

TITLE

Unravelling data for rapid evidence-based response to COVID-19: a summary of the unCoVer protocol

AUTHOR

Penalvo, J; Mertens, E; Ademović, E; et al.

JOURNAL

BMJ Open

DATE DEPOSITED

1 November 2022

This version available at

<https://research.stmarys.ac.uk/id/eprint/5516/>

COPYRIGHT AND REUSE

Open Research Archive makes this work available, in accordance with publisher policies, for research purposes.

VERSIONS

The version presented here may differ from the published version. For citation purposes, please consult the published version for pagination, volume/issue and date of publication.

BMJ Open Unravelling data for rapid evidence-based response to COVID-19: a summary of the unCoVer protocol

José L. Peñalvo ¹, Elly Mertens ¹, Enisa Ademović,² Seval Akgun,³ Ana Lúcia Baltazar,⁴ Dora Buonfrate ⁵, Miran Čoklo,⁶ Brecht Devleesschauwer,^{7,8} Paula Andrea Diaz Valencia,⁹ João C. Fernandes,¹⁰ Enrique Javier Gómez,^{11,12,13} Paul Hynds,¹⁴ Zubair Kabir,¹⁵ Jörn Klein ¹⁶, Polychronis Kostoulas,¹⁷ Lucía Llanos Jiménez,¹⁸ Lucia Maria Lotrean,¹⁹ Marek Majdan,²⁰ Ernestina Menasalvas,¹² Paul Nguewa,²¹ In-Hwan Oh,²² Georgie O'Sullivan,¹⁵ David M. Pereira,²³ Miguel Reina Ortiz ²⁴, Silvia Riva,²⁵ Gloria Soriano,¹ Joan B. Soriano,^{26,27,28} Fernando Spilki,²⁹ Mary Elizabeth Tamang,³⁰ Antígona Carmen Trofor,^{31,32} Michel Vaillant ³³, Sabrina Van Ierssel,³⁴ Jakov Vuković,³⁵ José M. Castellano,^{36,37,38} on behalf of the unCoVer network

To cite: Peñalvo JL, Mertens E, Ademović E, *et al*. Unravelling data for rapid evidence-based response to COVID-19: a summary of the unCoVer protocol. *BMJ Open* 2021;**11**:e055630. doi:10.1136/bmjopen-2021-055630

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2021-055630>).

Received 18 July 2021
Accepted 08 October 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Professor José L. Peñalvo;
jpenalvo@itg.be

ABSTRACT

Introduction unCoVer—Unravelling data for rapid evidence-based response to COVID-19—is a Horizon 2020-funded network of 29 partners from 18 countries capable of collecting and using real-world data (RWD) derived from the response and provision of care to patients with COVID-19 by health systems across Europe and elsewhere. unCoVer aims to exploit the full potential of this information to rapidly address clinical and epidemiological research questions arising from the evolving pandemic.

Methods and analysis From the onset of the COVID-19 pandemic, partners are gathering RWD from electronic health records currently including information from over 22 000 hospitalised patients with COVID-19, and national surveillance and screening data, and registries with over 1 900 000 COVID-19 cases across Europe, with continuous updates. These heterogeneous datasets will be described, harmonised and integrated into a multi-user data repository operated through Opal-DataSHIELD, an interoperable open-source server application. Federated data analyses, without sharing or disclosing any individual-level data, will be performed with the objective to reveal patients' baseline characteristics, biomarkers, determinants of COVID-19 prognosis, safety and effectiveness of treatments, and potential strategies against COVID-19, as well as epidemiological patterns. These analyses will complement evidence from efficacy/safety clinical trials, where vulnerable, more complex/heterogeneous populations and those most at risk of severe COVID-19 are often excluded.

Ethics and dissemination After strict ethical considerations, databases will be available through a federated data analysis platform that allows processing of available COVID-19 RWD without disclosing identification information to analysts and limiting output to data aggregates. Dissemination of unCoVer's activities will be related to the access and use of dissimilar RWD, as well as the results generated by the pooled analyses. Dissemination will include training and educational

Strengths and limitations of this study

- The unCoVer (Unravelling data for rapid evidence-based response to COVID-19) network includes a sizeable number of partners that will exploit the full potential of real-world data derived from the provision of care to patients with COVID-19 by health systems across Europe and elsewhere.
- With the availability of international harmonised real-world data provided by the unCoVer network, a large amount of information on patients with COVID-19 will be studied to better understand aspects of pathophysiology, progression and treatment, and the epidemiological patterns of this novel disease, as well as to grasp the less understood, and potentially unearthed, risk factors associated with COVID-19 severity.
- The development and deployment of a federated data platform for combined analysis meet patients' data protection principles and comply with ethical standards, including general data protection regulation and national data privacy legislation, while allowing for advanced analytics.
- Continuous process evaluation will be carried out throughout the project life to identify limitations and barriers to the harmonised use of data, and simultaneously providing advice on improving data systems for rapid response to future public health crises.

activities, scientific publications and conference communications.

INTRODUCTION

The outbreak of COVID-19, caused by SARS-CoV-2, was declared a public health emergency of international concern by the WHO



on 30 January 2020 and a pandemic on 11 March 2020. Despite the deployment of public health measures, such as restrictions of movements and gathering, personal protection, as well as massive vaccination campaigns, the coronavirus is still largely affecting mortality and morbidity worldwide, including persisting symptoms after the infection, what has now been termed post-COVID-19 condition.¹ Early epidemiological data on COVID-19 showed a higher risk of severe disease among older individuals, in particular those with chronic respiratory, cardiometabolic and other chronic diseases, including mental disorders and immunosuppressed individuals.²⁻⁴ While the pathogenesis of certain chronic diseases predisposes to serious COVID-19 outcomes,⁵ other factors such as common chronic medications might also increase this risk due to interaction between SARS-CoV-2 infection and the complex metabolic pathways.^{6,7} Also, common disease risk factors such as smoking and overweight/obesity have been identified as key predictors of hospitalisation and critical illness, even in young adults with no underlying conditions.^{8,9}

While the pandemic is evolving and countries are adapting their health systems to new phases of preventive measures, the research community is trying to fully elucidate the transmission and progression of the infection, as well as the most effective ways of treating and preventing new cases in preparation for any new waves, particularly due to new variants of SARS-CoV-2. The multidimensional and dynamic nature of the inter-related factors associated with individual responses to the SARS-CoV-2 infection and the diversity of long-term complications require a multidisciplinary research approach to unravel the natural history of this pandemic. Responding to the COVID-19 pandemic in real time required a colossal effort from health systems worldwide and across Europe, where several countries have been severely affected. As a result, a wealth of data have been accumulated as part of health systems' efforts to fight COVID-19. These real-world data (RWD) reflect the impact of COVID-19 on patients' health and characterise the protocols of healthcare in different health system settings. These close-to-reality data allow for studies into patients' characteristics, determinants of disease prognosis and effectiveness of potential strategies against COVID-19 in real-world settings. They also complement findings from ongoing efficacy/safety clinical trials, where vulnerable/heterogeneous populations and those most at risk of severe COVID-19 are often excluded. Harmonisation of data from different sources allows for comparison across health systems and improves patients' characterisation using the wider heterogeneity of information. Still, to date, these RWD sources related to COVID-19 have been exploited in a limited way and for specific questions; hence, there is an untapped opportunity to exploit the full potential of these data through identification, harmonisation and big data analysis.

