



Article scientifique

Article

2022

Published version

Open Access

This is the published version of the publication, made available in accordance with the publisher's policy.

---

## Applying the FAIR4Health solution to identify multimorbidity patterns and their association with mortality through a frequent pattern growth association algorithm

---

Carmona-Pírez, Jonás; Poblador-Plou, Beatriz; Poncel-Falcó, Antonio; Rochat, Jessica; Alvarez-Romero, Celia; Martínez-García, Alicia; Angioletti, Carmen; Almada, Marta; Gencturk, Mert; Sinaci, A Anil; Ternero-Vega, Jara Eloisa; Gaudet-Blavignac, Christophe; Lovis, Christian; Liperoti, &nbsp;Rosa [and 5 more]

### How to cite

CARMONA-PÍREZ, Jonás et al. Applying the FAIR4Health solution to identify multimorbidity patterns and their association with mortality through a frequent pattern growth association algorithm. In: International journal of environmental research and public health, 2022, vol. 19, n° 4, p. 2040. doi: 10.3390/ijerph19042040

This publication URL: <https://archive-ouverte.unige.ch//unige:167055>

Publication DOI: [10.3390/ijerph19042040](https://doi.org/10.3390/ijerph19042040)



Article

# Applying the FAIR4Health Solution to Identify Multimorbidity Patterns and Their Association with Mortality through a Frequent Pattern Growth Association Algorithm

Jonás Carmona-Pírez <sup>1,2,3,4,\*</sup>, Beatriz Poblador-Plou <sup>1,2,4,†</sup>, Antonio Poncel-Falcó <sup>1,2,4,5</sup>, Jessica Rochat <sup>6,7</sup>, Celia Alvarez-Romero <sup>8</sup>, Alicia Martínez-García <sup>8</sup>, Carmen Angioletti <sup>9</sup>, Marta Almada <sup>10</sup>, Mert Gencturk <sup>11</sup>, A. Anil Sinaci <sup>11</sup>, Jara Eloisa Ternero-Vega <sup>12</sup>, Christophe Gaudet-Blavignac <sup>6,7</sup>, Christian Lovis <sup>6,7</sup>, Rosa Liperoti <sup>9</sup>, Elisio Costa <sup>10</sup>, Carlos Luis Parra-Calderón <sup>8</sup>, Aida Moreno-Juste <sup>1,2,4,5</sup>, Antonio Gimeno-Miguel <sup>1,2,4,‡</sup> and Alexandra Prados-Torres <sup>1,2,4,‡</sup>



**Citation:** Carmona-Pírez, J.; Poblador-Plou, B.; Poncel-Falcó, A.; Rochat, J.; Alvarez-Romero, C.; Martínez-García, A.; Angioletti, C.; Almada, M.; Gencturk, M.; Sinaci, A.A.; et al. Applying the FAIR4Health Solution to Identify Multimorbidity Patterns and Their Association with Mortality through a Frequent Pattern Growth Association Algorithm. *Int. J. Environ. Res. Public Health* **2022**, *19*, 2040. <https://doi.org/10.3390/ijerph19042040>

Academic Editor: Paul B. Tchounwou

Received: 13 January 2022  
Accepted: 10 February 2022  
Published: 11 February 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

- <sup>1</sup> EpiChron Research Group, Aragon Health Sciences Institute (IACS), IIS Aragón, Miguel Servet University Hospital, 50009 Zaragoza, Spain; bpoblador.iacs@aragon.es (B.P.-P.); aponcel@salud.aragon.es (A.P.-F.); aidamorenog@gmail.com (A.M.-J.); agimenomi.iacs@aragon.es (A.G.-M.); sprados.iacs@aragon.es (A.P.-T.)
  - <sup>2</sup> Health Services Research on Chronic Patients Network (REDISSEC), ISCIII, 28029 Madrid, Spain
  - <sup>3</sup> Delicias-Sur Primary Care Health Centre, Aragon Health Service (SALUD), 50009 Zaragoza, Spain
  - <sup>4</sup> Red de Investigación en Cronicidad, Atención Primaria y Promoción de la Salud (RICAPPS), ISCIII, 28029 Madrid, Spain
  - <sup>5</sup> Aragon Health Service (SALUD), 50017 Zaragoza, Spain
  - <sup>6</sup> Division of Medical Information Sciences, Geneva University Hospitals, 1205 Geneva, Switzerland; Jessica.Rochat@hcuge.ch (J.R.); Christophe.Gaudet-Blavignac@hcuge.ch (C.G.-B.); christian.lovis@unige.ch (C.L.)
  - <sup>7</sup> Department of Radiology and Medical Informatics, University of Geneva, 1205 Geneva, Switzerland
  - <sup>8</sup> Group of Research and Innovation in Biomedical Informatics, Biomedical Engineering and Health Economy, Institute of Biomedicine of Seville (IBiS), Virgen del Rocío University Hospital/CSIC/University of Seville, 41013 Seville, Spain; celia.alvarez@juntadeandalucia.es (C.A.-R.); alicia.martinez.garcia@juntadeandalucia.es (A.M.-G.); carlos.parra.sspa@juntadeandalucia.es (C.L.P.-C.)
  - <sup>9</sup> Department of Geriatric and Orthopedic Sciences, Catholic University of Sacred Heart, 00168 Rome, Italy; carmen.angioletti92@gmail.com (C.A.); rosa.liperoti@policlinicogemelli.it (R.L.)
  - <sup>10</sup> Ucbio Requite, Faculty of Pharmacy, University of Porto, Porto4Ageing, 4050-313 Porto, Portugal; martassalmada@gmail.com (M.A.); emcosta@ff.up.pt (E.C.)
  - <sup>11</sup> SRDC Software Research & Development and Consultancy Corporation, Ankara 06800, Turkey; mert@srdc.com.tr (M.G.); anil@srdc.com.tr (A.A.S.)
  - <sup>12</sup> Internal Medicine Department, Virgen del Rocío University Hospital, 41013 Seville, Spain; jaraeloisa@hotmail.com
- \* Correspondence: jcarmona@iisaragon.es; Tel.: +34-976-765500 (ext. 5371/5375)  
† These authors contributed equally to this work.  
‡ These authors contributed equally to this work.

