Journal of Clinical Epidemiology 185 (2025) 111867

Journal of Clinical Epidemiology

ORIGINAL RESEARCH

Clusters of post-acute COVID-19 symptoms: a latent class analysis across 9 databases and 7 countries

Kim López-Güell[a,#], Martí Català[a,#], Daniel Dedman[b], Talita Duarte-Salles[c,d], Raivo Kolde[e], Raúl López-Blasco[f], Álvaro Martínez[g], Gregoire Mercier[h,i], Alicia Abellan[c], Johnmary T. Arinze[d], Theresa Burkard[a], Edward Burn[a], Zara Cuccu[b], Antonella Delmestri[a], Dominique Delseny[h], Sara Khalid[a], Chungsoo Kim[j], Ji-woo Kim[k], Kristin Kostka[a,l], Cora Loste[m,n,o,p], Miguel A. Mayer[q], Jaime Meléndez-Cardiel[f], Núria Mercadé-Besora[a,c], Mees Mosseveld[d], Akihito Nishimura[r], Hedvig ME. Nordeng[s,t], Jessie O. Oyinlola[b], Roger Paredes[m,n,p,u,v,w,x], Laura Pérez-Crespo[c], Marta Pineda-Moncusí[a], Juan Manuel Ramírez-Anguita[q], Nhung T.H. Trinh[s], Anneli Uusküla[y], Bernardo Valdivieso[g,z], Daniel Prieto-Alhambra[a,d,*], Junqing Xie[a], Lourdes Mateu[m,n,o,p,v,†], Annika M. Jödicke[a,†]

[a]*Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK*
[b]*CPRD, Medicines and Healthcare products Regulatory Agency, London, UK*
[c]*Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain*
[d]*Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands*
[e]*Institute of Computer Science, University of Tartu, Tartu, Estonia*
[f]*Biocomputing Unit, Aragon Health Sciences Institute (IACS), Zaragoza, Spain*
[g]*The Health Research Institute Hospital La Fe, Avenida Fernando Abril Martorell, 106 Torre A 7a planta, Valencia 46026, Spain*
[h]*Public Health Department, University Hospital of Montpellier, Montpellier 34295, France*
[i]*IDESP, Université de Montpellier, INSERM, Montpellier 34000, France*
[j]*Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, Republic of Korea*
[k]*Big Data Department, Health Insurance Review and Assessment Service, Wonju, Republic of Korea*
[l]*The OHDSI Center at the Roux Institute, Northeastern University, Portland, ME, USA*
[m]*Department of Infectious Diseases & irsiCaixa AIDS Research Institute, Hospital Universitari Germans Trias i Pujol, Badalona, Catalonia, Spain*
[n]*University of Vic- Central University of Catalonia (UVic-UCC), Vic, Spain*
[o]*REICOP (Red de Investigación Covid Persistente), Madrid, Spain*
[p]*Fundació Lluita Contra les Infeccions, Badalona, Catalonia, Spain*
[q]*Parc de Salut Mar, Hospital del Mar Medical Research Institute, Barcelona, Spain*
[r]*Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA*
[s]*Department of Pharmacy, Pharmacoepidemiology and Drug Safety Research Group, Faculty of Mathematics and Natural Sciences, University of Oslo, Oslo, Norway*
[t]*Department of Child Health and Development, Norwegian Institute of Public Health, Oslo, Norway*
[u]*Department of Pathology, Center for Global Health and Diseases, Case Western Reserve University School of Medicine, Cleveland, OH*
[v]*Universitat Autònoma de Barcelona, Catalonia, Spain*

Funding: This work has been funded by the European Health Data Evidence Network through an Evidence Generation Fund Grant. The European Health Data Evidence Network has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 806968. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation program and EFPIA. The project was supported by the National Institute for Health and Care Research Oxford Biomedical Research Centre. Daniel Prieto-Alhambra is funded through a National Institute for Health and Care Research Senior Research Fellowship (Grant number SRF-2018-11-ST2-004). The views expressed in this publication are those of the author(s) and not necessarily those of the National Institute for Health and Care Research or the Department of Health and Social Care. Kim López-Güell was supported by the Medical Research Council (grant number MR/W006731/1) and Bayer AG. Raivo Kolde was supported by the European Regional Development Fund (RITA 1/02-120) and the Estonian Research Council grant (PRG1844). Laura Pérez-Crespo was supported by a Sara Borrell fellowship awarded by the Spanish Institute of Health Carlos III (CD23/00223).

Joint First Author.
† Joint Senior Author.
* Corresponding author. University of Oxford Nuffield, Department of Orthopaedics Rheumatology and Musculoskeletal Sciences, Headington Botnar Research Centre, Windmill Road, Oxford, UK.
E-mail address: daniel.prietoalhambra@ndorms.ox.ac.uk (D. Prieto-Alhambra).

https://doi.org/10.1016/j.jclinepi.2025.111867

[w]CIBER Infectious Diseases (CIBERINFEC), Institute of Health Carlos III (ISCIII), Madrid, Spain
[x]IrsiCaixa AIDS Research Institute, Germans Trias i Pujol Research Institute (IGTP), Can Ruti Campus, Badalona, Catalonia, Spain
[y]Institute of Family Medicine and Public Health, University of Tartu, Tartu, Estonia
[z]The University and Polytechnic La Fe Hospital of Valencia, Avenida Fernando Abril Martorell, 106 Torre H 1a planta, Valencia 46026, Spain

Accepted 10 June 2025; Published online 13 June 2025

Abstract

Objective: Prior evidence has suggested the multisystem symptomatic manifestations of post-acute COVID-19 condition (PCC). Here we conducted a network cluster analysis of 24 World Health Organization−proposed symptoms to identify potential latent subclasses of PCC.

Study Design and Setting: Individuals with a positive test of or diagnosed with SARS-CoV-2 after September 2020 and with at least 1 symptom within ≥90 to 365 days following infection were included. Subanalyses were conducted among people with ≥3 different symptoms. Summary characteristics were provided for each cluster. All analyses were conducted separately in 9 databases from 7 countries, including data from primary care, hospitals, national health claims and national health registries, allowing to compare clusters across the different healthcare settings.

Results: This study included 787,078 persons with PCC. Single-symptom clusters were common across all databases, particularly for joint pain, anxiety, depression and allergy. Complex clusters included anxiety-depression and abdominal-gastrointestinal symptoms.

Conclusion: Substantial heterogeneity within and between PCC clusters was seen across health-care settings. Current definitions of PCC should be critically reviewed to reflect this variety in clinical presentation.