Objectives

The unCoVer (Unravelling data for rapid evidence-based response to COVID-19) network aims to provide a research platform for the expert use of RWD by bringing together complementary data and medical and scientific expertise to address the still urgent questions regarding the determinants of COVID-19 prognosis to inform more effective medical and public health strategies. Specifically, the network aims to facilitate access to otherwise scattered RWD sources, and hereby provide opportunities for enhanced risk characterisation and robust risk prediction algorithms, tackling the current pandemic and eventually any future epidemics. This approach should lead to control measures that will eventually relieve the pressure on health systems, improve patient prognosis especially among those more vulnerable (eg, chronic patients, immunosuppressed individuals and population subgroups with limited access to healthcare, among others) and mitigate the burden of COVID-19. The following are the specific objectives of unCoVer:

- ▶ To bring together European and international expertise and data to synchronise collaborative research efforts in addressing the ongoing COVID-19 pandemic in a common platform.
- ▶ To continuously monitor, identify and facilitate access and use of COVID-19-related RWD to fully exploit the potential of this routinely collected information, as a reflection of common medical practices.
- ▶ To identify data gaps and marginalised populations to proactively seek synergies with complementary existing and planned clinical databases related to COVID-19.
- ▶ To provide a platform for the use of dissimilar data sources capable of streamlining ethical and legal aspects and anticipating the needs for data harmonisation by innovative computational resources and integrated information for enhanced impact.
- ▶ To bring together expertise on the use of advanced computational, epidemiological and biostatistical methods to handle heterogeneous and multilayered information to facilitate rapid queries and data outputs related to SARS-CoV-2 infection, underlying drivers of COVID-19 prognosis, the safety and effectiveness of treatments, and the sequelae as well as the impact of COVID-19 on health system resources.
- ▶ To broadcast the use and results of the platform to attract new partners and to pursue complementarity with existing similar networks in Europe and internationally to save lives and optimise resources.

METHODS AND ANALYSIS

unCoVer is conceptualised as a functional network of partners capable of harvesting and analysing RWD derived from the response and provision of care to patients by the health systems across Europe and other countries, such as Brazil and Colombia, during the COVID-19 pandemic.

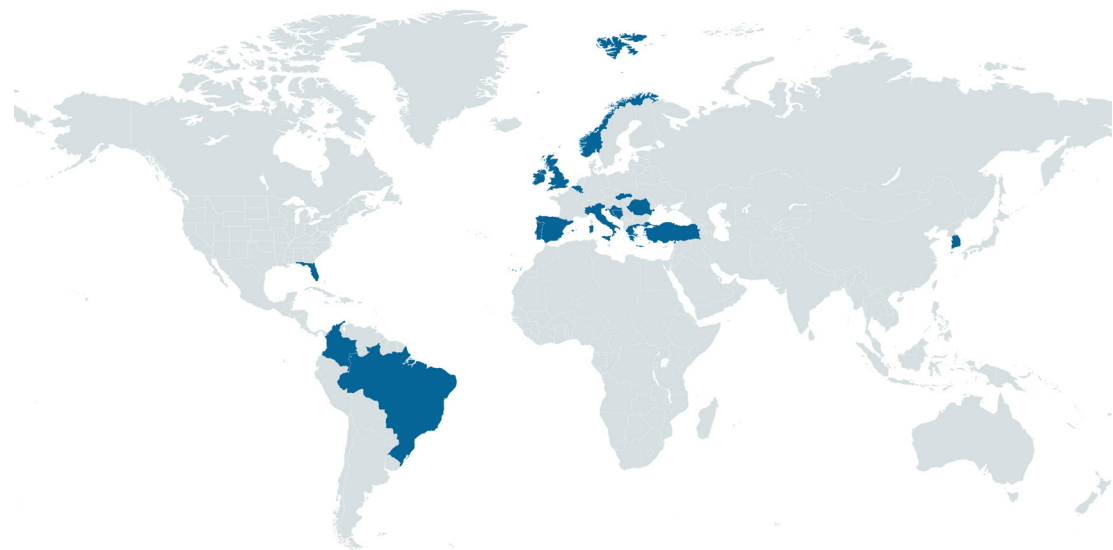


Figure 1 Geographical distribution of the unCoVer network (Belgium, Bosnia and Herzegovina, Brazil, Colombia, Croatia, Greece, Ireland, Italy, Luxembourg, Norway, Portugal, Romania, Slovakia, South Korea, Spain, Turkey, UK and USA). unCoVer, Unravelling data for rapid evidence-based response to COVID-19.

Setting

unCoVer comprises 29 partners from 25 institutions in the European Union (EU) and 4 non-EU partners representing 18 countries, namely Belgium, Bosnia and Herzegovina, Brazil, Colombia, Croatia, Greece, Ireland, Italy, Luxembourg, Norway, Portugal, Romania, Slovakia, South Korea, Spain, Turkey, UK and USA (figure 1). Partners provide data mostly from front-line hospitals but also from national health agencies, registries and investigator-led observational studies, and represent complementary scientific and medical fields, as well as expertise in research ethics, data management and statistical modelling. unCoVer, thus, works as a functional network bridging clinical expertise and data analytics, intending to exploit the full potential of routine healthcare data already collected from patients during the pandemic. The set-up of the network therefore relies on a continuous iteration process between (1) clinical partners, who will guide the development of research questions needed to improve patients' care and inform public health strategies; and (2) epidemiologists and analytical experts, who will operationalise the research questions with advanced data processing, analysis and simulation tools capable of generating innovative solutions. The work of the network is further supported by three external advisory boards (figure 2) that provide expert counselling concerning the relevance of medical research and findings (External Advisory Board), data protection (Data Protection and Ethics Advisory Board, DP-EAB) and stakeholder involvement (Societal and Regulatory Advisory Board, SRAB).

