

Original Paper

Predicting 30-Day Readmission Risk for Patients With Chronic Obstructive Pulmonary Disease Through a Federated Machine Learning Architecture on Findable, Accessible, Interoperable, and Reusable (FAIR) Data: Development and Validation Study

Celia Alvarez-Romero¹, MSc; Alicia Martinez-Garcia¹, PhD; Jara Ternero Vega², MSc; Pablo Díaz-Jiménez², MSc; Carlos Jiménez-Juan², MSc; María Dolores Nieto-Martín², PhD; Esther Román Villarán¹, MSc; Tomi Kovacevic^{3,4}, PhD; Darijo Bokan³, PhD; Sanja Hromis^{3,4}, PhD; Jelena Djekic Malbasa^{3,4}, PhD; Suzana Beslač³, MD; Bojan Zaric^{3,4}, PhD; Mert Gencturk⁵, MSc; A Anil Sinaci⁵, PhD; Manuel Ollero Baturone², PhD; Carlos Luis Parra Calderón¹, MSc

¹Computational Health Informatics Group, Institute of Biomedicine of Seville, Virgen del Rocío University Hospital, Consejo Superior de Investigaciones Científicas, University of Seville, Seville, Spain

²Internal Medicine Department, Virgen del Rocío University Hospital, Seville, Spain

³Institute for Pulmonary Diseases of Vojvodina, Sremska Kamenica

⁴Medical Faculty, University of Novi Sad, Novi Sad

⁵Software Research & Development and Consultancy Corporation, Ankara, Turkey

Corresponding Author:

Celia Alvarez-Romero, MSc

Computational Health Informatics Group

Institute of Biomedicine of Seville, Virgen del Rocío University Hospital

Consejo Superior de Investigaciones Científicas, University of Seville

Avda Manuel Siurot s/n

Seville

Spain

Phone: 34 955013313

Email: celia.alvarez@juntadeandalucia.es

Abstract

Background: Owing to the nature of health data, their sharing and reuse for research are limited by legal, technical, and ethical implications. In this sense, to address that challenge and facilitate and promote the discovery of scientific knowledge, the Findable, Accessible, Interoperable, and Reusable (FAIR) principles help organizations to share research data in a secure, appropriate, and useful way for other researchers.

Objective: The objective of this study was the FAIRification of existing health research data sets and applying a federated machine learning architecture on top of the FAIRified data sets of different health research performing organizations. The entire FAIR4Health solution was validated through the assessment of a federated model for real-time prediction of 30-day readmission risk in patients with chronic obstructive pulmonary disease (COPD).

Methods: The application of the FAIR principles on health research data sets in 3 different health care settings enabled a retrospective multicenter study for the development of specific federated machine learning models for the early prediction of 30-day readmission risk in patients with COPD. This predictive model was generated upon the FAIR4Health platform. Finally, an observational prospective study with 30 days follow-up was conducted in 2 health care centers from different countries. The same inclusion and exclusion criteria were used in both retrospective and prospective studies.

Results: Clinical validation was demonstrated through the implementation of federated machine learning models on top of the FAIRified data sets from different health research performing organizations. The federated model for predicting the 30-day hospital readmission risk was trained using retrospective data from 4.944 patients with COPD. The assessment of the predictive model was performed using the data of 100 recruited (22 from Spain and 78 from Serbia) out of 2070 observed (records viewed) patients during the observational prospective study, which was executed from April 2021 to September 2021. Significant accuracy

(0.98) and precision (0.25) of the predictive model generated upon the FAIR4Health platform were observed. Therefore, the generated prediction of 30-day readmission risk was confirmed in 87% (87/100) of cases.

Conclusions: Implementing a FAIR data policy in health research performing organizations to facilitate data sharing and reuse is relevant and needed, following the discovery, access, integration, and analysis of health research data. The FAIR4Health project proposes a technological solution in the health domain to facilitate alignment with the FAIR principles.

(*JMIR Med Inform* 2022;10(6):e35307) doi: [10.2196/35307](https://doi.org/10.2196/35307)

KEYWORDS

FAIR principles; research data management; clinical validation; chronic obstructive pulmonary disease; privacy-preserving distributed data mining; early predictive model

Introduction

Overview

FAIR4Health is a project that received funding from the European Union's (EU) Horizon 2020 research and innovation program under grant 824666. This project started in December 2018 and ended in November 2021. The main objective of this European project was to promote and encourage the EU health research community to apply the Findable, Accessible, Interoperable, and Reusable (FAIR) principles [1] in their data sets derived from publicly funded research initiatives through the implementation of an effective outreach strategy at the EU level, the production of a set of guidelines to set the foundations for a FAIR data certification road map, the development of an intuitive platform, and the demonstration of the potential impact on health research and health outcomes through the validation of 2 pathfinder case studies. At a high level, this project aimed to facilitate health research data sharing and reuse. This project brought together expertise from the key stakeholders involved in properly addressing this main objective: health research, data managers, medical informatics, software developers, standards, and lawyers. The FAIR4Health Consortium accounted for 17 partners from 11 EU and non-EU countries.

Despite strong concerns and challenges regarding data sharing in health research [2,3] and following efforts to distinguish between the concepts of *open data* [4,5] and *FAIR data* [6,7], it is evident that data sharing is one of the pillars of scientific progress. Cooperation between different countries and cultures is the fastest way to gather valuable knowledge and address challenges such as the current pandemic [8,9]. Given the strong global focus on scientific research and international cooperation, the adoption and implementation of a FAIR data policy in health research organizations is a strong requirement. Therefore, the implementation of FAIR data initiatives and lessons learned in the FAIRification process in the health field is paramount to support evidence-based clinical practice and research transparency in the era of big data and open research publishing [10].

The purpose of the FAIR4Health project [11] was to design a workflow [12] and develop a framework to reach the FAIRification of health research data sets addressing the relevant legal, technical, and ethical considerations and requirements of sensitive data. For that, FAIR4Health FAIRification tools were implemented and deployed in different health research performing organizations of the FAIR4Health Consortium.

Then, 2 pathfinder case studies were carried out to demonstrate the potential impact of the application of a FAIR strategy on health outcomes and health and social care research, making use of a privacy-preserving distributed data mining (PPDDM) architecture implemented on the FAIR4Health platform. The PPDDM architecture used a federated machine learning approach in which health research data do not leave its premises while the models travel between the data-hosting sites. The performance and validation of the FAIR4Health use case that was focused on the development of an early predictive model for 30-day readmission risk in patients with chronic obstructive pulmonary disease (COPD) is described in this paper.

Background

FAIR Data Principles

The aim of the FAIR data principles [1] is to ensure that data are shared in a way that enables and enhances reuse by humans and machines. Although FAIR data emerged from a workshop for the life science community, the FAIR principles are intended to be applied to data and metadata from all disciplines.