Keywords: Post-acute COVID-19 condition; Long COVID; Real-world data; Clustering; Latent class analysis

1. Introduction

With more than 200 post-acute COVID-19 symptoms being reported in the context of ''long COVID'', the clinical presentation of post-acute COVID-19 condition (PCC) is highly variable among affected individuals [1−6]. While studies estimate that around 1 in 10 [7,8] people with COVID-19 develop post-acute symptoms, this new condition is still poorly understood. PCC can lead to substantial disruptions in daily functioning, work capacity, and quality of life. Its symptoms are diverse and often unspecific, including fatigue, cognitive impairment, and respiratory complaints. Moreover, with as many as 60 million people potentially affected worldwide, PCC presents not only a huge burden to affected individuals, but also poses enormous health-care, economic, and sociopolitical challenges to the wider society [9].

To enable targeted clinical care and facilitate research, clinical experts selected a list of 25 relevant symptoms during a Delphi process to develop a clinical case definition of PCC, subsequently published by the World Health Organization (WHO) [10].

Machine learning methods including clustering techniques have been used to identify patterns of associated symptoms in the approach to identify clinical subgroups of PCC [11−15,16−18], which can subsequently facilitate targeted research into potential treatments and optimized care for affected individuals. However, previous clustering studies primarily relied on self-reported survey data or prospective cohorts, which, while rich in clinical detail, are often context-specific and may not be directly comparable across populations or health-care settings. [11−15,16−18]. Moreover, the methodological heterogeneity of these studies in terms of setting, design, symptom definition, and statistical approach prevents general comparison of clusters across studies. However, recent studies have illustrated the validity and relevance of real-world data to study PCC in large populations [19].

Our study therefore aims to identify and describe symptom clusters of PCC across geographies and healthcare settings through the use of standardized methods and by leveraging harmonized data mapped to a common format to facilitate external comparison.

2. Materials and methods

2.1. Data sources

This multinational study used routinely collected, deidentified health-care records from 7 countries and 9 databases, covering a broad range of healthcare settings. Primary care records were retrieved from the Netherlands (Integrated Primary Care Information (IPCI) database [20]) and the UK (Clinical Practice Research Datalink (CPRD) GOLD [21] and CPRD Aurum [22]). Hospital records were analyzed from The Health Research Institute Hospital La Fe (HULAFE, Valencia, Spain), the Parc de Salut Mar Barcelona Information System (IMASIS [23], Barcelona, Spain) and the Entrepôt de Données de Santé

What is new?

Key findings

- Symptom clusters of post-acute COVID-19 condition (PCC) were consistently identified across multiple countries and health-care settings. The most recurrent patterns included (1) anxiety and depression; (2) gastrointestinal symptoms and abdominal pain; and (3) respiratory and systemic symptoms including cough, dyspnea, and fatigue.

What this adds to what is known

- Previous clustering studies on PCC were largely based on small, self-reported surveys. This study is the first to use harmonized, large-scale, real-world data from diverse health-care settings, mapped to the Observational Medical Outcomes Partnership Common Data Model, to compare the presence of similar symptom clusters internationally.

What is the implication and what should change now?

- Our findings support previous suggestions to refine the World Health Organization PCC definition by considering typical combinations of multiple symptoms to define clinically meaningful subgroups of PCC. Our work provides a scalable framework for identifying PCC phenotypes that may support targeted health-care strategies and future research.

du Languedoc of the Centre Hospitalier Universitaire de Montpellier (eDOL, Montpellier, France). Adjudicated medical and pharmacy claims from the US (PharMetrics® Plus for Academics [24], US) and national health insurance claims from Estonia (CORIVA, Estonia). Finally, linked nationwide healthcare registries from Norway (NLHR@UiO [25], Norway) covering both primary and secondary care were included.

All contributing databases were previously mapped to the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM), allowing for federated network analyses without sharing patient-level data. More detail on the participating databases can be found in Supplementary Tables 1 to 3.

The RECORD guidelines were used for reporting [26].

2.2. Study population

All individuals registered in the participating databases for at least 365 days, and with a recorded COVID-19 diagnosis or positive SARS-CoV-2 test (polymerase chain reaction or Nasopharyngeal Lateral Flow Assay rapid antigen test) after September 1, 2020 comprised the source population. Recording of a COVID-19 infection via diagnosis or test varied across countries, as seen in Supplementary Table 2. Those with prior COVID-19 or influenza in the 90 days before entry were excluded. Participants with $<$ 120 days of follow-up were excluded. Follow-up ended at the earliest of end of observation, death or end of national testing. Children were not excluded.

Among those, we included people with any of the predefined 25 symptoms listed by the WHO clinical case definition for PCC [9], including abdominal pain, allergy, altered smell and/or taste, anxiety, blurred vision, chest pain, cognitive dysfunction, cough, depression, dizziness, dyspnea, fatigue or malaise, gastrointestinal issues, headache, intermittent fever, joint pain, memory issues, menstrual problems, muscles spasms or pain, neuralgia, pins and needles sensation, sleep disorder, tachycardia, postexertional fatigue and tinnitus and hearing problems. Code lists and computable phenotypes were developed for each symptom (using the 'observation' and 'condition occurrence' domains within the OMOP CDM) and reviewed by 3 clinical experts from PCC clinics [27].

Patients with PCC were identified by at least one record of those 25 symptoms between 90 and 365 days after the date of COVID-19 diagnosis or SARS-CoV-2 positive test. To avoid misclassification of pre-existing conditions, individuals were only included if the symptom had also not been recorded in the 180 days prior to their COVID-19 diagnosis.

Secondary analyses included people with $\geq$2 or $\geq$3 symptoms to capture a narrower definition of PCC. Supplementary Figure 1 shows database-specific attrition.

2.3. Statistical analyses

2.3.1. Descriptive analyses

Among all people with at least one post-acute COVID-19 symptom, key characteristics including age, sex and the number of different symptoms per person were summarized, and the frequency of each individual post-acute COVID-19 symptom calculated.

No missingness was present for covariates. Absence of a recorded symptom was considered absence of that symptom.

2.3.2. Clustering analyses

To retrieve clusters of associated symptoms, we performed latent class analysis (LCA) on individuals with at least 1 post-acute COVID-19 symptom. LCA is a soft clustering algorithm, which calculates probabilities of belonging to different clusters and assigns patients to the one they are most likely to belong to, based on posterior probabilities [28,29].

We used LCA to derive clusters based on symptoms occurrence, considering sex and age as additional covariates. We created binary variables for each of the symptoms

in question among all the patients belonging to the post-acute COVID-19 cohorts.

Models with 2 to 7 classes and 100 random starts were fitted. Class selection was based on Bayesian or consistent akaike information criterion (i.e., BIC or CAIC), classification accuracy (ie, mean posterior probability) and clinical relevance of the cluster characteristics, favoring solutions with fewer classes where possible [28,30,31]. We implemented the algorithm using the poLCA package [32] in R, v. 4.2.1.

We chose the LCA algorithm as it allows the testing of different hypotheses for the distribution of the data using goodness of fit tests, while accounting for covariates in the model [33,34].

