Existing Data Sources in Clinical Epidemiology: The Danish COVID-19 Cohort

Anton Pottegård, Kasper Bruun Kristensen, Mette Reilev, Lars Christian Lund, Martin Thomsen Ernst, Jesper Hallas, Reimar Wernich Thomsen, Christian Fynbo Christiansen, Henrik Toft Sørensen, Nanna Borup Johansen, Henrik Stevring, Steffen Christensen, Marianne Kragh Thomsen, Anders Husby, Marianne Voldstedlund, Jesper Kjær, Nikolai C Brun

1Clinical Pharmacology and Pharmacy, Department of Public Health, University of Southern Denmark, Odense, Denmark; 2Department of Clinical Biochemistry and Clinical Pharmacology, Odense University Hospital, Odense, Denmark; 3Department of Clinical Epidemiology, Aarhus University Hospital, Aarhus, Denmark; 4Department of Epidemiology and Population Health, Stanford University, Stanford, CA, USA; 5Department of Medical Evaluation and Biostatistics, Danish Medicines Agency, Copenhagen, Denmark; 6Department of Public Health – Biostatistics, Aarhus University, Aarhus, Denmark; 7Department of Anaesthesia and Intensive Care Medicine, Aarhus University Hospital, Aarhus, Denmark; 8Department of Clinical Microbiology, Aarhus University Hospital, Aarhus, Denmark; 9Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark; 10Infection Preparedness, Statens Serum Institut, Copenhagen, Denmark; 11Data Analytics Center, Danish Medicines Agency, Copenhagen, Denmark

Background: To facilitate research on severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a prospective cohort of all Danish residents tested for SARS-CoV-2 in Denmark is established.

Data Structure: All Danish residents tested by reverse transcriptase polymerase chain reactions (RT-PCR) for SARS-CoV-2 in Denmark are included. The cohort is identified using the Danish Microbiology Database. Individual-level record linkage between administrative and health-care registries is facilitated by the Danish Civil Registration System. Information on outcomes related to SARS-CoV-2 infection includes hospital admission, intensive care unit admission, mechanical ventilation, and death and is retrieved from the five administrative regions, the Danish National Patient Registry, and the Danish Register of Causes of Death. The Patient Registry further provides a complete hospital contact history of somatic and psychiatric conditions and procedures. Data on all prescriptions filled at community pharmacies are available from the Danish National Prescription Registry. Health-care authorization status is obtained from the Danish Register of Healthcare Professionals. Finally, selected laboratory values are obtained from the Register of Laboratory Results for Research. The cohort is governed by a steering committee with representatives from the Danish Medicines Agency, Statens Serum Institut, the Danish Health Authority, the Danish Health Data Authority, Danish Patients, the Institutes of Health Sciences at the Danish universities, and Danish regions. The steering committee welcomes suggestions for research studies and collaborations. Research proposals will be prioritized based on timeliness and potential clinical and public health implications. All research protocols assessing specific hypotheses for medicines will be made publicly available using the European Union electronic Register of Post-Authorisation Studies.

Conclusion: The Danish COVID-19 cohort includes all Danish residents with an RT-PCR test for SARS-CoV-2. Through individual-level linkage with existing Danish health and administrative registries, this is a valuable data source for epidemiological research on SARS-CoV-2.

Keywords: Covid-19, SARS-CoV-2, epidemiology, follow-up, database, prognosis, prospective cohort