**Abstract:** The current availability of electronic health records represents an excellent research opportunity on multimorbidity, one of the most relevant public health problems nowadays. However, it also poses a methodological challenge due to the current lack of tools to access, harmonize and reuse research datasets. In FAIR4Health, a European Horizon 2020 project, a workflow to implement the FAIR (findability, accessibility, interoperability and reusability) principles on health datasets was developed, as well as two tools aimed at facilitating the transformation of raw datasets into FAIR ones and the preservation of data privacy. As part of this project, we conducted a multicentric retrospective observational study to apply the aforementioned FAIR implementation workflow and tools to five European health datasets for research on multimorbidity. We applied a federated frequent pattern growth association algorithm to identify the most frequent combinations of chronic diseases and their association with mortality risk. We identified several multimorbidity patterns clinically plausible and consistent with the bibliography, some of which were strongly associated with mortality. Our results show the usefulness of the solution developed in FAIR4Health to overcome the difficulties in data management and highlight the importance of implementing a FAIR data policy to accelerate responsible health research.

**Keywords:** FAIR principles; multimorbidity; mortality; research data management; pathfinder case study; privacy-preserving distributed data mining

---

## 1. Introduction

Chronic conditions are responsible for most health problems in older people [1], in which multimorbidity, or coexistence of multiple chronic diseases, is the norm rather than the exception. There is growing scientific evidence that chronic conditions tend to cluster into specific non-random disease patterns, commonly referred to as multimorbidity patterns [2]. Nonetheless, most health systems and clinical practice guidelines are still designed to respond to specific diseases independently. Consequently, the needs of people with multimorbidity are often not adequately met by health care services and health professionals, resulting in avoidable negative effects on health and healthcare costs for patients and health systems, respectively [3]. Multimorbidity is commonly followed by polypharmacy or the prescription and use of multiple (five or more) medications by the patient. Multimorbidity and polypharmacy are associated with increased mortality and cognitive impairment and decreased quality of life and functional ability [3–5].

Different initiatives have tried to face the challenge of managing multimorbidity in clinical practice in recent years. Some of them are conceptual and propose new models of care based on the comprehensive assessment of the patient and shared decision making, such as the Ariadne principles [6] or the Integrated Multimorbidity Care Model [7], which have even been tested in clinical trials and real-life situations. Another feasible and less expensive option for obtaining scientific evidence on multimorbidity is to carry out studies based on real-world data. The current availability of health data, such as those contained in patients' electronic health records (EHR) represents an excellent opportunity for health research. However, it also presents some limitations, such as difficulties in performing cross-regional studies due to the lack of interoperability among datasets, problems related to data access and privacy and other challenges associated with the current lack of tools to harmonize and integrate health datasets.

In this context, the FAIR4Health project was born as a Horizon 2020 project aimed at facilitating and encouraging the European health research community to share, reuse and integrate publicly-funded research datasets [8] based on the FAIR (findability, accessibility, interoperability and reusability) principles that serve to guide scientific data management and stewardship [9]. A "FAIRification Workflow" to apply the FAIR principles to EHR and other health research data sources was designed and implemented [10] by adapting the GO FAIR process to health data's legal, ethical and technical requirements. In addition, a common data model was defined to allow federated data analysis through the integration of datasets from various health research organizations. Furthermore, different software tools were developed in FAIR4Health to implement the FAIRification Workflow. Among them, the Data Curation Tool [11,12] served to integrate raw health datasets by transforming them into FAIR datasets, whereas the Data Privacy Tool [13] preserved data privacy through data de-identification and anonymization methods. On top of these, the FAIR4Health Platform was developed to provide a set of services for the researchers in a user-friendly interface, aimed at allowing the application of federated machine learning algorithms on FAIR datasets.

