Predicting Avoidable Hospitalizations in Catalonia among Primary Care Teams: A Machine Learning Approach

Carlos Gallego

August 2025

Abstract

Introduction. Ambulatory Care Sensitive Conditions (ACSC) are widely used as a proxy for the performance of the primary-care production function: they capture hospitalizations that should be avoidable given timely, effective ambulatory management. We aim to develop and validate machine-learning models that forecast next-year rates of ACSC hospitalisations per 100,000 inhabitants at the Primary Care Team (PCT) level in Catalonia, using heterogeneous routine data and providing interpretable, management-oriented outputs.

Methods. Retrospective panel of PCT–years with a complete-year temporal split (2018–2022 train; 2023 test) and expanding-origin cross-validation to mimic prospective deployment. We implemented penalised linear models (Lasso, Ridge), decision tree, random forest, gradient-boosted trees (XGBoost), support vector regression (RBF), and MARS within a leakage-controlled pipeline. Feature selection aggregated three signals

(Boruta, Elastic Net, SHAP) to retain stable and high-contribution predictors. Models were trained under two conditions—using either the full original feature set (581 predictors) or a multi-signal selected (259 predictors) subset—and benchmarked against a naïve lag-1 baseline; grouped SHAP was used for interpretability. Performance was evaluated on the 2023 hold-out with RMSE (primary), MAE, and R^2.

Results. The best model was XGBoost with all features (RMSE $= 92.08$, MAE $= 73.18$, $R^2 = 0.91$), improving on the naïve lag-1 baseline (RMSE $= 118.00$) by an absolute margin of 25.92 per 100,000. Lasso (all features) was close (RMSE $= 93.30$), as was Random Forest on the selected set (RMSE $= 93.70$). Across paired comparisons, all-features variants generally matched or outperformed selected-feature counterparts, with Random Forest and Ridge as exceptions favouring selection (stability gains). SHAP explanations indicated a concentrated importance structure: the top 25 features accounted for $\approx 55\%$ of total mean absolute SHAP, with prominent roles for healthcare utilisation, clinical history, provider resources and quality, and a team identifier capturing residual between-team heterogeneity.

Conclusions. Routine, system-wide data can produce accurate, interpretable forecasts of ACSC rates at the team level. The small performance premium of the full feature set must be weighed against interpretability, stability, and computational costs of high dimensionality. Outputs enable risk-adjusted benchmarking, targeted resource allocation, and proactive operational monitoring; however, models are predictive—not causal—and should complement judgement and prospective evaluation when guiding interventions.

Keywords: Artificial Intelligence, Health Economics, Healthcare Utilisation Outcomes, Resource Allocation, Review.

1. Introduction

The rate of preventable hospitalisations—Ambulatory Care Sensitive Conditions (ACSCs)—is a core indicator of primary care performance, capturing access and clinical quality in the prevention, timely diagnosis, and management of chronic and acute problems. ACSCs are admissions that effective primary care should avert (e.g., diabetes complications, COPD exacerbations, hypertensive crises, early infections) [1–4].

For this reason, major health agencies such as AHRQ, OECD, WHO/PAHO, and NHS routinely employ ACSC admission rates—operationalised in instruments such as the AHRQ Prevention Quality Indicators—as proxy indicators of primary-care resolutive capacity and population access to the health system; these metrics are widely used in monitoring reports and to inform resource-allocation and policy decisions [3, 5–8].

In Catalonia, ACSC monitoring is embedded within the governance of the regional health system. The Catalan Health Service (CatSalut) systematically tracks ACSC admission rates, while the Agency for Health Quality and Evaluation of Catalonia (AQuAS) includes potentially avoidable hospitalisations in its composite socioeconomic index used to guide the allocation of primary care resources [9–11]. Early research in Catalonia validated ACSC as an indicator of primary care resolutive capacity across 161 basic health areas [12], and subsequent analyses of over 1.3 million discharges confirmed that a small set of diagnostic categories accounted for most preventable admissions and identified primary prevention, early diagnosis, and ambulatory management as priority strategies [13].

Variation in ACSC admissions is influenced by a wide array of factors operating at multiple levels. At the individual level, age, multimorbidity, and socioeconomic status are consistent predictors of risk [14–16]. Area-level determinants such as socioeconomic deprivation,

population density, and community disease burden are also strongly associated with ACSC rates [3, 17, 18].

Although much of the literature has emphasised demand-side predictors derived from electronic health data, evidence from small-area studies shows that supply and service characteristics also matter: rurality and distance to primary care, physician density, hospital bed availability, and service organisation explain part of the variation in different settings [19–21].

Despite this evidence, most predictive models of ACSC hospitalisations developed to date rely almost exclusively on patient-level variables, often omitting provider- and service-level covariates. This limits their usefulness for system-level planning and resource allocation [22, 23]. Conceptual frameworks such as Andersen's healthcare utilisation model offer a structured foundation for variable selection in predictive models, broadening the scope beyond individual risk stratification [21]. Rosella et al. emphasise that predictive models designed for decision support should explicitly include determinants relevant to health-system planning, rather than focusing solely on patient-level risk [24].

Methodological debates further complicate this landscape. Machine learning (ML) methods are well suited to high-dimensional administrative data and capture complex, nonlinear relationships, but often at the expense of interpretability. Classical statistical models, such as multivariable regression, are more transparent and allow for clear estimation of effect sizes, and in some contexts perform nearly as well. For example, a German study of 6.4 million insured individuals found logistic regression and Random Forest achieved comparable accuracy for predicting ACSC hospitalisations (c-statistics ≈ 0.78), with Random Forest slightly superior but regression offering greater interpretability [23]. This underscores the

need to balance predictive performance with transparency and practical usability.

There is no robust prediction of ACSC hospitalisations at the primary care team (Equip d'Atenció Primària, EAP) level that integrates both demand- and supply-side determinants. This gap is particularly relevant in Catalonia, where a mixed model of service provision and harmonised, system-wide administrative data—comparable across providers—enable large-scale, team-level analyses.

We aim to address these gaps by developing and validating Machine Learning models to predict the number of avoidable hospitalisations (ACSCs) for the adult population in Catalonia between 2018 and 2023 at the Primary Care Team (PCT) level.

By incorporating both demand- and supply-side determinants, this study also seeks factors associated with higher predicted numbers of avoidable hospitalisations, including the influence of provider characteristics and management models, to inform organisational strategies and resource allocation within primary care and to provide actionable insights for primary care management. By integrating demand- and supply-side determinants, we seek to identify which factors—such as workforce, workload, resource endowment, and management model—drive elevated predicted ACSC rates. Our models can be updated periodically to flag PCTs with performance concerns and surface interpretable drivers to guide targeted organisational strategies, prioritise resource allocation, and support operational decision-making in primary care. Specifically, the models will help identify PCTs at high risk of avoidable hospitalisations, distinguish whether inter-team variation stems from differences in population needs, performance or from resource constraints, and inform equitable, evidence-based decisions on resource allocation and intervention prioritisation. In doing so, the research contributes to the development of a learning health system in Catalonia, supports

strategic objectives to reduce avoidable hospitalisations and territorial inequities, and advances the methodological agenda for predictive modelling in primary care.