Introduction

Since the first cases of coronavirus disease 2019 (COVID-19) were identified in Wuhan, China in December 2019,1 the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has spread worldwide.2 In mid-March, Europe was named the epicenter of the pandemic with more reported cases and deaths than the rest of the world combined.3 The first COVID-19 case in Denmark was reported on February 27, 2020, and SARS-CoV-2 was hereafter transmitted widely in the Danish community despite efforts to prevent virus spread.4
The COVID-19 crisis has led to a surge of research activities, from laboratory studies of vaccine candidates, epidemiologic assessment and modeling of the spread of the disease, and effects of non-pharmaceutical interventions such as school closures and physical distancing, to clinical vaccine trials, and randomized clinical trials of drugs to treat COVID-19. Several hypotheses related to the use of medicines of major clinical and public health importance have appeared. As one example, reports that have been widely cited in the media suggest that the use of non-steroidal anti-inflammatory drugs (NSAIDs) may increase the risk of severe or fatal COVID-19. The European Medicines Agency stated that there is no clinical evidence to establish a link between NSAIDs and worsening of COVID-19 and highlighted that there is a “[...] need for epidemiological studies to be conducted in a timely manner to provide adequate evidence of any effects of NSAIDs on disease prognosis of COVID-19.” Another example is the concern that the use of angiotensin converting enzyme inhibitors or angiotensin receptor blockers leads to a worse prognosis in COVID-19. Again, the European Medicines Agency and the European Society of Cardiology have stressed the lack of evidence and recommended strongly against interrupting RAAS inhibitor treatment on the current evidence base.

Given the rapid spread of SARS-CoV-2 globally, these and other questions on the use and safety of medicines need to be examined in a timely manner. Thus, by an initiative of the Danish Medicines Agency’s Data Analysis Centre, a working group was asked to establish a nationwide prospective cohort of all SARS-CoV-2 PCR positive cases as well as those with negative SARS-CoV-2 tests in Denmark and link this cohort to established data sources on eg, drug use and hospital diagnoses. In this paper, we describe the construction of this cohort, the registry data that is linked to it, data governance availability, the Danish health-care system and its early response to COVID-19, and the perspectives for use of the Danish COVID-19 cohort for research.

Materials and Methods

The Danish COVID-19 cohort is a prospective cohort of all individuals in Denmark with confirmed COVID-19 from SARS-CoV-2 detection using reverse transcriptase polymerase chain reactions (RT-PCR) as well as those with negative PCR for SARS-CoV-2. Data on patient demographics, vital status, current and previous drug treatment, and coexisting medical conditions are obtained from the existing Danish nationwide administrative and health registries (Table 1). Individual and unambiguous record linkage is possible using the unique Danish civil registration number assigned to all Danish residents at birth or date of residency. The cohort and linked data are updated twice weekly. As it is likely that research questions will continue to rise, additional data from other sources will be added as needed. Changes in the governance and setting of the COVID-19 cohort may also be deemed necessary as the COVID-19 epidemic evolves. Any such, major changes to data structure or governance will be provided at a dedicated website hosted by the Danish Medicines Agency, which can be found at www.DKMA.dk/DACCovid.

Study Population

The cohort consists of all Danish individuals tested by RT-PCR for SARS-CoV-2 from February 27 and onwards, identified using the national Danish Microbiology Database which includes individuals tested in both primary and secondary health-care settings and governmental prevalence studies. Both positive and negative cases are included.

Data Sources

Basic demographic data on age and sex as well as migrations are obtained from the Danish Civil Registration System.

The Danish Microbiology Database (MiBa) is a national database established in 2010 to enable real-time surveillance of communicable diseases. The database collects an electronic copy of all microbiological test reports, and hence covers microbiology testing in the country, including public and private hospitals as well as general practitioners and private practicing specialists. All records on samples analyzed by RT-PCR for SARS-CoV-2 are retrieved from the Danish Microbiology Database and analyzed using automated search queries and data algorithms and transferred to a COVID-19 surveillance database at Statens Serum Institut, where data are combined with data on COVID related outcomes (described below). Samples for RT-PCR are mainly naso- or oro-pharyngeal swabs or lower respiratory tract specimens.

Data on outcomes related to COVID-19 are based on reporting twice daily from the five Danish regions to Statens Serum Institut. Accordingly, there is minimal delay in the registration of outcome data based on this source. The data include the date of hospital admission and discharge, intensive care unit admission, mechanical ventilation, and renal replacement therapy (see Table 1). An alternative source with more detailed data regarding hospital treatment of COVID-19 is the Danish National
Table 1 Currently Available Data in the Danish COVID-19 Cohort