To demonstrate the potential usefulness for health research of this FAIRification strategy, two pathfinder case studies were performed by using federated machine learning algorithms implemented upon the FAIR4Health Platform. In this paper, we present the results obtained in the FAIR4Health pathfinder case study aimed at identifying multimorbidity patterns in older adults from different healthcare settings and analyzing their impact on mortality through a frequent pattern growth association algorithm.

## 2. Materials and Methods

We conducted a multicentric retrospective observational study that included five European cohorts from different healthcare settings (i.e., hospital, primary care, nursing homes) and health research organizations: Université de Genève (UNIGE, Switzerland), Università Cattolica del Sacro Cuore (UCSC, Italy), University of Porto (UP, Portugal), Instituto Aragonés de Ciencias de la Salud (IACS, Spain) and Andalusian Health Service (SAS, Spain). Each of these organizations provided a database from a publicly funded research project for the purposes of the study.

UNIGE provided anonymized healthcare data from the EHR of the University Hospitals of Geneva, the largest one in Switzerland, and met the needs of around 0.5 M residents ( $n = 244$ ). UCSC provided health data from two research studies carried out within the SHELTER project [14,15] that aimed to implement a tool to assess and collect data about nursing home residents ( $n = 331$ ). UP provided a health research dataset based on the FRAILSURVEY study [16], which aimed to test the reliability of the FRAILSURVEY phone app for self-assessment of frailty in older adults ( $n = 861$ ). IACS provided a health research dataset based on the EpiChron Cohort [17], which investigates the clinical epidemiology and health outcomes associated with chronic diseases and multimorbidity in the Spanish region of Aragon ( $n = 3786$ ). SAS provided health care data from the EHR of Virgen del Rocío University Hospital of Seville, one of the biggest hospitals in Spain, and covering a population of more than 0.5 M inhabitants ( $n = 5812$ ).

The study population included patients over 65 years of age with multimorbidity (i.e., at least two chronic diseases). During the study, researchers from each institution carried out a secondary use of retrospectively collected data using federated machine learning algorithms. Ethical approval for this study was obtained in all countries based on local regulations (UNIGE, 2020-02683; UCSC, 1066/20-12/05/2020; UP, PARECER A-13/2020; SAS, 1269-M1-20; and IACS, 1269-M1-20).

### 2.1. Study Variables

UNIGE, UCSC, IACS and SAS shared a similar data structure and information on the same disease variables, allowing to analyze the four datasets together and, in this way, identifying multimorbidity patterns. We studied the following variables for each patient at cohort entry: age, gender, nationality, smoking status, institutionalization, polypharmacy (i.e., use of  $\geq 5$  drugs), number of prescribed drugs for those with polypharmacy and the presence of 47 chronic baseline conditions registered in patients' EHR. The selection of the 47 conditions analyzed was based on clinical consensus. Additionally, SAS and IACS datasets were used to analyze the impact of multimorbidity on mortality at six months due to their shared structure regarding this outcome.

The UP dataset included the following variables: age, gender, nationality, memory complaints, vision/hearing difficulties, unintentional weight loss, feeling depressed lately, feeling anxious lately, Groningen Frailty Index Frailty Score and domiciliary care. Therefore, it was analyzed independently to identify multimorbidity patterns.

### 2.2. FAIRification Workflow and Tools Developed

FAIR4Health extended the FAIRification process adopted by the GO FAIR initiative [18] for the health domain, considering specific technical, ethical and legal requirements. As a result, a FAIRification Workflow [10] consisting of the following 10 steps was introduced: (1) raw data analysis, (2) data curation & validation, (3) data de-identification & anonymization, (4) semantic modeling, (5) making data linkable, (6) license attribution, (7) data versioning, (8) indexing, (9) metadata aggregation and (10) publishing.

In order to achieve the objectives of the FAIRification Workflow, specific software tools were developed and utilized to enable data managers to make their raw health research data FAIR. The Data Curation Tool [11,12] represented the entry point to the FAIRification workflow. Its main goal is to annotate clinical datasets with medical terminologies and define mappings to the FAIR4Health Common Data Model [19], implemented following

the HL7 FHIR profiling approach [20]. This tool wrote data into a HL7 FHIR repository instance by processing the raw source data according to the mapping rules defined by the data manager. It has been proven to be an effective tool that meets the fundamental requirements of raw data analysis, curation and validation [21]. On the other hand, the Data Privacy Tool [13] was responsible for handling the privacy challenges on sensitive health data by applying several data de-identification and anonymization techniques. Once the curation process was finished, data managers used this tool to de-identify data before making it available to other systems/components as FAIR data.

The high-performance secure health data repository onFHIR.io [22], which is totally compliant with the HL7 FHIR specifications, was utilized as the HL7 FHIR Repository. This repository stored and maintained the data made FAIR according to the FAIR4Health Common Data Model, satisfying the objectives of the FAIRification workflow, such as licensing, versioning, indexing and publishing.