Since its formal release via the FORCE11 community [13], the FAIR data principles have been adopted by several funders and governments worldwide. The European Commission's data management guidelines were updated in 2017 to introduce the FAIR principles. Similarly, following the summit in June 2017, the European Open Science Cloud Declaration was launched [14]. In contrast, the recent staff working document proposed an implementation road map for the European Open Science Cloud [15]. These 2 relevant documents emphasize the central role of FAIR data.

FAIR principles are being adopted by a diverse range of research disciplines, such as economics, semantic web, and environment. Several groups have assessed the uptake to date and the challenges encountered. FAIR4Health [11] and other projects add to the state-of-the-art by documenting good practices and applying them to other domains, where possible, such as the medical domain.

FAIR4Health adds to the analysis and experience of the application of FAIR principles in the health research field, specifically in health research data sets on COPD.

COPD and Readmissions

COPD is a respiratory disease characterized by persistent symptoms and chronic limitation of airflow. This disease is known to be underdiagnosed even though it affects almost 10%

of adults worldwide [16] and its prevalence continues to increase with the aging of the population. The study by Mannino et al [17] showed that >50% of adults with impaired lung function were unaware that they were diagnosed with COPD [17]. COPD frequently presents itself with other comorbidities, such as cardiovascular disease, hypertension, and diabetes [18,19]. It has been shown that other comorbidities present in patients with COPD are observed at a younger age [20]. The cross-sectional studies conducted by Anechino et al [21] and Holguin et al [22] showed that 68% of patients admitted for COPD had at least one comorbidity, 16% had 2 or more comorbidities, and 30% had 4 or more comorbidities. It is also the third leading cause of death in the world [23]. This implies a significant need for the use of health services [24,25]. Therefore, the need and importance of using a FAIR strategy would facilitate data sharing and, thus, scientific discovery, in line with the objectives addressed in FAIR4Health.

Previous studies have shown that there are several risk factors associated with readmission in patients with COPD, such as significant deterioration of lung function, low oxygen saturation in pulse oximetry, decreased activity levels, comorbidities, and the absence of medication reconciliation during hospitalization [26]. Hospital readmissions usually have a negative impact on the quality of life of patients and their families and present a considerable economic burden for health care systems. Furthermore, previous findings support the recognition of high readmission risks associated with patients who have been hospitalized frequently in the past, along with other assessments that may be useful in better predicting readmission risk over the course of a patient's stay [27].

Regarding the comorbidities, it is noted that several studies agree that the greater the number of comorbidities, the greater the risk of readmission for patients with COPD [28,29]. The rate of readmission within 30 days of discharge is used on many occasions to judge the quality of hospital care received. Using data from Medicare beneficiaries, it is estimated that approximately 1 in 5 patients discharged from the hospital because of COPD are readmitted within 30 days [30,31]. A recently published study by Gershon et al [24] analyzed 252,756 individuals hospitalized for COPD and showed that the risk factors for readmission during this period were the number of previous admissions, the modified Medical Research Council dyspnea score, age, and chronic heart failure (both right and left).

Therefore, COPD is a major health problem that must be addressed and analyzed [32]. Several studies have evaluated the risk of readmission rate for these patients [30,33,34], but few studies have considered this risk for a period of 30 days. In addition, few studies have succeeded in considering the comorbidities and functional and care data of these patients. For all these reasons, the FAIR4Health pathfinder case study included the development of an early predictive model for 30-day readmission risk in patients with COPD. This study was carried out to understand the impact of these data on the rate of readmission within 30 days of discharge. Addressing these aspects, which are of high risk during the planning of hospital discharge, could help prevent readmission and develop a model

that helps predict which patients demonstrate greater frailty and, therefore, a higher risk of hospital readmission.

Goal of This Study

In this paper, the clinical validation of the FAIR4Health solution is described, including the development and selection of the most appropriate model for predicting 30-day readmission risk in patients with COPD and the assessment of such a model. This study builds upon the FAIRification of health research data sets of different health research performing organizations and a federated machine learning architecture on top of the FAIRified data sets of different organizations. The entire FAIR4Health solution was validated in real-world settings with the clinical use case described in this paper.

Methods

Study Design and Recruitment

The use case that was designed in this study to validate the FAIR4Health solution was composed of two phases: (1) a retrospective multicenter observational study, including the training of the predictive models in the FAIR4Health platform, and (2) an observational prospective study with a 30-day follow-up.

Retrospective Study

In the retrospective study, the population included patients aged >18 years diagnosed with COPD, considering that COPD-related comorbidities are observed at a younger age [20]. Patients with programmed admission in any hospital department within 30 days of discharge, patients with psychiatric disease, and patients with neurodegenerative diseases were excluded from the study. Following the clinical protocol defined in this study, this first phase covered retrospective data collection from the relevant data sources specified below.

In the first phase, which is to train the federated machine learning models, three different organizations participated with their health care (hospital, primary care, and nursing homes) and health research data sets: (1) Universite De Geneve from Switzerland provided health care data from the electronic health record (EHR) of the University Hospitals of Geneva; (2) Virgen del Rocío University Hospital as part of the Andalusian Health Service (Servicio Andaluz de Salud [SAS]) from Spain provided health care data from the EHR of the Virgen del Rocío University Hospital in Seville; and (3) Instituto Aragonés de Ciencias de la Salud and Instituto de Investigación Sanitaria Aragón from Spain provided a health research data set based on the EpiChron Cohort [20,35], a study carried out by the Instituto Aragonés de Ciencias de la Salud.

For organizations contributing with health research data sets from previous research projects, the sample size was defined by taking into account the original size of the data sets in the previous research, whereas for organizations contributing with health care data sets from the EHRs, it was defined from the number of patients fulfilling the inclusion and exclusion criteria.

The variables for the training and prediction processes were related to demographic, multimorbidity, comorbidities, polypharmacy, laboratory, and hospitalization data. The

principal dependent variable was readmission, defined as unplanned hospitalization for any cause related to COPD within 30 days of hospital discharge.

Prospective Study

Following the clinical protocol defined in the study, an observational prospective study with a 30 days follow-up was carried out after the retrospective study to assess the impact of the early predictive model by collecting data from a cohort of recruited patients. Patients aged ≥18 years with a diagnosis of COPD who were admitted to the hospital for this disease (unplanned hospitalization) and who signed the informed consent form (ICF) were included in the observational prospective study, complying with the same inclusion and exclusion criteria as described for the retrospective study.

