BMJ Open

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

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-055630
Article Type:	Protocol
Date Submitted by the Author:	18-Jul-2021
Complete List of Authors:	Penalvo, Jose; Institute of Tropical Medicine, Unit of Non-Communicable Diseases Mertens, Elly; Institute of Tropical Medicine, Ademović, Enisa; University of Sarajevo, Department of Epidemiology and Biostatistics Akgun, Seval; Baskent University Faculty of Medicine, Public Health Department Baltazar, Ana Lúcia; Polytechnic Institute of Coimbra, Scientific-Pedagogical Unit of Dietetics and Nutrition Buonfrate, Dora; IRCCS Ospedale Sacro Cuore Don Calabria, Centre for Tropical Diseases Čoklo, Miran; Institut za antropologiju, Centre for Applied Bioanthropology Devleesschauwer, Brecht; Sciensano, Epidemiology and public health; Ghent University Faculty of Veterinary Medicine, Department of Veterinary Public Health and Food Safety Diaz Valencia, Paula Andrea; University of Antioquia, Epidemiology Group, National College of Public Health Fernandes, João C; Universidade Católica Portuguesa Escola Superior de Biotecnologia, Centro de Biotecnologia e Química Fina (CBQF) Gómez, Enrique Javier; Universidad Politécnica de Madrid, Center for Biomedical Technology; Universidad Politécnica de Madrid, Biomedical Engineering and Telemedicine Centre Hynds, Paul; Technological University Dublin, Environmental Sustainability & Health Institute Kabir, Zubair; University College Cork, School of Public Health Klein, Jörn; University of South-Eastern Norway, Kostoulas , P; University of Thessaly Llanos Jiménez, Lucía; Instituto de Investigacion Sanitaria de la Fundacion Jimenez Diaz, Clinical Research Unit Lotrean, Lucia Maria; Iuliu Hagieganu University of Medicine and Pharmacy Faculty of Medicine Majdan , Marek ; Trnava University in Trnava, Menasalvas, Ernestina; Universidad Politecnica de Madrid Escuela Universitaria de Ingenieria Tecnica de Telecomunicacion, Biomedical Engineering and Telemedicine Centre Nguewa, paul; Universidad de Navarra Oh, In-Hwan; Kyung Hee University O'Sullivan, Georgie; University College Cork, School of Public Health Pereira, David M.; Universidade do Porto, 23 REQUIMTE/LAQV, Laboratório de Farmacognosia, Departamento de Química

	Reina Ortiz, Miguel; University of South Florida, Riva, Silvia; St Mary's University Twickenham, Department of Public Health Soriano, Gloria; Institute of Tropical Medicine, Department of Public Health Soriano, Joan; Universidad Autonoma de Madrid, Servicio de Neumología, Hospital Universitario de La Princesa Spilki, Fernando; FEEVALE University Tamang, Mary Elizabeth; Azienda ULSS 6 Euganea Trofor, Antigona; Universitatea de Medicina si Farmacie Gr T Popa Iasi, Pulmonary Diseases I Vaillant, Michel; Luxembourg Institute of Health, Competence Centre for Methodology and Statistics Van Ierssel, Sabrina; University Hospital Antwerp, Department of General Internal Medicine, Infectious diseases and Tropical Medicine Vukovic, Jakov; Croatian Institute of Public Health Castellano, Jose Maria; Hospital Universitario HM Montepríncipe, Cardiology Department
Keywords:	COVID-19, STATISTICS & RESEARCH METHODS, Public health < INFECTIOUS DISEASES

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

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], Antigona Carmen Trofor[31,32], Michel Vaillant[33], Sabrina Van Ierssel[34], Jakov Vukovic[35], and José M. Castellano[36,37,38] **on behalf of the unCoVer Network**

1 Department of Public Health, Institute of Tropical Medicine, Antwerp, Belgium.

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

3 Public Health Department, Baskent University School of Medicine, Turkey

4 Scientific-Pedagogical Unit of Dietetics and Nutrition, Coimbra Health School, Polytechnic Institute of Coimbra, Portugal

5 Department of Infectious Tropical diseases and Microbiology, IRCCS Sacro Cuore Don Calabria hospital, Verona, Italy

6 Centre for Applied Bioanthropology, Institute for Anthropological Research, Zagreb, Croatia

7 Department of Epidemiology and Public Health, Sciensano, Belgium

8 Department of Veterinary Public Health and Food Safety, Faculty of Veterinary Medicine, Ghent University, Belgium

9 Epidemiology group, National College of Public Health, University of Antioquia, Medellin, Colombia.

10 Centro de Biotecnologia e Química Fina (CBQF), Escola Superior de Biotecnologia da Universidade Católica Portuguesa, Porto, Portugal

11 Center for Biomedical Technology, Universidad Politécnica de Madrid, Pozuelo de Alarcón, Spain.

12 Biomedical Engineering and Telemedicine Centre, ETSI Telecomunicación, Universidad Politécnica de Madrid, Madrid, Spain.

13 Centro de Investigación Biomédica en Red de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Madrid, Spain.

14 Environmental Sustainability & Health Institute, Technological University Dublin, Ireland

15 School of Public Health, University College Cork, Ireland

16 University of South-Eastern Norway, Department of Nursing and Health Sciences, Norway

17 Faculty of Public Health, University of Thessaly, Greece

18 Clinical Research Unit, Instituto Investigación Sanitaria Fundación Jiménez Díaz (IIS-FJD, UAM), Spain

19 Iuliu Hatieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania

20 Institute for Global Health and Epidemiology, Trnava University, Trnava, Slovakia

21 Instituto de Salud Tropical (ISTUN), Department of Microbiology and Parasitology, Navarra Institute for Health Research (IdiSNA), University of Navarra, Pamplona, Spain.

22 Department of preventive Medicine, Kyung Hee University, Seoul, Republic of Korea

23 REQUIMTE/LAQV, Laboratório de Farmacognosia, Departamento de Química, Faculdade de Farmácia, Universidade do Porto, Portugal

24 College of Public Health, University of South Florida, MI, USA

25 Department of Psychology and Pedagogic Science, St Mary's University Twickenham, London, UK

26 Servicio de Neumología, Hospital Universitario de La Princesa, Universidad Autónoma de Madrid, Madrid, Spain

27 Centro de Investigación en Red de Enfermedades Respiratorias (CIBERES), Instituto de Salud Carlos III (ISCIII), Madrid, Spain.

28 COVID-19 Clinical Management Team, WHO Health Emergency Programme, World Health Organization HQ, Geneva, Switzerland

29 Feevale University, Novo Hamburgo, Brazil

30 Azienda ULSS6 Euganea, Padova, Italy

31 University of Medicine and Pharmacy Grigore T. Popa Iasi, Romania

32 Clinical Hospital of Pulmonary Diseases Iasi, Romania

33 Translational Medicine Operations Hub, Competence Center for Methodology and Statistics, Luxembourg Institute of Health, Luxembourg

34 Department of General Internal Medicine, Infectious diseases and Tropical Medicine, Antwerp University Hospital, Edegem, Belgium

35 Croatian Institute of Public Health, Zagreb, Croatia

36 Cardiology Department, Hospital Universitario HM Montepríncipe, HM Hospitales, Madrid, Spain.

37 Centro Nacional de Investigaciones Cardiovasculares (CNIC), Instituto de Salud Carlos III, Madrid, Spain

38 Fundación de Investigación HM Hospitales, Madrid, Spain

*Corresponding Author: José L. Peñalvo, PhD, Department of Public Health, Institute of Tropical Medicine, Antwerp (Belgium) Tel: +32(0)32476251 jpenalvo@itg.be | @JosePenalvo

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 utilizing real-world data (RWD) derived from the response and provision of care to COVID-19 patients by health systems across Europe and elsewhere. unCoVer aims to exploit the full potential of these information to rapidly address clinical and epidemiological research questions arising from the evolving pandemic.

Methods and analysis. From the onset of COVID-19 pandemic, partners are gathering RWD from electronic health records currently including information from over 22,000 COVID-19 hospitalized patients, 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 allows processing 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 activities, scientific publications, and conference communications.

Article Summary

- The unCoVer network includes a sizable number of partners that will exploit the full potential of real-world data derived from the provision of care to COVID-19 patients by health systems across Europe and elsewhere.
- With the availability of international harmonized RWD provided by the unCoVer network, a large amount of information on COVID-19 patients will be studied to better understand aspects of the pathophysiology, progression and treatment, and 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 GDPR 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.

.

INTRODUCTION

The outbreak of the coronavirus disease (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was declared a Public Health Emergency of International Concern by the World Health Organisation on 30 January 2020, and a pandemic, on 11 March 2020. Despite the deployment of public health measures such as restrictions of movements, gathering, and 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 condition[1]. Early epidemiological data of 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 hospitalization 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 multi-dimensional and dynamic nature of the inter-related factors associated with individual responses to 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 has been accumulated as part of

the health systems' efforts to fight COVID-19. These RWD reflect the impact of COVID-19 in patient's health and characterize the protocols of health care 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, those most at risk of severe COVID-19, are often excluded. Harmonization of data from different sources allows for comparison across health systems and improves patients' characterization using the wider heterogeneity of the 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 network aims to provide a research platform for the expert use of RWD, by bringing together complementary data, medical and scientific expertise to address the still urgent questions regarding 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 the health systems, improve patient prognosis especially among those more vulnerable (e.g., chronic patients, immunosuppressed individuals, and population subgroups with limited access to health care, among others), and mitigate the burden of COVID-19. The specific objectives of unCoVer are:

1. To bring together European and international expertise and data to synchronize collaborative research efforts in addressing the ongoing COVID-19 pandemic in a common platform.