2.3.3. Summary characteristics for clusters

For each cluster, we reported demographics, vaccination status, selected comorbidities and healthcare utilization (outpatient visits and hospitalizations) in the year before and the month after SARS-CoV-2 infection.

2.3.4. Secondary analyses

We repeated cluster analyses for those individuals with $\geq$2 and $\geq$ 3 symptoms for all databases that had sufficiently large sample sizes (ie, $\geq$500 patients) in the final cohorts to be clustered.

In an additional subgroup analysis, a record of ''post-acute COVID-19'' (OMOP Extension 705076) was included as an additional covariate to explore its correlation with individual symptoms.

We wrote the analytical code for this study in R version 4.2.3 and is available at [https://github.com/oxford-pharmacoepi/LongCovidStudyathon_W3]. All analyses were conducted separately for each database in a distributed manner and compiled by K.L. for this manuscript.

3. Results

3.1. Study population characteristics

A total of 787,078 individuals with PCC were included out of 3,209,411 eligible COVID-19 participants from nine databases from 6 European countries and the United States, with 36,779 people included from Estonia (database CORIVA), 344 from France (database eDOL), 258,995 from Norway (database NLHR@UiO), 84,284 from the Netherlands (database IPCI), 38,798 from Spain (databases HULAFE and IMASIS), 234,736 from the United Kingdom (databases CPRD GOLD and CPRD Aurum), and 133,675 from the United States (database Pharmetrics® Plus for Academics). Table 1 summarizes participant characteristics in each database. Participants with PCC were predominantly female (except in eDOL), and of median age between 35 and 52 years old, except for hospital databases IMASIS and eDOL where median age was $\geq$60 years old. Most participants only had one post-acute COVID-19 symptom recorded, whereas 21% to 51% of participants had at least two symptoms and 4% to 25% had at least 3 symptoms.

Proportions of individual symptoms in people with at least 1 recorded symptom (Figure 1), and in a subgroup of people with at least 2 and at least 3 different symptoms (Supplementary Figure 2) were estimated in all databases.

While the most common symptoms were generally different across databases, consistently common symptoms included abdominal pain, anxiety, gastrointestinal issues or joint pain.

3.2. Clustering

LCA was used to classify individuals with PCC into groups based on symptom patterns, for the 3 cohorts of patients with $\geq$1, $\geq$2, and $\geq$3 recorded symptoms. Most databases had the best model fit with 4 clusters (Supplementary Table 4). However, expert clinical review indicated that the 3-cluster solution yielded more interpretable and clinically meaningful groupings without a substantial loss of model fit. Therefore, those were chosen as the main results for discussion. Results for all other numbers of clusters are available in the interactive web application: https://dpa-pde-oxford.shinyapps.io/WP3_Clustering/.

3.3. Clustering among people with PCC

Results for the clusters among people with one or more post-acute symptoms are presented in Figure 2. The heatmap with all additional characterization features can be seen in Supplementary Figure 3. With most people having only 1 recorded post-acute COVID-19 symptom, many of the clusters were characterized by single symptoms or at most a combination of 2 highly prevalent symptoms. Single-symptom clusters defined by highly prevalent symptoms were seen across multiple databases, including ''joint pain'' in NLHR@UiO and CPRD GOLD, and ''allergy'' in IMASIS, NLHR@UiO and CORIVA.

An ''anxiety, depression'' cluster was the only multisymptom cluster seen consistently across more than one database (CPRD Aurum and eDOL). Another multisymptom cluster included the combination of ''dyspnea, chest pain, fatigue, gastrointestinal symptoms, tachycardia'', but this was only seen in US health claims (PharMetrics® Plus for Academics).

3.4. Clustering among people with at least 3 symptoms

When restricting to the subgroup of people with at least 3 symptoms, eDOL and IMASIS did not have the required more than 500 individuals to contribute to the analyses.

Results from clustering among people with at least 3 different symptoms are illustrated in Figure 3. The heatmap with all additional characterization features can be seen in Supplementary Figure 3. Clusters combining ''anxiety, depression'', ''abdominal pain, gastrointestinal issues'' and ''cough, dyspnea, fatigue, joint pain, chest pain'' were

Table 1. Summary characteristics of people with PCC

Database name	Country, region	Health-care setting	*N*	≥2 symptoms	≥3 symptoms	Median age	Sd age	Sex female
CORIVA	Estonia, nationwide	Claims, prescriptions	36,779	10,414 (28.3%)	2791 (7.6%)	42.9	13.6	23,747 (64.6%)
CPRD Aurum	UK, nationwide	Primary care	217,236	61,486 (28.3%)	16,853 (7.8%)	40.4	19.9	139,467 (64.2%)
CPRD GOLD	UK, nationwide	Primary care	17,500	4153 (23.7%)	1013 (5.8%)	44.4	14.8	11,371 (65.0%)
eDOL	France, Montpellier	Secondary care	344	NA	NA	62.2	24.4	145 (42.0%)
HULAFE	Spain, Valencia	Secondary care	35,587	12,773 (35.9%)	4293 (12.1%)	42.4	15.3	22,159 (62.3%)
IMASIS	Spain, Barcelona	Secondary care	3211	911 (28.4%)	NA	61.0	12.7	1698 (52.9%)
IPCI	Netherlands, nationwide	Primary care	84,284	18,061 (21.4%)	3759 (4.3%)	40.4	13.6	52,864 (62.7%)
NLHR@UiO	Norway, nationwide	National health Registries (primary secondary care)	258,995	62,337 (24.1%)	14,534 (5.6%)	35.1	19.1	154,688 (59.7%)
Pharmetrics® Plus for Academics	USA, nationwide	Claims	133,675	68,202 (51.0%)	33,701 (25.2%)	52.2	13.1	78,725 (58.9%)

PCC, post-acute COVID-19 condition; CPRD, clinical practice research datalink; IPCI, Integrated Primary Care Information; HULAFE, The Health Research Institute Hospital La Fe; IMASIS, Parc de Salut Mar Barcelona Information System; NA, nonavailable data.

Clinical characteristics and demographics of study population by database.

identified across almost all databases. Table 2 summarizes multisymptom clusters with their baseline characteristics.

Clusters were labeled based on the most prevalent symptom patterns. In Figure 3, harmonized labels such as "Anxiety−Depression" (AD), "Gastrointestinal−Abdominal" (GA), and "Respiratory−Fatigue−Pain" (RFP) were used to facilitate comparability across databases.

The cluster characterized by "anxiety, depression" had the highest percentage of women, was seen most often in the third decade of life, and sometimes included symptoms like headache, joint pain (HULAFE), fatigue malaise and headache (IPCI) and fatigue malaise, headache and allergy (NLHR@UiO).