Two health care organizations participated in the observational prospective study in which the trained predictive model was tested: (1) Internal Medicine Department of the Virgen del Rocío University Hospital in Seville as part of the Andalusian Health Service (SAS) from Spain and (2) Clinic for Obstructive Pulmonary Diseases and Acute Pneumopathies of the Institute for Pulmonary Diseases of Vojvodina (IPBV) from Serbia. In both cases, the sample size was defined by considering the number of patients admitted to the hospital during the prospective study period, thus fulfilling the inclusion and exclusion criteria.

Regarding the study variables, the same variables were collected at the time of inclusion of each patient during the prospective study as in the retrospective study. As a monitoring variable, aiming to assess the prediction performance of the model on the patient’s risk of readmission, it was analyzed whether the patient with COPD had a readmission within the 30 days of discharge.

Ethics Approval

Ethical approval for this study was obtained from all participating health research organizations based on regional regulations before involving them in the execution of the case studies (Universite De Geneve and University Hospitals of

Geneva from Switzerland: 2020-02683; Virgen del Rocío University Hospital as part of the Andalusian Health Service from Spain: 1269-M1-20; and Instituto Aragonés de Ciencias de la Salud and Instituto de Investigación Sanitaria Aragón from Spain , 1269-M1-20).

Technical and organizational measures were defined to safeguard the rights and freedoms of the data participants, including the data minimization principle. Informed consent procedures were defined, including informed consent and information sheets. A data protection officer was appointed at each data owner institution. To reinforce the appropriate coverage of these ethical aspects, at the beginning of the study, an external ethics advisory board was made up, which involved reviewing deliverables, generating reports, and performing presentations to support the FAIR4Health Consortium.

FAIRification Workflow and Tools

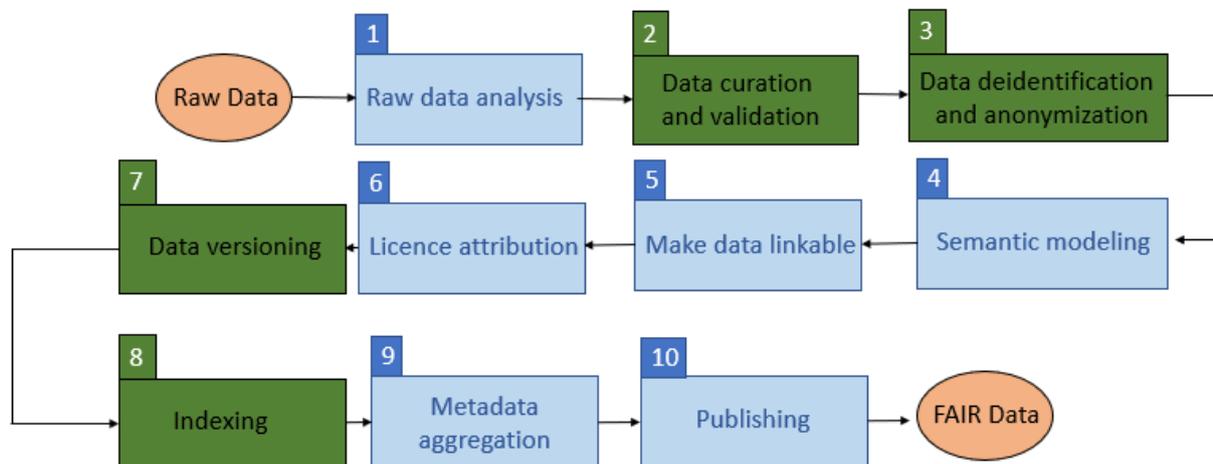
Making health data FAIR opens up new horizons, especially for the secondary use of health care and reuse of health research data sets. The FAIR4Health project proposed a FAIRification workflow [12] to be used for making existing health data sets FAIR. This workflow includes a series of actionable steps and a technological design and implementation guide for each step.

To address the challenges of the health domain, the proposed workflow adapted the generic FAIRification process defined by GO FAIR [36]. First, this workflow contextualizes the generic steps. Second, the FAIR4Health workflow introduced new steps with a strong consideration of the legal, technical, and ethical implications that reusing health data sets may have.

These steps were (1) raw data analysis, (2) data curation and validation, (3) data deidentification and anonymization, (4) semantic modeling, (5) making data linkable, (6) license attribution, (7) data versioning, (8) indexing, (9) metadata aggregation, and (10) publishing.

Steps 2, 3, 7, and 8 were newly introduced in the FAIR4Health FAIRification workflow. Figure 1 shows a visual representation of this workflow.

Figure 1. The FAIR4Health FAIRification workflow (redefined from the study by Sinaci et al [12]). FAIR: Findable, Accessible, Interoperable, and Reusable.



The FAIRification workflow was based on the HL7 Fast Healthcare Interoperability Resources (FHIR) [37]. Making data FAIR by using a well-established standard such as HL7 FHIR not only contributed to FAIRification but also helped the data owner organizations conform to a widely adopted standard. The FAIR4Health project developed a set of software tools around the HL7 FHIR as an implementation of the FAIRification workflow, the so-called FAIRification tools. In addition to the methodology and FHIR usage, these tools, namely, onFHIR.io repository [38], data curation tool (DCT) [39], and data privacy tool (DPT) [40], were deployed and used at each of the 3 data source organizations for the retrospective study. A set of FHIR profiles to serve as the common data model [41] of the FAIR4Health project was developed to cover the data requirements of the use cases. The onFHIR.io installations of the FAIR4Health project were shipped with the FAIR4Health profiles; hence, the FAIR4Health design led to uniform, interoperable, and reusable data sets once FAIRification was completed at each retrospective study organization.

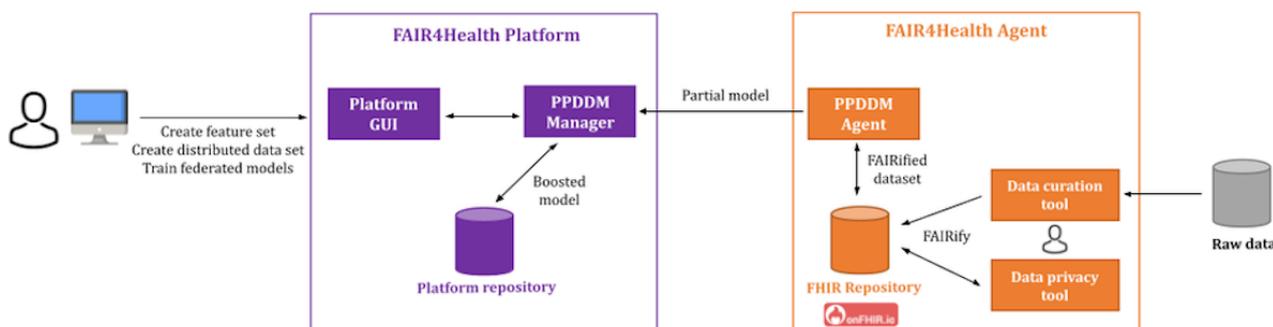
Along with the onFHIR.io repositories, at each organization, a DCT and a DPT were installed, and these tools were used by the data managers and FAIR4Health researchers to FAIRify their existing data sets, collaborating to appropriately treat the databases. Following the FAIRification workflow, the raw data were first transformed into HL7 FHIR by creating the associated FHIR resources through the DCT. It was shown that the DCT is a valid software tool that meets the challenges of raw data analysis, curation, and validation steps [42]. Once the data were migrated into the onFHIR.io repository, the DPT was used to deidentify the resources with respect to the policy requirements