2. To continuously monitor, identify and facilitate the 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.
3. To identify data gaps, and marginalized populations to proactively seek synergies with complementary existing and planned clinical databases related to COVID-19.
4. To provide a platform for the use of dissimilar data sources capable of streamlining ethical and legal aspects and anticipating the needs for data harmonization by innovative computational resources and integrated information for enhanced impact.
5. To bring together expertise on the use of advanced computational, epidemiological and biostatistical methods to handle heterogeneous, and multi-layered 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 sequelae, as well as the impact of COVID-19 in health system resources.
6. 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 optimize resources.

METHODS AND ANALYSIS

unCoVer is conceptualized 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.

Setting

unCoVer comprises 29 partners from 25 institutions in the European Union 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, United

Kingdom, and the United States of America (Figure 1). Partners provide data mostly from front-line hospitals but also national health agencies, registries, and investigator-lead 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 the routine healthcare data already collected from patients during the pandemic. The set-up of the network therefore relies on a continuous iteration process between a) clinical partners, who will guide the development of research questions needed to improve patient's care and inform public health strategies; and b) epidemiologists and analytical experts, who will operationalize 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 the medical research and findings (External Advisory Board, EAB), data protection (Data Protection and Ethics Advisory Board, DP-EAB), and stakeholder involvement (Societal and Regulatory Advisory Board, SRAB).

Databases

unCoVer hosts longitudinal observational secondary-level data, which are largely collected for non-research purposes, RWD, and refer to data generated during patient encounters with the health system 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 has information on hospitalization of COVID-19 patients with at least two-time points of data collection, at admission and discharge. In addition to demographics, and clinical/epidemiological data, other data types such as bio-specimens, imaging data, social network-/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 (CRF) at patient's admission (i.e. date of symptom onset and/or admission, signs and symptoms at admission, laboratory results, pre-admission medication, comorbidities & risk factors), during hospitalization (i.e. signs and symptoms, laboratory results, supportive treatment, admission to ICU) and at discharge (i.e. date of discharge, outcome, as well as medication and complications) (Supplementary Table 1).

STRUCTURE AND ACTIVITIES

Three blocks of activities, grouping work packages (WPs), were designed to build a functional network, 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 a standardised way, including a common codebook that specifies the harmonized variables, with standardised variable names and data format and labels, in preparation for data harmonization processes *i.e.* 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 (e.g. clinical data, hospital records or publicly available data, personal data, data collected in vulnerable groups, availability of follow-up data), informed consent (e.g. explicit consent or assent obtained), data protection (e.g. data protection officer identified and data protection impact assessment completed), ethical approval (obtained or pending), data privacy protection (e.g. 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 GDPR or equivalent guidelines are adhered to during the data processing activity. Once a dataset is categorised as low risk, it will be available to proceed with subsequent steps (**Supplemental Table 2**).

For the secure multi-party computation of unCoVer data, WP3 'Data Harmonization', developed an infrastructure based on Opal 4.1 (OBiBa suite, Maelstrom Research, Montreal, Canada) to facilitate interoperability of the data, including data management, harmonization 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 identifiers management and user authentication/authorization for access via a rights and roles management with username/password. Steps to achieve data harmonisation and secured data sharing and use are: 1) Set-up the Opal server for each data provider and import relevant datasets, 2) Configure a harmonized data description in each Opal server, and 3) Run distributed queries on harmonized datasets through the DataSHIELD application that 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 anonymized and harmonized 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 the WP4 'Data analysis and Outcomes' dedicated to unCoVer findings from the data acquired to maximize their use in informing COVID-19 response. Activities of 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 harmonized RWD within the unCoVer network, a large amount of hospitalised and discharged COVID-19 patients 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 characterization of COVID-19 patients, 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 optimization 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 (*i.e.* data analysts) as well as the lessons learned 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 the 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 utilizing best practices in project coordination, as outlined in WP6 'Coordination'. Concerning this, the unCoVer organization 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. A General Assembly (GA), involving WP leaders manages the network, and coordinates WP and tasks leaders, assisted also by a Management Support Team (MST) 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 the external advisory boards.

ETHICS AND DISSEMINATION

Ethical aspects are of utmost importance in unCoVer. Upon the project start date, the roles, and responsibilities of the independent DP-EAB were 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 General Data Protection Regulation (GDPR) and meet the ethics requirements, the unCoVer network will follow the data processing steps represented in **Figure 4,** in the following sequence:

1. 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 Assessment (DPIA). This information is processed by the research ethics team, responsible for categorizing the datasets into three different categories: low, medium, or high data privacy risk.
2. Datasets categorized as ''low-risk'' will be available to proceed with the harmonization 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 ID database that

stores the patient identifiers accessible by the data provider only, and (2) the study database that stores pseudo-anonymised patient's data to be used for data analyses accessible, through code only, by data analysts. The "medium-" and "high-risk" datasets will be subject to further review and requirements before harmonization processes.

3. Finally, the installation of the servers will allow the consortium to analyse the available RWD through an anonymisation layer to answer the pre-identified 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 the SRAB, i.e., a non-executive consulting substructure composed of several key stakeholders from the regulatory, governance, civil society level, and patient's 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 enhanced outreach to COVID-19 stakeholders included in external advisory boards, as well as a prominent work package on dissemination activities, that include but are not limited to:

- *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, *i.e.* for implementation and/or application of the unCoVer platform as a whole, and the unCoVer-partners publication, *i.e.* 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 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.
- *Policy makers.* 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 impacts of the project in 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 the stakeholder's needs and expectations.
- *General public.* To maximize 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, lecture, and infographics as made available on the project website, among others.

CONCLUSION

unCoVer brings a new global perspective of disease management in which protection of health requires not only looking at regional or national aspects but also the development of data networks; A worldwide outbreak needs to be tackled by a worldwide research network. The impact of the COVID-19 pandemic in low-resource settings, and other vulnerable populations is a clear reminder of the importance of equity and solidarity, also in medical research. As we move forward, it is vital that we explore the drivers of the pandemic, learn from the global response, and become more prepared for the similar future situations through global networks aiming at transcontinental data integration and analysis. The impact of the COVID-19 pandemic in low-resource settings, and other vulnerable populations is a clear reminder of the importance of equity and solidarity, also in medical research. unCoVer wants to underpin some of the major global issues arising from the pandemic, including the geography of health, the role of social and lifestyle determinants, and other bases of vulnerability as prognostic factors of severity, and effectiveness of treatments and guidelines recommendations,

leveraging the use of big data within the remit of GDPR or its equivalent, for maximal benefit to population's health.

AUTHORS CONTRIBUTION

José L Peñalvo, Elly Mertens, Gloria Soriano, Zubair Kabir, Ernestina Menasalva, José M Castellano, David M Pereira, Silvia Riva, and Joan B Soriano wrote the manuscript. All authors have critically reviewed the manuscript, and approved the final version.

FUNDING STATEMENT

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

DATA AVAILABILITY

unCoVer data will be available via a tiered-access web-based interface and complying with data protection requirements.

COMPETING INTERESTS

The authors declare no conflicts of interest.

ACKNOWLEDGMENTS

The unCoVer network: Institute of Tropical Medicine and Antwerp University Hospital, Belgium (José L. Peñalvo, Elly 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 Iași, 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 Lucía 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, Arnoldo 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 Devleesschauwer, 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 Universitesi Vakfi, Turkey (Seval Akgun, Tugba Gürgen Erdogan, Süleyman Çetinkünar, Cenk Belibağlı, Kübra Demir, Mustafa Görür, Turgut Bulut, and K.R. Nayar); St Mary's University Twickenham, United Kingdom (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, United States (Miguel Reina Ortiz); Universidad de Antioquía, Colombia (Paula Andrea Diaz Valencia, Lina Ruíz, Juan Pablo Pérez Bedoya, Oscar Ignacio Mendoza, Camilo Hincapie, Boris Rodriguez, and Noël Barengo); Feevale University, Associacao 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).

REFERENCES

1. World Health Organization (WHO). Expanding our understanding of post COVID-19 condition: report of a WHO webinar, 9 February 2021., 2021.
2. 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. *The Lancet* 2020;395(10229):1054-62. doi: 10.1016/S0140-6736(20)30566-3
3. 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. *Critical Care* 2020;24(1):108. doi: 10.1186/s13054-020-2833-7
4. World Health Organization (WHO). COVID-19 Clinical management: living guidance, 2021.
5. Jin Y, Ji W, Yang H, et al. Endothelial activation and dysfunction in COVID-19: from basic mechanisms to potential therapeutic approaches. *Signal Transduction and Targeted Therapy* 2020;5(1):293. doi: 10.1038/s41392-020-00454-7
6. 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. *Journal of Virology* 2020;94(7):e00127-20. doi: 10.1128/jvi.00127-20
7. 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(4):e21. doi: 10.1016/S2213-2600(20)30116-8 [published Online First: 2020/03/15]
8. Kassir R. Risk of COVID-19 for patients with obesity. *Obesity Reviews*;21(6):e13034. doi: 10.1111/obr.13034
9. 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;360:m1966. doi: 10.1101/2020.04.08.20057794
10. 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(5):1372-78. doi: 10.1093/ije/dyx180 [published Online First: 2017/10/13]

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, United Kingdom, and the United States of America)