The "abdominal pain, gastrointestinal issues" cluster was seen most often in the third and fourth decades of life and accompanied by additional symptoms such as intermittent fever, cough and allergy in one case (HULAFE).

The last cluster, characterized by "cough, dyspnea, fatigue malaise, joint pain, chest pain", was more prevalent in males, most common in patients in their fifth or sixth decade of life, and accompanied by tachycardia in Pharmetrics® Plus for Academics and the PCC diagnosis code in CPRD Aurum.

A description of the different clusters per database for people with at least 3 symptoms can be found in Supplementary Table 5, and for people with at least 2 different symptoms in Supplementary Figure 4. All results are available in an interactive web application [https://dpa-pde-oxford.shinyapps.io/WP3_Clustering/].

4. Discussion

4.1. Summary of key results

This is to our knowledge the first and largest international study reported to date on post-COVID-19 symptoms. While others have endeavored to describe PCC symptom clusters before [11−15,16−18], this is the first paper to offer results across multiple databases and countries and with a harmonized framework. Here, we identified different patterns of symptoms associated with PCC using latent class models and subsequently characterized the derived clusters across a large network of nine different databases from 7 countries and over 700,000 patients.

When clustering among people with PCC, as defined by WHO, those clusters were dominated by single isolated symptoms, for instance, joint pain or clinically related symptoms such as anxiety and depression. These are prevalent in the general population and might not necessarily be characteristic or differential for PCC. Among these individuals, observed clusters may reflect artifacts of coding conventions, documentation frequency, and symptom reporting, rather than true clinical subtypes.

Clusters identified among people with multiple symptoms; however, showed substantially more complex multisymptom patterns. Given that the majority of PCC patients present multiple symptoms, these are also richer and more meaningful from a clinical viewpoint. Among those, some specific clinical subgroups were observed across multiple databases, and were characterized by mental health-related symptoms, gastrointestinal issues or respiratory symptoms, sometimes associated with fatigue. These patterns of symptoms align with the previous literature and clinical presentation of people seeking specialist care for PCC [35].

4.2. Results in context

While our study is the first to directly compare subgroups of people with PCC across countries and health-care settings, previous studies also applied clustering to identify symptom patterns. Gottlieb et al [11] used latent class analyses to identify phenotypes of PCC based on survey data among 4505 people with COVID-19 in the United States during 3- and 6 months following their SARS-COV-2 infection. Four distinct clusters were identified: 1 with almost no symptoms, 1 with tiredness/headache/musculoskeletal issues, 1 with loss of taste and smell, and 1 with "many symptoms across multiple systems". Similar studies were conducted using survey data from prospective cohort studies

Figure 1. Post-acute COVID-19 symptoms and OMOP Extension post-COVID condition recorded after COVID-19 infection per database. The concept "OMOP Extension post-COVID condition diagnosis" refers to the OMOP Extension code 705076. CPRD, Clinical Practice Research Datalink; IPCI, Integrated Primary Care Information; HULAFE, The Health Research Institute Hospital La Fe; IMASIS, Parc de Salut Mar Barcelona Information System; NLHR@UiO, Norwegian Linked Health Registries at University of Oslo; OMOP, Observational Medical Outcomes Partnership; PCC, post-acute COVID-19 condition.

among 233 and 497 people presenting at PCC clinics in Ireland and Japan. Using a hierarchical clustering approach, Kenny et al [13] identified 3 clusters of symptoms dominated by pain (joint pain, myalgia, headache), cardiovascular symptoms (eg, chest pain, shortness of breath and palpitations), and a "substantially smaller number of concomitant symptoms", while Tsuchida and colleagues [12] described 5 clusters based on self-reported symptoms dominated by "fatigue", "fatigue/cardiovascular symptoms/forgetfulness", "fatigue/headache/mental health

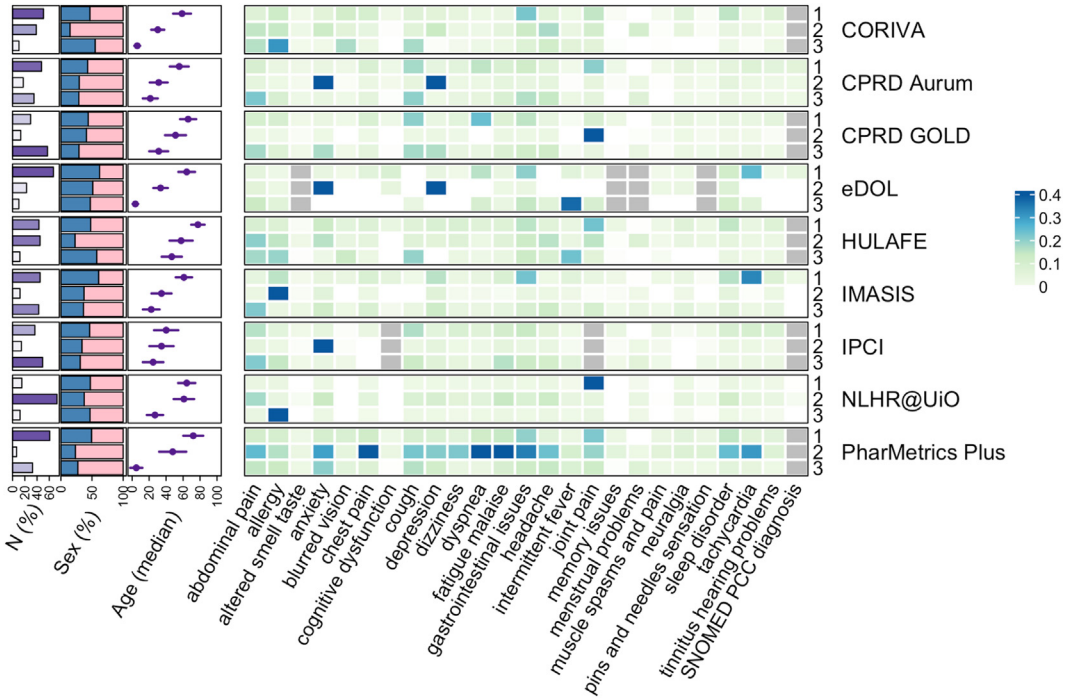

Figure 2. Proportion of symptoms per cluster (right) and their characteristics (left) for the contributing databases, for the cohorts of people with one or more symptoms postinfection. Color gradients represent relative frequency of each symptom per cluster, standardized row-wise to highlight differences in pattern rather than raw frequency. For sex, blue represents male percentage and pink represents female percentage. CPRD, Clinical Practice Research Datalink; IPCI, Integrated Primary Care Information; HULAFE, The Health Research Institute Hospital La Fe; IMASIS, Parc de Salut Mar Barcelona Information System; NLHR@UiO, Norwegian Linked Health Registries at University of Oslo; OMOP, Observational Medical Outcomes Partnership; PCC, post-acute COVID-19 condition. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

issues/forgetfulness'', ''hair loss'' and ''taste and smell disorders''. Other studies included additional markers on top of persisting symptoms for their clustering approach, for instance, COVID-19 disease severity [14], comorbidity burden [14,16], symptoms at inclusion [14,16], hospitalization data [16] and quality of life [14] to characterize clusters by the number of symptoms and disease severity [14,16], and impact on daily living activities [16].