In order to show the potential impact of the FAIRification strategy, a Privacy-Preserving Distributed Data Mining (PPDDM) architecture was implemented to build machine learning models in a federated way. The architecture consisted of two main components: the aforementioned FAIR4Health Platform and the FAIR4Health Agents, which were a suite of software applications installed locally at each participating sites' own system that not only provide the FAIRification tools, but also host the PPDDM Agent responsible for running machine learning algorithms on FAIR data and exchanging the results with the FAIR4Health Platform. Thus, no health data were shared among participating sites or with the FAIR4Health Platform.

The FAIR4Health Platform, on the other hand, provided a set of services with an elaborate Graphical User Interface (GUI) on top in charge of interacting with the agents and orchestrating the whole process.

### 2.3. Analysis

We applied a federated frequent pattern growth association (FP-Growth) algorithm [23], used for mining association rules, to identify the most frequent patterns among the set of variables studied. FP-Growth is an efficient, scalable and fast algorithm implemented by Han et al. [23] for mining frequent patterns, especially when the size of data and/or the number of variables are large. We implemented the FP-Growth algorithm in a federated manner in line with the PPDDM objectives so that no real data were shared between the participants. Given a dataset in a number of agents, the association rules were identified in two steps.

In the first step, each PPDDM agent calculated the item frequencies on their own data through the construction of a FP-tree and sent the results to the PPDDM Manager in the FAIR4Health Platform. The PPDDM Manager merged the results of all agents, found frequencies at the global level, and removed the ones below a minimum threshold value referred to as support that ranged from 0.0 to 1.0. For a disease, support could also be considered as its prevalence. For example, if an item appeared in 5 out of 10 records, it had a support of  $5/10 = 0.5$ . We considered 0.3 as the default minimum support value. The lower the minimum support, the more variables were included in the next step.

In the second step, the PPDDM Manager sent the global itemset to each agent and asked agents to find association rules containing items from this itemset. For each item, the conditional FP-tree was built, and the association rules were generated. Then, the confidence (i.e., how often an association was observed in the dataset) was calculated for each rule. For instance, if itemset X appeared five times, and X and Y appeared together three times, the confidence for the association rule  $X \geq Y$  was  $3/5 = 0.6$ .

Once association rules were generated, they were sent to the PPDDM Manager. Similar to the process in the first step, the PPDDM Manager combined the association rules retrieved from all the participating agents, calculated the global confidence values, and eliminated the ones having a confidence lower than the minimum value allowed, which was 0.8 by default. As a result, the remaining association rules constituted the frequent

patterns discovered in the datasets. The FAIR4Health Platform presented the association rules in an “Antecedent  $\geq$  Consequent” format, as shown in Figure 1.

**FP Growth**

**Minimum support:** 0.3

**Minimum confidence:** 0.6

Association boosted model table

Antecedent	Consequent	Confidence	Lift
Gender_male			
Heartfailure			
Hyperlipidemia	Diabetesmellitus	0.656	1.663
Hypertension			
Polypharmacy			

**Figure 1.** Example of presentation of the association rules in the FAIR4Health Platform.

The “antecedent” column represented the left-hand side of an association rule, while the “consequent” column represented the right-hand side. The “confidence” column defined the probability that a patient had the “consequent” conditions given that he/she already had the “antecedent” ones. For example, the association rule shown in Figure 1 indicates that a male patient with heart failure, hyperlipidemia, hypertension and polypharmacy presents a likelihood of suffering diabetes mellitus of 65.6% (0.656). For an association rule  $A \geq C$ , confidence was calculated as the ratio between patients having A and C, and patients having A.

The “lift” or correlation, on the other hand, indicates whether having the “antecedent” conditions (A) actually increases the probability to have the “consequent” conditions (C). It was calculated as the ratio between the confidence of  $A \geq C$  and support of C, and could be interpreted as a measure of the strength of association. In cases where “A” actually led to “C” (i.e., positive correlation), the lift value was greater than 1. In other words, the greater the lift value, the more likely the patient to have “C” given that he/she already has “A”. However, if the confidence was high but the value of lift was less than 1 (i.e., negative correlation), then we concluded that having “A” for a patient did not increase the likelihood of presenting “C”. In the example in Figure 1, the lift value of 1.663 indicates a strong association.

Following this methodology, different models were created adjusting the minimum support or frequency a variable should have to be included (i.e., the prevalence in the case of a disease), and the minimum confidence or frequency an association rule should have to be presented. The lower the minimum support and minimum confidence were, the higher the number of diseases and associations included was. As a result, hundreds of rules for each model can be obtained. However, only those association rules with positive correlation and highest clinical relevance and confidence will be presented in the results section in order to show the potentiality of the tools developed.

### 3. Results

The demographic characteristics of the 11,034 individuals included in the datasets used in the study are summarized in Table 1. The mean age of the population was 82.1 years and women represented almost 51% of the individuals studied.