of the organizations. The use of FHIR resources and the onFHIR.io repository helped us to successfully cover the other workflow steps such as versioning, indexing, and license attribution. At the end of the FAIRification process for each organization, the FAIR data were ready to be consumed by the federated machine learning algorithms so that predictive models could be built on top of the retrospective data.

Federated Machine Learning Models

The FAIR4Health project implemented the PPDDM philosophy by designing and implementing a federated machine learning architecture. The ultimate aim of this architecture is to address the challenging security and privacy concerns of health data owners. The PPDDM architecture does not allow data to leave their servers. Partial machine learning models were trained on each FAIRified data set at each organization, and then these partial models were used to develop a boosted machine learning model on the central FAIR4Health platform. The platform provides a web-based graphical user interface to the researchers so that they can define their features, create distributed data sets, and then train federated models. The PPDDM architecture was composed of the agent implementation. Then, the agents were deployed at each data source organization on top of their FAIRified data sets. These agents communicated with their associated onFHIR.io repositories at each deployment site. A manager was deployed as a backend to the FAIR4Health platform graphical user interface so that these agents can be orchestrated to build distributed data sets and federated predictive models on top of those distributed data sets. Figure 2 shows a graphical representation of the FAIR4Health federated architecture.

Figure 2. The FAIR4Health federated architecture. GUI: graphical user interface; PPDDM: privacy-preserving distributed data mining.



During the retrospective study, the researchers of the data owner organizations used the platform to train federated machine learning models on the retrospective data sets that were previously made FAIR using the FAIRification tools. The PPDDM implementation provided a set of machine learning algorithms to the researchers to be executed in a federated manner. These algorithms were grouped as (1) support vector machine, (2) logistic regression, (3) decision trees, (4) random forest, and (5) gradient-boosted trees.

Results

Model Generation and Adjustment

During the retrospective study, a number of machine learning models were generated by using the prediction algorithms listed

above as well as trying out various values for different parameters (eg, imputation strategy, classification threshold, maximum depth of a tree, and feature subset strategy). More focus was given to the tree-based algorithms because the data in the agents were skewed in one direction, and tree-based methods produced better results than the others when the data were unbalanced. In addition, k-fold cross-validation was used to split the data into a set of nonoverlapping training and test sets to obtain more accurate results.

In the experiments, better results were obtained with the predictive models generated using the random forest algorithm. An example screenshot of the platform is shown in Figure 3. While creating the model, different values for the maximum depth of a tree (range 5-15), minimum information gain

(between 0.0 and 0.5), impurity (gini or entropy), and number of trees (range 25-100) were provided. The FAIR4Health platform tried all these values with the grid search functionality to determine the best combination. Therefore, considering the knowledge of the FAIR4Health researchers with an expert background in this kind of algorithm, the best model with an accuracy of 98.6% was generated and selected with the following values:

- 3-fold cross-validation with area under the curve of the receiver operating characteristic evaluation metric

- Imputation strategy: median—replaces the missing values using the approximate median value of the feature
- Maximum depth of a tree: 5
- Minimum information gain: 0.0
- Impurity: gini
- Number of trees: 50
- Feature subset strategy: auto-calculates the number of features at each tree node as the square root of the total number of features in the classification algorithm.

Figure 3. Result of a random forest model.



Clinical Validation

After the parameters of the algorithm were selected, the predictive model was generated using retrospective data sets of 4,944 patients with COPD. Subsequently, an observational prospective study was conducted to validate and evaluate an

early predictive model for 30-day readmission risk in patients with COPD.

In total, 100 patients were recruited and included in the observational prospective study with a 30-day follow-up, from April 2021 to September 2021, including recruitment and

follow-up. During that period, the study participants were recruited by performing weekly prevalence cuts in which all patients hospitalized because of COPD conditions were systematically evaluated, offering inclusion in this study to all those who met the inclusion criteria and did not meet any exclusion criteria.

Clinicians and researchers performed functional and clinical validations of the FAIR4Health solution during the observational prospective study. As this was a multicenter observational study, the recruitment and inclusion of patients in the study were carried out as mentioned below.

For SAS, the clinical team reviewed 711 hospitalized patients during the study period, and 53 (7.5%) of them fulfilled the inclusion criteria and did not meet any exclusion criteria. Finally, 22 patients with COPD signed the ICF and were included in this observational prospective study. Out of the total recruited patients in SAS, 18% (4/22) were female and 82% (18/22) were male.

In the case of IPBV, out of 2070 hospitalized patients, 113 (5.46%) patients were hospitalized because of COPD exacerbation, and 83 (73.5%) patients met all inclusion criteria and did not meet any exclusion criteria and signed the ICF. A total of 78 patients were included in this observational prospective study.

Of the total patients recruited during the study period, 47% (37/78) were female and 53% (41/78) were male.

All data gathered from patients with COPD were entered into the FAIR4Health platform to obtain the prediction generated by the predictive model for 30-day readmission risk and to assess its performance.

Evaluation Outcomes

When the prediction was obtained using the FAIR4Health platform, a concordance analysis was performed to compare the real data with the predicted values. Concerning the reality of readmissions among the 100 patients recruited, in both cases, the patients were followed up during hospitalization, and the follow-up was performed during the following 30 days. Out of a total of 22 patients recruited from SAS, 3 (14%) were readmitted within 30 days of discharge (ie, during the follow-up period). Out of a total of 78 patients recruited from IPBV, 10 (15%) were readmitted during the follow-up period. Finally, from the 100 recruited patients, (1) the accuracy of predictions generated by the FAIR4Health platform was confirmed in 87% (87/100) of the cases; that is, either the patient was readmitted to the hospital because of COPD in real life and the algorithm predicted that there was early 30-day hospital readmission risk or the patient was not readmitted and the algorithm predicted that there was no early 30-day hospital readmission risk and (2) the prediction generated was not confirmed in 13% (13/100) of the cases; that is, in real life, the patient was readmitted within 30 days and the platform predicted that there was no early 30-day hospital readmission risk or the patient was not readmitted and the platform predicted that there was early 30-day hospital readmission.