Few studies have assessed cluster reproducibility in the context of PCC. Reese et al [17] used k-means clustering on data from the National COVID Cohort Collaborative (*N*3C) in 6 US hospitals and identified 6 clusters: ''multisystem/laboratory'', ''pulmonary'', ''neuropsychiatric'', ''cardiovascular'', ''pain/fatigue'', and ''Multisystem/pain''. Generalizability of clusters was subsequently assessed by assigning patients from other health-care centers to clusters based on high semantic similarity of symptoms recorded in original patients, and results were reproducible across systems. Similarly, Zhang et al [17] validated phenotype subgroups among people with PCC, suggesting 4 clusters consistent across 2 US health databases: ''cardiac/renal'', ''respiratory/sleep/anxiety'', ''musculoskeletal/nervous system'', and ''digestive/respiratory system''. More recent studies report again similar clusters with different hierarchical clustering methodologies [15]. However, no comparison has been conducted across different countries and health-care settings.

During the pandemic, health-care policies and access varied across countries, with substantial differences between hospital and outpatient populations. In 2021, a large meta-analysis [35] comprising 1.2 million people from 22 countries suggested that clinical symptoms for as much as 51%, 60% and 35% of people with post-COVID-19 conditions fit one of 3 PCC clusters: persistent fatigue with body pain and mood swings, ongoing respiratory problems, or cognitive problems. However, those clusters were defined based on reporting frequency in previous studies, and not directly derived from data.

Several clusters identified in our study are consistent with previous literature, with ''pain'' and ''mental health issues'' being commonly reported, and many clusters being associated with some degree of fatigue or tiredness [36,37]. Moreover, our subgroup analysis including the ''post-acute COVID-19 condition'' diagnosis code as an additional symptom showed its association with a cluster characterized by cough, dyspnea and fatigue in CPRD Aurum. It is likely that this relates to the most common presentation in specialized clinics.

4.3. Strengths and limitations

This study has several strengths. We conducted a multidatabase, international network study including data from different health-care settings and including a large number

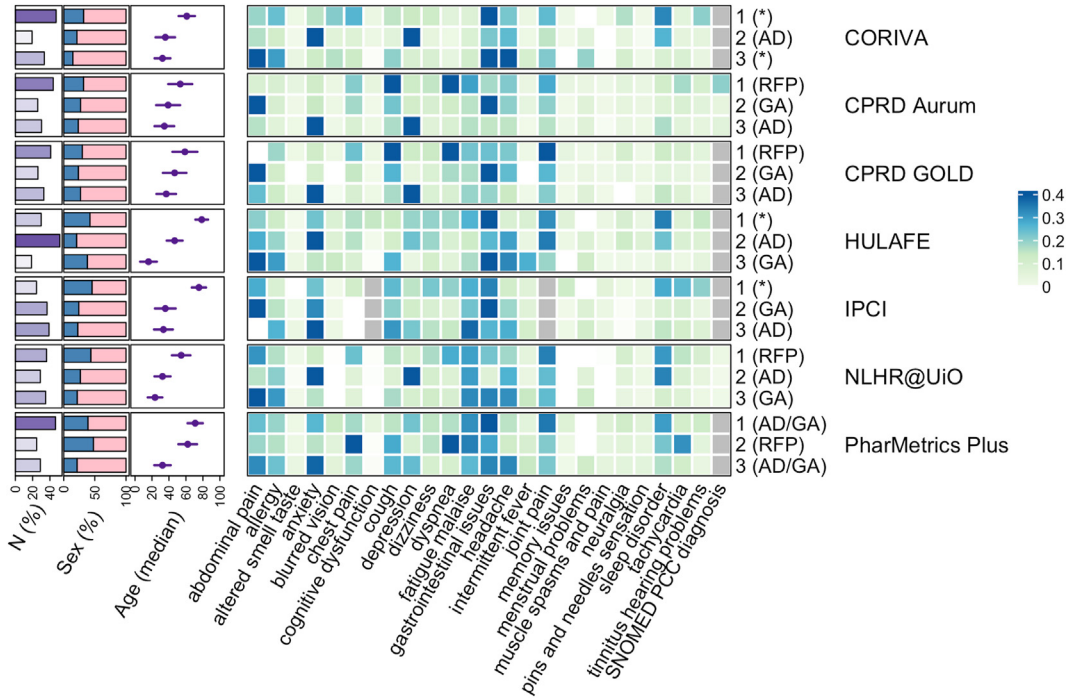

Figure 3. Proportion of symptoms per cluster (right) and their characteristics (left) for the contributing databases, for the cohorts of people with 3 or more symptoms postinfection. Color gradients represent relative frequency of each symptom per cluster, standardized row-wise to highlight differences in pattern rather than raw frequency. ''AD'' refers to the Anxiety-Depression cluster, ''GA'' to the Gastrointestinal issues-Abdominal pain one, and ''RFP'' to the set of Respiratory-Fatigue-Pain symptoms. For sex, blue represents male percentage and pink represents female percentage. CPRD, Clinical Practice Research Datalink; IPCI, Integrated Primary Care Information; HULAFE, The Health Research Institute Hospital La Fe; IMASIS, Parc de Salut Mar Barcelona Information System; NLHR@UiO, Norwegian Linked Health Registries at University of Oslo; OMOP, Observational Medical Outcomes Partnership; PCC, post-acute COVID-19 condition. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

of people with PCC. The main differences between previous studies might be explained by the methodological heterogeneity, including study populations, settings, post-acute COVID-19 definition and clustering methods. We used a single standardized methodological approach including a consistent definition of PCC, and statistical analyses, increasing the robustness and comparability across datasets. The use of a recognized common data model, OMOP CDM [38,39], allowed for federated analyses using the same code, shared in line with open science practises. This minimized further the bias in the process and maximized reproducibility. The outstanding similarity between the results of the similar CPRD GOLD and CPRD Aurum, in terms of the cluster solutions attained, databases reassure the robustness of the results.

There is, however, some heterogeneity across databases, which can be explained by underlying heterogeneity between geographies and health-care settings. A further limitation is that routinely collected data do not include information on the severity of individual symptoms. Similarly, the validity of recording of symptoms in such data is subject to misclassification and incompleteness, as information is collected in actual health-care settings by multiple health-care professionals. Furthermore, the motivation to report PCC might vary across settings. This makes it harder to compare clusters both within and across databases. Additionally, given the nonspecific nature of many symptoms (eg, fatigue, headache, anxiety), some observed patterns may reflect common health-care-seeking behavior rather than distinct PCC subtypes. The absence of a control population limits causal attribution. This study is, therefore, descriptive in nature. Finally, we did not assess clustering patterns specifically among children. However, the observed age distributions, in terms of median and standard deviation per cluster and country, suggest that the number of children included in the analysis was very small.