**Table 1.** Demographic characteristics of the populations from the five agents included in the study.

Institutions	Population (n, %)	Age (Mean)	Sex, Women (%)
Université de Genève	244 (2.2)	81.8	47.1
Università Cattolica del Sacro Cuore	331 (3.0)	95.5	71.6
University of Porto	861 (7.8)	76.6	57.5
Instituto Aragonés de Ciencias de la Salud	3786 (34.3)	82.1	49.9
Andalusian Health Service	5812 (52.7)	82.2	49.4
Total	11,034 (100)	82.1	50.8

### 3.1. Identification of Multimorbidity Patterns

The most frequently identified chronic conditions were cardio-metabolic (i.e., diabetes mellitus, hyperlipidemia, hypertension and obesity), cardiovascular (i.e., heart failure and chronic kidney disease) and mental (i.e., depression and anxiety). The most relevant multimorbidity patterns found, based on combinations of the parameters used in the models, are presented in Table 2.

**Table 2.** Multimorbidity patterns found in the study population based on the selected combinations of parameters used in the models.

Parameters Used		Generated Patterns			Institutions Providing Datasets in Each Model	
Minimum Support	Minimum Confidence	Antecedent (A)	Consequent (C)	Confidence		Correlation (Lift)
0.2	0.5	Atrial fibrillation Chronic anemia Chronic kidney disease Coronary heart disease Hypertension Polypharmacy	Heart failure	0.86	2.80	UNIGE, UCSC, IACS, and SAS
0.2	0.5	Atrial fibrillation Chronic anemia Chronic kidney disease Coronary heart disease Diabetes Mellitus Heart failure Hyperlipidemia Polypharmacy	Hypertension	1.00	1.33	UNIGE, UCSC, IACS, and SAS
0.3	0.5	Gender male Age 70–80 Feeling down or depressed lately Feeling nervous or anxious lately Memory complaints Vision difficulties	Hearing difficulties	0.909	2.52	UP
0.3	0.5	Gender male Age 80 and older Feeling down or depressed lately Feeling nervous or anxious lately Hearing difficulties Memory complaints Vision difficulties	Polymedicated	1.00	1.65	UP

UNIGE: Université de Genève; UCSC: Università Cattolica del Sacro Cuore; IACS: Instituto Aragonés de Ciencias de la Salud; SAS: Andalusian Health Service; UP: University of Porto.

The pattern with the highest strength of association (2.796) consisted in the presence of atrial fibrillation, chronic anemia, chronic kidney disease, coronary heart disease, hypertension and polypharmacy, which also resulted in the appearance of heart failure (probability of confidence, 0.86 out of 1). We also found a multimorbidity pattern consisting of a polymedicated patient with atrial fibrillation, chronic anemia, chronic kidney

disease, coronary heart disease, diabetes mellitus, heart failure and hyperlipidemia, who also develops hypertension (probability of confidence, 1; lift, 1.33).

As explained in the methodology, UP dataset was analyzed independently due to its different data structure. In this case, the pattern with the highest lift (2.52) consisted of a male patient, aged 70–80 years, feeling down or depressed and nervous or anxious lately and with memory complaints and vision difficulties, who also develops hearing difficulties (confidence, 0.91). We found some patterns with perfect confidence, such as one consisting of a male patient, aged 80 and older, feeling down or depressed and nervous or anxious lately and with hearing and vision difficulties, who was then polymedicated (lift, 1.65).

### 3.2. Impact of Multimorbidity Patterns on Mortality

The multimorbidity pattern with the highest positive correlation with mortality consisted of chronic anemia, chronic kidney disease, coronary heart disease, diabetes mellitus and heart failure, which was then associated with mortality with a probability of confidence of 0.58 out of 1 and a lift of 1.96 (Table 3).

**Table 3.** Impact of multimorbidity patterns on mortality based on the selected combinations of the parameters used in the models.

Parameters Used		Generated Patterns			
Minimum Support	Minimum Confidence	Antecedent (A)	Consequent (C)	Confidence	Correlation (Lift)
0.2	0.8	Chronic anemia Chronic kidney disease Coronary heart disease Diabetes mellitus Heart failure	Mortality	0.58	1.96
0.2	0.8	Chronic anemia Chronic kidney disease Coronary heart disease Diabetes mellitus Heart failure Hyperlipidemia Hypertension	Mortality	0.55	1.85
0.2	0.8	Chronic anemia Chronic kidney disease Coronary heart disease Diabetes mellitus Heart failure Hyperlipidemia Polypharmacy	Mortality	0.54	1.82

In patients with polypharmacy in its antecedents, the highest correlation with mortality was presented in those with chronic anemia, chronic kidney disease, coronary heart disease, diabetes mellitus, heart failure and hyperlipidemia (probability of confidence, 0.54 out of 1, correlation, 1.82).