<table>
<thead>
<tr>
<th>Type of Information</th>
<th>Variables</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographic data</td>
<td>Sex</td>
<td>The Danish Civil Person Registry</td>
</tr>
<tr>
<td></td>
<td>Date of birth</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dates of immigration/emigration</td>
<td></td>
</tr>
<tr>
<td>COVID-19 test</td>
<td>Sampling date</td>
<td>The Danish Microbiology Database</td>
</tr>
<tr>
<td></td>
<td>Test result</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Test type</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sample material</td>
<td></td>
</tr>
<tr>
<td>COVID-19 outcomes</td>
<td>Date of hospital admission</td>
<td>Direct reporting from the Danish regions to Statens Serum Institut</td>
</tr>
<tr>
<td></td>
<td>Date of hospital discharge</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Date of intensive care unit admission</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Date of mechanical ventilation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Date of renal replacement therapy</td>
<td></td>
</tr>
<tr>
<td>Medical conditions and procedures</td>
<td>Diagnosis code (ICD-10)</td>
<td>The Danish National Patient Registry</td>
</tr>
<tr>
<td></td>
<td>Date of diagnosis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Surgery code (NCSP)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Date of surgery</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Other procedures, eg BFCA0: implantation of pacemaker</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Date of procedure</td>
<td></td>
</tr>
<tr>
<td>Vital status</td>
<td>Date of death</td>
<td>The Danish Causes of Death Registry and the Danish Civil Person Registry</td>
</tr>
<tr>
<td>Drug use</td>
<td>Date of prescription fill</td>
<td>The Danish National Prescription Registry</td>
</tr>
<tr>
<td></td>
<td>Drug dispensed (ATC codes)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pack size</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Number of packs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Strength of tables</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Drug administration and formulation</td>
<td></td>
</tr>
<tr>
<td>Authorized health care worker</td>
<td>Health care authorization and profession</td>
<td>The Danish Register of Healthcare Professionals</td>
</tr>
<tr>
<td>Laboratory test results</td>
<td>Laboratory values for selected biochemical markers</td>
<td>Register of Laboratory Results for Research</td>
</tr>
</tbody>
</table>

Abbreviations: ICD-10, International Classification of Diseases version 10; NCSP, Nordic Classification of Surgical Procedures; ATC, Anatomical Therapeutic Chemical Classification System.

Patient Registry, which includes information on intensive care unit admission, mechanical ventilation, dialysis, extracorporeal membrane oxygenation, and other non-invasive and invasive procedures. The data are partly reported at the time of hospital admission and partly at discharge, and may thus be delayed several weeks for admissions of a long duration.

Data on deaths are retrieved from the Danish Register of Causes of Death that is based on mandatory physician-filled certificates for all deaths in Denmark.

The Danish National Patient Registry also provides nationwide data on previous somatic in- and out-patient clinic diagnoses and procedures as well as psychiatric in- and out-patient contacts. No data on in-hospital drug treatment are available, except for treatments coded as procedures (eg, chemotherapy or biological treatments). The reporting structure of the National Patient Registry has been changed and modernized in a transition process. Hospitals have switched to Patient Registry version 3.0 reporting from February 1, 2019, which is expected to have led to some data breaches. Importantly, these data breaches are expected to be of small magnitude during the COVID-19 pandemic, as more than one year has passed since the transition to version 3.0. As of May 5, 2020, the
Danish Health Data Authority declared that the registry is considered to be of suitable quality for research purposes (also retrospectively).

The Danish National Prescription Registry is used to collect information on the use of drugs. The Danish National Prescription Registry has recorded all prescriptions filled at any community pharmacy in Denmark since 1995, including prescriptions issued by private practicing specialists and hospital prescribers. The type of drug (categorized according to the Anatomic Therapeutic Chemical index), date of filling, and quantity as expressed by the number of defined daily doses and the strength of dosing unit are recorded. The registry does not record dosing information or indication for prescribing and does not contain information on drugs administered at hospitals or provided directly to the patients by hospitals.

The Danish Register of Healthcare Professionals is used to identify individuals with COVID-19, who are registered as authorized health-care professionals in Denmark at the time of diagnosis, and include data on the profession of the health-care worker (eg, physician or nurse).

Data on vital status, residency, and migrations are retrieved from the Danish Civil Registration System. The Register of Laboratory Results for Research is an emerging data source in Danish epidemiology. While nationwide coverage has not yet been achieved, four out of five Danish health regions have contributed biochemical test results to this database from 2015 onwards. From this register we obtained laboratory test results for a range of biochemical tests, eg, hematological parameters, creatinine, electrolytes, liver function tests, and C-reactive protein, taken in and out of hospital.