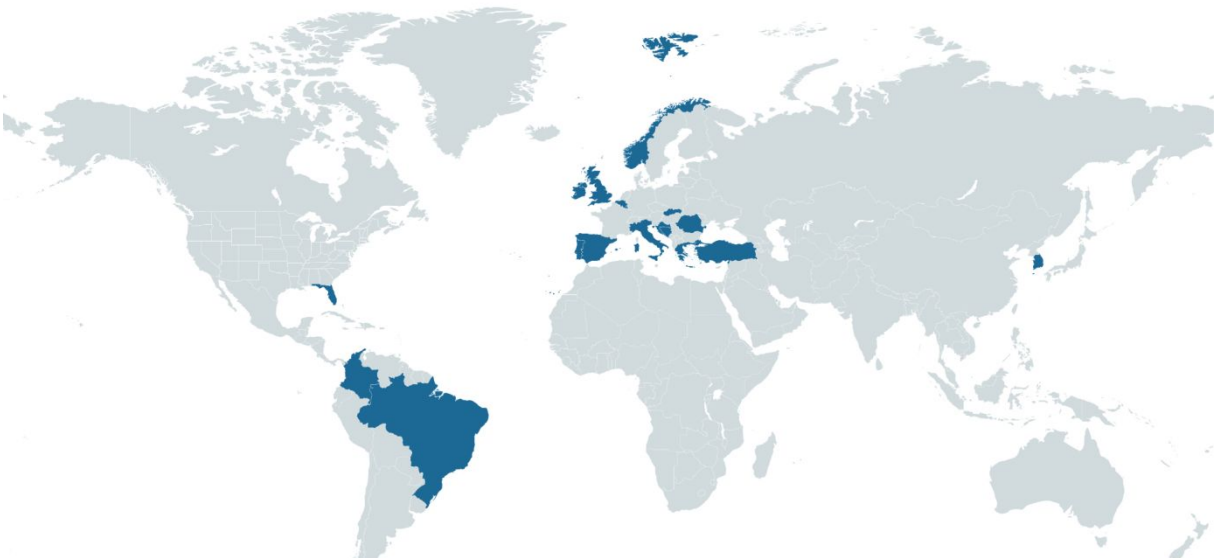


Figure 2. Management structure of the unCoVer network

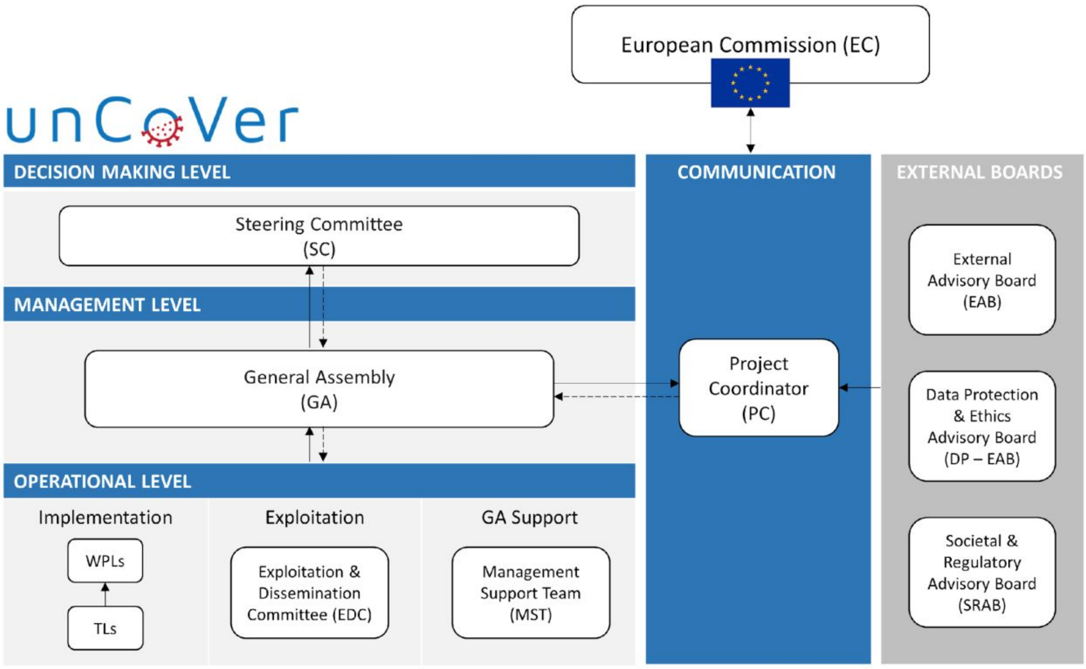


Figure 3. Secure multi-party computation of unCoVer data based on Opal/DataShield infrastructure

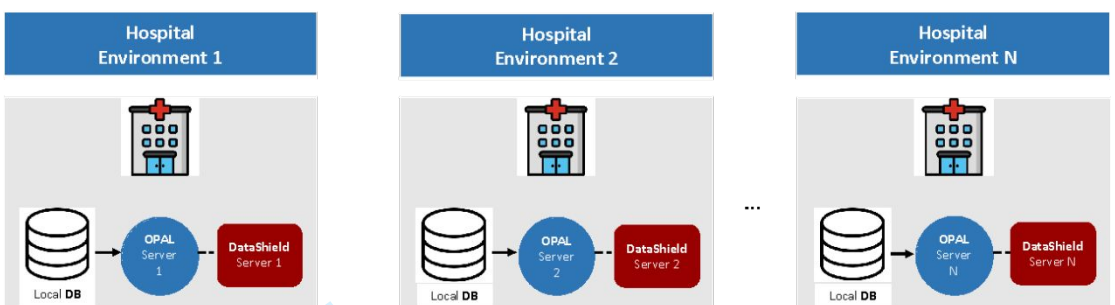


Figure 4. Data management process of the COVID-19 related data available within the unCoVer network

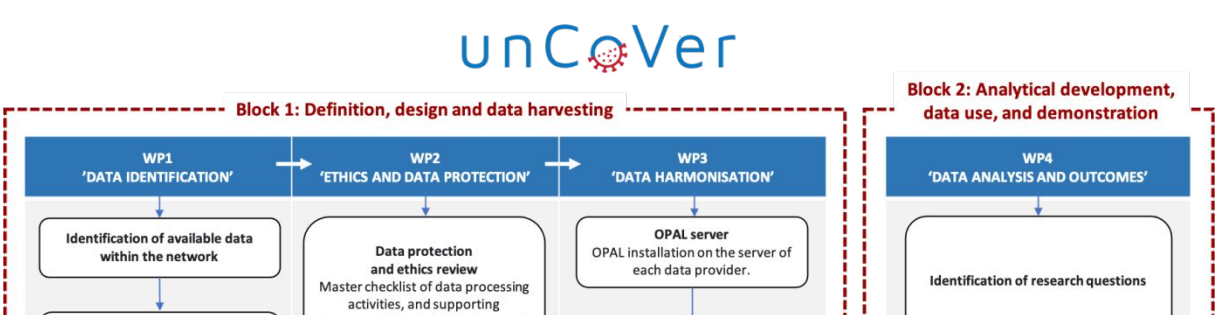


Table 1. Overview of the COVID-19 real-world data among the unCoVer network

Country	Institution[1]	Num. centres[2]	Start date	End date	Num. patients[2]	Study Population[3]	Num. time points	Variable type												
								Demographics	Clinical/Epidemiology	Human OMICs	Pathogen OMICs	Bio-specimens	Imaging	Social network	Movement-related	Mental health	Economic	Diet	Screening
Patient data from Electronic Medical Records																				
BA	UNSA	10	17/03/2020	20/05/2021	2,000	1, 2, 3	2	X	X				X						
BE	ITM-UZA	1	22/04/2020	Ongoing	187	1	2	X	X				X			X			
BR	ASPEUR	40	26/03/2020	Ongoing	2,000	1, 3	15	X	X		X	X					X		
ES	FIHM	12	01/03/2020	15/02/2021	4,480	1	2	X	X				X						
ES	HULPr	1	01/01/2020	29/09/2020	2,217	1	2	X	X										
ES	IIS-FJD	1	07/03/2020	31/05/2020	1,861	1	2	X	X				X						
ES	UNAV	2	01/04/2020	01/01/2021	100	1	1	X	X										
HR	INANTRO	2	01/08/2020	Ongoing	200	1	1	X	X			X	X			X			
IT	IRCCS - DB1	1	03/03/2020	Ongoing	200	1	1	X	X				X						
IT	IRCCS - DB4	1	01/03/2020	09/05/2020	355	4	1	X	X										
IT	ULSS6	4	01/03/2020	Ongoing	1,000	1	3	X	X				X						
RO	UMF Cluij	1			100	1	4	X	X				X						
RO	UMF IASI	1	01/03/2020	Ongoing	150	1	2	X	X				X						
SK	TU	1	01/10/2020	Ongoing	800	1	1	X	X										
TR	BU - DB1	6	01/03/2020	Ongoing	7,000	1	4	X	X										
Public Health Surveillance data and Registers																				
SK	TU	240	01/03/2020	Ongoing	776,000	2	1	X	X										X
BE	Sciensano	98	14/03/2020	Ongoing	50,000	1	2	X	X				X						
CO	UdeA	1,314	14/03/2020	Ongoing	3,997,021	2	1	X	X										
HR	CIPH	NA	25/02/2020	Ongoing	220,000	2	1	X	X										
IE	TUDublin	48	29/02/2020	30/11/2020	74,000	2	1	X	X						X		X		
NO	USN	40	31/03/2020	Ongoing	2,313	1	10	X	X							X			
PT	UPORTO	NA	03/03/2020	Ongoing	830,000	2	1	X	X										

