Machine & Learning number 1 - SpringerLink. *Mach. Learn.* **45**, 45 5–32 (2001).
39. Jiang, H. et al. Joint analysis of two microarray gene-expression data sets to select lung adenocarcinoma marker genes. *BMC Bioinform.* **5**, 81 (2004).
40. Andersen, R. & Newman, J. F. Societal and individual determinants of medical care utilization in the united States. *Milbank Mem. Fund Q. Health Soc.* **51** (1), 95–124 (1973).
41. Doyal, L. Sex, gender, and health: the need for a new approach. *BMJ 3 Nov*. **323** (7320), 1061–1063 (2001).
42. Thompson, A. E. et al. The influence of gender and other patient characteristics on health care-seeking behaviour: a QUALICOPC study. *BMC Fam Pract.* **17**, 38 (2016).
43. Bertakis, K. D., Azari, R., Helms, L. J., Callahan, E. J. & Robbins, J. A. Gender differences in the utilization of health care services. *J. Fam Pract. Févr*. **49** (2), 147–152 (2000).
44. NAGELKERKE NJD. A note on a general definition of the coefficient of determination. *Biometrika 1 Sept*. **78** (3), 691–692 (1991).
45. Youden, W. J. Index for rating diagnostic tests. *Cancer* **3** (1), 32–35 (1950).
46. Le Cossec, C. et al. Frailty, and Multimorbidity: prevalences and associated characteristics from two French National surveys. *J. Nutr. Health Aging*. **20** (8), 860–869 (2016).
47. Constantinou, P. et al. Two morbidity indices developed in a nationwide population permitted performant outcome-specific severity adjustment. *J. Clin. Epidemiol. 1 Nov*. **103**, 60–70 (2018).
48. Kuhn, M. Building predictive models in R using the caret package. *J. Stat. Softw.* **28**, 1–26 (2008).
49. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33** (1), 1–22 (2010).
50. Copeland, K. T., Checkoway, H., McMichael, A. J. & Holbrook, R. H. Bias due to misclassification in the estimation of relative risk. *Am. J. Epidemiol.* **105** (5), 488–495 (1977).
51. Schmidhuber, J. Deep learning in neural networks: an overview. *Neural Networks 1 Janv*. **61**, 85–117 (2015).
52. Blanes-Selva, V., Doñate-Martínez, A., Linklater, G. & García-Gómez, J. M. Complementary frailty and mortality prediction models on older patients as a tool for assessing palliative care needs. *Health Inf. J.* **28** (2), 14604582221092592 (2022).
53. Aponte-Hao, S. et al. Machine learning for identification of frailty in Canadian primary care practices. *Int. J. Popul. Data Sci.* **6** (1), 1650 (2021).
54. Heins, S. E., Agniel, D., Mann, J., Sorbero, M. E. & Reviewing refining, and validating claims-based algorithms of frailty and functional impairment: final report [Internet]. RAND Corporation; 2023 nov [cité 29 janv 2024]. Disponible sur: https://www.rand.org/pubs/research_reports/RRA1871-1.html
55. Segal, J. B., Huang, J., Roth, D. L. & Varadhan, R. External validation of the claims-based frailty index in the National health and aging trends study cohort. *Am. J. Epidemiol. 15 Sept*. **186** (6), 745–747 (2017).
56. Rodríguez-Mañas, L. & Rodriguez-Sánchez, I. Research on frailty: where we stand and where we need to go. *J. Am. Med. Dir. Association 1 Mars*. **22** (3), 520–523 (2021).
57. Carcaillon-Bentata, L., Soullier, N., Beltzer, N. & Coste, J. Alteration in perceived health status of those aged 55 to 65 between 2010 and 2017 in France: role of socioeconomic determinants. *BMC Public. Health*. **7** (1), 1804 (oct 2021).
58. Welsh, C. E., Matthews, F. E. & Jagger, C. Trends in life expectancy and healthy life years at birth and age 65 in the UK, 2008–2016, and other countries of the EU28: An observational cross-sectional study. Lancet Reg Health Eur. 4 janv. ;2:100023. (2021).
59. Boissonneault, M. & Rios, P. Changes in healthy and unhealthy working-life expectancy over the period 2002-17: a population-based study in people aged 51–65 years in 14 OECD countries. *Lancet Healthy Longev. Oct.* **2** (10), e629–e638 (2021).
60. Buta, B. et al. Agreement between standard and self-reported assessments of physical frailty syndrome and its components in a registry of community-dwelling older adults. *BMC Geriatr. 25 Août*. **22** (1), 705 (2022).
61. Rodríguez-Mañas, L. et al. Searching for an operational definition of frailty: a Delphi method based consensus statement: the frailty operative definition-consensus conference project. J Gerontol A Biol Sci Med Sci. janv. ;68(1):62–7. (2013).
62. Soong, J. T. Y., Ng, S. H. X., Tan, K. X. Q., Gammall, J. & Hopper, A. Variation in coded frailty syndromes in secondary care administrative data: an international retrospective exploratory study. *BMJ Open. 27 Janv*. **12** (1), e052735 (2022).
63. Street, A. et al. The use of linked routine data to optimise calculation of the hospital frailty risk score on the basis of previous hospital admissions: a retrospective observational cohort study. *Lancet Healthy Longev. 1 Mars*. **2** (3), e154–e162 (2021).
64. Le dispositif d'enquêtes « Autonomie ». –2025) | Direction de la recherche, des études, de l'évaluation et des statistiques [Internet]. [cité 29 janv 2024]. Disponible sur: (2021). https://drees.solidarites-sante.gouv.fr/sources-outils-et-enquetes/le-dispositif-denquetes-autonomie-2021-2025
65. Hucteau, E., Noize, P., Pariente, A., Helmer, C. & Pérès, K. ADL-dependent older adults were identified in medico-administrative databases. *J. Clin. Epidemiol.* **139**, 297–306 (2021).

Acknowledgements

We would like to thank the French Secure Access Data Centre (CASD) for providing access to the data used in this study.

Author contributions

HL: Methodology, Data curation, Formal analysis, Software, Validation, Writing - Original draft, Writing - Review & Editing.LMB: Methodology, Data curation, Software, Writing - Review & Editing.SG, VW: Software, Formal analysis, Writing - Review & Editing. CF, CH, KP, MT: Methodology, Writing - Review & Editing.LCB: Conceptualization, Methodology, Writing - Original draft, Writing - Review & Editing, Supervision, Project administration, Validation.

Declarations

Competing interests
The authors declare no competing interests.

Ethics approval
Access by Santé Publique France to the ESPS data linked to the SNDS received ethical approval from the relevant institutional review board (DE-2017-361).

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-95629-z.

Correspondence and requests for materials should be addressed to L.C.-B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.