Databases

unCoVer facilitates access to observational data for secondary analyses. These data are largely collected for non-research purposes, RWD, and refer to data generated during patient encounters with the health systems

which have established information technology protocols and tools for retrieving and storing information about the healthcare provided. To date, the unCoVer network incorporates 16 databases of electronic medical records from 10 different countries, 6 national registries, 4 observational cohorts and 2 databases on population screening (table 1). The data available to unCoVer have information on hospitalisation of patients with COVID-19 with at least two time points of data collection: at admission and at discharge. In addition to demographics and clinical/epidemiological data, other types of data such as biospecimens, imaging data, social network-related/contact-tracing-related data, movement-related data and mental health data are also available but with limited access. Clinical/epidemiological data include case report forms at the patient's admission (ie, date of symptom onset and/or admission, signs and symptoms at admission, laboratory results, preadmission medication, comorbidities and risk factors), during hospitalisation (ie, signs and symptoms, laboratory results, supportive treatment, admission to intensive care unit) and at discharge (ie, date of discharge, outcome, as well as medication and complications) (online supplemental table 1).

Structure and activities

Three blocks of activities, grouping work packages (WPs), were designed to build a functional network over a period of 36 months from November 2020 onwards, and aim to facilitate the flow of information for rapid assessment of research questions.

Block 1: definition, design and data harvesting

This first block forms the architectural foundation and the core of the unCoVer network. In order to provide a comprehensive repository of available data, WP1 'Data Identification' collects and catalogues all data in

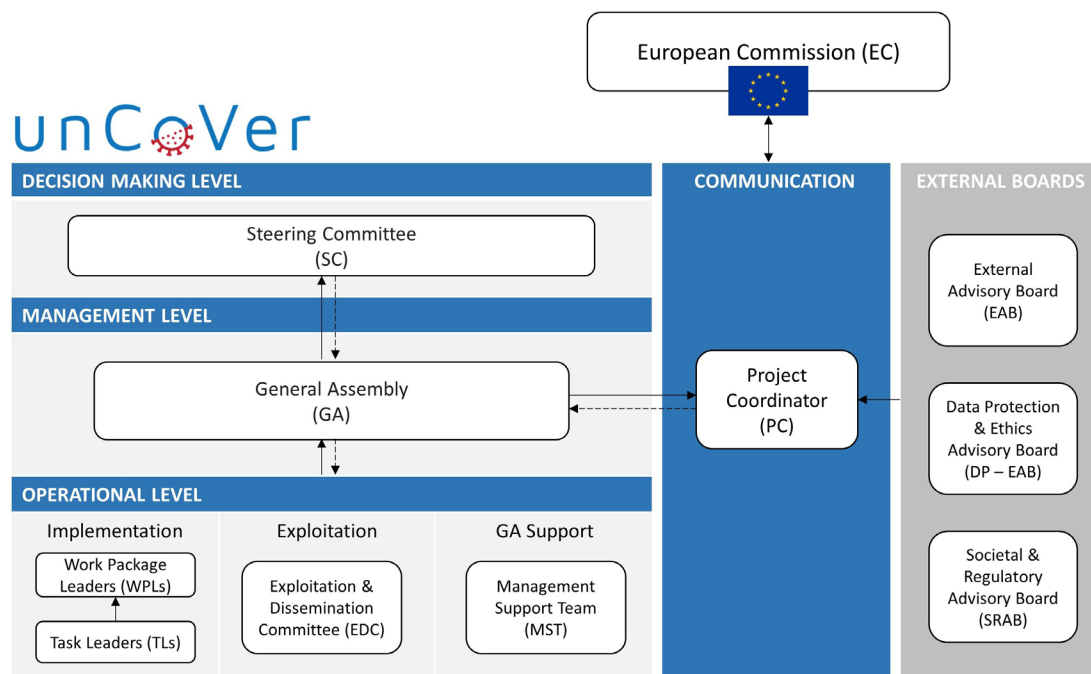


Figure 2 Management structure of the unCoVer network. unCoVer, Unravelling data for rapid evidence-based response to COVID-19.

a standardised way, including a common codebook that specifies the harmonised variables, with standardised variable names and data format and labels, and range of plausible values, in preparation for data harmonisation and validation processes, that is, the key for the development of a unified pool of data for analyses. In parallel, and acknowledging the sensitive nature of health data, and personal information compliance with ethical and legal aspects, a checklist for assessing the risks involved in data processing is implemented by WP2 ‘Ethics and Data Protection’, with due consideration to legal and regulatory issues concerning data protection, privacy and information security. The checklist includes questions on the nature of the data (eg, clinical data, hospital records or publicly available data, personal data, data collected in vulnerable groups, availability of follow-up data), informed consent (eg, explicit consent or assent obtained), data protection (eg, data protection officer identified and data protection impact assessment completed), ethical approval (obtained or pending), data privacy protection (eg, anonymisation vs pseudonymisation, data minimisation), and data transfer and use (data transfer agreement needed, name of the data controller, data processor, joint data controller if applicable, and whether international transfer outside EU). Each data provider within the network populates this checklist, which is reviewed on a case-by-case basis by the unCoVer research ethics expert team together with the independent DP-EAB, and accordingly informed decisions on risk mitigation are taken. In short, WP2 ensures that general data protection regulation (GDPR) or equivalent guidelines are adhered to during the data processing activity. Once a data set is

categorised as low risk, it will be available to proceed with subsequent steps (online supplemental table 2).

For the secure multi-party computation of unCoVer data, WP3 ‘Data Harmonization’ developed an infrastructure based on Opal V.4.1 (OBiBa suite; Maelstrom Research, Montreal, Canada) to facilitate interoperability of the data, including data management, harmonisation and dissemination in a secured environment.¹⁰ The Opal server application provides the necessary key features for data encryption and decryption managed through public key infrastructure, as well as participant identifier management and user authentication/authorisation for access via a rights and roles management with username/password. The following are the steps to achieve data harmonisation and secured data sharing and use: (1) set up the Opal server for each data provider and import relevant data sets; (2) configure a harmonised data description in each Opal server; and (3) run distributed queries on harmonised data sets through the DataSHIELD application, which enables individual-level data analyses across multiple Opal servers without sharing and disclosing any individual-level data (figure 3). Thus, by using computational power and standardising dissimilar information, while complying with ethical and legal requirements, a data repository of anonymised and harmonised COVID-19 RWD will be made available for secured data analyses.