Country	Institution[1]	Num. centres[2]	Start date	End date	Num. patients[2]	Study Population[3]	Num. time points	Variable type											
								Demographics	Clinical/Epidemiology	Human OMICs	Pathogen OMICs	Bio-specimens	Imaging	Social network	Movement-related	Mental health	Economic	Diet	Screening
Observational Research data																			
IT	IRCCS - DB3	1	22/04/2020	25/05/2020	1,515	5	1	X	X										
LU	LIH	1	15/04/2020	01/04/2021	900	2	5	X	X			X							
PT	IPC	1	01/03/2020	01/07/2020	550	5	1	X										X	
UK	SMUC	1	01/07/2020	01/01/2021	200	2	1	X								X			
Screening data																			
IT	IRCCS - DB2	1	31/03/2020	12/05/2020	1,635	6	1	X	X										X
TR	BU – DB2	1	01/03/2020	Ongoing	18,000	5,7	1	X											X

[1]Institution's acronyms: ITM-UZA, Institute of Tropical Medicine and Antwerp University Hospital; FIHM, Fundación Investigación HM Hospitales; UNAV, Universidad de Navarra; UPORTO, Universidade Do Porto; TUDublin, Technological University Dublin; UMF Cluij, Universitatea De Medicina Si Farmacie Iuliu Hatieganu Cluj-Napoca; UMF IASI, Universitatea De Medicina Si Farmacie Grigore T Popa Din Iași; *LIH, Luxembourg Institute of Health; TU, Trnavska Univerzita V Trnave; IPC, Instituto Politécnico de Coimbra; HULPr, Hospital Universitario de La Princesa; IIS-FJD, Instituto Investigación Sanitaria Fundación Jiménez Díaz; USN, Universitetet I Sorost-Norge; IRCCS, Istituto Don Calabria; CIPH, Croatian Institute of Public Health; INANTRO, Institut Za Antropologiju; BU, Baskent Universitesi Vakfi; SMUC, St Mary's University Twickenham; ULSS6, Azienda ULSS6 Euganea; UdeA, Universidad de Antioquía; ASPEUR, Associacao Pro Ensino Superior Em Novo Hamburgo; UNSA, Univerzitet U Sarajevu. [2] Number of centres providing information, and number of patients/individuals in the datasets could change due to the continuous update of the information; NA, Not available at the moment. [3] Study population refers to, 1) COVID-19 hospitalized patients, 2) COVID-19 cases, 3) COVID-19 patients attending primary care services, 4) COVID-19 patients attending emergency services, 5) general population, 6) health care workers, and 7) tourists.*

Supplemental Table 1. Clinical and epidemiological data available in unCoVer network databases containing hospitalized patient data.

Country	Institution[1]	Patient admission CFR						Daily CFR during hospitalization						Patient outcome CRF			
								At the ward			In ICU						
		Demographics	Onset and/or Admission date	Signs and symptoms	Laboratory results	Pre-admission medication	Co-morbidities & risk factors	Signs and symptoms	Laboratory results	Supportive Treatment	ICU admission date	Signs and symptoms	Laboratory results	Medication treatment	Complications	Discharge date	Outcome status
	Patient data from Electronic Medical Records																
BA	UNSA	X	X	X	X	X	X			X	X		X	X		X	X
BE	ITM-UZA	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
BR	ASPEUR	X	X	X	X	X	X	X	X		X	X					X
ES	FIHM	X	X	X[2]	X[2]	X[2]	X[2]	X		X[2]	X			X[2]		X	X
ES	HULPr	X	X	X[2]			X		X	X[2]	X[3]			X		X[4]	X
ES	IIS-FJD	X	X	X	X	X	X		X		X[4]	X		X	X	X	X
ES	UNAV	X	X		X												
HR	INANTRO	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X
IT	IRCCS - DB1	X	X	X	X	X	X	X		X	X[3]	X		X	X	X	X
IT	IRCCS - DB4	X	X	X	X		X[3]										
IT	ULSS6	X	X	X[2]	X	X	X	X	X	X	X	X	X	X		X	X
RO	UMFCluij	X	X	X			X		X	X				X		X[4]	X
RO	UMFIASI	X	X	X		X	X		X	X	X			X		X	X
SK	TU	X	X	X	X	X	X	X	X	X	X		X	X		X	X
TR	BU – DB1	X	X	X			X							X			
TR	BU – DB2	X	X	X		X	X	X	X	X	X	X		X		X	X
	Patient data from Public Health Surveillance and National Registers																
SK	TU	X	X	X	X	X	X	X	X	X	X		X	X		X	X
BE	Sciensano	X	X	X	X		X			X	X	X		X	X	X	X
CO	UdeA	X														X	X
HR	CIPH	X	X				X									X	X
IE	TUDublin	X	X				X										
NO	USN	X	X	X	X		X							X	X	X	X

PT	UPORTO	X	X				X[2]			X[3]	X						X		
	Patient data from Observational Research																		
IT	IRCCS - DB2	X		X			X[3]												
IT	IRCCS - DB3	X		X			X[3]												
LU	LIH	X	X	X	X	X	X	X	X	X	X[3]	X	X	X	X	X	X	X	X

[1]Acronyms: ITM-UZA, Institute of Tropical Medicine and Antwerp University Hospital; FIHM, Fundación Investigación HM Hospitales; UNAV, Universidad de Navarra; UPORTO, Universidade Do Porto; TUDublin, Technological University Dublin; UMF Cluij, Universitatea De Medicina Si Farmacie Iuliu Hatieganu Cluj-Napoca; UMF IASI, Universitatea De Medicina Si Farmacie Grigore T Popa Din Iași; LIH, Luxembourg Institute of Health; TU, Trnavska Univerzita V Trnave; IPC, Instituto Politécnico de Coimbra; HULPr, Hospital Universitario de La Princesa; IIS-FJD, Instituto Investigación Sanitaria Fundación Jiménez Díaz; USN, Universitetet I Sorost-Norge; IRCCS, Istituto Don Calabria; CIPH, Croatian Institute of Public Health; INANTRO, Institut Za Antropologiju; BU, Baskent Universitesi Vakfi; SMUC, St Mary's University Twickenham; ULSS6, Azienda ULSS6 Euganea; UdeA, Universidad de Antioquía; ASPEUR, Associacao Pro Ensino Superior Em Novo Hamburgo; UNSA, Univerzitet U Sarajevu.[2]Data needs to be extracted from text variables; [3]Coded as yes/no without any further specifications; [4]Length of stay in days instead of exact date of admission and/or discharge from hospital and/or ICU

Supplemental Table 2. Data protection Risk Assessment checklist

> Was personal data/sensitive collected? Yes: What permissions and safeguarding are in place for the individual (evaluated through question below)? No: What data was collected and what are the risks?
> Was Explicit Consent or Assent received (if necessary)? Yes: Method of requesting individual consent or assuming assent must be provided to unCoVer No: Should it have been collected?
> Is Ethical Approval necessary? Yes: Ethical Approval must be provided to unCoVer No: If Ethical Approval is not necessary, personal data must not be identifiable (GDPR does not apply) and the data should be 'public' to some extent
> Has the partner received Ethical Approval? Yes: Does unCoVer need any additional information following Ethical Approval? No: Hold until Ethical Approval is provided and assess
> Anonymisation / Pseudonymization: Is the process used to anonymise or pseudonymise the data to be shared with uncover known? Who undertook this process? Yes: Transparent and appropriate methods used were clear and acceptable. Parties with data access are identified. Still ensure no personal data is passed onto unCoVer No: if these processes were undertaken, is there follow-up required?
> Is there a letter from the Data Protection Officer (DPO) confirming the standardisation of Ethical Approval? Yes: Letter has been signed and returned by local DPO No: Follow-up with partner is needed

Low risk	Medium risk (a combination of some of the below)	High risk (a combination of some of the below)
Ethical approval received or not applicable	Ethical approval received or pending	Ethical approval not received
Explicit consent & assent received or not applicable	Explicit consent & assent received or waiting confirmation	Unclear details of consent / assent
Anonymisation and/or pseudonymization transparent & appropriate	Waiting confirmation of anonymization/ pseudonymization process(es)	Anonymization or pseudonymization incomplete. Or process unexplained
Data transfer agreement in place if necessary	Data transfer pending or in place if necessary	Non-EU partners without appropriate Data transfer agreements
DPO letter received	DPO letter pending/received	DPO letter pending

The RECORD statement – checklist of items, extended from the STROBE statement, that should be reported in observational studies using routinely collected health data.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found		RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	Abstract, page 3
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported			Pag 5, 6
Objectives	3	State specific objectives, including any prespecified hypotheses			Pag 6, 7
Methods					
Study Design	4	Present key elements of study design early in the paper			Pag 7
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection			Pag 7, 8, and Figure 1
Participants	6	*(a) Cohort study* - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up		RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided.	Pag 7, 8, 9, and Table 1 and Supplemental Table 1

		Case-control study - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case		RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.		RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	Table 1 and Supplemental Table 1
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group			Table 1 and Supplemental Table 1
Bias	9	Describe any efforts to address potential sources of bias			Pag 9, 10, 11
Study size	10	Explain how the study size was arrived at			Pag 9, 10, 11
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why			Pag 9, 10, 11, and Figure 3

Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses			Pag 9, 10, 11, and Figure 3
Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population. RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	Figure 3
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	Figure 3
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the		RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.,* study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in	NA (protocol paper not including results)

		study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram		the text and/or by means of the study flow diagram.	
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)			NA (protocol paper not including results)
Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report numbers in each exposure category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures			NA (protocol paper not including results)
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period			NA (protocol paper not including results)

Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses			NA (protocol paper not including results)
Discussion					
Key results	18	Summarise key results with reference to study objectives			NA (protocol paper not including results)
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias		RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	Pag 15
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence			Pag 15
Generalisability	21	Discuss the generalisability (external validity) of the study results			Pag 15
Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based			Pag 16
Accessibility of protocol, raw data, and programming code		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	Pag 17

*Reference: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.