1.1. Institutional Catalan Healthcare System Background

Established by the 1990 Health System Organisation Law (LOSC), the Catalan system follows a purchaser–provider split: CatSalut, a public institutional body attached to the Department of Health, serves as the public insurer, planning, financing, purchasing, and evaluating health services while contracting a mixed network of public, non-profit, for-profit, and consortial providers [25]. Primary care illustrates this model. The territory is organised into basic health units ($\approx 5,000$–25,000 residents), each served by at least one primary care center and a multidisciplinary Primary Care Team (PCT) responsible for promotion, prevention, basic care, rehabilitation, and coordination with specialised care. Catalonia has $\approx 370 PCTs$, with each citizen assigned a reference professional [26].

Management is heterogeneous: the Catalan Health Institute (ICS) is the main Public Company of the Catalan Government and operates $\approx 78\%$ of PCTs; other public non-ICS entities (e.g., consortia) manage ≈ 12–13%; and private contractors ≈ 9–10%, including professional-led EBAs funded by capitation under public contracts. For users, services are uniform with a common portfolio and no co-payment.

System-wide planning by the public insurer, CatSalut, standardises objectives and indicators, producing harmonised, comparable data to assess preventive effectiveness (e.g., ACSCs) across providers [27]. This setting supports analyses of how provider characteristics influence healthcare utilisation and health outcomes.

2. Methods

2.1. Study design and setting

A retrospective panel study was conducted with annual observations for each entity. The data spans multiple years (2018-2023) and is structured by entity and year. The outcome of interest is a numeric rate of ambulatory-care-sensitive condition events per PCT. Model development used earlier years as training data and last year as a hold-out test set, reflecting a prospective prediction scenario and avoiding any use of future data in training.

2.2. Data sources

This study primarily used two secondary data sources from the Catalan health system: PADRIS (*Programa d'Analítica de Dades per a la Recerca i la Innovació en Salut*) and SISAP (*Sistema d'Informació dels Serveis d'Atenció Primària*).

2.2.1 PADRIS:

The PADRIS program, led by the Agency for Health Quality and Assessment of Catalonia (AQuAS) under the Ministry of Health, is a pivotal initiative designed to facilitate health research, innovation, and evaluation through the reuse and linkage of healthcare data generated by the integrated public health system of Catalonia (SISCAT) [28]. PADRIS prioritizes making health data available to the scientific community, ensuring adherence to legal and ethical frameworks, and promoting transparency. Its mission is to leverage extensive health information to advance research, innovation, and evaluation, thereby improving public health and strengthening Catalonia's position as a reference in health information society and high value-added services.

2.2.2 SISAP:

SISAP, initiated in 2006, is a fundamental information system designed to provide critical data primarily for clinical management to healthcare professionals and management structures within primary care [29]. It achieves this by collecting data from various sources and processing them to construct meaningful indicators. The Sisap-eCap is the web application linked to the eCAP electronic health record system, allowing primary care professionals to consult their own performance indicators and patient lists that do not meet specific criteria. This system is crucial for continuous quality improvement and decision-making in primary care.

2.3. Variables: Outcome and predictors

This combined dataset comprised a total of 789 variables. After excluding 208 SISAP-derived variables according to a multi-year missingness assessment (variables with >45% missingness in all years; >45% missingness in four or more years; or >55% missingness in three or more years), the final predictor set comprised 581 variables (see Supplemental Material 1 for the complete list). The key variable groups included:

- **Accessibility indicators** (n = 24): Reflecting ease of access to primary care services.
- **Administrative/Geographic** (n = 4): Variables describing administrative functioning and the geographic context of Primary Care Teams (PCTs).
- **Longitudinality indicators** (n = 10): Measures of continuity of care and patient–provider relationships.
- **Medical history / clinical record** (n = 264): Aggregated information on patient morbidities, including CIAP (International Classification of Primary Care), GMA

(Adjusted Morbidity Groups), and PCCMACA (Adjusted Clinical Risk Grouping System in Primary Care for Chronic Patients with Multiple Conditions).

- **Population and socioeconomic (assigned and attended)** (n = 6): Characteristics of the population *assigned to* and *attended by* each PCT, including socioeconomic attributes.
- **Provider resources** (n = 124): Staffing and financial resources available to providers.
- **Quality indicators** (n = 49): Indicators of service quality, including medication use and the quality of pharmaceutical prescribing (e.g., pharmacy invoicing charged to the Catalan Health Institute, ICS).
- **Service utilization** (n = 80): Extent to which patients engage with healthcare services (e.g., visits, activity).

The outcome is the continuous count of avoidable hospitalizations per 100,000 inhabitants at the PCT-year level, defined as the aggregate of admissions for standard ACSC diagnoses (e.g., COPD, hypertension, heart failure, dehydration, bacterial pneumonia, urinary tract infection, angina, asthma, and diabetes-related complications).

For categorical predictors, very low-frequency levels in the training data were merged: categories appearing in less than 0.5% of training instances were replaced by an "other" level. (In the final modeling stage, a 1% frequency threshold was used similarly.) If a category appeared in validation or test sets that was unseen in training, it was labeled as "new level"). These steps prevented sparse or novel factor levels from destabilizing the modeling. All modeling steps respected the temporal ordering of data, with no leakage of future information into training folds or final models.

2.4. Preprocessing

Data preprocessing followed a prespecified pipeline estimated exclusively on the training data to prevent information leakage. The main steps were as follows. **Imputation**: missing numeric values were replaced with the training-set median, and missing categorical values with the training-set mode; in model-specific recipes, missing categories were assigned to an explicit "unknown" level. **Categorical handling**: rare categories were pooled, and previously unseen categories at assessment were flagged; for algorithms requiring numeric inputs, categorical predictors were one-hot encoded. **Normalization**: numeric predictors were standardized (zero mean, unit variance) for scale-sensitive algorithms (e.g., linear models, SVMs). Tree-based and boosting models did not require scaling, and for these models factor encodings were retained rather than one-hot encoded. **Variance filtering**: predictors with zero variance in the training set were removed. All preprocessing parameters (e.g., imputation values, standardization moments, level maps) were learned from the training data within each resampling fold and then applied to the corresponding validation or test data; no information from future or held-out observations was used at any stage.