Discussion

Principal Findings

The application of the FAIR principles in health research data sets of health research performing organizations from different countries allowed the federated data analysis to accelerate the discovery of scientific outputs. Therefore, the analysis of legal, technical, and ethical requirements of health research data were addressed during data FAIRification. Furthermore, a clinical decision support model for predicting 30-day readmission risk in patients with COPD at discharge based on the risk factors uncovered previously, using data mining approaches, was implemented, deployed, and validated. Finally, through a multicenter study in which the rate of readmission of patients with COPD within 30 days after hospital discharge was analyzed, clinical partners could reach use case objectives and obtain an early 30-day hospital readmission risk predictive model. Further details of the FAIR4Health pathfinder case studies can be found in the FAIR4Health public report on the demonstrators' performance [43].

It is important to highlight that the FAIR4Health solution was implemented following a practical extensibility capacity, so that other research questions can be covered using the solution without the need to perform adaptations. Furthermore, to improve the reusability capacity of the study, using both the open-source code and the generated metadata freely available in GitHub [44], the study can be reproduced.

Limitations

First, significant cross-cutting data-related challenges were addressed during data collection. Data extraction from EHRs and other types of health care sources aligning this extraction with a FAIR4Health common data model was not trivial and required a lot of conceptual and technical efforts because of (1) the complexity of the raw data (the sources of EHRs are commonly very complex including the information in several tables in the source databases), (2) free text used in some fields in the raw data sources, and (3) differences between the types of raw data sources. To address the complexity of the raw data, each health research organization from different countries that participated in the data extraction involved colleagues who were experts in each source data model. To handle the information in free text fields, natural language processing techniques were assessed. Finally, in some cases, manual natural language processing to extract structured information from unstructured information was performed. To manage the differences between the nature of the raw data sources, each raw data set was analyzed in depth in a collaborative effort between each clinical partner and the technical partners to reach the required configuration in the FAIR4Health solution, achieving the FAIRification of all raw data and finally achieving the PPDDM models' generation using all sources.

Second, concerning the predictive model generated in this study, it can be stated that it is possible to generate more efficient prediction parameters (with better accuracy, precision, and recall values) if the distribution of the readmission variable in the data sets is better adjusted. The readmission variable, which was the dependent variable, was not balanced in the data sets of the

retrospective studies (data sets used to generate the predictive model for this prospective study), which resulted in the generated results being good but not perfect as desired. For more effective models, in the future, a better adjustment of the distribution of the readmission variable using data sets with more patients will be addressed to boost the application of predictive models in clinical practice. Most studies of predictive models based on machine learning show poor methodological quality and are at a high risk of bias. The small study size, poor management of missing data, and failure to address overfitting are factors that contribute to the risk of bias [45].

In contrast, it is crucial to add that this study was carried out while these 2 health care organizations were experiencing the consequences of the COVID-19 pandemic, and clinical researchers had to make significant efforts to properly conclude the prospective study:

- IPBV as a health care institution was included in the national COVID-19 system of health care institutions caring for COVID-19 positive patients with severe clinical difficulties. Owing to this reorganization of the Serbian health care system, the likelihood of hospitalization of patients with COPD has been reduced since March 2020. Many of the researchers responsible for patient recruitment in the prospective study were engaged in COVID-19 departments, and the remaining researchers were overworked during the study period.
- On the side of SAS, this health care institution was involved in the care of patients with suspicion of COVID-19 and COVID-19-positive patients with severe clinical difficulties. All health professionals in SAS had a higher workload in health care. In fact, different clinical researchers participating in this observational study were transferred during the project to the COVID-19 Emergency Hospital in Seville (Spain), relieving each other, with an essential health care priority and looking after patients who did not meet the inclusion criteria of this study and could not be recruited. The clinical researchers identified a low use of health care services (both urgencies and consultancies) by patients with COPD; presumably, the patients waited for more severe symptoms to go to the health care centers because of the fear of having contact with COVID-19-positive patients. In addition, hospitalizations of patients with COPD were restricted, similar to what has happened in other pathologies, to avoid patient flow through health care centers.

Next Steps

Considering the final version of the FAIR4Health solution and the main outcomes of this study, some future advances can be taken into account:

- Both the FAIRification tools and the FAIR4Health platform were validated using the FAIR4Health common data model. The solution has been designed and developed by considering the extensive capacity of other data models, so it is appropriate to continue the validation and testing with other data models in future clinical validations.
- The whole FAIR4Health solution covers alignment with relevant standards: HL7 FHIR, International Classification

of Diseases, SNOMED Clinical Terms, Logical Observation Identifiers Names and Codes, and the Anatomical Therapeutic Chemical classification system. Other standards such as other HL7 standards, epidemiological standards, and W3C standards could be considered to be integrated if viable.

- The FAIR4Health platform was validated using the following machine learning algorithms: frequent pattern growth, support vector machine, logistic regression, decision trees, random forest, and gradient-boosted trees. Deep learning algorithms such as neural networks can be considered in future studies to improve the capabilities of the FAIR4Health platform.

From a scientific point of view, some researchers of the FAIR4Health Consortium contribute to the application of the FAIR principles in the health research field, being involved in international working groups part of the European Open Science Cloud, the European Federation for Medical Informatics, the Research Data Alliance, the GO FAIR initiative, and HL7 International.

Conclusions

Despite the limitations mentioned above, the objective of this study was achieved: to validate the FAIR4Health solution through the assessment of a federated model that was generated by applying a federated machine learning architecture on top of the FAIRified data sets of different health research performing organizations for real-time prediction of 30-day readmission risk in patients with COPD.

The clinical, technical, and functional validation of the FAIR4Health solution was achieved through (1) the application of FAIR principles through the FAIR4Health FAIRification tools in health research data sets of different health research performing organizations and FAIRifying data from 4,944 patients with COPD; (2) development and use of federated machine learning architecture on top of the FAIRified data sets; and (3) clinical, technical, and functional development and assessment of a federated model for predicting 30-day readmission risk in patients with COPD, with an accuracy of 0.98, a precision of 0.25, and a confirmed prediction in 87% (87/100) of the cases.

In the retrospective study where 3 different organizations participated with their health care (hospital, primary care, and nursing homes) and health research data sets, the federated model was generated with an accuracy of 98.6% and a precision of 25%. In the observational prospective study in which 2 health care organizations participated, 100 patients were recruited for the federated model to predict their readmission risk to the hospital within 30 days because of COPD. Therefore, the accuracy of predictions generated by the model, and hence the FAIR4Health platform, was confirmed in 87% (87/100) of the cases.

