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 Byunfrate,5 Miran Ćoklo,6 Brecht Devleesschauwer,7,8 Paula Andrea Diaz Valencia,9 João C. Fernandes,10 Enrique Javier Gómez,11,12,13 Paul Hyndes,14 Zubair Kabir,15 Jörn Klein,16 Polychronis Kostoulas,17 Lucia 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 Vuković,35 José M. Castellano,36,37,38 on behalf of the unCoVer network

ABSTRACT

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

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

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

INTRODUCTION

The outbreak of COVID-19, caused by SARS-CoV-2, was declared a public health emergency of international concern by the WHO...
on 30 January 2020 and a pandemic on 11 March 2020. Despite the deployment of public health measures, such as restrictions of movements and gathering, personal protection, as well as massive vaccination campaigns, the coronavirus is still largely affecting mortality and morbidity worldwide, including persisting symptoms after the infection, what has now been termed post-COVID-19 condition. Early epidemiological data on COVID-19 showed a higher risk of severe disease among older individuals, in particular those with chronic respiratory, cardiometabolic and other chronic diseases, including mental disorders and immunosuppressed individuals. While the pathogenesis of certain chronic diseases predisposes to serious COVID-19 outcomes, 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. Also, common disease risk factors such as smoking and overweight/obesity have been identified as key predictors of hospitalisation and critical illness, even in young adults with no underlying conditions.

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

Objectives

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

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

METHODS AND ANALYSIS

UnCoVer is conceptualised as a functional network of partners capable of harvesting and analysing RWD derived from the response and provision of care to patients by the health systems across Europe and other countries, such as Brazil and Colombia, during the COVID-19 pandemic.
unCoVer comprises 29 partners from 25 institutions in the European Union (EU) and 4 non-EU partners representing 18 countries, namely Belgium, Bosnia and Herzegovina, Brazil, Colombia, Croatia, Greece, Ireland, Italy, Luxembourg, Norway, Portugal, Romania, Slovakia, South Korea, Spain, Turkey, UK and USA (figure 1). Partners provide data mostly from front-line hospitals but also from national health agencies, registries and investigator-led observational studies, and represent complementary scientific and medical fields, as well as expertise in research ethics, data management and statistical modelling. unCoVer, thus, works as a functional network bridging clinical expertise and data analytics, intending to exploit the full potential of routine healthcare data already collected from patients during the pandemic. The set-up of the network therefore relies on a continuous iteration process between (1) clinical partners, who will guide the development of research questions needed to improve patients’ care and inform public health strategies; and (2) epidemiologists and analytical experts, who will operationalise the research questions with advanced data processing, analysis and simulation tools capable of generating innovative solutions. The work of the network is further supported by three external advisory boards (figure 2) that provide expert counselling concerning the relevance of medical research and findings (External Advisory Board), data protection (Data Protection and Ethics Advisory Board, DP-EAB) and stakeholder involvement (Societal and Regulatory Advisory Board, SRAB).

Databases

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

Structure and activities

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

Block 1: definition, design and data harvesting

This first block forms the architectural foundation and the core of the unCoVer network. In order to provide a comprehensive repository of available data, WP1 ‘Data Identification’ collects and catalogues all data in
Figure 2  Management structure of the unCoVer network. unCoVer, Unravelling data for rapid evidence-based response to COVID-19.

a standardised way, including a common codebook that specifies the harmonised variables, with standardised variable names and data format and labels, and range of plausible values, in preparation for data harmonisation and validation processes, that is, the key for the development of a unified pool of data for analyses. In parallel, and acknowledging the sensitive nature of health data, and personal information compliance with ethical and legal aspects, a checklist for assessing the risks involved in data processing is implemented by WP2 ‘Ethics and Data Protection’, with due consideration to legal and regulatory issues concerning data protection, privacy and information security. The checklist includes questions on the nature of the data (eg, clinical data, hospital records or publicly available data), personal data, data collected in vulnerable groups, availability of follow-up data), informed consent (eg, explicit consent or assent obtained), data protection (eg, data protection officer identified and data protection impact assessment completed), ethical approval (obtained or pending), data privacy protection (eg, anonymisation vs pseudonymisation, data minimisation), and data transfer and use (data transfer agreement needed, name of the data controller, data processor, joint data controller if applicable, and whether international transfer outside EU). Each data provider within the network populates this checklist, which is reviewed on a case-by-case basis by the unCoVer research ethics expert team together with the independent DP-EAB, and accordingly informed decisions on risk mitigation are taken. In short, WP2 ensures that general data protection regulation (GDPR) or equivalent guidelines are adhered to during the data processing activity. Once a data set is categorised as low risk, it will be available to proceed with subsequent steps (online supplemental table 2).