2.5. Train/test split and temporal resampling

The data were partitioned by calendar year: five years (2018-2022) were used for training, and the most recent year (2023) were held out for testing. Thus, the model was trained on an initial set of years and evaluated on completely later years. Within the training period, an expanding-window temporal cross-validation was performed. Folds were defined such that each fold's training set consisted of all data up to year Y–1 and the validation set was the data from year Y. In practice, two folds were used: one validating on the second-to-last training year and another on the last training year. This ensured that each validation year came after its

training data. No data from a given year were ever used to train a model evaluated on that same year, preventing look-ahead bias. This temporal resampling was used for model tuning and feature selection, providing a realistic estimate of performance on "next-year" predictions.

2.6. Feature selection strategy (multi-signal)

Feature selection was driven by three methods applied on each training fold:

Boruta (random forest wrapper): Boruta was run on the training fold's predictors (after preprocessing) to identify important features, using a significance threshold of $p < 0.01$ (max 100 iterations). It returned a subset of important dummy-coded features, which were then mapped back to their original variable names.

XGBoost SHAP values: An XGBoost model (regression with squared-error loss, using fixed hyperparameters such as learning rate 0.05 and maximum depth 6) was trained on the training fold. SHAP values were then computed on the fold's validation set; the mean absolute SHAP contributions of dummy features were summed by original variable to quantify each predictor's importance.

Elastic Net (regularized regression): An Elastic Net linear model ($\alpha = 0.5$) was fit on the training fold, with regularization λ selected via cross-validation (using the one-standard-error rule in cv.glmnet). Predictors with non-zero coefficients in the resulting model were recorded (after mapping any dummy features back to their original names).

Each method produced a set of selected or ranked variables for a fold. To aggregate results across folds, stable features were defined as those chosen in 60% of the folds by Boruta or by Elastic Net. Separately, top SHAP features were determined by averaging the SHAP importance across folds and selecting the set of variables that accounted for $\approx 90\%$ of the total

mean SHAP contribution (cumulative). The union of these sets was then taken to define the final selected feature set. In other words, a variable was included if it was consistently selected in multiple folds or if it was among the most important predictors by SHAP contribution. This multi-signal approach combined the strengths of different selection criteria to yield a robust predictor set.

2.7. Model training, hyperparameter tuning and baseline comparator

Using the selected predictors together with the lagged outcome, we trained and tuned a suite of models: penalised linear regressions (Lasso and **Ridge**), a **decision tree**, a **random forest**, **gradient–boosted trees**, **support vector machine** regression with a radial–basis kernel, and multivariate adaptive regression splines (**MARS**). All models were embedded in a fixed preprocessing–modeling pipeline estimated on the training data only. Hyperparameters were optimised via grid search within k-fold cross–validation using RMSE as the objective. Specifically, we tuned the L1/L2 penalty (λ) for Lasso/Ridge; tree depth and minimum node size for the decision tree; the number of candidate predictors per split (with the number of trees set large and fixed) for the random forest; learning rate and tree parameters (e.g., depth) for gradient boosting; the cost and kernel bandwidth (γ) for support vector regression; and the number of basis functions for MARS. The configuration minimising the mean validation RMSE across folds was selected, and the corresponding model was then refit on the full training set prior to final evaluation.

Each algorithm was trained twice—on the multi-signal selected feature set and on the full original feature set—using the same hyperparameter search space and the same expanding-origin cross-validation protocol (RMSE selection). Optimal hyperparameter values were re-estimated separately for each algorithm and feature set. This design isolates the effect

of feature selection from that of hyperparameter tuning: the tuning protocol is identical across conditions, while optimal values are allowed to differ. This served to benchmark the performance of the feature-selected models against the same models using all available predictors.

As a baseline benchmark, a naïve lag-1 model was used: for each entity i and year t, the prediction is the outcome of the previous year ($\hat{Y}_{i,t} = Y_{i,t-1}$). A feature `y_lag1` was created to carry the lagged outcome value for each entity (the value of the outcome from the prior year). This simple model assumes the best prediction for the current year is the last year's value. For test instances where no prior year was available (e.g., an entity's first appearance in the data), the baseline could not generate a prediction; those cases were omitted from baseline error calculations. The lag-1 model provides a floor of performance that the more complex models should surpass.

2.8. Performance evaluation

Model performance was evaluated using Root Mean Squared Error (RMSE) as the primary metric, with Mean Absolute Error (MAE) and R-squared (R^2) reported as secondary metrics. During cross-validation, RMSE and R^2 were computed for each fold's validation set predictions, and the mean (± standard deviation) across folds was calculated to summarize performance. For final evaluation, each model was evaluated on the hold-out test set (the unseen years). RMSE, MAE, and R^2 were calculated on the test data to assess out-of-sample accuracy. The same metrics were also computed for the naïve lag-1 baseline on the test set (for those instances where a lagged prediction existed) to provide a point of reference. The expectation was that the trained models would achieve lower error and higher R^2 than the baseline if they captured predictive patterns.

2.9. Reproducibility and software

The analysis was conducted in a scripted, reproducible workflow (R, v4.4.3). Full implementation details—including the complete list of software packages and exact version numbers—are available in the project's GitHub repository (see: ¡notyetavailablelink¿). Random seeds were set to ensure reproducibility of results. These practices allow the analysis to be audited and reproduced exactly using the same code and data.

3. Results

3.1. Descriptive results

Table 1 summarizes the evolution of ambulatory care–sensitive condition (ACSC) rates per 100,000 inhabitants across provider ownership types from 2018 to 2023. For each ownership category, the total number of entity–year observations (N total) is shown together with the annual mean ± standard deviation. The final row aggregates all providers. These descriptive statistics characterize the panel over complete calendar years and contextualize the subsequent modeling results.

Table 1: ACSC per 100,000 inhabitants by ownership type, 2018–2023 (mean ± SD).