Block 2: analytical development, data use and demonstration

The broad range of medical, public health and research expertise available within the unCoVer network is at the heart of WP4 ‘Data analysis and Outcomes’, dedicated to unCoVer findings from the data acquired to maximise their use in informing COVID-19 response. Activities of

Table 1 Overview of the COVID-19 real-world data among the unCoVer network

Country	Institution	Centres* (n)	Start date	End date	Patients* (n)	Study population†	Number of time points	Variable type									
								Demographics	Clinical/epidemiology	Human Pathogen OMICs	Biospecimens	Imaging	Social network	Movement-related	Mental health	Economic	Diet
Patient data from electronic medical records																	
BA	UNSA	10	17 March 2020	20 May 2021	2000	1, 2, 3	≥2	X	X			X					
BE	ITM-UZA	1	22 April 2020	Ongoing	187	1	≥2	X	X			X			X		
BR	ASPEUR	40	26 March 2020	Ongoing	2000	1, 3	15	X	X	X		X					X
ES	FIHM	12	1 March 2020	15 February 2021	4480	1	≥2	X	X			X					
ES	HULPr	1	1 January 2020	29 September 2020	2217	1	≥2	X	X								
ES	IIS-FJD	1	7 March 2020	31 May 2020	1861	1	≥2	X	X			X					
ES	UNAV	2	1 April 2020	1 January 2021	100	1	≥1	X	X								
HR	INANTRO	2	1 August 2020	Ongoing	200	1	≥1	X	X	X		X			X		
IT	IRCCS-DB1	1	3 March 2020	Ongoing	200	1	≥1	X	X			X					
IT	IRCCS-DB4	1	1 March 2020	9 May 2020	355	4	≥1	X	X								
IT	ULSS6	4	1 March 2020	Ongoing	1000	1	≥3	X	X			X					
RO	UMF Cluj	1	1 March 2020	Ongoing	100	1	≥4	X	X			X					
RO	UMF IASI	1	1 March 2020	Ongoing	150	1	≥2	X	X			X					
SK	TU	1	1 October 2020	Ongoing	800	1	≥1	X	X								
TR	BU-DB1	6	1 March 2020	Ongoing	7000	1	4	X	X								
Public health surveillance data and registers																	
SK	TU	240	1 March 2020	Ongoing	776 000	2	≥1	X	X								X
BE	Sciensano	98	14 March 2020	Ongoing	50 000	1	≥2	X	X			X					
CO	UdeA	1314	14 March 2020	Ongoing	1 027 021	2		X	X								
HR	CIPH	NA	25 February 2020	Ongoing	220 000	2	≥1	X	X								
IE	TUDublin	48	29 February 2020	30 November 2020	74 000	2	≥1	X	X				X		X		X

Continued



Table 1 Continued

Country	Institution (n)	Centres* Start date	End date	Patients* (n)	Study population†	Number of time points	Variable type										
							Demographics	Clinical/epidemiology	Human Pathogen OMICs	Biospecimens	Imaging	Social network	Movement-related	Mental health	Economic	Diet	Screening
NO	USN	31 March 2020	Ongoing	2313	1	10	X	X									X
PT	UPORTO	3 March 2020	Ongoing	830 000	2	≥1	X	X									
Observational research data																	
IT	IRCCS-DB3	22 April 2020	25 May 2020	1515	5	≥1	X										
LU	LIH	15 April 2020	1 April 2021	900	2	5	X	X			X						
PT	IPC	1 March 2020	1 July 2020	550	5	≥1	X										X
UK	SMUC	1 July 2020	1 January 2021	200	2		X										X
Screening data																	
IT	IRCCS-DB2	31 March 2020	12 May 2020	1635	6		X										X
TR	BU-DB2	1 March 2020	Ongoing	18 000	5, 7		X										X

*Number of centres providing information and number of patients/individuals in the data sets could change due to the continuous update of information.

†Study population refers to (1) hospitalised patients with COVID-19, (2) COVID-19 cases, (3) patients with COVID-19 attending primary care services, (4) patients with COVID-19 attending emergency services, (5) general population, (6) healthcare workers and (7) tourists. ASPEUR, Associação Pro Ensino Superior Em Novo Hamburgo; BA, Bosnia and Herzegovina; BE, Belgium; BR, Brazil; BU, Baskent Universitesi Vakfi; CIPH, Croatian Institute of Public Health; CO, Colombia; DB, Database; ES, Spain; FHMM, Fundación Investigación HM Hospitales; HR, Croatia; HULP, Hospital Universitario de La Princesa; IE, Ireland; IIS-FJD, Instituto Investigación Sanitaria Fundación Jiménez Díaz; IPC, Instituto Politécnico de Coimbra; IRCCS, Istituto Don Calabria; IT, Italy; ITM-LUZA, Institute of Tropical Medicine and Antwerp University Hospital; LIH, Luxembourg Institute of Health; LU, Luxembourg; NA, not available at the moment; NO, Norway; PT, Portugal; RO, Romania; SK, Slovakia; SMUC, St Mary's University Twickenham; TR, Turkey; TU, Trnavska Univerzita V Trnave; TUDublin, Technological University Dublin; UdaA, Universidad de Antioquia; UK, United Kingdom; ULSS6, Azienda ULSS6 Euganea; UMF Cluj, Universitatea De Medicina Si Farmacie Iuliu Hatieganu Cluj-Napoca; UMF IASI, Universitatea De Medicina Si Farmacie Grigore T Popa Din Iasi; UNAV, Universidad de Navarra; unCoVer, Unravelling data for rapid evidence-based response to COVID-19; UNSA, Universitet U Sarajevo; UPORTO, Universidade Do Porto; USN, Universitetet I Sørst-Norge.