BMJ Open

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

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-055630.R1
Article Type:	Protocol
Date Submitted by the Author:	20-Sep-2021

Complete List of Authors:	Penalvo, Jose; Institute of Tropical Medicine, Unit of Non-Communicable Diseases Mertens, Elly; Institute of Tropical Medicine, Ademović, Enisa; University of Sarajevo, Department of Epidemiology and Biostatistics Akgun, Seval; Baskent University Faculty of Medicine, Public Health Department Baltazar, Ana Lúcia; Polytechnic Institute of Coimbra, Scientific-Pedagogical Unit of Dietetics and Nutrition Buonfrate, Dora; IRCCS Ospedale Sacro Cuore Don Calabria, Centre for Tropical Diseases Čoklo, Miran; Institut za antropologiju, Centre for Applied Bioanthropology Devleesschauwer, Brecht; Sciensano, Epidemiology and public health; Ghent University Faculty of Veterinary Medicine, Department of Veterinary Public Health and Food Safety Diaz Valencia, Paula Andrea; University of Antioquia, Epidemiology Group, National College of Public Health Fernandes, João C; Universidade Católica Portuguesa Escola Superior de Biotecnologia, Centro de Biotecnologia e Química Fina (CBQF) Gómez, Enrique Javier; Universidad Politécnica de Madrid, Center for Biomedical Technology; Universidad Politécnica de Madrid, Biomedical Engineering and Telemedicine Centre Hynds, Paul; Technological University Dublin, Environmental Sustainability & Health Institute Kabir, Zubair; University College Cork, School of Public Health Klein, Jörn; University of South-Eastern Norway, Kostoulas , P; University of Thessaly Llanos Jiménez, Lucía; Instituto de Investigacion Sanitaria de la Fundacion Jimenez Diaz, Clinical Research Unit Lotrean, Lucia Maria; Iuliu Hagieganu University of Medicine and Pharmacy Faculty of Medicine Majdan , Marek ; Trnava University in Trnava, Menasalvas, Ernestina; Universidad Politecnica de Madrid Escuela Universitaria de Ingenieria Tecnica de Telecomunicacion, Biomedical Engineering and Telemedicine Centre Nguewa, paul; Universidad de Navarra Oh, In-Hwan; Kyung Hee University O'Sullivan, Georgie; University College Cork, School of Public Health Pereira, David M.; Universidade do Porto, 23 REQUIMTE/LAQV, Laboratório de Farmacognosia, Departamento de Química

	Reina Ortiz, Miguel; University of South Florida, Riva, Silvia; St Mary's University Twickenham, Department of Public Health Soriano, Gloria; Institute of Tropical Medicine, Department of Public Health Soriano, Joan; Universidad Autonoma de Madrid, Servicio de Neumología, Hospital Universitario de La Princesa Spilki, Fernando; FEEVALE University Tamang, Mary Elizabeth; Azienda ULSS 6 Euganea Trofor, Antigona; Universitatea de Medicina si Farmacie Gr T Popa Iasi, Pulmonary Diseases I Vaillant, Michel; Luxembourg Institute of Health, Competence Centre for Methodology and Statistics Van Ierssel, Sabrina; University Hospital Antwerp, Department of General Internal Medicine, Infectious diseases and Tropical Medicine Vukovic, Jakov; Croatian Institute of Public Health Castellano, Jose Maria; Hospital Universitario HM Montepríncipe, Cardiology Department
Primary Subject Heading:	Research methods
Secondary Subject Heading:	Ethics, Epidemiology, Infectious diseases, Public health
Keywords:	COVID-19, STATISTICS & RESEARCH METHODS, Public health < INFECTIOUS DISEASES

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

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], Antigona Carmen Trofor[31,32], Michel Vaillant[33], Sabrina Van Ierssel[34], Jakov Vukovic[35], and José M. Castellano[36,37,38] **on behalf of the unCoVer network**

1 Department of Public Health, Institute of Tropical Medicine, Antwerp, Belgium.

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

3 Public Health Department, Baskent University School of Medicine, Turkey

4 Scientific-Pedagogical Unit of Dietetics and Nutrition, Coimbra Health School, Polytechnic Institute of Coimbra, Portugal

5 Department of Infectious Tropical diseases and Microbiology, IRCCS Sacro Cuore Don Calabria hospital, Verona, Italy

6 Centre for Applied Bioanthropology, Institute for Anthropological Research, Zagreb, Croatia

7 Department of Epidemiology and Public Health, Sciensano, Belgium

8 Department of Veterinary Public Health and Food Safety, Faculty of Veterinary Medicine, Ghent University, Belgium

9 Epidemiology group, National College of Public Health, University of Antioquia, Medellin, Colombia.

10 Centro de Biotecnologia e Química Fina (CBQF), Escola Superior de Biotecnologia da Universidade Católica Portuguesa, Porto, Portugal
11 Center for Biomedical Technology, Universidad Politécnica de Madrid, Pozuelo de Alarcón, Spain.
12 Biomedical Engineering and Telemedicine Centre, ETSI Telecomunicación, Universidad Politécnica de Madrid, Madrid, Spain.
13 Centro de Investigación Biomédica en Red de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Madrid, Spain.
14 Environmental Sustainability & Health Institute, Technological University Dublin, Ireland
15 School of Public Health, University College Cork, Ireland
16 University of South-Eastern Norway, Department of Nursing and Health Sciences, Norway
17 Faculty of Public Health, University of Thessaly, Greece
18 Clinical Research Unit, Instituto Investigación Sanitaria Fundación Jiménez Díaz (IIS-FJD, UAM), Spain
19 Iuliu Hatieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania
20 Institute for Global Health and Epidemiology, Trnava University, Trnava, Slovakia
21 Instituto de Salud Tropical (ISTUN), Department of Microbiology and Parasitology, Navarra Institute for Health Research (IdiSNA), University of Navarra, Pamplona, Spain.
22 Department of preventive Medicine, Kyung Hee University, Seoul, Republic of Korea
23 REQUIMTE/LAQV, Laboratório de Farmacognosia, Departamento de Química, Faculdade de Farmácia, Universidade do Porto, Portugal
24 College of Public Health, University of South Florida, MI, USA
25 Department of Psychology and Pedagogic Science, St Mary's University Twickenham, London, UK
26 Servicio de Neumología, Hospital Universitario de La Princesa, Universidad Autónoma de Madrid, Madrid, Spain
27 Centro de Investigación en Red de Enfermedades Respiratorias (CIBERES), Instituto de Salud Carlos III (ISCIII), Madrid, Spain.

28 COVID-19 Clinical Management Team, WHO Health Emergency Programme, World Health Organization HQ, Geneva, Switzerland

29 Feevale University, Novo Hamburgo, Brazil

30 Azienda ULSS6 Euganea, Padova, Italy

31 University of Medicine and Pharmacy Grigore T. Popa Iasi, Romania

32 Clinical Hospital of Pulmonary Diseases Iasi, Romania

33 Translational Medicine Operations Hub, Competence Center for Methodology and Statistics, Luxembourg Institute of Health, Luxembourg

34 Department of General Internal Medicine, Infectious diseases and Tropical Medicine, Antwerp University Hospital, Edegem, Belgium

35 Croatian Institute of Public Health, Zagreb, Croatia

36 Cardiology Department, Hospital Universitario HM Montepríncipe, HM Hospitales, Madrid, Spain.

37 Centro Nacional de Investigaciones Cardiovasculares (CNIC), Instituto de Salud Carlos III, Madrid, Spain

38 Fundación de Investigación HM Hospitales, Madrid, Spain

*Corresponding Author: José L. Peñalvo, PhD, Department of Public Health, Institute of Tropical Medicine, Antwerp (Belgium) Tel: +32(0)32476251 jpenalvo@itg.be | @JosePenalvo

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 utilizing real-world data (RWD) derived from the response and provision of care to COVID-19 patients by health systems across Europe and elsewhere. unCoVer aims to exploit the full potential of these information to rapidly address clinical and epidemiological research questions arising from the evolving pandemic.

Methods and analysis. From the onset of COVID-19 pandemic, partners are gathering RWD from electronic health records currently including information from over 22,000 COVID-19 hospitalized patients, 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 allows processing 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 activities, scientific publications, and conference communications.

Article Summary

- The unCoVer network includes a sizable number of partners that will exploit the full potential of real-world data derived from the provision of care to COVID-19 patients by health systems across Europe and elsewhere.
- With the availability of international harmonized RWD provided by the unCoVer network, a large amount of information on COVID-19 patients will be studied to better understand aspects of the pathophysiology, progression and treatment, and 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 GDPR 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.

.

INTRODUCTION

The outbreak of the coronavirus disease (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was declared a Public Health Emergency of International Concern by the World Health Organisation on 30 January 2020, and a pandemic, on 11 March 2020. Despite the deployment of public health measures such as restrictions of movements, gathering, and 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 condition[1]. Early epidemiological data of 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 hospitalization 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 multi-dimensional and dynamic nature of the inter-related factors associated with individual responses to 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 has been accumulated as part of

the health systems' efforts to fight COVID-19. These RWD reflect the impact of COVID-19 in patient's health and characterize the protocols of health care 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, those most at risk of severe COVID-19, are often excluded. Harmonization of data from different sources allows for comparison across health systems and improves patients' characterization using the wider heterogeneity of the 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 network aims to provide a research platform for the expert use of RWD, by bringing together complementary data, medical and scientific expertise to address the still urgent questions regarding 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 the health systems, improve patient prognosis especially among those more vulnerable (e.g., chronic patients, immunosuppressed individuals, and population subgroups with limited access to health care, among others), and mitigate the burden of COVID-19. The specific objectives of unCoVer are:

1. To bring together European and international expertise and data to synchronize collaborative research efforts in addressing the ongoing COVID-19 pandemic in a common platform.