Ownership type	N total	2018	2019	2020	2021	2022	2023
		Mean ± SD	Mean ± SD	Mean ± SD	Mean ± SD	Mean ± SD	Mean ± SD
Catalan Health Institute	1731	1129.4 ± 322.8	1103.4 ± 338.6	980.0 ± 320.6	834.8 ± 251.8	916.7 ± 273.9	1000.1 ± 292.2
Consortium	142	986.4 ± 305.2	1002.9 ± 261.9	873.5 ± 192.5	727.5 ± 166.6	781.0 ± 197.2	873.0 ± 193.0
Private Beneficent Foundation	90	941.8 ± 287.2	951.6 ± 282.1	858.1 ± 292.9	723.7 ± 273.3	757.7 ± 231.3	877.8 ± 259.3
Public Enterprise	78	906.1 ± 231.8	943.1 ± 262.8	810.3 ± 236.0	676.5 ± 214.1	764.5 ± 209.3	908.2 ± 186.4
EBA (Associative Base Entity)	66	941.7 ± 280.0	1043.0 ± 305.0	1002.8 ± 273.4	714.7 ± 252.4	775.0 ± 273.1	910.1 ± 287.5
Other Private	48	921.1 ± 265.9	914.4 ± 266.8	824.1 ± 236.6	689.2 ± 206.7	713.5 ± 224.7	827.9 ± 268.8
Municipal	42	1606.5 ± 241.7	1651.3 ± 275.9	1400.8 ± 233.8	1431.0 ± 237.7	1526.5 ± 269.1	1538.5 ± 297.0
Other Public	18	551.9 ± 105.7	580.6 ± 160.0	620.3 ± 224.2	650.0 ± 317.7	678.8 ± 368.8	604.4 ± 405.2
Society	18	1205.3 ± 116.2	1298.5 ± 272.6	1022.7 ± 272.1	661.3 ± 320.5	732.7 ± 353.1	913.7 ± 242.7
Social Welfare Mutual	6	931.2 ± —	993.4 ± —	896.8 ± —	700.7 ± —	769.4 ± —	814.7 ± —
TOTAL	2441	1076.0 ± 326.8	1063.5 ± 338.3	951.2 ± 314.3	819.7 ± 258.4	889.3 ± 281.9	962.8 ± 296.7

3.2. Sample and Temporal Split

A total of 2441 — entity-year observations were analyzed. The modeling pipeline was trained on annual data from 2018 through 2022, with 2023 held out for testing. The expanding-origin cross-validation (CV) scheme used 4 folds, each using all data up to year Y for training and evaluating on year Y+1: Fold 1 trained on 2018 and validated on 2019; Fold 2 trained on 2018–2019 and validated on 2020; Fold 3 trained on 2018–2020 and validated on 2021; and Fold 4 trained on 2018–2021 and validated on 2022.

3.3. Variable Selection

The multi-method feature selection process resulted in a final set of 259 predictor variables. These variables represented the union of features identified as important by the Boruta and Elastic Net stability methods. In the Boruta stability analysis, 139 variables achieved a selection frequency of at least 60%, while the Elastic Net selection frequency criterion (60%) was met by 141 variables. In total, 259 unique predictors met the 0.60 stability threshold in at least one of the two methods, confirming a substantial agreement between Boruta and Elastic Net on the informative features.

The stability of feature importance was further characterized by each method. Many predictors showed high selection stability (frequency near 1.0) in Boruta, indicating they were consistently confirmed as important across CV resamples. Similarly, the Elastic Net identified a subset of predictors with high coefficient stability. The overlap between methods was considerable, as reflected in the union count above, though each method also contributed some unique variables below the 0.60 threshold in the other method.

Aggregated SHAP importance values were calculated for the selected features using the final

model. Table 3 lists the top 25 features ranked by mean absolute SHAP value, along with their individual contribution percentages and cumulative contribution. The most influential predictor had the highest mean SHAP value, contributing the largest single share of explained model output variability. The top 25 features together accounted for 55% of the total importance. Because this cumulative contribution did not reach 90%, the importance analysis was extended. Beyond the top predictors, the importance values dropped off gradually, indicating a long tail of features with small contributions.

Table 2: Top 25 predictor variables ranked by mean absolute SHAP value (grouped by original variable).

Description	mean shap	sd shap	share	cumulative share	variable group
Prescriptions per person	60.947	6.375	0.108	0.108	Quality indicators
Medication users: Oral antidiabetics (%)	34.167	2.932	0.060	0.168	Service utilization
Primary Care Team code	28.123	5.433	0.050	0.218	Administrative/Geographic
Allergic rhinitis	22.648	18.281	0.040	0.258	Medical history / clinical record
Medication users: Antipsychotics (%)	21.684	3.000	0.038	0.296	Service utilization
Digital health portal usage (% users)	21.538	27.757	0.038	0.334	Service utilization
Non–insulin–dependent diabetes	16.141	2.868	0.029	0.363	Medical history / clinical record
Medication users: Antibiotics (%)	15.529	7.133	0.027	0.390	Service utilization
Medication users: Antidepressants (%)	13.282	3.651	0.023	0.414	Service utilization
Public pharmaceutical expenditure (per assigned patient)	10.520	2.563	0.019	0.432	Provider resources
% Recommended medicines / HBP	7.564	6.541	0.013	0.446	Quality indicators
Medication users: Lipid-lowering agents (%)	7.127	2.866	0.013	0.458	Service utilization
Trauma / unspecified injury	5.884	8.322	0.010	0.469	Medical history / clinical record
Appointment supply per professional per month	5.734	0.683	0.010	0.479	Service utilization
Reagents and analogues	4.867	2.961	0.009	0.487	Provider resources
Percentage of scheduled visits completed	4.744	3.848	0.008	0.496	Service utilization
Primary care visits: All	4.615	1.899	0.008	0.504	Service utilization
Primary care visits: Online	4.432	3.610	0.008	0.512	Service utilization
Social workers expenditures	3.892	1.287	0.007	0.519	Provider resources
FreeStyle Lite-TR blood glucose	3.595	3.026	0.006	0.525	Provider resources
GMA33 – Patients with chronic disease in 4 or more systems	3.415	0.204	0.006	0.531	Medical history / clinical record
Cataract	3.326	1.316	0.006	0.537	Medical history / clinical record
Incomes: Third-party billing	3.272	0.187	0.006	0.543	Provider resources
Referrals observed	3.247	1.341	0.006	0.548	Referrals
GMA332 – Patients with chronic disease in 4 or more systems – CMPLX 2	2.981	2.057	0.005	0.554	Medical history / clinical record

Notes. ‘mean shap‘ and ‘sd shap‘ are the mean and standard deviation of absolute SHAP contributions aggregated by original variable across validation folds. ‘share‘ denotes the variable’s fraction of the total mean absolute SHAP sum; ‘cumulative share‘ is the running total. Values are shown with three decimal places. Groups follow the variable grouping used in the analysis.

3.4. Model tuning and hyperparameter contrast: selected features vs. all features

Each algorithm was tuned under both conditions using the same search space and the same expanding–origin cross–validation protocol (RMSE selection), while allowing the optimal

values to differ. The best configuration for each model balanced flexibility and generalization. Table 3 contrasts the resulting hyperparameters: (i) Lasso and Ridge retained essentially the same penalties in both conditions; (ii) the decision tree shifted to a much smaller *min_n* and larger cost–complexity when all features were used; (iii) Random Forest increased *mtry* substantially with all features; (iv) XGBoost moved to a shallower but much larger ensemble with a smaller learning rate and stronger row/column subsampling under all features; (v) SVM RBF increased cost and reduced kernel width on the full–feature space; and (vi) MARS kept a parsimonious specification in both settings. These tuned hyperparameters were used to train the final models on the entire 2018–2022 training data.