## 4. Discussion

In this study, we tested the usefulness of the FAIR4Health solution to apply the FAIR principles in health research by developing a pathfinder case study aimed at identifying multimorbidity patterns and their impact on mortality based on a federated data analysis on five datasets from different European health research organizations using PPDDM methodologies and a frequent pattern growth algorithm.

The objectives proposed by the project’s clinical researchers were satisfactorily addressed in the pathfinder case study. Cardiometabolic and mental health patterns were



identified among the most frequent and relevant ones in our study, a result consistent with previous studies [2,24]. The systematic review by Busija et al. in 2019 [24] concluded that the only replicable and clinically meaningful multimorbidity profiles are the cardiometabolic and mental health; a previous systematic review by Prados-Torres et al. in 2014 [2], described three main multimorbidity patterns, cardiometabolic, mental health and musculoskeletal. These results largely coincide with our findings and support the existence of the multimorbidity patterns identified, besides a strong association between multimorbidity with mortality was described, demonstrating the potentiality of our FAIRification strategy on health research and, hopefully, on patients' health outcomes.

Another potentiality and novelty of our study is that we can analyze the antecedents and consequences of the patterns detected. This approach can help to identify key associations that lead to specific consequences, analyzing the clinical impact of the patterns using the diseases as the study unit. From a clinical point of view, this is relevant as it can help to develop preventive actions based on the disease associations and the frequency of their appearance. However, we must be careful about the clinical results obtained in this study that should be interpreted with caution.

The FAIRification workflow and tools developed allowed us to analyze heterogeneous datasets and to increase the variability of studied datasets (i.e., more detailed clinical, demographic, environmental and social information) compared with studies not applying FAIR principles and always in a secure way.

However, we had to face some challenges regarding data collection, which, at the same time, helped us to create cross-cutting solutions in the process. First, the data extraction of EHR and other health research data sources had to be aligned with the FAIR4Health Common Data Model and which required relevant efforts. Each participating organization in the data extraction involved experts in their source data model in tackling these problems, which improved the communication between different specialists from different areas, an essential element for research dynamic. In some cases, the application of natural language processing (NLP) techniques to handle the information in free text fields was required, developing human-machine interaction skills fundamental for this project and in future ones. Finally, to deal with the differences between the types of raw data sources (e.g., research and clinical-administrative datasets), we analyzed each source raw dataset in-depth in a collaborative effort between clinical and technical researchers. All this led us to reach the precise configuration to apply FAIR principles within the FAIR4Health solution, making all raw data FAIR and then generating PPDDM models using all data sources. A pathfinder project like this probably can help to build multidisciplinary teams essential in health research to face new challenges. The application of FAIR principles and the tools developed in this project have great potentiality in different health research contexts; they can be applied to different types of datasets and can help to answer different research questions, which can help us to guide scientific data management and drive scientific discovery to a new paradigm.

Regarding the limitations of the study, some of them were related to the association patterns obtained. It would be possible to generate more efficient association rules if we could better adjust the mortality variable distribution in our datasets, including a larger number of patients and from other regions, and, in this way, control the risk of bias. We also have to consider the computational limitations of the frequent pattern growth association algorithm applied in this study. When clinical researchers decreased the minimum confidence and support values to include diseases with low prevalence, the number of combinations increased, and the model could not get the results. To address this challenge, other types of associative methods, such as factor, cluster and network analysis [25–27] could be explored and implemented in future works.

## 5. Conclusions

The use of the FAIR4Health solution enabled us to identify multimorbidity patterns and their association with mortality in older adults using complex and heterogeneous

FAIR databases from different European countries. Our results show the potential of implementing a FAIR data policy in health research and support the usefulness of the FAIR4Health solution, encouraging the scientific community to use the tools developed to test and validate their performance in different research contexts.

**Author Contributions:** Conceptualization, A.P.-T., A.G.-M. and B.P.-P.; methodology, B.P.-P., A.G.-M., J.C.-P., J.E.T.-V., J.R., C.A.-R., A.M.-G., C.A., M.A., R.L., M.G. and A.A.S.; formal analysis, B.P.-P., J.C.-P., C.A.-R., A.M.-G., M.A., M.G., J.R. and C.G.-B.; data curation, A.P.-F., B.P.-P., C.A.-R., A.M.-G., M.A., C.A., J.R. and C.G.-B.; writing—original draft preparation, J.C.-P.; writing—review and editing, A.G.-M., J.C.-P., B.P.-P., A.P.-T., M.G., J.R., C.A.-R., A.M.-G., C.A., C.G.-B., C.L., R.L., E.C., A.A.S., A.P.-F., M.A., J.E.T.-V., A.M.-J. and C.L.P.-C.; supervision, A.P.-T. and A.G.-M.; funding acquisition, C.L.P.-C., J.C.-P. and A.P.-T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was performed in the framework of FAIR4Health, a project that has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement number 824666. Also, this research has been co-supported by the Carlos III National Institute of Health, through the IMPaCT Data project (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 co-funded by European Regional Development Fund (FEDER) ‘A way of making Europe’, and by REDISSEC (RD16/0001/0005) and RICAPPS (RD21/0016/0019) from Carlos III National Institute of Health. This work was also supported by Instituto de Investigación Sanitaria Aragón and Carlos III National Institute of Health [Río Hortega Program, grant number CM19/00164].