4.4. Implications and recommendations for future research

PCC is a debilitating illness that afflicts individuals in their forties, incurring a substantial economic cost [40]. Health-care systems must be well-prepared to address the needs of millions of patients. Accurately estimating the number of affected individuals is essential for effective resource planning.

While the WHO definition of PCC proves useful, it is inevitably subject to develop over time given new research findings and does not consider a minimum number of symptoms. The identification of large single-symptom clusters characterized by highly common symptoms suggests

Table 2. Summary of the 3-cluster solution for the ≥3 symptoms cohorts across contributing databases.

CLUSTER DESCRIPTION	CORIVA	CPRD Aurum	CPRD GOLD	HULAFE	IPCI	Pharmetrics® Plus for Academics	NLHR@UiO
Anxiety **Depression**	**N** 530 (19%) **Age** 37 (sd 15) **Male** 13%	**N** 5056 (30%) **Age** 36 (sd 16) **Male** 23%	**N** 334 (33%) **Age** 38 (sd 15) **Male** 27%	**N** 2182 (51%) **Age** 47 (sd 13) **Male** 20% *+ headache, joint pain*	**N** 1396 (39%) **Age** 35 (sd 15) **Male** 23% *+ fatigue malaise, headache*	These two clusters are approximately merged in this database.	**N** 4197 (29%) **Age** 34 (sd 13) **Male** 27% *+ fatigue malaise, headache, allergy*
Gastrointestinal issues **Abdominal pain**		**N** 4382 (26%) **Age** 40 (sd 19) **Male** 27%	**N** 263 (26%) **Age** 47 (sd 19) **Male** 24%	**N** 816 (19%) **Age** 19 (sd 13) **Male** 38% *+ intermittent fever, cough, allergy*	**N** 1324 (37%) **Age** 37 (sd 17) **Male** 24%		**N** 5091 (35%) **Age** 26 (sd 12) **Male** 22%
Cough **Dyspnoea** **Fatigue** **Joint pain** **Chest pain**		**N** 7415 (44%) **Age** 53 (sd 19) **Male** 32% + OMOP Extension *post-COVID-19 code*	**N** 416 (41%) **Age** 58 (sd 20) **Male** 30%			**N** 8425 (25%) **Age** 61 (sd 15) **Male** 48% *+ tachycardia*	**N** 5246 (36%) **Age** 54 (sd 15) **Male** 44%

CPRD, clinical practice research datalink; IPCI, integrated primary care information; HULAFE, Hospital Universitario La Fe; IMASIS, Parc de Salut Mar Barcelona Information System; NLHR@UiO, Norwegian Linked Health Registries at University of Oslo; OMOP, observational medical outcomes partnership.

In green, clusters found in the respective database. In orange otherwise.

that the WHO definition of long COVID needs refining, with the inclusion of more than 1 co-occurring symptom.

Identifying symptom patterns can support tailored treatments and provide insights into identifying potential diverse pathogenic mechanisms.

5. Conclusion

This large-scale international study included more than 787,000 persons with PCC from 7 countries, and found both single and complex symptom clusters, mainly ''anxiety and depression''; ''gastrointestinal issues and abdominal pain''; and ''cough, dyspnea, fatigue, joint, and chest pain''.

Single-symptom clusters likely reflect symptoms with high prevalence in the general population as well as common complications of COVID-19. Current definitions for PCC should therefore be critically reviewed and refined where needed to reflect the multiplicity of symptoms and variety in clinical presentations of PCC. While descriptive in nature, the multisymptom patterns found support the design of targeted management strategies.

Ethics statement

This study is based in part on data from the Clinical Practice Research Datalink (CPRD) obtained under license from the UK Medicines and Healthcare Products Regulatory Agency. We thank the patients who provided these data and the NHS who collected the data as part of their care and support. Use of Clinical Practice Research Datalink (CPRD) data for this study were approved via the Research Data Governance (RDG) Process of the UK Medicines and Healthcare Products Regulatory Agency (protocol 23_002603). Ethical approval for NHR@UiO in this study was obtained from The Regional Committee for Research Ethics (approval number 155294) and the Data Protection Officer at the University of Oslo (approval number 523275). Ethical approval for CORIVA data was obtained from the Research Ethics Committee of the University of Tartu (No. 351/M-8). Ethical approval for HULAFE was obtained from the Comité de ética de la investigación con medicamentos (number 2023-232-1). Ethical approval for IMASIS was obtained by the Parc de Salut Mar Research Ethics Committee CEIm-Parc de Salut Mar (number 2021/9975). Ethical approval for IPCI was obtained by the Integrated Primary Care Information review board (registration number 9/2023). For eDOL, no ethical approval was required according to French law for this study. All patients admitted to the hospital are provided with general information about the collection and secondary use of their data, and an opt-out option is offered. PharMetrics® Plus for Academics needed no approval for use of pseudoanonymized secondary data.

CRediT authorship contribution statement

Kim López-Güell: Writing – review & editing, Writing – original draft, Software, Formal analysis, Conceptualization. **Martí Català:** Writing – review & editing, Writing – original draft, Software, Formal analysis, Conceptualization. **Daniel Dedman:** Writing – review & editing, Formal analysis. **Talita**

Duarte-Salles: Writing – review & editing, Formal analysis. **Raivo Kolde:** Writing – review & editing. **Raúl López-Blasco:** Writing – review & editing, Formal analysis. **Álvaro Martínez:** Writing – review & editing, Formal analysis. **Gregoire Mercier:** Writing – review & editing, Formal analysis. **Alicia Abellan:** Writing – review & editing, Formal analysis. **Johnmary T. Arinze:** Writing – review & editing, Formal analysis. **Theresa Burkard:** Writing – review & editing, Formal analysis. **Edward Burn:** Writing – review & editing, Formal analysis. **Zara Cuccu:** Writing – review & editing, Formal analysis. **Antonella Delmestri:** Writing – review & editing, Data curation. **Dominique Delseny:** Writing – review & editing, Formal analysis. **Sara Khalid:** Writing – review & editing, Formal analysis. **Chungsoo Kim:** Writing – review & editing, Formal analysis. **Ji-woo Kim:** Writing – review & editing, Formal analysis. **Kristin Kostka:** Writing – review & editing, Formal analysis, Conceptualization. **Cora Loste:** Writing – review & editing, Formal analysis. **Miguel A. Mayer:** Writing – review & editing, Formal analysis. **Jaime Meléndez-Cardiel:** Writing – review & editing, Formal analysis. **Núria Mercadé-Besora:** Writing – review & editing, Formal analysis. **Mees Mosseveld:** Writing – review & editing. **Akihito Nishimura:** Writing – review & editing, Formal analysis. **Hedvig ME. Nordeng:** Writing – review & editing, Formal analysis. **Jessie O. Oyinlola:** Writing – review & editing, Formal analysis. **Roger Paredes:** Writing – review & editing, Formal analysis. **Laura Pérez-Crespo:** Writing – review & editing. **Marta Pineda-Moncusí:** Writing – review & editing. **Juan Manuel Ramírez-Anguita:** Writing – review & editing, Formal analysis. **Nhung T.H. Trinh:** Writing – review & editing, Formal analysis. **Anneli Uusküla:** Writing – review & editing, Formal analysis. **Bernardo Valdivieso:** Writing – review & editing, Formal analysis. **Daniel Prieto-Alhambra:** Writing – review & editing, Writing – original draft, Funding acquisition, Conceptualization. **Junqing Xie:** Writing – review & editing, Formal analysis, Conceptualization. **Lourdes Mateu:** Writing – review & editing, Writing – original draft, Formal analysis. **Annika M. Jödicke:** Writing – review & editing, Writing – original draft, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