The Danish Health Care System
Denmark had 5,822,763 inhabitants in the first quarter of 2020, all of whom have access to universal, tax-funded health care including general practice, practicing specialists, and hospital care. Prescription drugs are eligible for reimbursement from the government, with a maximal annual patient co-payment of approximately 560 EUR. The government-funded health-care system combined with individual-level administrative record keeping enables longitudinal research using the entire Danish population as a cohort. The access to secondary health care in Denmark is through the general practices acting as gatekeepers to the rest of the health-care system. During the COVID-19 pandemic, the general practices continue to act as gatekeepers by referring patients with suspected COVID-19 symptoms to hospitals. Primary care physicians are advised to triage and refer patients with suspected COVID-19 using telemedicine instead of seeing patients in consultation.

Danish COVID-19 Testing Strategies
In the initial phase of observed COVID-19 clusters in Denmark in late February and early March, efforts were made to prevent the disease by testing as many suspected cases of COVID-19 as possible, with a focus on cases with a relevant travel history (mainly from China and Italy). As the disease began to spread more widely in the community, the focus changed on March 12 from containment of the disease to a mitigation strategy, with a focus on preventing spread within hospitals and optimizing the use of the limited available testing capacity. Thus, before March 12, SARS-CoV-2 tests were offered to all individuals suspected to have COVID-19, based on relevant symptoms together with a travel history to certain geographical regions with reported high prevalence (mainly China and Northern Italy) or close contact with confirmed COVID-19 cases. On March 12, the testing strategy changed to prioritize tests of patients who had suspected COVID-19 requiring hospital admission, and from March 18 extended to include symptomatic health-care workers in critical functions (eg, emergency room and intensive care unit personnel). Testing capacity has, however, subsequently been gradually upscaled in Denmark, and testing indications were widened late March 2020 to include patients with moderate and mild symptoms of COVID-19, as well as broader screening of health-care personnel. During April and May, tests were made available to the whole population, including those with mild symptoms and those with no symptoms but contact with a confirmed COVID-19 case from April 21. From May 24, all Danes can be tested even without a prescription from the general practitioner or other health-care professionals.

These changes in testing strategy have introduced variation in patient characteristics and disease severity in the different phases of the epidemic, unrelated to the disease itself. This variation needs to be handled in studies using these data. If testing strategies are changed further, a description of these changes will be made available at the project’s website (www.DKMA.dk/DACCOVID).

Collaboration and Governance
The Danish COVID-19 cohort is established by the initiative of the Danish Medicines Agency’s Data Analysis Centre in order to leverage the Danish registry data to provide evidence in a timely manner. The project is managed by a steering
committee with representatives from the Danish Medicines Agency, Statens Serum Institut, the Danish Health Authority, the Danish Health Data Authority, Danish Patients, the Faculty of Health Sciences of the Danish Universities, and Danish Regions. An epidemiological working group consisting of researchers from the University of Southern Denmark, Aarhus University, Statens Serum Institut, and the Danish Medicines Agency has been established to manage the data and assist with design and data analysis. Studies using the Danish COVID-19 cohort will aim to optimize timely research dissemination, transparency, and reproducibility by publishing protocols online in the EU PAS registry (www.encepp.eu) and by designing and reporting studies according to relevant guidance such as the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance Guide on Methodological Standards in Pharmacoepidemiology and Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement.32-35

Data Access and Study Propositions
Study ideas and suggestions for collaborations can be directed to the Danish Medicines Agency at daccovid@dkma.dk. The steering committee prioritizes study suggestions according to a predefined and published set of criteria including, among others, relevance to public health, timeliness, and feasibility. For transparency, all requirements for submitting a new proposal, already prioritized study proposals, and current steering and expert group members are listed on www.DKMA.dk/DACCovid.

Data Availability, Ethics Approval and Informed Consent
The Danish COVID-19 cohort data are kept at the Danish Health Data Authority (record no 00004874) and approved by the Data Protection office at the University of Southern Denmark (record no 10.960). Data are pseudonymized centrally at the Danish Health Data Authority. According to Danish law, ethical permission is not required for registry-based research. Individual-level data will not be made publicly available in accordance with Danish law.