Health research performing organizations are aware of the need to implement a FAIR data policy to facilitate data sharing and reuse following the discovery, access, integration, and analysis of health research data. One obvious example would be the COVID-19 pandemic, where international cooperation allowed

the rapid sequencing and epidemiological studies to be carried out, thus demonstrating the need and importance of data sharing to accelerate health research [46,47]. For this purpose, organizations are usually making efforts to align themselves with the FAIR principles. This is the real and practical consequence of the FAIR4Health project in terms of patient management and health planning: to improve health research in specific pathologies through the findability-, accessibility-,

interoperability-, and reusability-enhanced features in the case of health data.

The FAIR4Health project proposes a technological solution in the health domain to facilitate the use of larger and more heterogeneous data sets, thus increasing the variability of the data and the size of the data sets. Therefore, an increase in the scope of the research will be obtained and a significant improvement in the ability to generate more accurate predictive models.

Acknowledgments

This work was supported by the FAIR4Health project [10], which received funding from the European Union's Horizon 2020 research and innovation program under grant 824666. This research has also been cosupported by the Carlos III National Institute of Health through the Programa de Ciencia de Datos de la Infraestructura de Medicina de Precisión asociada a la Ciencia y la Tecnología program (IMPACT-Data, code IMP/00019) and through the Platform for Dynamization and Innovation of the Spanish National Health System industrial capacities and their effective transfer to the productive sector (code PT20/00088), both cofunded by the European Regional Development Fund Fondo Europeo de Desarrollo Regional "A way of making Europe." The authors would like to thank the clinical researchers of the project, coming from the organizations that are part of the FAIR4Health Consortium: Universite De Geneve (Switzerland), University Hospitals of Geneva (Switzerland), Università Cattolica Del Sacro Cuore (Italy), Universidade Do Porto (Portugal), Instituto Aragonés de Ciencias de la Salud (Spain), Institut Za Plucne Bolesti Vojvodine (Serbia), and Servicio Andaluz de Salud (Spain).

Conflicts of Interest

None declared.

References

1. Wilkinson M, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 2016 Mar 15;3:160018-160019 [FREE Full text] [doi: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18)] [Medline: [26978244](https://pubmed.ncbi.nlm.nih.gov/26978244/)]
2. Parra-Calderón CL, Sanz F, McIntosh LD. The challenge of the effective implementation of FAIR principles in biomedical research. *Methods Inf Med* 2020 Aug 22;59(4-05):117-118. [doi: [10.1055/s-0040-1721726](https://doi.org/10.1055/s-0040-1721726)] [Medline: [33618419](https://pubmed.ncbi.nlm.nih.gov/33618419/)]
3. Delgado J, Llorente S. Security and privacy when applying FAIR principles to genomic information. *Stud Health Technol Inform* 2020 Nov 23;275:37-41. [doi: [10.3233/SHTI200690](https://doi.org/10.3233/SHTI200690)] [Medline: [33227736](https://pubmed.ncbi.nlm.nih.gov/33227736/)]
4. Dijkers MP. A beginner's guide to data stewardship and data sharing. *Spinal Cord* 2019 Mar 5;57(3):169-182. [doi: [10.1038/s41393-018-0232-6](https://doi.org/10.1038/s41393-018-0232-6)] [Medline: [30723254](https://pubmed.ncbi.nlm.nih.gov/30723254/)]
5. Couture JL, Blake RE, McDonald G, Ward CL. A funder-imposed data publication requirement seldom inspired data sharing. *PLoS One* 2018 Jul 6;13(7):e0199789 [FREE Full text] [doi: [10.1371/journal.pone.0199789](https://doi.org/10.1371/journal.pone.0199789)] [Medline: [29979709](https://pubmed.ncbi.nlm.nih.gov/29979709/)]
6. Almada M, Midão L, Portela D, Dias I, Núñez-Benjumea FJ, Parra-Calderón CL, et al. [A new paradigm in health research: FAIR data (Findable, Accessible, Interoperable, Reusable)]. *Acta Med Port* 2020 Dec 02;33(12):828-834 [FREE Full text] [doi: [10.20344/amp.12910](https://doi.org/10.20344/amp.12910)] [Medline: [33496252](https://pubmed.ncbi.nlm.nih.gov/33496252/)]
7. Holub P, Kohlmayer F, Prasser F, Mayrhofer MT, Schlünder I, Martin GM, et al. Enhancing reuse of data and biological material in medical research: from FAIR to FAIR-health. *Biopreserv Biobank* 2018 Apr;16(2):97-105 [FREE Full text] [doi: [10.1089/bio.2017.0110](https://doi.org/10.1089/bio.2017.0110)] [Medline: [29359962](https://pubmed.ncbi.nlm.nih.gov/29359962/)]
8. Mello MM, Lieou V, Goodman SN. Clinical trial participants' views of the risks and benefits of data sharing. *N Engl J Med* 2018 Jun 07;378(23):2202-2211 [FREE Full text] [doi: [10.1056/NEJMs1713258](https://doi.org/10.1056/NEJMs1713258)] [Medline: [29874542](https://pubmed.ncbi.nlm.nih.gov/29874542/)]
9. Rios R, Zheng KI, Zheng MH. Data sharing during COVID-19 pandemic: what to take away. *Expert Rev Gastroenterol Hepatol* 2020 Dec;14(12):1125-1130. [doi: [10.1080/17474124.2020.1815533](https://doi.org/10.1080/17474124.2020.1815533)] [Medline: [32842793](https://pubmed.ncbi.nlm.nih.gov/32842793/)]
10. Inau E, Sack J, Waltemath D, Zeleke AA. Initiatives, concepts, and implementation practices of FAIR (findable, accessible, interoperable, and reusable) data principles in health data stewardship practice: protocol for a scoping review. *JMIR Res Protoc* 2021 Feb 02;10(2):e22505 [FREE Full text] [doi: [10.2196/22505](https://doi.org/10.2196/22505)] [Medline: [33528373](https://pubmed.ncbi.nlm.nih.gov/33528373/)]
11. FAIR4Health key outputs for the scientific community. FAIR4Health. URL: <https://www.fair4health.eu/> [accessed 2022-05-11]
12. Sinaci A, Núñez-Benjumea FJ, Gencturk M, Jauer ML, Deserno T, Chronaki C, et al. From raw data to FAIR data: the FAIRification workflow for health research. *Methods Inf Med* 2020 Jun;59(S 01):e21-e32 [FREE Full text] [doi: [10.1055/s-0040-1713684](https://doi.org/10.1055/s-0040-1713684)] [Medline: [32620019](https://pubmed.ncbi.nlm.nih.gov/32620019/)]
13. The FAIR data principles. FORCE11. URL: <https://www.force11.org/group/fairgroup/fairprinciples> [accessed 2022-05-11]
14. EOSC Declaration. URL: https://eosc-portal.eu/sites/default/files/eosc_declaration.pdf [accessed 2022-05-11]