Table 3: Best hyperparameters per model: selected features vs. all features (values rounded to 3 decimals).

Model	**Hyperparameter(s)**	**Selected features (values)**	**All features (values)**
Lasso (glmnet)	penalty (λ)	10.000	10.000
Ridge (glmnet)	penalty (λ)	0.000[a]	0.000[a]
Decision Tree (rpart)	cost_complexity; tree_depth; min_n	0.000[b]; 6; 32	0.002; 6; 2
Random Forest (ranger)[c]	mtry; min_n; trees	155; 18; 1000	255; 5; 1000
XGBoost (xgboost)	trees; tree_depth; min_n; learn_rate; loss_reduction (γ); sample_size; mtry	448; 10; 5; 0.046; 0.000[d]; 0.569; 5	1170; 2; 11; 0.008; 0.001; 0.972; 555
SVM (RBF, kernlab)	cost; rbf_sigma	4.632; 0.001	12.435; 0.001
MARS (earth)	num_terms; prod_degree; prune_method	5; 2; backward	5; 2; backward

[a] Exact value $\lambda = 10^{-6}$ (rounds to 0.000 at three decimals).
[b] Exact value 10^{-10} for cost_complexity (rounds to 0.000).
[c] Number of trees fixed at 1000 in training (not tuned).
[d] Exact value 4.431697×10^{-5} for XGBoost's γ (rounds to 0.000).

3.5. Test Set Performance and Baseline Comparison

Each model's predictive performance was evaluated on the held-out 2023 test set and compared with a naïve lag-1 baseline. Table 3 reports test RMSE, MAE, and R^2 for all

algorithms under both conditions (selected features vs. all features) and the baseline. The best overall model was XGBoost (all features) (RMSE $= 92.08$, MAE $= 73.18$, $R^2 = 0.91$). Its RMSE improved upon the baseline (naïve lag-1: RMSE $= 118.00$, MAE $= 94.10$, $R^2 = 0.91$) by an absolute margin of 25.92 units (i.e., $92.08 - 118.00 = -25.92$). The second-best result was Lasso (all features) (RMSE $= 93.30$, MAE $= 74.30$, $R^2 = 0.90$), 1.22 units above XGBoost (all features). The best model trained on the selected-feature set was Random Forest (selected features) (RMSE $= 93.70$, MAE $= 73.60$, $R^2 = 0.91$), 1.62 units above the top performer.

Across paired comparisons (same algorithm with vs. without feature selection), all-features variants generally matched or outperformed their selected-feature counterparts, with two notable exceptions. For Random Forest, the selected-feature model outperformed the all-features version (RMSE $= 93.70$ vs. 97.01), whereas for Ridge the selected-feature model was markedly superior to the all-features variant (RMSE $= 122.00$ vs. 1364.57; $R^2 = 0.84$ vs. 0.03). For the remaining algorithms the all-features configurations were equal or better: XGBoost (selected features) (RMSE $= 98.30$) vs. XGBoost (all features) (92.08), Lasso (selected features) (95.10) vs. Lasso (all features) (93.30), MARS (selected features) (97.50) vs. MARS (all features) (97.13), SVM (RBF) (selected features) (117.00) vs. SVM (RBF) (all features) (113.71), and Decision Tree (selected features) (111.00) vs. Decision Tree (all features) (104.12).

Ranking by RMSE on the test set was therefore led by XGBoost (all features) (92.08), followed by Lasso (all features) (93.30) and Random Forest (selected features) (93.70). Ensemble methods (boosting and bagging) and regularized linear models showed the strongest performance overall, whereas the single decision tree and SVM variants yielded higher errors. The naïve baseline produced RMSE $= 118.00$ and MAE $= 94.10$ (both worse than the top

models) with $R^2 = 0.91$. Absolute test metrics for every model and the baseline are summarized in Table 3.

Table 4: Test set performance (2023) for all models (selected features vs. all features) and the naïve baseline.

Model	Feature set	RMSE	MAE	R^2
XGBoost	All	92.08	73.18	0.91
Lasso	All	93.30	74.30	0.90
Random Forest	Selected	93.70	73.60	0.91
Lasso	Selected	95.10	76.40	0.90
Random Forest	All	97.01	75.32	0.90
MARS	All	97.13	75.57	0.91
MARS	Selected	97.50	75.60	0.91
XGBoost	Selected	98.30	78.30	0.90
Decision Tree	All	104.12	82.54	0.88
Decision Tree	Selected	111.00	86.30	0.86
SVM (RBF)	All	113.71	93.97	0.88
SVM (RBF)	Selected	117.00	97.60	0.88
Naive lag1 (baseline)	Baseline	118.00	94.10	0.91
Ridge	Selected	122.00	90.20	0.84
Ridge	All	1364.57	149.36	0.03

Notes: "Selected features" denotes the multi-signal feature set; "All features" denotes the full original feature set. Metrics were computed on the 2023 held-out test set. Using elastic columns (one `Z` for text and three `Y` for numbers) distributes width more evenly across columns.

4. Discussion

4.1. Principal Findings

The machine learning models demonstrated high out-of-sample accuracy in predicting avoidable hospitalizations at the Primary Care Team level. The best-performing model was the XGBoost regressor using the full set of features, which achieved a test RMSE of 92.08 admissions per 100,000 (MAE 73.18) and $R^2 = 0.91$. This outperformed a naive baseline by an absolute RMSE margin of 25.92 (baseline RMSE = 118.00; MAE 94.10). It means that the XGBoost model reduced prediction error by $\approx 22\%$ compared to carrying forward last year's rate. The second-best model was a Lasso regression with all features (RMSE 93.30, $R^2 = 0.90$), only 1.22 points higher in RMSE. The top model using a feature-selected subset was

the Random Forest, with RMSE 93.70 ($R^2 = 0.91$), about 1.6 points above the XGBoost benchmark. Thus, several approaches (boosted trees, penalized linear regression, and bagged trees) yielded comparably strong results, all explaining roughly 90% of the variance in next-year ACSC rates. In contrast, less flexible models like a single decision tree or SVM were markedly less accurate (e.g. decision tree RMSE 104–111; SVM RMSE 113–117). All top models comfortably outperformed the naive lag-1 baseline in error terms, although the baseline's R^2 was also $\approx 91\%$ due to high autocorrelation in team-level rates.

Comparing models trained on all 581 predictors with those using the pruned, multi-signal subset shows a consistent pattern. For most algorithms, access to the full feature set yielded equal or lower RMSE; for instance, the full XGBoost model reduced RMSE by approximately 6% relative to its feature-selected counterpart, while Lasso gained about 2%. Two exceptions are informative: Random Forest and Ridge performed better with feature reduction, indicating greater susceptibility to overfitting or noise in the high-dimensional setting. Hence, although the full specification offers a modest accuracy edge for the strongest learners, a carefully curated subset performs nearly as well (often within $\approx 2\%$ in RMSE) and can outperform in specific cases, suggesting diminishing returns beyond the most informative predictors.