2. To continuously monitor, identify and facilitate the 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.
3. To identify data gaps, and marginalized populations to proactively seek synergies with complementary existing and planned clinical databases related to COVID-19.
4. To provide a platform for the use of dissimilar data sources capable of streamlining ethical and legal aspects and anticipating the needs for data harmonization by innovative computational resources and integrated information for enhanced impact.
5. To bring together expertise on the use of advanced computational, epidemiological and biostatistical methods to handle heterogeneous, and multi-layered 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 sequelae, as well as the impact of COVID-19 in health system resources.
6. 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 optimize resources.

METHODS AND ANALYSIS

unCoVer is conceptualized 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.

Setting

unCoVer comprises 29 partners from 25 institutions in the European Union 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, United

Kingdom, and the United States of America (Figure 1). Partners provide data mostly from front-line hospitals but also national health agencies, registries, and investigator-lead 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 the routine healthcare data already collected from patients during the pandemic. The set-up of the network therefore relies on a continuous iteration process between a) clinical partners, who will guide the development of research questions needed to improve patient's care and inform public health strategies; and b) epidemiologists and analytical experts, who will operationalize 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 the medical research and findings (External Advisory Board, EAB), 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 system 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 has information on hospitalization of COVID-19 patients with at least two-time points of data collection, at admission and discharge. In addition to demographics, and clinical/epidemiological data, other data types such as bio-specimens, imaging data, social network-/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 (CRF) at patient's admission (i.e. date of symptom onset and/or admission, signs and symptoms at admission, laboratory results, pre-admission medication, comorbidities & risk factors), during hospitalization (i.e. signs and symptoms, laboratory results, supportive treatment, admission to ICU) and at discharge (i.e. date of discharge, outcome, as well as medication and complications) (**Supplementary 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 a standardised way, including a common codebook that specifies the harmonized variables, with standardised variable names and data format and labels, and range of plausible values, in preparation for data harmonization and validation processes *i.e.* 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 (e.g. clinical data, hospital records or publicly available data, personal data, data collected in vulnerable groups, availability of follow-up data), informed consent (e.g. explicit consent or assent obtained), data protection (e.g. data protection officer identified and data protection impact assessment completed), ethical approval (obtained or pending), data privacy protection (e.g. 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 GDPR or equivalent guidelines are adhered to during the data processing activity. Once a dataset is categorised as low risk, it will be available to proceed with subsequent steps (**Supplemental Table 2**).

For the secure multi-party computation of unCoVer data, WP3 'Data Harmonization', developed an infrastructure based on Opal 4.1 (OBiBa suite, Maelstrom Research, Montreal, Canada) to facilitate interoperability of the data, including data management, harmonization 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 identifiers management and user authentication/authorization for access via a rights and roles management with username/password. Steps to achieve data harmonisation and secured data sharing and use are: 1) Set-up the Opal server for each data provider and import relevant datasets, 2) Configure a harmonized data description in each Opal server, and 3) Run distributed queries on harmonized datasets through the DataSHIELD application that 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 anonymized and harmonized 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 the WP4 'Data analysis and Outcomes' dedicated to unCoVer findings from the data acquired to maximize their use in informing COVID-19 response. Activities of 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 harmonized RWD within the unCoVer network, a large amount of hospitalised and discharged COVID-19 patients 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 characterization of COVID-19 patients, 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 optimization 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 (*i.e.* data analysts) as well as the lessons learned 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 the 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 utilizing best practices in project coordination, as outlined in WP6 'Coordination'. Concerning this, the unCoVer organization 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. A General Assembly (GA), involving WP leaders manages the network, and coordinates WP and tasks leaders, assisted also by a Management Support Team (MST) 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 the 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. Upon the project start date, the roles, and responsibilities of the independent DP-EAB were 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 General Data Protection Regulation (GDPR) and meet the ethics requirements, the unCoVer network will follow the data processing steps represented in Figure 4, in the following sequence:

1. 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 Assessment (DPIA). This information is processed by the research ethics team, responsible for categorizing the datasets into three different categories: low, medium, or high data privacy risk.

2. Datasets categorized as ''low-risk'' will be available to proceed with the harmonization 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 ID database that stores the patient identifiers accessible by the data provider only, and (2) the study database that stores pseudo-anonymised patient's data to be used for data analyses accessible, through code only, by data analysts. The "medium-" and "high-risk" datasets will be subject to further review and requirements before harmonization processes.
3. Finally, the installation of the servers will allow the consortium to analyse the available RWD through an anonymisation layer to answer the pre-identified 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 the SRAB, i.e., a non-executive consulting substructure composed of several key stakeholders from the regulatory, governance, civil society level, and patient's 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 enhanced outreach to COVID-19 stakeholders included in external advisory boards, as well as a prominent work package on dissemination activities, that include but are not limited to:

- *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, *i.e.* for implementation and/or application of the unCoVer platform as a whole, and the unCoVer-partners publication, *i.e.* 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 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.
- *Policy makers.* 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 impacts of the project in 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 the stakeholder's needs and expectations.
- *General public.* To maximize 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, lecture, 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 recognized 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 synchronize 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 two years, efforts of the first year were mainly focussed 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 datasets 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 provide valuable input for the use of complementary RWD with robust methodologies at hand for the still ongoing current pandemic and future pandemics.

AUTHORS CONTRIBUTION

JLP coordinates the unCoVer network. GS manages overall project's activities. JLP, EM, EMEN, EJC, ZK, JMC, and DMP lead project's work packages. JLP, EM, EA, SA, ALB, DB, MC, BD, PADV, PH, JK, LLJ, LML, MM, PN, IHO, DMP, SR, JBS, FS, MET, ACT, MV, SVI, JV, and JMC characterized the data available to the network. JLP, EM, EJG, EMEN, GS designed the data harmonization system and access to the federated infrastructure. JLP, ZK, EMEN, GOS, SR and GS supervised compliance with legal and ethical requirements of the data use. JLP, JCF, JK, PK, MM, EM, DMP, MRO, and GS are responsible for the dissemination of the network's activities. JLP, EM, GS, ZK, EMEN, JMC, DMP, SR, and JBS drafted the manuscript and designed the figures. All authors have critically reviewed and approved the final version of the manuscript. The unCoVer network consist of all individual partner institutions and investigators who are responsible for the identification, characterization, compliance with legal and ethical requirements of the data use, as well as dissemination and coordination activities.

FUNDING STATEMENT

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

DATA AVAILABILITY

unCoVer data will be available via a tiered-access web-based interface and complying with data protection requirements.

COMPETING INTERESTS

The authors declare no conflicts of interest.

COLLABORATORS

The unCoVer network: Institute of Tropical Medicine and Antwerp University Hospital, Belgium (José L. Peñalvo, Elly 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 Iași, 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 Lucía 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, Arnoldo 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 Devleesschauwer, 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 Universitesi Vakfi, Turkey (Seval Akgun, Tugba Gürgen Erdogan, Süleyman Çetinkünar, Cenk Belibağlı, Kübra Demir, Mustafa Görür, Turgut Bulut, and K.R. Nayar); St Mary's University Twickenham, United Kingdom (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, United States (Miguel Reina Ortiz); Universidad de Antioquía, Colombia (Paula Andrea Diaz Valencia, Lina Ruíz, Juan Pablo Pérez Bedoya, Oscar Ignacio Mendoza, Camilo Hincapie, Boris Rodriguez, and Noël Barengo); Feevale University, Associacao 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).

REFERENCES

1. World Health Organization (WHO). Expanding our understanding of post COVID-19 condition: report of a WHO webinar, 9 February 2021., 2021.

2. 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. *The Lancet* 2020;395(10229):1054-62. doi: 10.1016/S0140-6736(20)30566-3

3. 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. *Critical Care* 2020;24(1):108. doi: 10.1186/s13054-020-2833-7

4. World Health Organization (WHO). COVID-19 Clinical management: living guidance, 2021.

5. Jin Y, Ji W, Yang H, et al. Endothelial activation and dysfunction in COVID-19: from basic mechanisms to potential therapeutic approaches. *Signal Transduction and Targeted Therapy* 2020;5(1):293. doi: 10.1038/s41392-020-00454-7

6. 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. *Journal of Virology* 2020;94(7):e00127-20. doi: 10.1128/jvi.00127-20

7. 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(4):e21. doi: 10.1016/S2213-2600(20)30116-8 [published Online First: 2020/03/15]

8. Kassir R. Risk of COVID-19 for patients with obesity. *Obesity Reviews*;21(6):e13034. doi: 10.1111/obr.13034

9. 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;360:m1966. doi: 10.1101/2020.04.08.20057794

10. 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(5):1372-78. doi: 10.1093/ije/dyx180 [published Online First: 2017/10/13]