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

Block 2: analytical development, data use and demonstration
The broad range of medical, public health and research expertise available within the unCoVer network is at the heart of WP4 ‘Data analysis and Outcomes’, dedicated to unCoVer findings from the data acquired to maximise their use in informing COVID-19 response. Activities of
Table 1 Overview of the COVID-19 real-world data among the unCoVer network

<table>
<thead>
<tr>
<th>Country</th>
<th>Institution</th>
<th>Centres*</th>
<th>Start date</th>
<th>End date</th>
<th>Patients*</th>
<th>Study population†</th>
<th>Number of time points</th>
<th>Demographics</th>
<th>Clinical/epidemiology</th>
<th>Human OMICs</th>
<th>Pathogen OMICs</th>
<th>Biospecimens</th>
<th>Imaging</th>
<th>Social network</th>
<th>Movement-related</th>
<th>Mental health</th>
<th>Economic</th>
<th>Diet</th>
<th>Screening</th>
</tr>
</thead>
<tbody>
<tr>
<td>BA</td>
<td>UNSA</td>
<td>10</td>
<td>17 March 2020</td>
<td>20 May 2021</td>
<td>2000</td>
<td>1, 2, 3</td>
<td>≥2</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BE</td>
<td>ITM-UZA</td>
<td>1</td>
<td>22 April 2020</td>
<td>Ongoing</td>
<td>187</td>
<td>1</td>
<td>≥2</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BR</td>
<td>ASPEUR</td>
<td>40</td>
<td>26 March 2020</td>
<td>Ongoing</td>
<td>2000</td>
<td>1, 3</td>
<td>15</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ES</td>
<td>FIHM</td>
<td>12</td>
<td>1 March 2020</td>
<td>15 February 2021</td>
<td>4480</td>
<td>1</td>
<td>≥2</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ES</td>
<td>HULPr</td>
<td>1</td>
<td>1 January 2020</td>
<td>29 September 2020</td>
<td>2217</td>
<td>1</td>
<td>≥2</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ES</td>
<td>IIS-FJD</td>
<td>1</td>
<td>7 March 2020</td>
<td>31 May 2020</td>
<td>1861</td>
<td>1</td>
<td>≥2</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ES</td>
<td>UNAV</td>
<td>2</td>
<td>1 April 2020</td>
<td>1 January 2021</td>
<td>100</td>
<td>1</td>
<td>≥1</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HR</td>
<td>INANTRO</td>
<td>2</td>
<td>1 August 2020</td>
<td>Ongoing</td>
<td>200</td>
<td>1</td>
<td>≥1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IT</td>
<td>IRCCS-DB1</td>
<td>1</td>
<td>3 March 2020</td>
<td>Ongoing</td>
<td>200</td>
<td>1</td>
<td>≥1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IT</td>
<td>IRCCS-DB4</td>
<td>1</td>
<td>1 March 2020</td>
<td>9 May 2020</td>
<td>355</td>
<td>4</td>
<td>≥1</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IT</td>
<td>ULS96</td>
<td>4</td>
<td>1 March 2020</td>
<td>Ongoing</td>
<td>1000</td>
<td>1</td>
<td>≥3</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RO</td>
<td>UMF Cluj</td>
<td>1</td>
<td>1 March 2020</td>
<td>Ongoing</td>
<td>100</td>
<td>1</td>
<td>≥4</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RO</td>
<td>UMF IASI</td>
<td>1</td>
<td>1 March 2020</td>
<td>Ongoing</td>
<td>150</td>
<td>1</td>
<td>≥2</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SK</td>
<td>TU</td>
<td>1</td>
<td>1 October 2020</td>
<td>Ongoing</td>
<td>800</td>
<td>1</td>
<td>≥1</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TR</td>
<td>BU-DB1</td>
<td>6</td>
<td>1 March 2020</td>
<td>Ongoing</td>
<td>7000</td>
<td>1</td>
<td>4</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Public health surveillance data and registers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SK</td>
<td>TU</td>
<td>240</td>
<td>1 March 2020</td>
<td>Ongoing</td>
<td>776 000</td>
<td>2</td>
<td>≥1</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BE</td>
<td>Sciensano</td>
<td>98</td>
<td>14 March 2020</td>
<td>Ongoing</td>
<td>50000</td>
<td>1</td>
<td>≥2</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CO</td>
<td>UdeA</td>
<td>1314</td>
<td>14 March 2020</td>
<td>Ongoing</td>
<td>3 997 021</td>
<td>2</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HR</td>
<td>CIPH</td>
<td>NA</td>
<td>25 February 2020</td>
<td>Ongoing</td>
<td>220 000</td>
<td>2</td>
<td>≥1</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IE</td>
<td>TUDublin</td>
<td>48</td>
<td>29 February 2020</td>
<td>30 November 2020</td>
<td>74 000</td>
<td>2</td>
<td>≥1</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Continued
Table 1 Continued