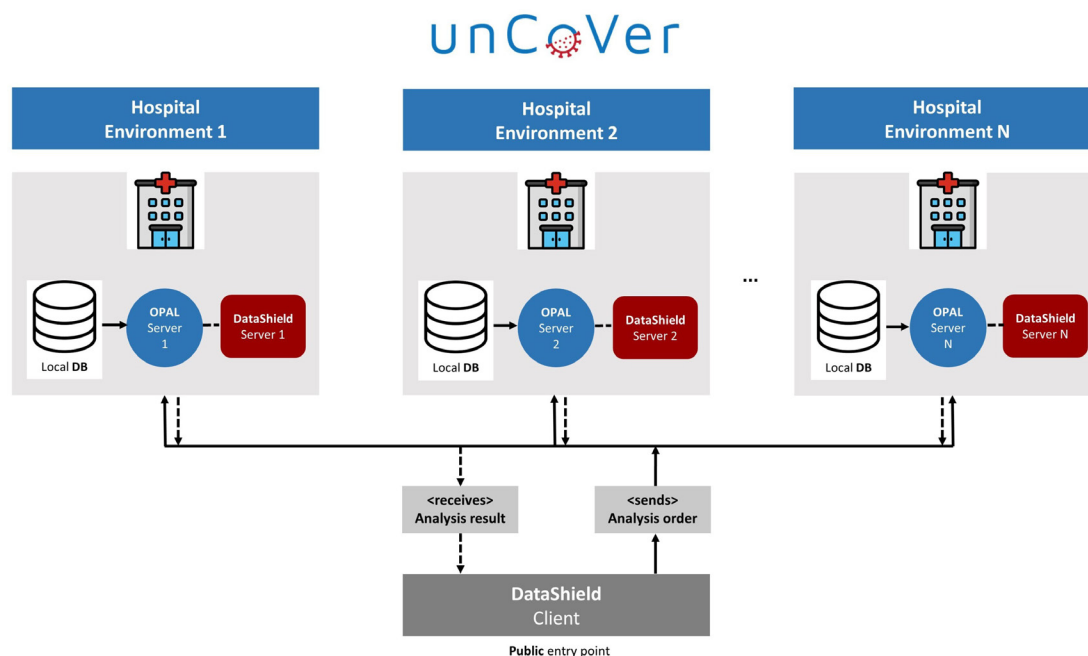


Figure 3 Secure multi-party computation of unCoVer data based on Opal-DataSHIELD infrastructure. unCoVer, Unravelling data for rapid evidence-based response to COVID-19.

Abbreviation: **DB**, database.

this block aim to facilitate and streamline rapid response to identified research gaps using the unCoVer infrastructure developed in block 1. With the availability of cross-national harmonised RWD within the unCoVer network, a large amount of hospitalised and discharged patients who had COVID-19 will be studied in depth, together with complementary epidemiological data, to understand the pathophysiology, progression, treatment (long-term) complications and (less frequent) risk factors for early prevention of this novel disease, as well as to grasp the cross-national heterogeneity of the COVID-19 burden. To this end, an analytical toolbox, including both traditional statistical methods and machine learning techniques and a Bayesian estimation framework, will be developed for identifying relationships between early clinical and diagnostic profiles and the future course of the infection, and for a detailed clinical and epidemiological characterisation of patients with COVID-19, being able to generate patients' risk classification and risk prediction. The application of this toolbox in real patient data, available within the unCoVer network, will then allow uncovering real-world insights that would support policies and protocols for optimisation of health resources of the hospital and critical care. Therefore, the activities of this block will be related to the iterative trial use of the Opal server and the toolbox by end users (ie, data analysts), as well as the lessons learnt and potentially preparing actions for the sustainability of the unCoVer platform, including the repository and toolbox.

Block 3: project management, communication and exploitation

The outputs of the network, including the scientific and technological knowledge and outcomes, provided by the

previous two blocks, are streamlined through scientific publications, training and educational activities, organisation and participation in events, among others, steered by WP5 'Communication, Exploitation & Dissemination'. This last block is also dedicated to maintaining the functional network both internally and externally, and with special attention to the management of intellectual property by using best practices in project coordination, as outlined in WP6 'Coordination'. Concerning this, the unCoVer organisation structure works under a consortium agreement signed by all partners and includes the following key bodies within the consortium and management structure (figure 2): a steering committee formed by the principal investigators of the 29 partner institutions at the decision-making level; and a general assembly (GA), involving WP leaders, which manages the network and coordinates WP and tasks leaders, assisted also by a management support team to reinforce partners' representation. An internal Exploitation and Dissemination Committee (EDC) also collaborates in the overall management of the GA. Finally, the project coordinators communicate with the sponsor and facilitate crosstalk between the network and external advisory boards.

ETHICS AND DISSEMINATION

The unCoVer study has been approved by the Institutional Review Board of the Institute of Tropical Medicine in Antwerp (IRB/RR/ac/151, protocol number 1524/21). Ethical aspects are of utmost importance in unCoVer. On the project start date, the roles and responsibilities of the independent DP-EAB were

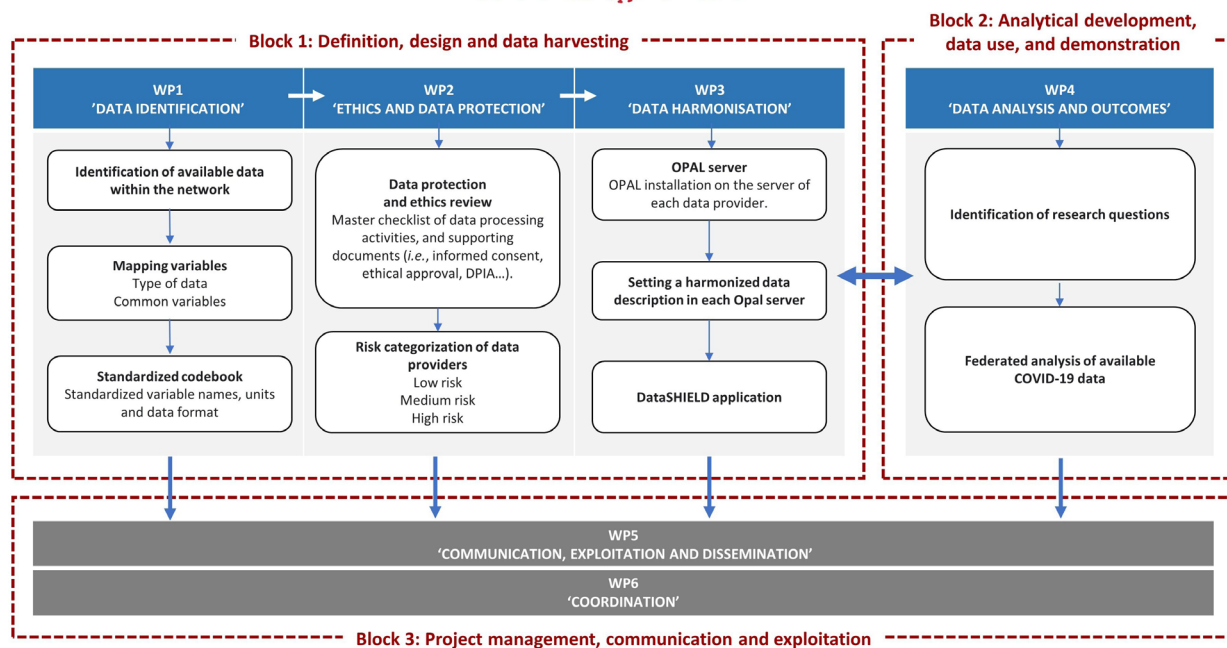


Figure 4 Data management process of the COVID-19-related data available within the unCoVer network. DPIA, data protection impact assessment; unCoVer, Unravelling data for rapid evidence-based response to COVID-19; WP, work package.