**Institutional Review Board Statement:** Ethical approval for this study was obtained in all countries based on local regulations (UNIGE, 2020-02683; UCSC, 1066/20-12/05/2020; UP, PARECER A-13/2020; SAS, 1269-M1-20; and IACS, 1269-M1-20).

**Informed Consent Statement:** Patient consent was waived due to nature of the study that consisted in the secondary use of anonymized data that were presented only at an aggregated level.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** This work was supported by the FAIR4Health project [8], which has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement number 824666. Also, this research has been co-supported by the Carlos III National Institute of Health, through the IMPaCT Data project (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 co-funded by European Regional Development Fund (FEDER) ‘A way of making Europe’. Special acknowledgements to the clinical researchers of the project, coming from the health research performing organizations that are part of the FAIR4Health consortium: Université de Genève (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:** The authors declare no conflict of interest.

## References

1. WHO. *WHO Global Strategy and Action Plan on Aging and Health*; WHO: Geneva, Switzerland, 2017; ISBN 9789241513500.
2. Prados-Torres, A.; Calderón-Larrañaga, A.; Hanco-Saavedra, J.; Poblador-Plou, B.; van den Akker, M. Multimorbidity patterns: A systematic review. *J. Clin. Epidemiol.* **2014**, *67*, 254–266. [[CrossRef](#)] [[PubMed](#)]
3. Barnett, K.; Mercer, S.W.; Norbury, M.; Watt, G.; Wyke, S.; Guthrie, B. Epidemiology of multimorbidity and implications for health care, research, and medical education: A cross-sectional study. *Lancet* **2012**, *380*, 37–43. [[CrossRef](#)]
4. Masnoon, N.; Shakib, S.; Kalisch-Ellett, L.; Caughey, G.E. What is polypharmacy? A systematic review of definitions. *BMC Geriatr.* **2017**, *17*, 230. [[CrossRef](#)]
5. Bradley, M.C.; Motterlini, N.; Padmanabhan, S.; Cahir, C.; Williams, T.; Fahey, T.; Hughes, C.M. Potentially inappropriate prescribing among older people in the United Kingdom. *BMC Geriatr.* **2014**, *14*, 72. [[CrossRef](#)] [[PubMed](#)]
6. Muth, C.; van den Akker, M.; Blom, J.W.; Mallen, C.D.; Rochon, J.; Schellevis, F.G.; Becker, A.; Beyer, M.; Gensichen, J.; Kirchner, H.; et al. The Ariadne principles: How to handle multimorbidity in primary care consultations. *BMC Med.* **2014**, *12*, 223. [[CrossRef](#)] [[PubMed](#)]