<table>
<thead>
<tr>
<th>Country</th>
<th>Institution</th>
<th>Centres*</th>
<th>Start date</th>
<th>End date</th>
<th>Patients*</th>
<th>Study population†</th>
<th>Number of time points</th>
<th>Demographics</th>
<th>Clinical/epidemiology</th>
<th>Human OMICs</th>
<th>Pathogen OMICs</th>
<th>Biospecimens</th>
<th>Imaging</th>
<th>Social network</th>
<th>Movement-related</th>
<th>Mental health</th>
<th>Economic</th>
<th>Diet</th>
<th>Screening</th>
</tr>
</thead>
<tbody>
<tr>
<td>NO</td>
<td>USN</td>
<td>40</td>
<td>31 March 2020</td>
<td>Ongoing</td>
<td>2313</td>
<td>1</td>
<td>10</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PT</td>
<td>UPORTO</td>
<td>NA</td>
<td>3 March 2020</td>
<td>Ongoing</td>
<td>830 000</td>
<td>2</td>
<td>1</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Observational research data

<table>
<thead>
<tr>
<th>Country</th>
<th>Institution</th>
<th>Centres</th>
<th>Start date</th>
<th>End date</th>
<th>Patients</th>
<th>Study population†</th>
<th>Number of time points</th>
<th>Demographics</th>
<th>Clinical/epidemiology</th>
<th>Human OMICs</th>
<th>Pathogen OMICs</th>
<th>Biospecimens</th>
<th>Imaging</th>
<th>Social network</th>
<th>Movement-related</th>
<th>Mental health</th>
<th>Economic</th>
<th>Diet</th>
<th>Screening</th>
</tr>
</thead>
<tbody>
<tr>
<td>IT</td>
<td>IRCCS-DB3</td>
<td>1</td>
<td>22 April 2020</td>
<td>25 May 2020</td>
<td>1515</td>
<td>5</td>
<td>1</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LU</td>
<td>LfH</td>
<td>1</td>
<td>15 April 2020</td>
<td>1 April 2021</td>
<td>900</td>
<td>2</td>
<td>5</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PT</td>
<td>IPC</td>
<td>1</td>
<td>1 March 2020</td>
<td>1 July 2020</td>
<td>550</td>
<td>5</td>
<td>1</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UK</td>
<td>SMUC</td>
<td>1</td>
<td>1 July 2020</td>
<td>1 January 2021</td>
<td>200</td>
<td>2</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Screening data