described, including the selection procedure of the board members and its final composition and mandate. A ‘scoping exercise’ was also conducted across the network to ensure that all partners are aware of the common obligations in terms of data processing activities using health or health-related data according to European and international guidelines. Moreover, to be compliant with the GDPR and meet the ethics requirements, the unCoVer network will follow the data processing steps represented in figure 4, in the following sequence:

- ▶ The unCoVer master checklist of data processing activities in network partners is circulated within the network to be completed by the data providers, and data providers are required to provide the supporting documentation of each indicator of this list, such as informed consents, ethical approvals and data protection impact assessments. This information is processed by the research ethics team, responsible for categorising the data sets into three different categories: low, medium or high data privacy risk.
- ▶ Data sets categorised as ‘low-risk’ will be available to proceed with the harmonisation process and therefore Opal-DataSHIELD servers will be installed. Within Opal, the patient identifiers will be separated from the patient study data by employing two databases: (1) the identifier (ID) database, which stores patient identifiers accessible to the data provider only; and (2) the study database, which stores pseudoanonymised patient data to be used for data analyses accessible, through code only, to data analysts. The ‘medium-’ and ‘high-risk’ datasets

will be subject to further review and requirements before harmonisation processes.

- ▶ Finally, the installation of the servers will allow the consortium to analyse the available RWD through an anonymisation layer to answer the preidentified research questions. The system also facilitates the definition of analytical projects and the specific databases and/or variables that will be used for a specific project. As a rule, all output of data analytics will be restricted to the presentation of data aggregates or to line listing deprived of personal identifiers so that the identity of the study patient cannot be deduced (no backward identification).

To maximise the unCoVer network’s output, dissemination and exploitation strategies, as planned by the EDC and advised by the SRAB, that is, a non-executive consulting substructure composed of several key stakeholders from the regulatory, governance, civil society level and patients’ public initiatives, will be segmented according to the network block activities, the potential users and the most adequate channels of dissemination and interaction with potential users. The website (uncover-eu.net/), social media accounts (Twitter @uncoverEU, LinkedIn, YouTube) and the project newsletter will be the channels to reach all partners and stakeholders of unCoVer, both devoted to providing regular updates on project activities and announcing upcoming milestones and events. The website will serve as a repository of the project goals and activities and deliverables in an easy-to-understand language, as well as publications, lectures

and expert documents hosted for access by the partners or stakeholders.

PATIENT AND PUBLIC INVOLVEMENT

The unCoVer network has been designed to facilitate interactions and enhance outreach to COVID-19 stakeholders included in external advisory boards, as well as prominent WPs on dissemination activities, which include but are not limited to the following:

- ▶ *Scientific community.* To contribute to the body of knowledge in the field, two types of publications in peer-reviewed journals under open-access schemes are foreseen: the unCoVer network publications, that is, for implementation and/or application of the unCoVer platform as a whole; and the unCoVer partners publication, that is, for specific collaborations between two or more unCoVer partners. In addition to both types of scientific publications, jointly organised workshops, virtual trainings and virtual conferences will be instrumental channels for the dissemination of the unCoVer activities and results to the scientific community. In such a yearly organised workshop, the application of the unCoVer repository and toolbox will be presented and expert feedback will be sought for further improvement. These activities of dissemination to the scientific community will result in overall awareness and international recognition of the unCoVer network, simultaneously strengthening the visibility and competitiveness of the institutions involved as centres of excellence.
- ▶ *European platforms and data-driven initiatives.* Cooperation with other European projects dedicated to COVID-19 data sharing such as Orchestra (orchestra-cohort.eu), Synchros (synchros.eu), Dragon (imi.europa.eu/projects-results/project-factsheets/dragon), RecodID (recodid.eu) and EC-COVID-19 Data Platform (covid19dataportal.org), as well as large networks such as the European Burden of Disease Network (burden-eu.net) and initiatives on data sharing infrastructures such as the Population Health Information Research Infrastructure (phiri.eu), will be established for the co-organisation of dissemination events along with seeking alignment and synergies to avoid duplication of efforts.
- ▶ *Policymakers.* The accumulated prior experience and contact networks in the regulatory, policymaking framework of several members of the unCoVer network will be used to ensure that the work and output created can reach regulatory entities and policymakers, thus contributing to the impact of the project on the decision-making process. This appears relevant as the unCoVer network is willing to merge different sources of medical data with social, economic, mental and geographical data with the potential to identify highly tailored profiles of risk for community prevention programmes and educational goals in different countries.

- ▶ *Engagement with key stakeholders.* Results from the project will be further disseminated through involvement with societal, regulatory and administrative partners, as ensured by the external advisory boards to the network, with the goal to warrant the project's impact according to stakeholders' needs and expectations.
- ▶ *General public.* To maximise awareness of unCoVer among the general public, the project activities and milestones will be broadcasted via social media.
- ▶ *Patient associations and clinicians.* Given the foreseen impact of unCoVer's output, patient associations and clinicians are identified as end users that benefit from the data-oriented results, and subsequently these can be translated to them via seminars, lectures and infographics as made available on the project website, among others.

DISCUSSION

During the early phases of the pandemic, the unCoVer network grew organically from initial partnerships accessing individual databases to answer pressing COVID-19 questions. These initial collaborations soon recognised the need for accessing extended information to develop more robust analytical models and outputs. To this end, the concept of unCoVer originates in the shared interest of its partners to synchronise the research efforts needed for exploiting and valorising the underutilised and inexpensive RWD in addressing the ongoing COVID-19 pandemic. Within a limited duration of 2 years, efforts of the first year were mainly focused on the organisation of the network and the set-up of a collaborative federated data infrastructure, accounting for local, national and international ethical and data protection guidelines and streamlining procedures for data processing activities, including data identification, sharing, harmonisation, validation and analytics. Entering the second year, the unCoVer network aims to demonstrate the usability of the developed infrastructure on the combined use of RWD to address clinical and epidemiological research questions related to the COVID-19 pandemic, both revising questions answered with limited data sets and new questions arising from the evolving pandemic. The unCoVer network, therefore, serves as a proof of concept for building a federated data infrastructure facilitating data interoperability in a secured environment, while complying with ethical and data protection guidelines. Moving forward, the established framework of unCoVer provides valuable input for the use of complementary RWD with robust methodologies at hand for the still ongoing current pandemic and future pandemics.