7. Palmer, K.; Marengoni, A.; Forjaz, M.J.; Jureviciene, E.; Laatikainen, T.; Mammarella, F.; Muth, C.; Navickas, R.; Prados-Torres, A.; Rijken, M.; et al. Multimorbidity care model: Recommendations from the consensus meeting of the Joint Action on Chronic Diseases and Promoting Healthy Ageing across the Life Cycle (JA-CHRODIS). *Health Policy* **2018**, *122*, 4–11. [[CrossRef](#)] [[PubMed](#)]
8. FAIR4Health FAIR4Health Project. Available online: <https://www.fair4health.eu/> (accessed on 12 January 2022).
9. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.W.; da Silva Santos, L.B.; Bourne, P.E.; et al. Comment: The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 1–9. [[CrossRef](#)] [[PubMed](#)]
10. Sinaci, A.A.; Núñez-Benjumea, F.J.; Gencturk, M.; Jauer, M.L.; Deserno, T.; Chronaki, C.; Cangiolli, G.; Cavero-Barca, C.; Rodríguez-Pérez, J.M.; Pérez-Pérez, M.M.; et al. From Raw Data to FAIR Data: The FAIRification Workflow for Health Research. *Methods Inf. Med.* **2020**, *59*, E21–E32. [[CrossRef](#)] [[PubMed](#)]
11. FAIR4Health Project. Data Curation Tool. Available online: <https://github.com/fair4health/data-curation-tool> (accessed on 12 January 2022).
12. Gencturk, M.; Teoman, A.; Alvarez-Romero, C.; Martinez-Garcia, A.; Parra-Calderon, C.L.; Poblador-Plou, B.; Löbe, M.; Sinaci, A.A. End user evaluation of the FAIR4Health data curation tool. In *Public Health and Informatics; Proc. MIE 2021*; IOS Press: Amsterdam, The Netherlands, 2021; pp. 8–12.
13. FAIR4Health Project Data Privacy Tool. Available online: <https://github.com/fair4health/data-privacy-tool> (accessed on 12 January 2022).
14. Onder, G.; Carpenter, I.; Finne-Soveri, H.; Gindin, J.; Frijters, D.; Henrard, J.; Nikolaus, T.; Topinkova, E.; Tosato, M.; Liperoti, R.; et al. Assessment of nursing home residents in Europe: The Services and Health for Elderly in Long TERM care (SHELTER) study. *BMC Health Serv. Res.* **2012**, *12*, 5. [[CrossRef](#)] [[PubMed](#)]
15. Onder, G.; Liperoti, R.; Fialova, D.; Topinkova, E.; Tosato, M.; Danese, P.; Gallo, P.F.; Carpenter, I.; Finne-Soveri, H.; Gindin, J.; et al. Polypharmacy in nursing home in Europe: Results from the SHELTER study. *J. Gerontol.-Ser. A Biol. Sci. Med. Sci.* **2012**, *67*, 698–704. [[CrossRef](#)] [[PubMed](#)]
16. Midao, L.; Sá, C.; Marques, E.; Duarte, M.; Paúl, C.; Viana, J.; Costa, E. Ehealth on Frailty: Frailsurvey, a Reliable Smartphone Application for Self-Assessment of Frailty. *Innov. Aging* **2019**, *3*, 336. [[CrossRef](#)]
17. Prados-Torres, A.; Poblador-Plou, B.; Gimeno-Miguel, A.; Calderón-Larrañaga, A.; Poncel-Falcó, A.; Gimeno-Feliú, L.A.; González-Rubio, F.; Laguna-Berna, C.; Marta-Moreno, J.; Clerencia-Sierra, M.; et al. Cohort Profile: The Epidemiology of Chronic Diseases and Multimorbidity. The EpiChron Cohort Study. *Int. J. Epidemiol.* **2018**, *47*, 382–384. [[CrossRef](#)] [[PubMed](#)]
18. GO FAIR Initiative. Available online: <https://www.go-fair.org/fair-principles/fairification-process/> (accessed on 12 January 2022).
19. FAIR4Health Project FAIR4Health Common Data Model. Available online: <https://github.com/fair4health/common-data-model> (accessed on 12 January 2022).
20. HL7\_FHIR HL7 FHIR. Available online: <http://hl7.org/fhir/> (accessed on 12 January 2022).
21. FAIR4Health D5.5. Report on the Demonstrators Performance; FAIR4Health Consortium. 2021. Available online: [https://www.fair4health.eu/storage/files/Resource/58/D55%20Report%20on%20the%20demonstrators%20performance\\_v2\\_vf.pdf](https://www.fair4health.eu/storage/files/Resource/58/D55%20Report%20on%20the%20demonstrators%20performance_v2_vf.pdf) (accessed on 12 January 2022).
22. OnFHIR.io Repository onFHIR.io Repository. Available online: <https://onfhir.io> (accessed on 12 January 2022).
23. Han, J.; Pei, J.; Yin, Y. Mining FrequentPatterns without Candidate Generation. *SIGMOD* **2000**, *29*, 1–12. [[CrossRef](#)]
24. Busija, L.; Lim, K.; Szoeki, C.; Sanders, K.M.; McCabe, M.P. Do replicable profiles of multimorbidity exist? Systematic review and synthesis. *Eur. J. Epidemiol.* **2019**, *34*, 1025–1053. [[CrossRef](#)] [[PubMed](#)]
25. Ioakeim-Skoufa, I.; Poblador-Plou, B.; Carmona-Pérez, J.; Díez-Manglano, J.; Navickas, R.; Gimeno-Feliu, L.A.; González-Rubio, F.; Jureviciene, E.; Dambrauskas, L.; Prados-Torres, A.; et al. Multimorbidity Patterns in the General Population: Results from the EpiChron Cohort Study. *Int. J. Environ. Res. Public Health* **2020**, *17*, 4242. [[CrossRef](#)] [[PubMed](#)]
26. Carmona-Pérez, J.; Poblador-Plou, B.; Díez-Manglano, J.; Morillo-Jiménez, M.J.; Marín Trigo, J.M.; Ioakeim-Skoufa, I.; Gimeno-Miguel, A.; Prados-Torres, A. Multimorbidity networks of chronic obstructive pulmonary disease and heart failure in men and women: Evidence from the EpiChron Cohort. *Mech. Ageing Dev.* **2021**, *193*, 111392. [[CrossRef](#)] [[PubMed](#)]
27. Carmona-Pérez, J.; Poblador-Plou, B.; Ioakeim-Skoufa, I.; González-Rubio, F.; Gimeno-Feliú, L.A.; Díez-Manglano, J.; Laguna-Berna, C.; Marin, J.M.; Gimeno-Miguel, A.; Prados-Torres, A. Multimorbidity clusters in patients with chronic obstructive airway diseases in the EpiChron Cohort. *Sci. Rep.* **2021**, *11*, 4784. [[CrossRef](#)] [[PubMed](#)]