<table>
<thead>
<tr>
<th>Country</th>
<th>Institution</th>
<th>Centres</th>
<th>Start date</th>
<th>End date</th>
<th>Patients</th>
<th>Study population†</th>
<th>Number of time points</th>
<th>Demographics</th>
<th>Clinical/epidemiology</th>
<th>Human OMICs</th>
<th>Pathogen OMICs</th>
<th>Biospecimens</th>
<th>Imaging</th>
<th>Social network</th>
<th>Movement-related</th>
<th>Mental health</th>
<th>Economic</th>
<th>Diet</th>
<th>Screening</th>
</tr>
</thead>
<tbody>
<tr>
<td>IT</td>
<td>IRCCS-DB2</td>
<td>1</td>
<td>31 March 2020</td>
<td>12 May 2020</td>
<td>1635</td>
<td>6</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TR</td>
<td>BU-DB2</td>
<td>1</td>
<td>1 March 2020</td>
<td>Ongoing</td>
<td>18 000</td>
<td>5, 7</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Number of centres providing information and number of patients/individuals in the data sets could change due to the continuous update of information.
†Study population refers to (1) hospitalised patients with COVID-19, (2) COVID-19 cases, (3) patients with COVID-19 attending primary care services, (4) patients with COVID-19 attending emergency services, (5) general population, (6) healthcare workers and (7) tourists.

ASPEUR, Asociaciones Pro Enseñanza Superior Em Nuevo Hamburgo; BA, Bosnia and Herzegovina; BE, Belgium; BR, Brazil; BU, Basler Universitätsklinikum; CO, Colombia; DB, Database; ES, Spain; FIMH, Fundacion Investigacion HM Hospitalaria; HR, Croatia; HULPr, Hospital Universitario de la Princesa; IE, Ireland; IIS-FJD, Instituto Investigation Sanitaria Fundacion Jimenez Diaz; INANTRO, Instituto Za Antrapoligiju; IPC, Instituto Politecnico de Colombia; IRCCS, Istituto Don Calabria; IT, Italy; ITM-UZA, Institute of Tropical Medicine and Antwerp University Hospital; LU, Luxembourg Institute of Health; LU, Luxembourg; NA, not available at the moment; NO, Norway; PT, Portugal; RO, Romania; SK, Slovakia; SMUC, St Mary’s University Twickenham; TR, Turkey; TU, Treska Universita V Tirana; TUDublin, Technological University Dublin; UdeA, Universidad de Antioquia; UK, United Kingdom; ULISS6, Asociación ULISS6; UMF Cluj, Universitatea de Medicina si Farmacie Iuliu Hatieganu Cluj-Napoca; UMF IAS, Universitatea de Medicina si Farmacie Groara T Poppa Din Iasi; UNAI, Universidad de Navarra; unCoVer, Unraveling data for rapid evidence-based response to COVID-19; UNSA, Univerzitet U Sarajevu; UPORTO, Universidade Do Porto; USA, Universitetet I Sorø-Norge.
Figure 3 Secure multi-party computation of unCoVer data based on Opal-DataSHIELD infrastructure. unCoVer, Unravelling data for rapid evidence-based response to COVID-19.

Abbreviation: DB, database.

Block 3: project management, communication and exploitation

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

ETHICS AND DISSEMINATION

The unCoVer study has been approved by the Institutional Review Board of the Institute of Tropical Medicine in Antwerp (IRB/RR/ac/151, protocol number 1524/21). Ethical aspects are of utmost importance in unCoVer. On the project start date, the roles and responsibilities of the independent DP-EAB were
described, including the selection procedure of the board members and its final composition and mandate. A ‘scoping exercise’ was also conducted across the network to ensure that all partners are aware of the common obligations in terms of data processing activities using health or health-related data according to European and international guidelines. Moreover, to be compliant with the GDPR and meet the ethics requirements, the unCoVer network will follow the data processing steps represented in figure 4, in the following sequence:

- The unCoVer master checklist of data processing activities in network partners is circulated within the network to be completed by the data providers, and data providers are required to provide the supporting documentation of each indicator of this list, such as informed consents, ethical approvals and data protection impact assessments. This information is processed by the research ethics team, responsible for categorising the data sets into three different categories: low, medium or high data privacy risk.