These results imply a three-way trade-off. First, *accuracy*: the all-features models achieve the lowest errors in several instances. Second, *interpretability*: trimming predictors substantially improves transparency, shifting models from many diffuse, low-magnitude effects toward a tractable set of salient drivers. Third, *computational cost*: reducing dimensionality lowers latency and resource use. Concretely, XGBoost required 1,170 trees on the full set versus 448 with selection, paired with a much smaller learning rate and extensive subsampling in the full setting; Random Forest evaluated far more candidate splits per node with all features (e.g., $mtry = 255$ vs. 155), and tolerated smaller terminal nodes, both of which increase training

and inference burden. These adjustments reveal that the full-feature models must be heavily regularized to generalize, whereas selected-feature models can admit greater structural complexity (e.g., deeper trees) without comparable overfitting risk.

Stability considerations reinforce this view. The collapse of Ridge with all features (near-zero test R^2) versus its reasonable performance after selection illustrates how correlated, high-dimensional inputs can overwhelm certain estimators even under regularization. By focusing on predictors consistently important across folds (multi-signal selection), the reduced models likely improve out-of-sample robustness and attenuate spurious patterns.

In sum, including all available predictors confers a small accuracy premium at the expense of interpretability, computational efficiency, and (for some algorithms) stability. For implementation, decision-makers should weigh these marginal accuracy gains against the benefits of parsimony: leaner models are more transparent, faster to retrain and deploy, and potentially more portable across teams and periods, while retaining nearly all the predictive performance that matters for planning.

4.2. Interpreting model drivers

The inclusion of heterogeneous predictors spanning patient demographics, clinical history, service use, provider resources, quality indicators, and administrative/geographic attributes enabled the models to uncover multiple drivers of avoidable hospitalizations. SHAP analyses quantified each feature's contribution to predictions and revealed a concentrated structure of importance: the top 25 individual features accounted for $\approx 55\%$ of total mean absolute SHAP, indicating that a relatively small subset of strong predictors dominated model behavior. These top features drew from several domains, underscoring the multifactorial nature of ACSCs. Service-utilization signals were especially prominent: 10 of the top 25 features related to

healthcare use (e.g., medication prevalence, visit volumes, appointment supply), together contributing $\approx 24\%$ of the total. The single most influential predictor was *prescriptions per person* ($\approx 10.8\%$), and a guideline-adherence measure (*% recommended antihypertensive medications*) also appeared among the top features, suggesting that primary care processes and quality have measurable associations with ACSC outcomes.

Provider-side and organizational factors were also relevant, though each individual variable tended to have a modest share. *Public pharmaceutical expenditure per assigned patient* ranked within the top ten ($\approx 1.9\%$), and *social workers' expenditures* appeared among the top twenty, pointing to the salience of resource endowment. In addition, an administrative identifier—the Primary Care Team (PCT) code—was the third most important feature ($\approx 5\%$ of total SHAP), indicating persistent between-team differences not explained by observable covariates. Functionally, the PCT code acts as a unit-level fixed effect, proxying for unmeasured attributes (e.g., organizational culture, management practices, contextual factors, or residual population mix). Its high importance shows that, even after controlling for numerous measured predictors, some teams systematically exhibit higher or lower ACSC rates, consistent with latent performance or contextual effects. Substantively, this suggests that observable inputs do not fully capture performance gaps and motivates comparative audits of practice style, governance, and operational processes in high- versus low-performing teams to identify actionable drivers.

Aggregating importance by domain over the full feature space further clarifies these patterns. *Service utilization* features contributed $\approx 31.2\%$ of total SHAP, *medical history/clinical record* $\approx 23.7\%$, and *quality* $\approx 18.5\%$; *provider resources* accounted for $\approx 10.6\%$ and *administrative/geographic* $\approx 6.1\%$. In contrast, *accessibility* ($\approx 0.9\%$), *population and socioeconomic* ($\approx 1.0\%$), and *longitudinality* ($\approx 0.6\%$) added only marginal predictive signal.

Collectively, these distributions indicate that demand-side morbidity and utilization patterns, alongside care quality and resources, drive most of the explained variation, whereas access and continuity—as measured here—offer limited additional discrimination in this setting. This suggests that within the relatively integrated Catalan system (with universal coverage and geographically organized teams), variations in physical access or continuity (at least as quantified by available metrics) did not substantially differentiate ACSC rates. It may be that access to primary care is uniformly high (little variation in, say, travel time to clinics across urban vs. rural teams in this dataset), or that other variables (like utilization rates) already captured the effects of access and continuity indirectly. Consequently, the model implies that improving access or continuity alone might yield limited gains relative to other factors, although one should interpret this cautiously since a lack of importance in the model does not mean these factors are unimportant in absolute terms.

Overall, the interpretability analysis supports a coherent narrative: sicker, more intensively treated populations are associated with higher predicted ACSC rates; primary care processes and resource levels may attenuate or amplify that baseline risk; and residual, team-specific factors remain important. These findings are associative rather than causal. For example, a high prescriptions-per-capita value likely reflects underlying morbidity rather than a direct causal pathway from prescribing volume to admissions. Accordingly, the model's explanations should be used to guide investigation and targeted management actions, not as direct causal levers.

4.3. Implications for health system management and planning

Predicting avoidable hospitalizations at the Primary Care Team (PCT) level has direct implications for performance management, resource allocation, and equitable planning. First,

model-based *expected* ACSC rates enable risk-adjusted benchmarking and contract design. Rather than comparing raw admission rates—which penalise teams serving sicker populations—authorities can contrast observed outcomes with model-predicted values to identify over- and under-performance and to set realistic, risk-adjusted improvement targets. This approach operationalises accountable care with equity: teams that outperform their expected rate are recognised, while those falling short are flagged for support and review.

Second, forecasts support proactive resource allocation and targeted interventions. If a PCT is predicted to face persistently high ACSCs, managers can pre-emptively deploy care-management programmes, intensify follow-up for complex chronic patients, or reinforce nursing and social care capacity. Explanations (grouped SHAP) clarify *why* a team is high risk—for example, high chronic disease burden or process shortfalls—thereby directing action (e.g., strengthening geriatric management or continuity of care) and informing regional capacity planning when clusters of teams signal rising risk.

Third, the models facilitate equity audits and needs assessments. By incorporating socio-demographic and morbidity indicators, predictions can be decomposed to assess how much risk reflects population need versus modifiable care processes. This helps align resources with need, identify underserved areas for targeted investment, and surface potential structural issues. Scenario exploration (e.g., worsening local unemployment) can support intersectoral responses.

