

Construction, Quality Control and Regulatory Aspect of a Database of 30,423 COVID-19 Patients Cared for at the IHU Méditerranée Infection France

Philippe Brouqui* and Didier Raoult

MEPHI, Aix Marseille Université, 19-21 Boulevard Jean Moulin, 13005 Marseille, France

***Corresponding author:** Philippe Brouqui, MEPHI, Aix Marseille Université, 19-21 Boulevard Jean Moulin, 13005 Marseille, France

ARTICLE INFO

Received: 📅 August 17, 2023

Published: 📅 August 28, 2023

Citation: Philippe Brouqui and Didier Raoult. Construction, Quality Control and Regulatory Aspect of a Database of 30,423 COVID-19 Patients Cared for at the IHU Méditerranée Infection France. Biomed J Sci & Tech Res 52(3)-2023. BJSTR. MS.ID.008265.

ABSTRACT

In this paper, we describe the construction of a database of 30 423 COVID-19 patients cared for in our Institute (the Institut Hospitalo- Universitaire (IHU) Méditerranée Infection, Marseille, France) between 2 March 2020 and 31 December 2021. We wish to make this database findable, accessible, interoperable and reusable (FAIR) in order to allow any researcher to carry out their own analysis and drawn their own conclusions. The dataset was build using comprehensive data from independent administrative sources including hospital admission files, computerised pharmacy prescription files, and the French National Death Registry (INSEE). We present a full description of the raw data and the data processing. We also provide an external validation of the dataset. The construction of the database and quality control of the data was followed by a mandated by an independent adjudicator who verified and attested to the presence of all the traceability elements guaranteeing the quality of the data in the database. The database is accessible at <https://doi.org/10.57760/sciencedb.07803> and <https://doi.org/10.5061/dryad.ksn02v78v>.

Keywords: SARS-CoV-2; COVID-19; COVID-19 Drug Treatment; Open Data

Introduction

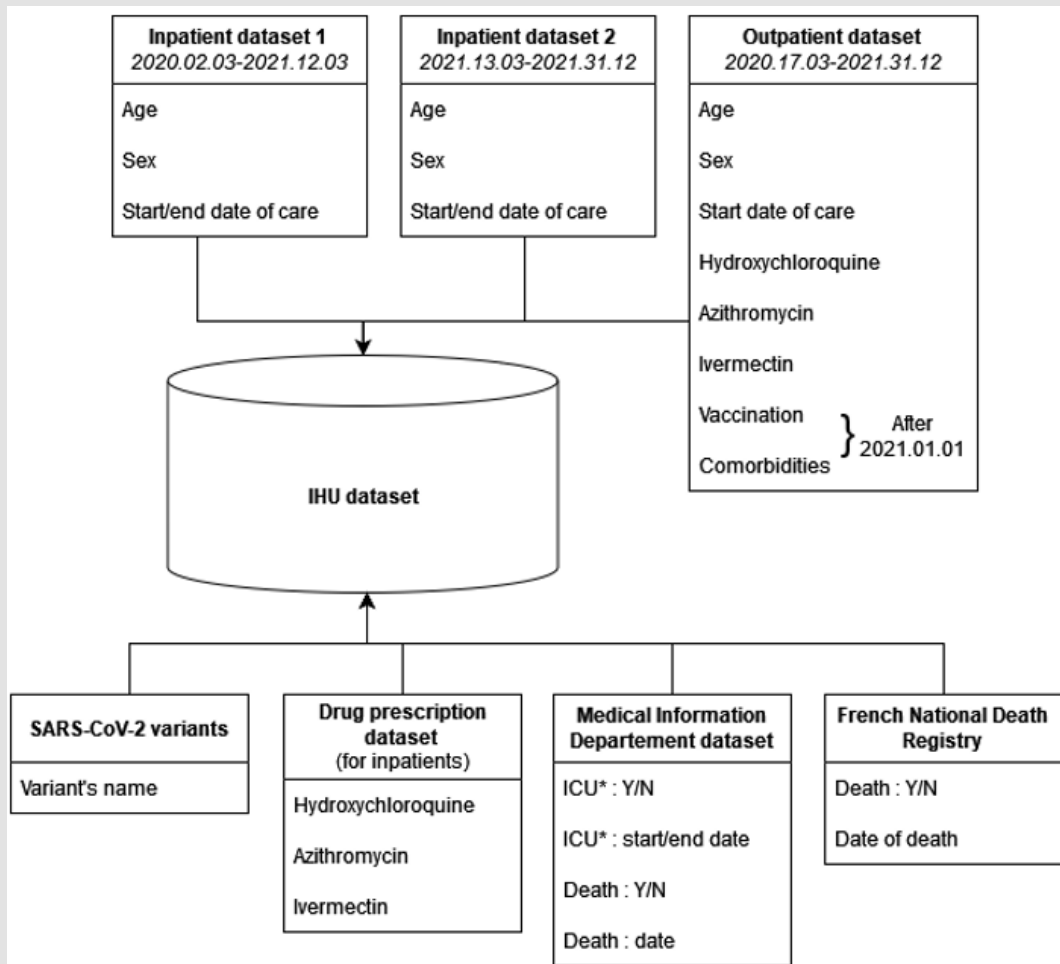
As of March 2023, the Coronavirus disease 2019 (COVID-19) pandemic had caused 677 million illnesses and 6.9 million fatalities [1]. In our Institute we treated more than 30 000 patients with COVID-19. During the pandemic, the transparency, availability and verifiability of raw data and data processing were identified as critical issues [2-4]. This inspired us to collect unbiased raw data and share it with the scientific community. In order to the database was findable, accessible, interoperable, and reusable (FAIR) and as transparent as possible, we used raw data from different administrative sources, including hospital admission files and pharmacy prescription files, while recorded deaths were obtained from the French National Death Registry held by the Institut National des Statistiques et des Études Économiques (INSEE) (12). With this information we generated a retrospective cohort dataset of 30423 COVID-19 patients. In this paper, we present a full description of the raw data and the data processing. We also provide an external validation of the dataset.