15. European Open Science Cloud (EOSC) Strategic Implementation Plan. European Commission. URL: https://ec.europa.eu/info/publications/european-open-science-cloud-eosc-strategic-implementation-plan_en [accessed 2022-05-11]
16. Adeloye D, Chua S, Lee C, Basquill C, Papan A, Theodoratou E, Global Health Epidemiology Reference Group (GHERG). Global and regional estimates of COPD prevalence: systematic review and meta-analysis. *J Glob Health* 2015 Dec;5(2):020415 [FREE Full text] [doi: [10.7189/jogh.05-020415](https://doi.org/10.7189/jogh.05-020415)] [Medline: [26755942](https://pubmed.ncbi.nlm.nih.gov/26755942/)]
17. Mannino D, Gagnon R, Petty T, Lydick E. Obstructive lung disease and low lung function in adults in the United States: data from the National Health and Nutrition Examination Survey, 1988-1994. *Arch Intern Med* 2000 Jun 12;160(11):1683-1689. [doi: [10.1001/archinte.160.11.1683](https://doi.org/10.1001/archinte.160.11.1683)] [Medline: [10847262](https://pubmed.ncbi.nlm.nih.gov/10847262/)]
18. Baty F, Putora P, Isenring B, Blum T, Brutsche M. Comorbidities and burden of COPD: a population based case-control study. *PLoS One* 2013;8(5):e63285 [FREE Full text] [doi: [10.1371/journal.pone.0063285](https://doi.org/10.1371/journal.pone.0063285)] [Medline: [23691009](https://pubmed.ncbi.nlm.nih.gov/23691009/)]
19. Divo M, Cote C, de Torres JP, Casanova C, Marin JM, Pinto-Plata V, et al. Comorbidities and risk of mortality in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2012 Jul 15;186(2):155-161. [doi: [10.1164/rccm.201201-0034oc](https://doi.org/10.1164/rccm.201201-0034oc)]
20. Divo MJ, Celli BR, Poblador-Plou B, Calderón-Larrañaga A, de-Torres JP, Gimeno-Feliu LA, EpiChron—BODE Collaborative Group. Chronic Obstructive Pulmonary Disease (COPD) as a disease of early aging: evidence from the EpiChron Cohort. *PLoS One* 2018 Feb 22;13(2):e0193143 [FREE Full text] [doi: [10.1371/journal.pone.0193143](https://doi.org/10.1371/journal.pone.0193143)] [Medline: [29470502](https://pubmed.ncbi.nlm.nih.gov/29470502/)]
21. Anecchino C, Rossi E, Fanizza C, De Rosa M, Tognoni G, Romero M, working group ARNO project. Prevalence of chronic obstructive pulmonary disease and pattern of comorbidities in a general population. *Int J Chron Obstruct Pulmon Dis* 2007;2(4):567-574 [FREE Full text] [Medline: [18268930](https://pubmed.ncbi.nlm.nih.gov/18268930/)]
22. Holguin F, Folch E, Redd SC, Mannino DM. Comorbidity and mortality in COPD-related hospitalizations in the United States, 1979 to 2001. *Chest* 2005 Oct;128(4):2005-2011. [doi: [10.1378/chest.128.4.2005](https://doi.org/10.1378/chest.128.4.2005)] [Medline: [16236848](https://pubmed.ncbi.nlm.nih.gov/16236848/)]
23. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012 Dec 15;380(9859):2095-2128. [doi: [10.1016/S0140-6736\(12\)61728-0](https://doi.org/10.1016/S0140-6736(12)61728-0)] [Medline: [23245604](https://pubmed.ncbi.nlm.nih.gov/23245604/)]
24. Gershon A, Thiruchelvam D, Aaron S, Stanbrook M, Vozoris N, Tan W, et al. Socioeconomic status (SES) and 30-day hospital readmissions for chronic obstructive pulmonary (COPD) disease: a population-based cohort study. *PLoS One* 2019;14(5):e0216741 [FREE Full text] [doi: [10.1371/journal.pone.0216741](https://doi.org/10.1371/journal.pone.0216741)] [Medline: [3112573](https://pubmed.ncbi.nlm.nih.gov/3112573/)]
25. About gold. Global Initiative for Chronic Obstructive Lung Disease. URL: <http://goldcopd.org/> [accessed 2022-05-11]
26. Coventry P, Gemmell I, Todd C. Psychosocial risk factors for hospital readmission in COPD patients on early discharge services: a cohort study. *BMC Pulm Med* 2011 Nov 04;11:49 [FREE Full text] [doi: [10.1186/1471-2466-11-49](https://doi.org/10.1186/1471-2466-11-49)] [Medline: [22054636](https://pubmed.ncbi.nlm.nih.gov/22054636/)]
27. Jiang W, Siddiqui S, Barnes S, Barouch LA, Korley F, Martinez DA, et al. Readmission risk trajectories for patients with heart failure using a dynamic prediction approach: retrospective study. *JMIR Med Inform* 2019 Sep 16;7(4):e14756 [FREE Full text] [doi: [10.2196/14756](https://doi.org/10.2196/14756)] [Medline: [31579025](https://pubmed.ncbi.nlm.nih.gov/31579025/)]
28. Brand C, Sundararajan V, Jones C, Hutchinson A, Campbell D. Readmission patterns in patients with chronic obstructive pulmonary disease, chronic heart failure and diabetes mellitus: an administrative dataset analysis. *Intern Med J* 2005 May;35(5):296-299. [doi: [10.1111/j.1445-5994.2005.00816.x](https://doi.org/10.1111/j.1445-5994.2005.00816.x)] [Medline: [15845113](https://pubmed.ncbi.nlm.nih.gov/15845113/)]
29. Kelly M. Self-management of chronic disease and hospital readmission: a care transition strategy. *J Nursing Healthcare Chronic Illness* 2011;3(1):4-11. [doi: [10.1111/j.1752-9824.2010.01075.x](https://doi.org/10.1111/j.1752-9824.2010.01075.x)]
30. The Lancet Respiratory Medicine. Reducing COPD readmissions—a personal and political priority. *Lancet Respiratory Med* 2013 Jul;1(5):347. [doi: [10.1016/s2213-2600\(13\)70153-x](https://doi.org/10.1016/s2213-2600(13)70153-x)]
31. Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the medicare fee-for-service program. *N Engl J Med* 2009 Apr 02;360(14):1418-1428. [doi: [10.1056/nejmsa0803563](https://doi.org/10.1056/nejmsa0803563)]
32. Vos T, Allen C, Arora M, Barber R, Bhutta Z, Brown A, et al. . [doi: [10.1016/S0140-6736\(16\)31678-6](https://doi.org/10.1016/S0140-6736(16)31678-6)]
33. Chan F, Wong F, Yam C, Cheung W, Wong E, Leung M, et al. Risk factors of hospitalization and readmission of patients with COPD in Hong Kong population: analysis of hospital admission records. *BMC Health Serv Res* 2011 Aug 10;11:186 [FREE Full text] [doi: [10.1186/1472-6963-11-186](https://doi.org/10.1186/1472-6963-11-186)] [Medline: [21831287](https://pubmed.ncbi.nlm.nih.gov/21831287/)]
34. Jacobs DM, Noyes K, Zhao J, Gibson W, Murphy TF, Sethi S, et al. Early hospital readmissions after an acute exacerbation of chronic obstructive pulmonary disease in the nationwide readmissions database. *Annals ATS* 2018 Jul;15(7):837-845. [doi: [10.1513/annalsats.201712-913oc](https://doi.org/10.1513/annalsats.201712-913oc)]
35. Prados-Torres A, Poblador-Plou B, Gimeno-Miguel A, Calderón-Larrañaga A, Poncel-Falcó A, Gimeno-Feliú LA, et al. Cohort profile: the epidemiology of chronic diseases and multimorbidity. The EpiChron cohort study. *Int J Epidemiol* 2018 Apr 01;47(2):382-34f [FREE Full text] [doi: [10.1093/ije/dyx259](https://doi.org/10.1093/ije/dyx259)] [Medline: [29346556](https://pubmed.ncbi.nlm.nih.gov/29346556/)]
36. GO FAIR initiative. GO FAIR. URL: <https://www.go-fair.org> [accessed 2022-05-11]
37. Welcome to FHIR. HL7 FHIR. URL: <http://hl7.org/fhir/> [accessed 2022-05-11]
38. HL7 FHIR® Based Secure Data Repository. onFHIR.io. URL: <https://onfhir.io> [accessed 2022-05-11]
39. FAIR4Health data curation and validation tool. GitHub. URL: <https://github.com/fair4health/data-curation-tool> [accessed 2022-05-11]