Author affiliations

¹Unit of Non-Communicable Diseases, Department of Public Health, Institute of Tropical Medicine, Antwerp, Belgium

²Department of Epidemiology and Biostatistics, Faculty of Medicine, University of Sarajevo, Sarajevo, Bosnia and Herzegovina

³Public Health Department, Baskent University Faculty of Medicine, Ankara, Turkey

- ⁴Scientific-Pedagogical Unit of Dietetics and Nutrition, Coimbra Health School, Polytechnic Institute of Coimbra, Coimbra, Portugal
- ⁵Department of Infectious Tropical Diseases and Microbiology, IRCCS Ospedale Sacro Cuore Don Calabria, Negrar, Italy
- ⁶Centre for Applied Bioanthropology, Institut za antropologiju, Zagreb, Croatia
- ⁷Epidemiology and Public Health, Sciensano, Brussels, Belgium
- ⁸Department of Veterinary Public Health and Food Safety, Ghent University Faculty of Veterinary Medicine, Merelbeke, Belgium
- ⁹Epidemiology Group, National College of Public Health, University of Antioquia, Medellin, Colombia
- ¹⁰Centro de Biotecnologia e Química Fina (CBQF), Escola Superior de Biotecnologia, Universidade Católica Portuguesa, Porto, Portugal
- ¹¹Center for Biomedical Technology, Universidad Politécnica de Madrid, Madrid, Spain
- ¹²Biomedical Engineering and Telemedicine Centre, Universidad Politécnica de Madrid, Madrid, Spain
- ¹³Centro de Investigación Biomédica en Red de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Madrid, Spain
- ¹⁴Environmental Sustainability and Health Institute, Technological University Dublin, Dublin, Ireland
- ¹⁵School of Public Health, University College Cork, Cork, Ireland
- ¹⁶Department of Nursing and Health Sciences, University of South-Eastern Norway, Kongsberg, Norway
- ¹⁷Faculty of Public Health, University of Thessaly, Volos, Greece
- ¹⁸Clinical Research Unit, Instituto de Investigación Sanitaria de la Fundación Jiménez Díaz, Madrid, Spain
- ¹⁹Faculty of Medicine, Iuliu Hatieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania
- ²⁰Institute for Global Health and Epidemiology, Trnava University in Trnava, Trnava, Slovakia
- ²¹Instituto de Salud Tropical (ISTUN), Department of Microbiology and Parasitology, Navarra Institute for Health Research (IdiSNA), Universidad de Navarra, Pamplona, Spain
- ²²Department of Preventive Medicine, Kyung Hee University, Seoul, Korea (the Republic of)
- ²³REQUIMTE/LAQV, Laboratório de Farmacognosia, Departamento de Química, Faculdade de Farmácia, Universidade do Porto, Porto, Portugal
- ²⁴College of Public Health, University of South Florida, Tampa, Florida, USA
- ²⁵Department of Psychology and Pedagogic Science, St Mary's University Twickenham, UK
- ²⁶Servicio de Neumología, Hospital Universitario de La Princesa, Universidad Autónoma de Madrid, Madrid, Spain
- ²⁷Centro de Investigación en Red de Enfermedades Respiratorias (CIBERES), Instituto de Salud Carlos III (ISCIII), Madrid, Spain
- ²⁸COVID-19 Clinical Management Team, WHO Health Emergency Programme, World Health Organization HQ, Geneva, Switzerland
- ²⁹Feevale University, Novo Hamburgo, Brazil
- ³⁰Azienda ULSS6 Euganea, Padova, Italy
- ³¹Clinical Hospital of Pulmonary Diseases Iasi, Clinical Hospital of Pulmonary Diseases Iasi, Iasi, Romania
- ³²University of Medicine and Pharmacy Grigore T. Popa Iasi, Iasi, Romania
- ³³Translational Medicine Operations Hub, Competence Centre for Methodology and Statistics, Luxembourg Institute of Health, Strassen, Luxembourg
- ³⁴Department of General Internal Medicine, Infectious Diseases and Tropical Medicine, University Hospital Antwerp, Edegem, Belgium
- ³⁵Croatian Institute of Public Health, Zagreb, Croatia
- ³⁶Cardiology Department, Hospital Universitario HM Montepíncipe, Boadilla del Monte, Spain
- ³⁷Centro Nacional de Investigaciones Cardiovasculares (CNIC), Instituto de Salud Carlos III, Madrid, Spain
- ³⁸Fundación de Investigación HM Hospitales, Madrid, Spain

Twitter José L. Peñalvo @JosePenalvo, Gloria Soriano @Glosorpal and on behalf of the unCoVer network @uncoverEU

Collaborators The unCoVer network: Institute of Tropical Medicine and Antwerp University Hospital, Belgium (José L. Peñalvo, Ely Mertens, Marianne Van der Sande, Patrick Soentjens, Diana Sagastume, James Cottam, Gloria Soriano, Sabrina Van Ierssel and Hanne Van Tiggelen); Fundación Investigación HM Hospitales, Spain (José M Castellano, Paula Villares, José Barberán, Mercedes Villareal, Justo