Materials and Methods

This database of 30 423 COVID-19 patients treated for COVID-19 in our Institute between 2 March 2020 and 31 December 2021, has been released to be used by independent researchers to analyse the outcomes (ICU admission and death) of patients with COVID-19 depending on the treatment received. The way variables have been obtained (institutional and recognised database) should avoid any doubt as to the validity of the data source.

Source of Data

The database was constructed by merging different databases created through medical records collected in the hospital information system (outpatient and inpatient datasets), treatments recorded in the hospital pharmacy database, virus genomes collected in the microbiology lab database, and mortality data from the INSEE national database (Figure 1). The final dataset was obtained by merging these datasets according to patient status (inpatient/outpatient) and time during the pandemic.



Note: *ICU: Intensive care unit.

Figure 1: IHU dataset - Data collection.

Data Collection: Between 2 March 2020 and 12 March 2021, data on inpatients were collected using the Electronic Patient Record (EPR). The EPR centralises all medical information about a patient with regards to their hospital stay at the Assistance Publique-Hôpitaux de Marseille (AP-HM). Between 13 March 2021 and 31 December 2021, data were extracted from the AP-HM administrative database (PASTEL). Medical data are not available in the PASTEL database, which only includes the following information: IPP (the “identifiant permanent du patient”, a unique identifier for each patient), age, gender, date of hospitalisation, and date of discharge. For outpatients, the medical records of all patients who attended the outpatient unit were completed by the medical staff. The content of these medical records evolved during the epidemic, particularly as new knowledge on the

disease emerged (prevention and treatment of thromboses, etc.), with regards to the collection of information on patients’ vaccination status, and risk factors. From the outset, the medical record contained at least the following information: IPP, gender, age, and date of hospital admission.

Inclusion Criteria: Data on all patients ≥ 18 years of age, with PCR-proven COVID-19, regardless of symptoms (asymptomatic or symptomatic), who received care in our Institute, and underwent a medical examination by one of the doctors as an outpatient (day clinic) or inpatient (hospitalised for at least one night) were included in the database (Figure 2). The studied period was 2 March 2020 to 31 December 2021.