D.P.A.'s department has received grant/s from Amgen, Chiesi–Taylor, Lilly, Janssen, Novartis, and UCB Biopharma. His research group has received consultancy fees from Astra Zeneca and UCB Biopharma. Amgen, Astellas, Janssen, Synapse Management Partners and UCB Biopharma have funded or supported training programs organized by D.P.A.'s department. R.P. reports serving on advisory boards for Gilead Sciences, Inc, Pfizer, Inc, Roche Therapeutics, MSD, GSK, ViiV Healthcare, Eli Lilly and Company, PharmaMar, and Atea Pharmaceuticals, Inc; and receiving research grants paid to his institution from MSD, ViiV Healthcare, Gilead Sciences, and PharmaMar. L.M. reports receiving grants from Grifols; receiving honoraria as a speaker from AstraZeneca, Gilead Sciences, GSK, and Pfizer; and participation in advisory boards for Gilead Sciences and Merck. M.M. works for a research group that in the past 3 years received unconditional research grants from Chiesi, UCB, Amgen, Johnson & Johnson, Innovative Medicines Initiative and the European Medicines Agency. D.Ded., Z.C. and J.O. are employees of the Medicines and Healthcare Products Regulatory Agency, which provides the CPRD research service. K.K. is a consortial author in the US National Institutes of Health National COVID Cohort Collaborative (funding expired in 2022 with no renewal or active impact on any current work). There are no competing interests for any other author.

Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jclinepi.2025.111867.

Data availability

Patient data used for the analyses is confidential, but analysis code is available on GitHub and shared in the Methods section.

References

[1] Lopez-Leon S, Wegman-Ostrosky T, Perelman C, Sepulveda R, Rebolledo P, Cuapio A, et al. More than 50 long-term effects of COVID-19: a systematic review and meta-analysis. Sci Rep 2021;11(1):16144.

[2] Nalbandian A, Sehgal K, Gupta A, Madhavan MV, McGroder C, Stevens JS, et al. Post-acute COVID-19 syndrome. Nat Med 2021;27:601–15. https://doi.org/10.1038/s41591-021-01283-z.

[3] Greenhalgh T, Knight M, A'Court C, Buxton M, Husain L. Management of post-acute covid-19 in primary care. BMJ 2020;370:m3026. https://doi.org/10.1136/bmj.m3026.

[4] Whitaker M, Elliott J, Chadeau-Hyam M, Riley S, Darzi A, Cooke G, et al. Persistent COVID-19 symptoms in a community study of 606,434 people in England. Nat Commun 2022;13(1):1957. https://doi.org/10.1038/s41467-022-29521-z.

[5] Subramanian A, Nirantharakumar K, Hughes S, Myles P, Williams T, Gokhale KM, et al. Symptoms and risk factors for long COVID in non-hospitalized adults. Nat Med 2022;28:1706–14. https://doi.org/10.1038/s41591-022-01909-w.

[6] Bowe B, Xie Y, Al-Aly Z. Postacute sequelae of COVID-19 at 2 years. Nat Med 2023;29:2347–57. https://doi.org/10.1038/s41591-023-02521-2.

[7] The Lancet. Long COVID: 3 years in. Lancet 2023;401(10379):795. ISSN 0140-6736 https://doi.org/10.1016/S0140-6736(23)00493-2.

[8] Fang Z, Ahrnsbrak R, Rekito A. Evidence mounts that about 7% of US adults have had long COVID. JAMA 2024;332:5–6. https://doi.org/10.1001/jama.2024.11370.

[9] University CaJH. COVID-19 dashboard. Available at: https://coronavirus.jhu.edu/map.html2023. Accessed March 13, 2023.

[10] WHO. A clinical case definition of post COVID-19 condition by a Delphi consensus. 2021. Available at: https://www.who.int/publications/i/item/WHO-2019-nCoV-Post_COVID-19_condition-Clinical_case_definition-2021. Accessed October 6, 2021.

[11] Gottlieb M, Spatz ES, Yu H, Wisk LE, Elmore JG, Gentile NL, et al. Long COVID clinical phenotypes up to 6 Months after infection identified by latent class analysis of self-reported symptoms.