- Data sets categorised as ‘low-risk’ will be available to proceed with the harmonisation process and therefore Opal-DataSHIELD servers will be installed. Within Opal, the patient identifiers will be separated from the patient study data by employing two databases: (1) the identifier (ID) database, which stores patient identifiers accessible to the data provider only; and (2) the study database, which stores pseudoanonymised patient data to be used for data analyses accessible, through code only, to data analysts. The ‘medium’- and ‘high-risk’ datasets will be subject to further review and requirements before harmonisation processes.

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

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

**PATIENT AND PUBLIC INVOLVEMENT**

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

- **Scientific community.** To contribute to the body of knowledge in the field, two types of publications in peer-reviewed journals under open-access schemes are foreseen: the unCoVer network publications, that is, for implementation and/or application of the unCoVer platform as a whole; and the unCoVer partners publication, that is, for specific collaborations between two or more unCoVer partners. In addition to both types of scientific publications, jointly organised workshops, virtual trainings and virtual conferences will be instrumental channels for the dissemination of the unCoVer activities and results to the scientific community. In such a yearly organised workshop, the application of the unCoVer repository and toolbox will be presented and expert feedback will be sought for further improvement. These activities of dissemination to the scientific community will result in overall awareness and international recognition of the unCoVer network, simultaneously strengthening the visibility and competitiveness of the institutions involved as centres of excellence.

- **European platforms and data-driven initiatives.** Cooperation with other European projects dedicated to COVID-19 data sharing such as Orchestra (orchestra-cohort.eu), Synchros (synchros.eu), Dragon (imi.europa.eu/projects-results/project-factsheets/dragon), RecodID (recodid.eu) and EC-COVID-19 Data Platform (covid19dataportal.org), as well as large networks such as the European Burden of Disease Network (burden-eu.net) and initiatives on data sharing infrastructures such as the Population Health Information Research Infrastructure (phiiri.eu), will be established for the co-organisation of dissemination events along with seeking alignment and synergies to avoid duplication of efforts.

- **Policymakers.** The accumulated prior experience and contact networks in the regulatory, policymaking framework of several members of the unCoVer network will be used to ensure that the work and output created can reach regulatory entities and policymakers, thus contributing to the impact of the project on the decision-making process. This appears relevant as the unCoVer network is willing to merge different sources of medical data with social, economic, mental and geographical data with the potential to identify highly tailored profiles of risk for community prevention programmes and educational goals in different countries.

- **Engagement with key stakeholders.** Results from the project will be further disseminated through involvement with societal, regulatory and administrative partners, as ensured by the external advisory boards to the network, with the goal to warrant the project’s impact according to stakeholders’ needs and expectations.

- **General public.** To maximise awareness of unCoVer among the general public, the project activities and milestones will be broadcasted via social media.

- **Patient associations and clinicians.** Given the foreseen impact of unCoVer’s output, patient associations and clinicians are identified as end users that benefit from the data-oriented results, and subsequently these can be translated to them via seminars, lectures and infographics as made available on the project website, among others.