From an operational perspective, forecasts can be integrated into dashboards for continuous surveillance, shifting governance from retrospective review to *anticipatory* management. Deviations between observed and expected trajectories can trigger supportive visits or rapid-response measures. Crucially, explanations must accompany scores to avoid

"black-box" perceptions and to foster constructive engagement with teams.

Economic considerations are central. Building and maintaining high-dimensional models entails ongoing data engineering, computation, and analytic capacity. The case for adoption rests on avoided hospitalisations and efficiency gains from targeted interventions. Concentrating resources on predicted hot spots can be more cost-effective than blanket policies and, over time, can inform budget allocation by anticipated cost avoidance.

These are predictive—not causal—tools. They should complement judgement and formal evaluation. Use requires structured monitoring for drift, periodic recalibration and retraining (e.g., on a rolling yearly basis), and governance that reviews global SHAP profiles to ensure credible, fair use. Successful implementation depends not only on accuracy but on an ecosystem of data maintenance, transparency, user training, and feedback loops that continuously refine the tool as system conditions evolve. Finally, responsible use requires structured monitoring for drift, periodic recalibration and retraining (e.g., on a rolling yearly basis), and governance that reviews global SHAP profiles to ensure credible, fair use. Successful implementation depends not only on accuracy but on an ecosystem of data maintenance, transparency, user training, and feedback loops that refine the tool as system conditions evolve.

4.4. Strengths

This study's methodological approach was designed to maximize validity and usefulness for health management through a rigorous temporal evaluation. Models were trained on five complete years (2018–2022) and tested on a strictly later year (2023), mirroring a prospective, year-ahead forecasting scenario. An expanding-origin cross-validation scheme tuned models on successively newer folds (training up to year $Y-1$, validating on year Y), thereby aligning

model selection with the intended use case and avoiding look-ahead bias. All preprocessing steps (imputation, scaling, encoding, rare-level pooling) were estimated within training folds and then applied to validation and test data, further reducing leakage. This calendar-year design captures temporal non-stationarity and reflects how health planners would deploy the tool in practice.

A second strength is the comprehensive yet controlled feature strategy. From an initial set of 581 candidate predictors spanning clinical, utilisation, resource, quality, and contextual domains, a multi-signal selection combined three complementary criteria: a Random-Forest wrapper (Boruta), SHAP-based importance from gradient boosting, and Elastic Net sparsity. Features were retained if they were consistently selected across folds or contributed materially to cumulative SHAP importance, yielding a subset that is both predictive and interpretable. Including the lagged outcome set a realistic yardstick; against this baseline, the best model reduced error by ≈ 26 admissions per 100,000 (RMSE), demonstrating value beyond simple persistence.

The pipeline also prioritised transparent benchmarking and explanation. Each algorithm was compared to the naive lag-1 forecast on the same held-out year, anchoring performance in a decision-relevant counterfactual. Post hoc grouped SHAP analyses then quantified how domains contributed to predictions, translating complex models into actionable insights (e.g., the relative roles of service utilisation, chronic disease burden, quality indicators, and provider resources). This coupling of high performance with domain-level explanations strengthens credibility and supports governance.

Finally, the breadth of inputs is a substantive advantage for health system planning. By integrating heterogeneous information—not only patient morbidity and demographics but also

provider capacity, care processes, administrative/geographic factors, and proxies for organisational scale—the models speak to levers available to managers, not just to risk stratification. In sum, the combination of temporal validation, leakage controls, multi-signal feature selection, explicit baseline benchmarking, and grouped SHAP interpretability offers a robust, transferable template for predictive tools that are accurate, transparent, and operationally ready.

4.5. Limitations

This study acknowledges several potential limitations. A primary challenge lies in the inherent complexity of predicting health outcomes, particularly at an aggregated level, where numerous unmeasured confounding factors could influence ACSC rates. While the study includes a wide array of determinants, some unmeasured patient-level behaviours, lifestyle factors, or specific clinical details not routinely captured in administrative data could still play a role.

The reliance on routinely collected administrative data, while offering broad coverage, may also be subject to issues of data quality, completeness, or coding accuracy, which could potentially introduce biases or inaccuracies into the analysis. Although data cleaning and preprocessing steps will be rigorously applied, inherent limitations in the source data cannot be entirely mitigated.

A significant methodological challenge arises from the unique and disruptive nature of the COVID-19 pandemic. The drastic alterations in healthcare seeking behavior, access to care, and the direct impact of the virus on health outcomes during this period [30] could significantly alter the patterns of avoidable hospitalisations. This temporal variability, makes the task of data partitioning for model training and validation particularly complex. A simple random split would not capture the temporal shifts, and while temporal validation is planned,

the distinct nature of the pandemic years (2020-2022) compared to pre- and post-pandemic periods (2018-2019, 2023) might make it difficult to develop a single model that performs optimally across all timeframes. This may necessitate the development of period-specific models or dynamic modelling approaches, potentially impacting the interpretability of overall trends.

Furthermore, while the study aims to extract knowledge regarding PCT-level factors, the aggregation of patient-level data to the PCT level inherently involves a loss of individual-level granularity. This might obscure some fine-grained relationships between individual patient characteristics and ACSC risk within a given team. The generalisability of findings to other health systems outside of Catalonia, which may have different organisational structures, funding models, or population health profiles, might also be limited. Finally, the ”avoidable” nature of ACSCs implies an ideal scenario; in reality, some hospitalisations may be unavoidable due to the severity or rapid progression of conditions, even with optimal primary care. This inherent ambiguity in the definition of ACSCs could subtly influence model performance and interpretation.

Finally, these models are predictive in nature and not causal. Associations—such as links between specific primary care indicators and lower ACSC rates—must not be interpreted as evidence that changing those indicators will *cause* hospitalisations to fall. The study did not establish causal effects, and unmeasured confounding may drive both predictors and outcomes. For example, a high value on a “quality” indicator may correlate with fewer ACSCs because of a third factor (e.g., stronger management or higher patient adherence), rather than a direct causal pathway. Policy should therefore avoid simplistic “fix-the-predictor” responses; instead, model insights should guide deeper diagnostic analysis, prospective evaluation, and evidence-based interventions.