Menéndez, Nerea Ruiz del Árbol and Alberto Estirado); Universidad Politécnica de Madrid, Spain (Ernestina Menasalvas, Enrique Javier Gómez, Alberto Blázquez Herranz, David Fernandez Lobón and Paloma Chausa); Universidad de Navarra, Spain (Paul Nguewa); Universidade do Porto, Portugal (David M Pereira and Morteza Hosseini); Technological University Dublin, Ireland (Paul Hynds, John Kelleher and Elizabeth Hunter); University College Cork, Ireland (Zubair Kabir, Ella Arensman, Brendan Palmer and Georgie O'Sullivan); Universitatea De Medicina Si Farmacie Iuliu Hatieganu Cluj-Napoca, Romania (Milena Man, Lucia Maria Lotrean, Mihaela Lupse and Mira Florea); Universitatea De Medicina Si Farmacie Grigore T Popa Din Iasi, Romania (Antigona Carmen Trofor, Andrei Tudor Cernomaz, Radu Adrian Crisan-Dabija and Cristina Grigorescu); Luxembourg Institute of Health (Michel Vaillant and Guy Fagherazzi); Universidade Católica Portuguesa, Portugal (João C Fernandes and João Silva); Trnava University, Slovakia (Marek Majdan, Daria Rabarova, Adriana Krsakova, Jaroslava Brnova, Janka Prnova, Jaroslav Slany and Dominika Plancikova); Instituto Politécnico de Coimbra, Portugal (Ana Lucia Baltazar); Hospital Universitario de La Princesa, Spain (Joan B Soriano, Julio Ancochea, Nisa Boukichou Abdelkader, Adrián Peláez, Elena Ávalos and Gorane Iturricastillo); Instituto Investigación Sanitaria Fundación Jiménez Díaz, Spain (Lucía Llanos, Miguel Górgolas, Olga Sánchez-Pernaute, Arnaldo Santos Oviedo, Sergio Luis Lima, Antonio Herrero and Pablo Minguez); Panepistimio Thessalias, Greece (Polychronis Kostoulas, Olympia Lioupi, Eleftherios Meletis, Konstantinos Pateras and Costas Tsiamis); Universitetet I Sorost-Norge, Norway (Jörn Klein and Mustafa Asfari); Istituto Don Calabria, Italy (Dora Buonfrate, Tamara Ursini and Nicoletta De Santis); Sciensano, Belgium (Brecht Devleeschauwer, Petronille Bogaert, Koen Blot, Miriam Saso and Mathil Vandromme); Croatian Institute of Public Health, Croatia (Jakov Vukovic, Ivan Pristas, Tamara Poljicanin, Jelena Dimnjakovic, Marko Brkic and Marija Svajda); Institut Za Antropologiju, Croatia (Miran Čoklo, Saša Missoni, Jelena Šarac, Natalija Novokmet, Luka Bočkor, Ivan Dolanc, Antonija Jonjić and Iva Šunić); Baskent Üniversitesi Vakfi, Turkey (Seval Akgun, Tugba Gürgen Erdogan, Süleyman Çetinkünar, Cenk Belibağlı, Kübra Demir, Mustafa Görür, Turgut Bulut and KR Nayar); St Mary's University Twickenham, UK (Silvia Riva); Azienda ULSS6 Euganea, Italy (Mary Elizabeth Tamang, Carlo Giordani and Petra Golin); Korea University, South Korea (In-Hwan Oh and Seok Jun Yoon); University of South Florida, USA (Miguel Reina Ortiz); Universidad de Antioquia, Colombia (Paula Andrea Diaz Valencia, Lina Ruiz, Juan Pablo Pérez Bedoya, Oscar Ignacio Mendoza, Camilo Hincapie, Boris Rodríguez and Noël Barengo); Feevale University, Associação Pro Ensino Superior Em Novo Hamburgo, Brazil (Fernando Spilki, Juliane Deise Fleck and Matheus Nunes Weber); Univerzitet U Sarajevu, Bosnia and Herzegovina (Enisa Ademović, Lejla Burnazović-Ristić, Semra Čavaljuga, Džan Ahmed Jesenković and Lejla Džananović).

Contributors JP coordinates the unCoVer network. GS manages overall project activities. JP, EM, ErM, EJG, ZK, JMC and DMP lead the project's work packages. JP, EM, EA, SA, ALB, DB, MC, BD, PADV, PH, JK, LLJ, LML, MM, PN, I-HO, DMP, SR, JS, FS, MET, ACT, MV, SVI, JV and JMC characterised the data available to the network. JP, EM, EJG, ErM and GS designed the data harmonisation system and access to the federated infrastructure. JP, ZK, ErM, GO'S, SR and GS supervised compliance with legal and ethical requirements of data use. JP, JCF, JK, PK, MM, EM, DMP, MRO and GS are responsible for the dissemination of the network's activities. JP, EM, GS, ZK, ErM, JMC, DMP, SR and JS drafted the manuscript and designed the figures. All authors have critically reviewed and approved the final version of the manuscript. The unCoVer network consists of all individual partner institutions and investigators who are responsible for the identification, characterisation, and compliance with legal and ethical requirements of data use, as well as dissemination and coordination activities.

Funding This project is funded by the European Union's Horizon 2020 Research and Innovation Programme (Grant Agreement No 101016216).

Map disclaimer The inclusion of any map (including the depiction of any boundaries therein), or of any geographic or locational reference, does not imply the expression of any opinion whatsoever on the part of BMJ concerning the legal status of any country, territory, jurisdiction or area or of its authorities. Any such expression remains solely that of the relevant source and is not endorsed by BMJ. Maps are provided without any warranty of any kind, either express or implied.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and

responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

José L. Peñalvo <http://orcid.org/0000-0003-0096-6238>

Elly Mertens <http://orcid.org/0000-0003-4957-4235>

Dora Buonfrate <http://orcid.org/0000-0003-0108-6822>

Jörn Klein <http://orcid.org/0000-0002-1803-4167>

Miguel Reina Ortiz <http://orcid.org/0000-0003-3309-6881>

Michel Vaillant <http://orcid.org/0000-0003-4714-8128>

REFERENCES

- World Health Organization. Expanding our understanding of post COVID-19 condition: report of a who webinar, 2021 [Accessed 09 Feb 2021].
- Zhou F, Yu T, Du R, *et al*. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 2020;395:1054–62.
- Shi Y, Yu X, Zhao H, *et al*. Host susceptibility to severe COVID-19 and establishment of a host risk score: findings of 487 cases outside Wuhan. *Crit Care* 2020;24:108.
- World Health Organization (WHO). COVID-19 clinical management: living guidance, 2021. Available: <https://www.who.int/publications/i/item/WHO-2019-nCoV-clinical-2021-1>
- Jin Y, Ji W, Yang H, *et al*. Endothelial activation and dysfunction in COVID-19: from basic mechanisms to potential therapeutic approaches. *Signal Transduct Target Ther* 2020;5:293.
- Wan Y, Shang J, Graham R, *et al*. Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. *J Virol* 2020;94:e00127–20.
- Fang L, Karakiulakis G, Roth M. Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? *Lancet Respir Med* 2020;8:e21.
- Kassir R. Risk of COVID-19 for patients with obesity. *Obes Rev* 2020;21:e13034.
- Petrilli CM, Jones SA, Yang J, *et al*. Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in New York City: prospective cohort study. *BMJ* 2020;369:m1966.
- Doiron D, Marcon Y, Fortier I, *et al*. Software application profile: opal and MICA: open-source software solutions for epidemiological data management, harmonization and dissemination. *Int J Epidemiol* 2017;46:1372–8.