**DISCUSSION**

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

**Author affiliations**

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

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

3Public Health Department, Baskent University Faculty of Medicine, Ankara, Turkey
Menéndez, Nerea Ruiز del Árbol y Alberto Estradó); Universidad Politécnica de Madrid, Spain (Ernestina Menasalvas, Enrique Javier Gómez, Alberto Blázquez Herranz, David Fernandez Lobín y Paloma Chausa); Universidad de Navarra, Spain (Paul Nguewa); Universidad de Porto, Portugal (David M Pereira y Mortezos Hossein); Technological University Dublin, Ireland (Paul Hynds, John Kelleher y Elizabeth Hunter); University College Cork, Ireland (Zubair Kabir, Ali Arensman, Brendan Palmer y Georgie O’Sullivan); Universidad De Medicina Si Farmacie Iuliu Hatieganu Cluj-Napoca, Romania (Milenia Mân, Lucía María Lotrean, Mihaela Lupescu și Mihaela Florea); Universidade De Medicina Si Farmacie Grigore T Popa din Iași, Romania (Angetina Carmen Trofor, Andrei Tudor Cernomaz, Radu Adrian Crisan-Dabija y Cristina Grigorescu); Luxembourg Institute of Health (Michel Vaillant y Gay Fagherazzi); Universidade Católica Portuguesa, Portugal (João C Fernandes y João Silva); Tsuva University, Slovakia (Marek Majdan, Daria Bararova, Adriana Krasova, Jaroslava Brnova, Janka Prnova, Jaroslav Slany y Dominika Plancikova); Instituto Politécnico de Coimbra, Portugal (Ana Lucia Baltazar); Hospital Universitario de La Princesa, Spain (Joa B Soriano, Julio Ancoceno, Nisa Bouchchou Abdelkader, Adrián Peláez, Elena Ávalos y Gorane Iturricastillo); Instituto Investigación Sanitaria Fundación Jiménez Díaz, Spain (Lucía Llanos, Miguel Góngolas, Olga Sánchez-Pérez, Antonio Natalo Antioquiano, Siddiquey Ntikinjaka, Gorki Belibaqy, Kúbra Demir, Mustafa Görür, Turgut Bulut y KR Nayar); St Mary’s University Twickenham, UK (Silvia Riva); Azienda ULSS6 Euganea, Italy (Mary Elizabeth Tamang, Carlo Giordani y Petra Golfin); Korea University, South Korea (In-Hwan Oh y Seok Jun Yoon); University of South Florida, USA (Michel Reina Ortiz); Universidad de Antioquia, Colombia (Paula Andrea Diaz Valencia, Lina Ruiz, Juan Pablo Pérez Bedoya, Oscar Ignacio Mendoza, Camilo Hincapié, Boris Rodríguez y Noël Bargeno); Feenave University, Asociacion Pro Ensino Superior Em Novo Hamburgo, Brazil (Fernando Spilkli, Juliane Deise Fleck y Matheus Nunes Weber); Universitar U Sarajevu, Bosnia and Herzegovina (Enisa Dzedovic, Ljubica Burnazovic-Ristic, Senca Cavajdža, Džan Ahmed Jesenkovic y Ljilja Džanovic).

Contributors
JP coordinates the unCoVer network. GS manages overall project activities. JP, EM, Erm, EJG, ZK, JMC and DMP lead the project’s work packages. JP, EM, EA, SA, ALB, DB, MC, SC, PAVY, PH, JK, LLJ, LML, MM, PN, I-HD, DMP, SR, JS, FET, MET, ACT, MJV, SJ, JV and JMC characterised the federated infrastructure. JP, EM, EJG, ZK and JMC designed the data harmonisation system and access to the federated infrastructure. JP, ZK, EM, GO’S, SR and GS supervised compliance with legal and ethical requirements of data use. JP, ICF, JK, PK, MM, EM, DMP, MRO and GS are responsible for the dissemination of the network’s activities. JP, EM, GS, ZK, Erm, JMC, DMP, SR and JS drafted the manuscript and designed the figures. All authors have critically reviewed and approved the final version of the manuscript.

The unCoVer network consists of all individual partner institutions and investigators who are responsible for the identification, characterisation, and compliance with legal and ethical requirements of data use, as well as dissemination and coordination activities.

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

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

Competing interests
None declared.

Patient consent for publication
Not required.

Provenance and peer review
Not commissioned; externally peer reviewed.

Supplemental material
This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and
references arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

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

ORCID iDs
José L. Peñalvo http://orcid.org/0000-0003-0096-6238
Elly Mertens http://orcid.org/0000-0003-4957-4235
Dora Buonfrate http://orcid.org/0000-0003-1080-6822
Jörn Klein http://orcid.org/0000-0002-1803-4167
Miguel Reina Ortiz http://orcid.org/0000-0003-3309-6881
Michel Vaillant http://orcid.org/0000-0003-4714-8128

REFERENCES