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

Country	Institution[1]	Num. centres[2]	Start date	End date	Num. patients[2]	Study Population[3]	Num. time points	Variable type												
								Demographics	Clinical/Epidemiology	Human OMICs	Pathogen OMICs	Bio-specimens	Imaging	Social network	Movement-related	Mental health	Economic	Diet	Screening
Patient data from Electronic Medical Records																			
BA	UNSA	10	17/03/2020	20/05/2021	2,000	1, 2, 3	≥2	X	X				X						
BE	ITM-UZA	1	22/04/2020	Ongoing	187	1	≥2	X	X				X			X			
BR	ASPEUR	40	26/03/2020	Ongoing	2,000	1, 3	15	X	X		X	X					X		
ES	FIHM	12	01/03/2020	15/02/2021	4,480	1	≥2	X	X				X						
ES	HULPr	1	01/01/2020	29/09/2020	2,217	1	≥2	X	X										
ES	IIS-FJD	1	07/03/2020	31/05/2020	1,861	1	≥2	X	X				X						
ES	UNAV	2	01/04/2020	01/01/2021	100	1	≥1	X	X										
HR	INANTRO	2	01/08/2020	Ongoing	200	1	≥1	X	X			X	X			X			
IT	IRCCS - DB1	1	03/03/2020	Ongoing	200	1	≥1	X	X				X						
IT	IRCCS - DB4	1	01/03/2020	09/05/2020	355	4	≥1	X	X										
IT	ULSS6	4	01/03/2020	Ongoing	1,000	1	≥3	X	X				X						
RO	UMF Cluij	1			100	1	≥4	X	X				X						
RO	UMF IASI	1	01/03/2020	Ongoing	150	1	≥2	X	X				X						
SK	TU	1	01/10/2020	Ongoing	800	1	≥1	X	X										
TR	BU - DB1	6	01/03/2020	Ongoing	7,000	1	4	X	X										
Public Health Surveillance data and Registers																			
SK	TU	240	01/03/2020	Ongoing	776,000	2	≥1	X	X										X
BE	Sciensano	98	14/03/2020	Ongoing	50,000	1	≥2	X	X				X						
CO	UdeA	1,314	14/03/2020	Ongoing	3,997,021	2	1	X	X										
HR	CIPH	NA	25/02/2020	Ongoing	220,000	2	≥1	X	X										
IE	TUDublin	48	29/02/2020	30/11/2020	74,000	2	≥1	X	X						X		X		
NO	USN	40	31/03/2020	Ongoing	2,313	1	10	X	X							X			
PT	UPORTO	NA	03/03/2020	Ongoing	830,000	2	≥1	X	X										

Country	Institution[1]	Num. centres[2]	Start date	End date	Num. patients[2]	Study Population[3]	Num. time points	Variable type											
								Demographics	Clinical/Epidemiology	Human OMICs	Pathogen OMICs	Bio-specimens	Imaging	Social network	Movement-related	Mental health	Economic	Diet	Screening
Observational Research data																			
IT	IRCCS - DB3	1	22/04/2020	25/05/2020	1,515	5	≥1	X	X										
LU	LIH	1	15/04/2020	01/04/2021	900	2	5	X	X			X							
PT	IPC	1	01/03/2020	01/07/2020	550	5	≥1	X										X	
UK	SMUC	1	01/07/2020	01/01/2021	200	2	1	X								X			
Screening data																			
IT	IRCCS - DB2	1	31/03/2020	12/05/2020	1,635	6	1	X	X										X
TR	BU – DB2	1	01/03/2020	Ongoing	18,000	5,7	1	X											X

[1]Institution's acronyms: ITM-UZA, Institute of Tropical Medicine and Antwerp University Hospital; FIHM, Fundación Investigación HM Hospitales; UNAV, Universidad de Navarra; UPORTO, Universidade Do Porto; TUDublin, Technological University Dublin; UMF Cluij, Universitatea De Medicina Si Farmacie Iuliu Hatieganu Cluj-Napoca; UMF IASI, Universitatea De Medicina Si Farmacie Grigore T Popa Din Iași; LIH, Luxembourg Institute of Health; TU, Trnavska Univerzita V Trnave; IPC, Instituto Politécnico de Coimbra; HULPr, Hospital Universitario de La Princesa; IIS-FJD, Instituto Investigación Sanitaria Fundación Jiménez Díaz; USN, Universitetet I Sorost-Norge; IRCCS, Istituto Don Calabria; CIPH, Croatian Institute of Public Health; INANTRO, Institut Za Antropologiju; BU, Baskent Universitesi Vakfi; SMUC, St Mary's University Twickenham; ULSS6, Azienda ULSS6 Euganea; UdeA, Universidad de Antioquía; ASPEUR, Associacao Pro Ensino Superior Em Novo Hamburgo; UNSA, Univerzitet U Sarajevu. [2] Number of centres providing information, and number of patients/individuals in the datasets could change due to the continuous update of the information; NA, Not available at the moment. [3] Study population refers to, 1) COVID-19 hospitalized patients, 2) COVID-19 cases, 3) COVID-19 patients attending primary care services, 4) COVID-19 patients attending emergency services, 5) general population, 6) health care workers, and 7) tourists.

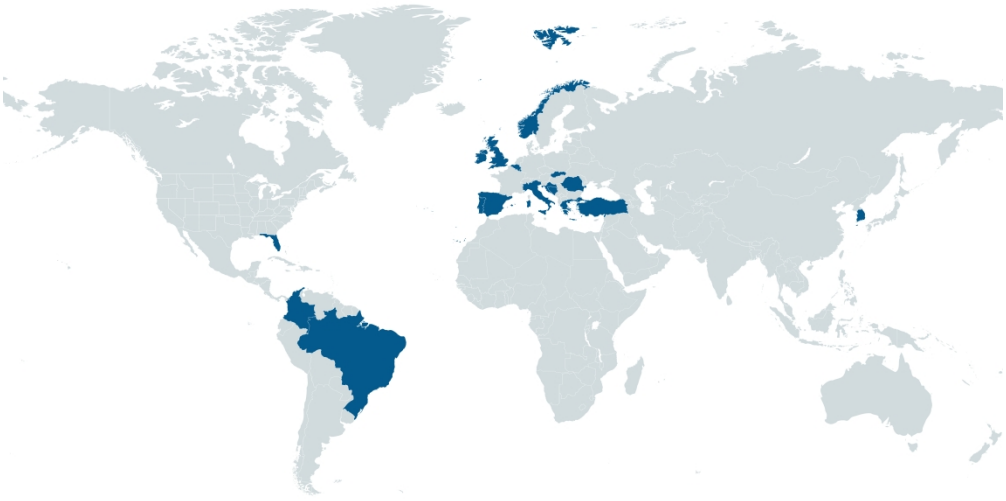
Figure legend

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, United Kingdom, and the United States of America)

Figure 2. Management structure of the unCoVer network

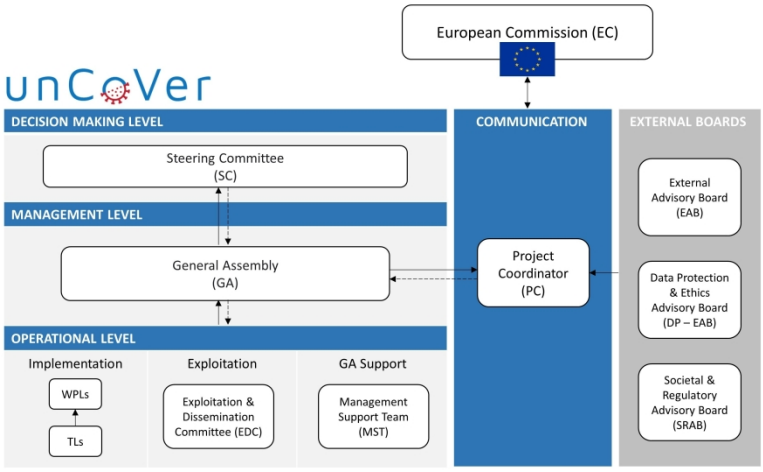
Figure 3. Secure multi-party computation of unCoVer data based on Opal/DataShield infrastructure

Figure 4. Data management process of the COVID-19 related data available within the unCoVer network



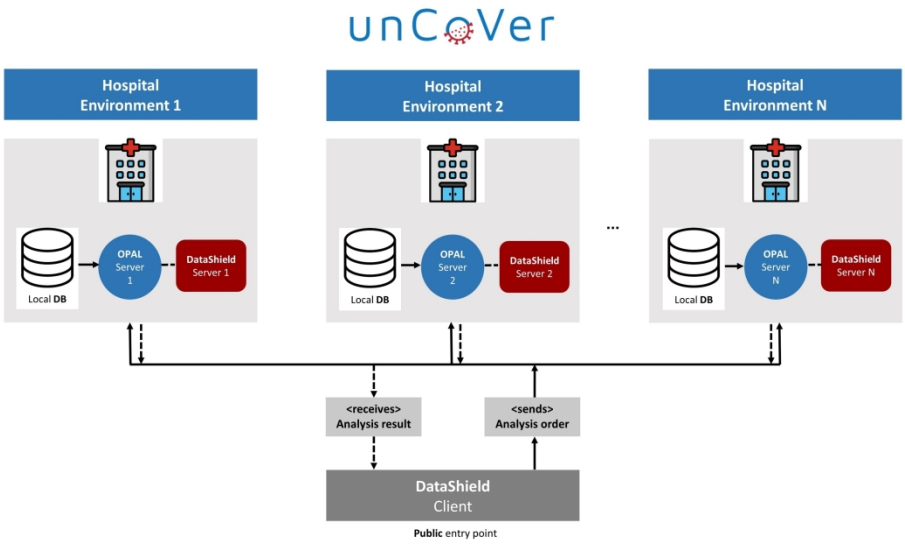
Geographical distribution of the unCoVer network (Belgium, Bosnia and Herzegovina, Brazil, Colombia, Croatia, Greece, Ireland, Italy, Luxembourg, Norway, Portugal, Romania, Slovakia, South Korea, Spain, Turkey, United Kingdom, and the United States of America)