Strengths and Weaknesses
The Danish COVID-19 cohort is well suited to quickly address a range of hypotheses given its size and complete geographical and temporal coverage of COVID-19 patients. An important strength of the cohort is the individual-level linkage with existing Danish registries with nationwide coverage that have been used and validated for research purposes. Data from additional registries and clinical databases can be added as necessary. This could include, for example, data from the Danish Intensive Care Database with more detailed information on intensive care treatment.36 The Danish COVID-19 cohort is updated twice weekly, allowing for a limited delay in data extraction. The uniform Danish health-care system with established access for testing and treatment of COVID-19 ensures a well-defined population with complete information. For example, the cohort is well suited to address hypotheses regarding the safety of ibuprofen use in patients with COVID-1914 or the influence of angiotensin converting enzyme inhibitors and angiotensin receptor blockers on the prognosis of COVID-19.16 Finally, the inclusion of individuals with negative SARS-CoV-2 tests allows for supplementary analyses to be made using this subpopulation as a control population, eg, in case-negative control analyses.37 In using these patients, however, care needs to be taken to address and handle the underlying reason for testing, eg, hospital admission for non-COVID-19-related reasons.

As in other countries, a restrictive testing scheme was implemented in Denmark at the beginning of the epidemic, where mainly patients with suspected moderate to severe COVID-19 were tested. Thus, a high number of milder infections were undetected and not included in the Danish COVID-19 cohort initially. As testing strategies continue to change in Denmark as well as in other countries, the characteristics of cases are subject to change over time, and it is important to consider which individuals have been offered testing at a given time. For example, researchers examining determinants for the transition from outpatient to inpatient status should carefully consider what patient categories are established as outpatient cases. Restriction to hospitalized patients in a study could ensure some degree of homogeneity regarding disease severity. An alternative strategy is to stratify by calendar time according to the testing strategy in place during a given period. Another limitation of the cohort is that data on in-hospital drug use are not available, which might hinder the assessment of some clinical questions where in-hospital drug treatment is of particular importance. Further, there are no detailed clinical data available from eg, medical charts or radiology reports for COVID-19 patients. Finally, the assessment of “health-care professional” status is based on the Danish Register of Healthcare Professionals, which only covers authorization status, and not whether the person currently works in the health care system.
Conclusions
With its population-based design, individual-level data linkage, and recency of data, the new Danish COVID-19 cohort has promising potential to aid researchers, decision-makers and clinicians evaluate and address some of the inevitable clinical problems and questions that will arise during the COVID-19 pandemic.

Acknowledgments
The Danish Health Data Authority, in particular Sophie Fisker Nielsen and Martin Lind Kelstrup, and Statens Serum Institut, in particular the people from Data Integration and Analysis, are acknowledged for extraordinary support with preparation and linkage of data. Further, the departments of Clinical Microbiology throughout Denmark are acknowledged for contributing to the national infectious disease surveillance.

Author Contributions
All authors contributed substantially to establishing the working group, acquisition of the data, and analysis plan of the data. All authors took part in drafting or revising the manuscript critically. All authors approve the final version of the manuscript for publication and agree to be accountable for all aspects of the work.

Funding
There is no external funding to report for this study.

Disclosure
KBK, NBJ, MAV, SC, NB, AH, JK, MTE, and MKT declare no conflicts of interest. RWT, CFC, and HTS declare no personal conflicts of interest. The Department of Clinical Epidemiology is involved in studies with funding from various companies as research grants to and administered by Aarhus University. None of these studies are related to the current study. HS reports personal fees from Bristol-Myers Squibb, Novartis, and Roche, outside the submitted work. AP and JH report participation in research funded by Alcon, Almirall, Astellas, AstraZeneca, Boehringer-Ingelheim, Novo Nordisk, Servier, and LEO Pharma, all with funds paid to the institution where they were employed (no personal fees) and with no relation to the work reported in this paper. MR reports participation in research projects funded by LEO Pharma, with funds paid to the institution where she was employed (no personal fees) and with no relation to the work reported in this paper.

References