References

[1] Billings J, Zeitel L, Lukomnik J, Carey TS, Blank AE, Newman L. Impact of Socioeconomic Status on Hospital Use in New York City. Health Affairs. 1993; 10.1377/hlthaff.12.1.162

[2] Purdy S, Griffin T, Salisbury C, Sharp D. Ambulatory care sensitive conditions: terminology and disease coding need to be more specific to aid policy makers and clinicians. Public Health. 2009; 10.1016/j.puhe.2008.11.001

[3] Organisation for Economic Co-operation and Development . Health at a Glance 2023: OECD Indicators – Avoidable Hospital Admissions.. 2023;

[4] Keskimäki I, Satokangas M, Lumme S, Partanen VM, Arffman M, Manderbacka K. Are ambulatory care sensitive conditions a valid indicator for quality of primary health care?. European Journal of Public Health. 2020; . 1 citations (Crossref) [2023-11-20]10.1093/eurpub/ckaa165.461

[5] Agency for Healthcare Research and Quality (AHRQ) . Guide to Prevention Quality Indicators: Hospital Admission for Ambulatory Care Sensitive Conditions. technical report, . 2001; . Accessed 2025-08-21.

[6] World Health Organization (WHO) . Assessing health services delivery performance through ambulatory care–sensitive condition hospitalisations. who regional office review / guidance, . 2016; . Accessed 2025-08-21.

[7] Pan American Health Organization (PAHO) . Compendium of Outcome Indicators (Primary care and avoidable hospitalisations).. 2019; . Accessed 2025-08-21.

[8] The King's Fund . Emergency hospital admissions for ambulatory care-sensitive conditions.. 2020; . Accessed 2025-08-21.

[9] Observatori del Sistema de Salut de Catalunya . Àmbit d'atenció primària. Dades 2017–2023 de la *Central de Resultats*.. 2024; .

[10] Generalitat de Catalunya, Departament de Salut, Servei Català de la Salut . Resolució per la qual s'aproven les condicions generals de contractació de serveis sanitaris per a 2024. tech. rep., . 2024; . Accessed: 2025-05-20.

[11] Colls C. Un indicador para una financiación de la atención primaria más justo. AQuAS Blog; . 2019; .

[12] Caminal J, Cols M, Mundet X, Ponsá J. Las hospitalizaciones evitables por "Ambulatory Care Sensitive Conditions". Una medida de la capacidad de resolución de la atención primaria en Catalunya.. 1999; . Accessed 2025-08-21. Pilot cross-sectional study of 161 ABS assessing ACSC as an indicator of primary-care resolutive capacity.

[13] Caminal Homar J, Morales Espinoza M, Sánchez Ruiz E, Cubells Larrosa MJ, Bustins Poblet M. Hospitalizaciones prevenibles mediante una atención primaria oportuna y efectiva. Atención Primaria. 2003; . Descriptive analysis of 1,376,632 discharges (1998–1999); 8.42% of discharges classified as ACSC.10.1016/S0212-6567(03)70653-2

[14] Loenen vT, Berg v. dMJ, Westert GP, Faber MJ. Organizational aspects of primary care related to avoidable hospitalization: a systematic review. Family Practice. 2014; 10.1093/fampra/cmu053

[15] Gao J, Moran E, Li Y, Almenoff PL. Predicting potentially avoidable hospitalizations.

Medical Care. 2014; . Accessed: 2025-05-2010.1097/MLR.0000000000000041

[16] Purdy S, Griffin T, Salisbury C, Sharp D. Ambulatory Care Sensitive Conditions: Terminology and Disease Coding Need to Be More Specific to Aid Policy Makers and Clinicians. Medical Care. 2015; 10.1097/MLR.0000000000000342

[17] Busby J, Purdy S, Hollingworth W. A systematic review of the magnitude and cause of geographic variation in unplanned hospital admission rates and length of stay for ambulatory care sensitive conditions. BMC Health Services Research. 2015; 10.1186/s12913-015-0964-3

[18] Laditka JN, Laditka SB. Associations of Socioeconomic Status and Rurality with Health Care Access, Health Behavior, and Outcomes in the United States: A Systematic Review. International Journal for Equity in Health. 2020; 10.1186/s12939-020-01189-2

[19] Ansari Z, Rowe S, Ansari H, Sindall C. Small area analysis of ambulatory care sensitive conditions in Victoria, Australia. Population Health Management. 2013; 10.1089/pop.2012.0047

[20] Satokangas M, Arffman M, Antikainen H, Leyland AH, Keskimäki I. Individual and Area-level Factors Contributing to the Geographic Variation in Ambulatory Care Sensitive Conditions in Finland: A Register-based Study. Medical Care. 2021; 10.1097/MLR.0000000000001454

[21] Falster MO, Jorm LR. Sociodemographic and Health Characteristics, Rather Than Primary Care Supply, Are Major Drivers of Geographic Variation in Preventable Hospitalizations in Australia. Medical Care. 2015; 10.1097/MLR.0000000000000435

[22] Yi SE, Harish V, Gutierrez J, Ravaut M, Kornas K, Watson Tea. Predicting

hospitalisations related to ambulatory care sensitive conditions with machine learning for population health planning: derivation and validation cohort study. BMJ Open. 2022; 10.1136/bmjopen-2021-051403

[23] Hoffmann F, Icks A, Goffrier B, et al. Hospitalizations for Ambulatory Care Sensitive Conditions: A Big Data Analysis of 6.4 Million People in Germany. Health Services Research. 2022; 10.1111/1475-6773.13873

[24] Rosella LC, Hurst M, O'Neill M, et al. A study protocol for a predictive model to assess population-based avoidable hospitalization risk: Avoidable Hospitalization Population Risk Prediction Tool (AvHPoRT). Diagnostic and Prognostic Research. 2024; 10.1186/s41512-024-00165-5

[25] Servei Català de la Salut . The Health care System in Catalonia: Evolution and Strategic Orientation From the Perspective of the Catalan Health Service. tech. rep., Generalitat de Catalunya, Departament de Salut; . 2010; . Accessed: 2025-05-20.

[26] Sales-Coll M, Manjón S, Salabert J, et al. A plan to transform primary and community care at Catalonia based on a process improvement methodology. Primary Health Care Research & Development. 2024; . Accessed: 2025-05-2010.1017/S1463423624000604

[27] Servei Català de la Salut . El SISCAT: sistema sanitari integral d'utilització pública de Catalunya. `https://catsalut.gencat.cat/ca/coneix-catsalut/presentacio/model-sanitari-catala/siscat/`; . Sin fecha. Accessed: 2025-05-20.

[28] Catalunya (AQuAS) d. Q. i. A. S. dA. Programa d'analítica de dades per a la recerca i la innovació en salut (PADRIS). tech. rep., . 2017; . Accessed on May 15, 2024.

[29] Balears d. C. M. i. d. l. S. d. C. i. dA. El Sistema d'Informació dels Serveis d'Atenció Primària (SISAP). Annals de Medicina. 2023; .

[30] Moynihan R, Sanders S, Michaleff ZA, Scott AM, Clark J, Jones Mea. Impact of COVID-19 pandemic on utilisation of healthcare services: a systematic review. BMJ Open. 2021; 10.1136/bmjopen-2020-045343