40. FAIR4Health data privacy tool. GitHub. URL: <https://github.com/fair4health/data-privacy-tool> [accessed 2022-05-11]
41. fair4health / common-data-model. GitHub. URL: <https://github.com/fair4health/common-data-model> [accessed 2022-05-11]
42. Gencturk M, Teoman A, Alvarez-Romero C, Martinez-Garcia A, Parra-Calderon CL, Poblador-Plou B, et al. End user evaluation of the FAIR4Health data curation tool. *Stud Health Technol Inform* 2021 May 27;281:8-12. [doi: [10.3233/SHTI210110](https://doi.org/10.3233/SHTI210110)] [Medline: [34042695](https://pubmed.ncbi.nlm.nih.gov/34042695/)]
43. D5.5. Report on the demonstrators performance_v2_vf.pdf. OSF Home. URL: <https://osf.io/tfnqa/> [accessed 2022-05-11]
44. FAIR4Health. GitHub. URL: <https://github.com/fair4health/> [accessed 2022-05-11]
45. Andaur Navarro CL, Damen JA, Takada T, Nijman SW, Dhiman P, Ma J, et al. Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. *BMJ* 2021 Oct 20;375:n2281 [FREE Full text] [doi: [10.1136/bmj.n2281](https://doi.org/10.1136/bmj.n2281)] [Medline: [34670780](https://pubmed.ncbi.nlm.nih.gov/34670780/)]
46. Kinsella C, Santos PD, Postigo-Hidalgo I, Folgueiras-González A, Passchier TC, Szillat KP, et al. Preparedness needs research: how fundamental science and international collaboration accelerated the response to COVID-19. *PLoS Pathog* 2020 Oct;16(10):e1008902 [FREE Full text] [doi: [10.1371/journal.ppat.1008902](https://doi.org/10.1371/journal.ppat.1008902)] [Medline: [33035262](https://pubmed.ncbi.nlm.nih.gov/33035262/)]
47. Besançon L, Peiffer-Smadja N, Segalas C, Jiang H, Masuzzo P, Smout C, et al. Open science saves lives: lessons from the COVID-19 pandemic. *BMC Med Res Methodol* 2021 Jun 05;21(1):117-118 [FREE Full text] [doi: [10.1186/s12874-021-01304-y](https://doi.org/10.1186/s12874-021-01304-y)] [Medline: [34090351](https://pubmed.ncbi.nlm.nih.gov/34090351/)]

Abbreviations

COPD: chronic obstructive pulmonary disease
DCT: data curation tool
DPT: data privacy tool
EHR: electronic health record
EU: European Union
FAIR: Findable, Accessible, Interoperable, and Reusable
FHIR: Fast Healthcare Interoperability Resources
ICF: informed consent form
IPBV: Institute for Pulmonary Diseases of Vojvodina
PPDDM: privacy-preserving distributed data mining
SAS: Servicio Andaluz de Salud

Edited by C Lovis; submitted 30.11.21; peer-reviewed by H Abaza, JJ Mira; comments to author 26.12.21; revised version received 16.03.22; accepted 21.04.22; published 02.06.22

Please cite as:

Alvarez-Romero C, Martinez-Garcia A, Ternero Vega J, Díaz-Jiménez P, Jiménez-Juan C, Nieto-Martín MD, Román Villarán E, Kovacevic T, Bokan D, Hromis S, Djekic Malbasa J, Beslac S, Zaric B, Gencturk M, Sinaci AA, Ollero Baturone M, Parra Calderón CL

Predicting 30-Day Readmission Risk for Patients With Chronic Obstructive Pulmonary Disease Through a Federated Machine Learning Architecture on Findable, Accessible, Interoperable, and Reusable (FAIR) Data: Development and Validation Study

JMIR Med Inform 2022;10(6):e35307

URL: <https://medinform.jmir.org/2022/6/e35307>

doi: [10.2196/35307](https://doi.org/10.2196/35307)

PMID:

©Celia Alvarez-Romero, Alicia Martinez-Garcia, Jara Ternero Vega, Pablo Díaz-Jiménez, Carlos Jiménez-Juan, María Dolores Nieto-Martín, Esther Román Villarán, Tomi Kovacevic, Darijo Bokan, Sanja Hromis, Jelena Djekic Malbasa, Suzana Beslac, Bojan Zaric, Mert Gencturk, A Anil Sinaci, Manuel Ollero Baturone, Carlos Luis Parra Calderón. Originally published in JMIR Medical Informatics (<https://medinform.jmir.org/>), 02.06.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <https://medinform.jmir.org/>, as well as this copyright and license information must be included.