1382x704mm (118 x 118 DPI)



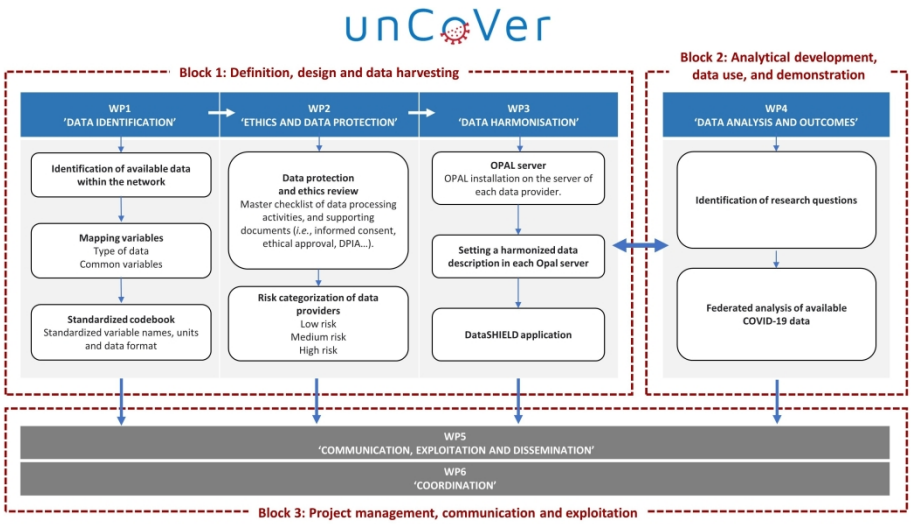
Management structure of the unCoVer network

338x190mm (300 x 300 DPI)



Secure multi-party computation of unCoVer data based on Opal/DataShield infrastructure

304x171mm (300 x 300 DPI)



Data management process of the COVID-19 related data available within the unCoVer network

338x190mm (300 x 300 DPI)

Supplemental Table 1. Clinical and epidemiological data available in unCoVer network databases containing hospitalized patient data.

Country	Institution[1]	Patient admission CFR						Daily CFR during hospitalization						Patient outcome CRF			
								At the ward			In ICU						
		Demographics	Onset and/or Admission date	Signs and symptoms	Laboratory results	Pre-admission medication	Co-morbidities & risk factors	Signs and symptoms	Laboratory results	Supportive Treatment	ICU admission date	Signs and symptoms	Laboratory results	Medication treatment	Complications	Discharge date	Outcome status
	Patient data from Electronic Medical Records																
BA	UNSA	X	X	X	X	X	X			X	X		X	X		X	X
BE	ITM-UZA	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
BR	ASPEUR	X	X	X	X	X	X	X	X		X	X					X
ES	FIHM	X	X	X[2]	X[2]	X[2]	X[2]	X		X[2]	X			X[2]		X	X
ES	HULPr	X	X	X[2]			X		X	X[2]	X[3]			X		X[4]	X
ES	IIS-FJD	X	X	X	X	X	X		X		X[4]	X		X	X	X	X
ES	UNAV	X	X		X												
HR	INANTRO	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X
IT	IRCCS - DB1	X	X	X	X	X	X	X		X	X[3]	X		X	X	X	X
IT	IRCCS - DB4	X	X	X	X		X[3]										
IT	ULSS6	X	X	X[2]	X	X	X	X	X	X	X	X	X	X		X	X
RO	UMFCluij	X	X	X			X		X	X				X		X[4]	X
RO	UMFIASI	X	X	X		X	X		X	X	X			X		X	X
SK	TU	X	X	X	X	X	X	X	X	X	X		X	X		X	X
TR	BU – DB1	X	X	X			X							X			
TR	BU – DB2	X	X	X		X	X	X	X	X	X	X		X		X	X
	Patient data from Public Health Surveillance and National Registers																
SK	TU	X	X	X	X	X	X	X	X	X	X		X	X		X	X
BE	Sciensano	X	X	X	X		X			X	X	X		X	X	X	X
CO	UdeA	X														X	X
HR	CIPH	X	X				X									X	X
IE	TUDublin	X	X				X										
NO	USN	X	X	X	X		X							X	X	X	X
PT	UPORTO	X	X				X[2]			X[3]	X						X
	Patient data from Observational Research																

IT	IRCCS - DB2	X		X			X[3]												
IT	IRCCS - DB3	X		X			X[3]												
LU	LIH	X	X	X	X	X	X	X		X	X[3]		X	X		X	X	X	X

[1]Acronyms: ITM-UZA, Institute of Tropical Medicine and Antwerp University Hospital; FIHM, Fundación Investigación HM Hospitales; UNAV, Universidad de Navarra; UPORTO, Universidade Do Porto; TUDublin, Technological University Dublin; UMF Cluij, Universitatea De Medicina Si Farmacie Iuliu Hatieganu Cluj-Napoca; UMF IASI, Universitatea De Medicina Si Farmacie Grigore T Popa Din Iași; LIH, Luxembourg Institute of Health; TU, Trnavska Univerzita V Trnave; IPC, Instituto Politécnico de Coimbra; HULPr, Hospital Universitario de La Princesa; IIS-FJD, Instituto Investigación Sanitaria Fundación Jiménez Díaz; USN, Universitetet I Sorost-Norge; IRCCS, Istituto Don Calabria; CIPH, Croatian Institute of Public Health; INANTRO, Institut Za Antropologiju; BU, Baskent Universitesi Vakfi; SMUC, St Mary's University Twickenham; ULSS6, Azienda ULSS6 Euganea; UdeA, Universidad de Antioquía; ASPEUR, Associacao Pro Ensino Superior Em Novo Hamburgo; UNSA, Univerzitet U Sarajevu.[2]Data needs to be extracted from text variables; [3]Coded as yes/no without any further specifications; [4]Length of stay in days instead of exact date of admission and/or discharge from hospital and/or ICU

Supplemental Table 2. Data protection Risk Assessment checklist

> Was personal data/sensitive collected?
Yes: What permissions and safeguarding are in place for the individual (evaluated through question below)?
No: What data was collected and what are the risks?

> Was Explicit Consent or Assent received (if necessary)?
Yes: Method of requesting individual consent or assuming assent must be provided to unCoVer
No: Should it have been collected?

> Is Ethical Approval necessary?
Yes: Ethical Approval must be provided to unCoVer
No: If Ethical Approval is not necessary, personal data must not be identifiable (GDPR does not apply) and the data should be 'public' to some extent

> Has the partner received Ethical Approval?
Yes: Does unCoVer need any additional information following Ethical Approval?
No: Hold until Ethical Approval is provided and assess

> Anonymisation / Pseudonymization: Is the process used to anonymise or pseudonymise the data to be shared with uncover known? Who undertook this process?
Yes: Transparent and appropriate methods used were clear and acceptable. Parties with data access are identified. Still ensure no personal data is passed onto unCoVer
No: if these processes were undertaken, is there follow-up required?

> Is there a letter from the Data Protection Officer (DPO) confirming the standardisation of Ethical Approval?
Yes: Letter has been signed and returned by local DPO
No: Follow-up with partner is needed

Low risk	Medium risk (a combination of some of the below)	High risk (a combination of some of the below)
Ethical approval received or not applicable	Ethical approval received or pending	Ethical approval not received
Explicit consent & assent received or not applicable	Explicit consent & assent received or waiting confirmation	Unclear details of consent / assent
Anonymisation and/or pseudonymization transparent & appropriate	Waiting confirmation of anonymization/ pseudonymization process(es)	Anonymization or pseudonymization incomplete. Or process unexplained
Data transfer agreement in place if necessary	Data transfer pending or in place if necessary	Non-EU partners without appropriate Data transfer agreements
DPO letter received	DPO letter pending/received	DPO letter pending

The RECORD statement – checklist of items, extended from the STROBE statement, that should be reported in observational studies using routinely collected health data.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found		RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	Abstract, page 4
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported			Pag 6, 7
Objectives	3	State specific objectives, including any prespecified hypotheses			Pag 7, 8
Methods					
Study Design	4	Present key elements of study design early in the paper			Pag 8
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection			Pag 8, 9, and Figure 1
Participants	6	*(a) Cohort study* - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up		RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided.	Pag 8, 9, 10, Table 1 and Supplemental Table 1

		Case-control study - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case		RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.		RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	Table 1 and Supplemental Table 1
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group			Table 1 and Supplemental Table 1
Bias	9	Describe any efforts to address potential sources of bias			Pag 10, 11, 12
Study size	10	Explain how the study size was arrived at			Pag 10, 11, 12
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why			Pag 10, 11, 12, and Figure 3

Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses			Pag 10, 11, 12, and Figure 3
Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population. RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	Figure 3
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	Figure 3
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the		RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.,* study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in	NA (protocol paper not including results)

		study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram		the text and/or by means of the study flow diagram.	
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)			NA (protocol paper not including results)
Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report numbers in each exposure category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures			NA (protocol paper not including results)
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period			NA (protocol paper not including results)

Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses			NA (protocol paper not including results)
Discussion					
Key results	18	Summarise key results with reference to study objectives			NA (protocol paper not including results)
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias		RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	Pag 16, 17
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence			Pag 16, 17
Generalisability	21	Discuss the generalisability (external validity) of the study results			Pag 16, 17
Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based			Pag 18
Accessibility of protocol, raw data, and programming code		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	Pag 18

*Reference: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.