Open Forum Infect Dis 2023;10(7):ofad277. https://doi.org/10.1093/ofid/ofad277.
[12] Tsuchida T, Yoshimura N, Ishizuka K, Katayama K, Inoue Y, Hirose M, et al. Five cluster classifications of long COVID and their background factors: a cross-sectional study in Japan. Clin Exp Med 2023;23(7):3663−70. https://doi.org/10.1007/s10238-023-01057-6.
[13] Kenny G, McCann K, O'Brien C, Savinelli S, Tinago W, Yousif O, et al. Identification of distinct long COVID clinical phenotypes through cluster analysis of self-reported symptoms. Open Forum Infect Dis 2022;9(4):ofac060. https://doi.org/10.1093/ofid/ofac060.
[14] Fischer A, Badier N, Zhang L, Elbéji A, Wilmes P, Oustric P, et al. Long COVID classification: findings from a ClusteringAnalysis in the predi-COVID cohort study. Int J Environ Res Public Health 2022;19(23):16018. https://doi.org/10.3390/ijerph192316018.
[15] Niewolik J, Mikuteit M, Klawitter S, Schröder D, Stölting A, Vahldiek K, et al. Cluster analysis of long COVID symptoms for deciphering a syndrome and its long-term consequence. Immunol Res 2024;72(4):605−13. https://doi.org/10.1007/s12026-024-09465-w.
[16] Fernandez-de-Las-Penas C, Martin-Guerrero JD, Florencio LL, Navarro-Pardo E, Rodríguez-Jiménez J, Torres-Macho J, et al. Clustering analysis reveals different profiles associating long-term post-COVID symptoms, COVID-19 symptoms at hospital admission and previous medical comorbidities in previously hospitalized COVID-19 survivors. Infection 2023;51(1):61−9. https://doi.org/10.1007/s15010-022-01822-x.
[17] Reese JT, Blau H, Casiraghi E, Bergquist T, Loomba JJ, Callahan TJ, et al. Generalisable long COVID subtypes: findings from the NIH N3C and RECOVER programmes. EBioMedicine 2023;87:104413. https://doi.org/10.1016/j.ebiom.2022.104413.
[18] Zhang H, Zang C, Xu Z, Zhang Y, Xu J, Bian J, et al. Data-driven identification of post-acute SARS-CoV-2 infection subphenotypes. Nat Med 2023;29(1):226−35. https://doi.org/10.1038/s41591-022-02116-3.
[19] Català M, Mercadé-Besora N, Kolde R, Trinh NTH, Roel E, Burn E, et al. The effectiveness of COVID-19 vaccines to prevent long COVID symptoms: staggered cohort study of data from the UK, Spain, and Estonia. Lancet Respir Med 2024;12(3):225−36. S2213-S2600(23)00414-9. https://doi.org/10.1016/S2213-2600(23)00414-9
[20] Vlug A, van der Lei J, Mosseveld B, van Wijk MA, van der Linden PD, Sturkenboom MC, et al. Postmarketing surveillance based on electronic patient records: the IPCI project. Methods Inf Med 1999;38:339−44.
[21] Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, et al. Data resource profile: clinical Practice research Datalink (CPRD). Int J Epidemiol 2015;44(3):827−36. https://doi.org/10.1093/ije/dyv098.
[22] Wolf A, Dedman D, Campbell J, Booth H, Lunn D, Chapman J, et al. Data resource profile: clinical Practice research Datalink (CPRD) Aurum. Int J Epidemiol 2019;48(6):1740g. https://doi.org/10.1093/ije/dyz034.
[23] JJ Sancho IP, Domènech D, Martín-Baranera M, Palau J, Sanz F. Imasis. A multicenter hospital information system - experience in Barcelona. Stud Health Technol Inform 1998;56:35−42. https://doi.org/10.3233/978-1-60750-900-4-35.
[24] IQVIA. IQVIA PharMetrics Plus for Academics enhanced with mortality data. Available at: https://www.iqvia.com/locations/united-states/library/fact-sheets/iqvia-pharmetrics-plus-for-academics-enhanced-with-mortality-data2023. Accessed May 20, 2023.
[25] Bakken IJ, Ariansen AMS, Knudsen GP, Johansen KI, Vollset SE. The Norwegian patient registry and the Norwegian registry for primary health care: research potential of two nationwide health-care registries. Scand J Public Health 2020;48(1):49−55. https://doi.org/10.1177/1403494819859737.
[26] Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, et al, RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS Med 2015;12(10):e1001885. https://doi.org/10.1371/journal.pmed.1001885.
[27] Prieto-Alhambra D, Kostka K, Roel E, Trinh N, Delmestri A, Mateu L, et al. The burden of long COVID: a multinational cohort analysis of Spanish and UK data including SARS-CoV-2 infections, reinfections, and matched contemporaneous test negative controls. 2023. Res Square. Available at: https://ora.ox.ac.uk/objects/uuid:101ad4fd-9973-46a8-a8d2-9b21937a653b. Accessed May 20, 2024.
[28] Lanza ST, Rhoades BL. Latent class analysis: an alternative perspective on subgroup analysis in prevention and treatment. Prev Sci 2013;14(2):157−68. https://doi.org/10.1007/s11121-011-0201-1.
[29] Bozdogan H. Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. Psychometrika 1987;52(3):345−70. https://doi.org/10.1007/BF02294361.
[30] Naldi L, Cazzaniga S. Research techniques made simple: latent class analysis. J Invest Dermatol 2020;140(9):1676−16780.e1. https://doi.org/10.1016/j.jid.2020.05.079.
[31] Weller BE, Bowen NK, Faubert SJ. Latent class analysis: a guide to best Practice. J Black Psychol 2020;46(4):287−311. https://doi.org/10.1177/0095798420930932.
[32] Drew A, Linzer JBL. poLCA: an R package for polytomous variable latent class analysis. J Stat Softw 2011;42(10):1−29. https://doi.org/10.18637/jss.v042.i10.
[33] Feuillet F, Bellanger L, Hardouin JB, Victorri-Vigneau C, Sébille V. On comparison of clustering methods for pharmacoepidemiological data. J Biopharm Stat 2015;25(4):843−56. https://doi.org/10.1080/10543406.2014.920855.
[34] Goodman LA. Latent class analysis: The empirical study of latent types, latent variables, and latent structures. In: Hagenaars JA, McCutcheon AL, editors. Applied latent class analysis. Cambridge, United Kingdom: Cambridge University Press; 2002:3−55.
[35] Global Burden of Disease Long CC, Wulf Hanson S, Abbafati C, Aerts JG, Al-Aly Z, Ashbaugh C, et al. Estimated global proportions of individuals with persistent fatigue, cognitive, and respiratory symptom clusters following symptomatic COVID-19 in 2020 and 2021. JAMA 2022;328(16):1604−15. https://doi.org/10.1001/jama.2022.18931.
[36] Mateu L, Tebe C, Loste C, Santos JR, Lladós G, López C, et al. Determinants of the onset and prognosis of the post-COVID-19 condition: a 2-year prospective observational cohort study. Lancet Reg Health Eur 2023;33:100724.
[37] Thaweethai T, Jolley SE, Karlson EW, Levitan EB, Levy B, McComsey GA, et al. Development of a definition of postacute sequelae of SARS-CoV-2 infection. JAMA 2023;329:1934−46. https://doi.org/10.1001/jama.2023.8823.
[38] Schuemie M, Kostka K, Madigan D, Suchard M, Huser V, Reps J, et al. The book of OHDSI: chapter 4 - the common data model. Available at: https://ohdsi.github.io/TheBookOfOhdsi/CommonDataModel.html. Accessed May 20, 2023.
[39] Burn E, Fernández-Bertolín S, Voss EA, Blacketer C, Aragón M, Recalde M, et al. Establishing and characterising large COVID-19 cohorts after mapping the information system for research in primary care in catalonia to the OMOP common data model. medRxiv. [preprint]. Available at: 2021. https://www.medrxiv.org/content/10.1101/2021.11.23.21266734v1. Accessed May 20, 2024.
[40] Gandjour A. Long COVID: costs for the German economy and health care and pension system. BMC Health Serv Res 2023;23(1):641. https://doi.org/10.1186/s12913-023-09601-6.