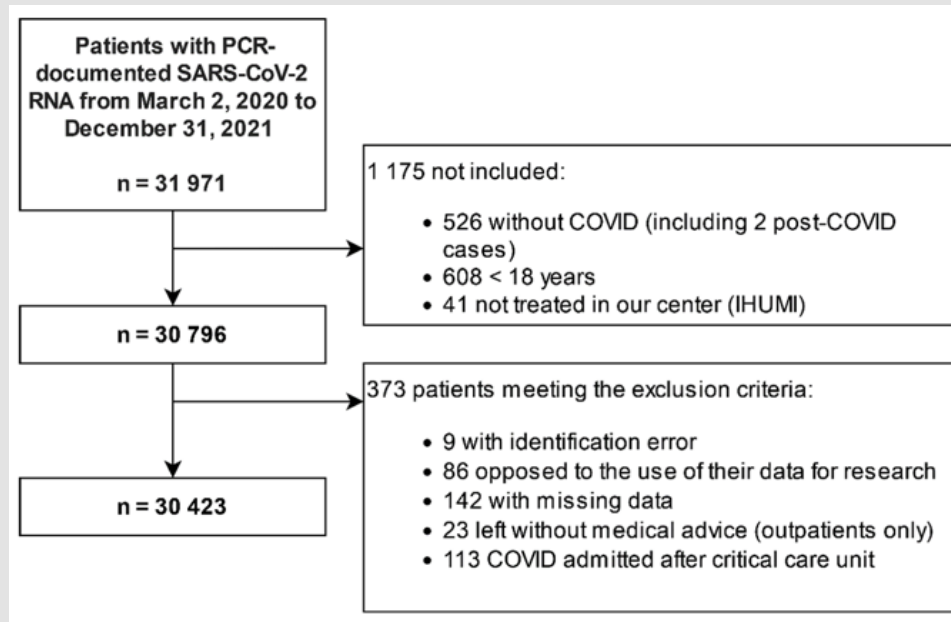


Figure 2: Flowchart of the retrospective Cohort (n=30 423).

Exclusion Criteria: Inaccurate patient identification (wrong patient ID and duplicates), lack of medical data, absence of COVID-19 after checking the medical record (including patients without COVID-19 consulting for a post-COVID-19 syndrome), or a statement of opposition to the use of their medical data for research purposes (in accordance with the European General Data Protection Regulation, see Ethics section below) were reasons for exclusion from the cohort. Inpatients treated in our Institute following a stay in the intensive care unit were also excluded, as well as outpatients who left the Institute without receiving any medical advice (Figure 2).

SARS-CoV-2 Variants: SARS-CoV-2 variants were retrieved from the microbiology laboratory database, which included all the samples and their associated genotypes with a unique patient identifier (Nex-Labs Technidata Medical Software extraction). Virus variants were characterised and named according to the Pangolin classification [5], as previously reported, with the exception of the first epidemic period. The letter 'W' is used here to designate all Wuhan-derived SARS-CoV-2 cases, the variant that circulated during the first epidemic period in our geographical area (from February to May 2020).

Treatment Data: For outpatients, data on treatment were extracted from the medical records. For inpatients, data on treatment were extracted from the PHARMA medico-administrative database (AP-HM drug prescription database). Treatments were identified using ATC codes (J01FA for azithromycin (AZ), D11AX for hydroxychloroquine (HCQ) and P02CF for ivermectin (IVM)).

Clinical Outcomes: To identify deaths (from all causes), two separate files were used:

- 1) An extraction from the Medical Information Department (DIM). If a patient died during their hospital stay, the information system retrieves this information from the "discharge" section.
- 2) The nominative file of deaths registered by INSEE [6]. By using these two sources of information, we minimised the risk of overlooking deaths, especially among outpatients. Data on transfer to an intensive care unit were retrieved from DIM data only.

Data Management: SAS software, version 9.4 (SAS Institute, Cary NC), was used to read the entered data files, check for potential errors, apply corrections, and merge datasets.

Final Dataset: The final dataset was the product of merging these datasets (after removing duplicate records). It included the following information: IPP, age, gender, outpatient treatment (Y/N), inpatient treatment (Y/N), date of start/end of treatment. If a patient had been treated for more than one episode of SARS-CoV-2 during the follow-up period, only the first episode was considered.

Anonymisation: In order to release the database in accordance with the General Data Protection Regulation, the IPP, the patients' first and surnames, and their dates of birth were removed from the database. Age was indicated by a range, making it impossible to identify patients from the database data.

Final Dataset Available in the Database: The following data were retained in the final database and are available in open source: age range, gender, pandemic period, outpatient, inpatient, HCQ, AZ, IVM, virus genomic variant, ICU treatment, time of death, vaccine status, obesity, asthma, cancer, immunodeficiency, chronic cardiac disease and auto-immune disease. The description of the data and the file structure are reported in detail in the “readme” file in the database folder.

Ethics Statement

Patient management and this conduct of retrospective study were performed in accordance with the Helsinki Declaration, as revised in 2013 [7], and the International Ethical Guidelines for Health-related Research Involving Humans [8]. This study does not constitute research involving humans within the meaning of Article R1121-1 of the French Public Health Code, because its purpose is in the public interest of research, study, or evaluation in the field of health “conducted exclusively from the exploitation of the processing of personal data”. This methodology, as well as the retrospective nature of the study, was approved by the Méditerranée Infection independent ethics committee (No. 2021-015). In accordance with European Regulation No. 2016/679, known as the General Data Protection Regulation

(GDPR), the protocols were registered in the hospital’s GDPR registry under numbers 2020-151 and 2020-152, and all patients were informed of the potential reuse of their data through the Institute’s information procedure, which informed them of their right to object via the MyAPHM online portal and/or by post or email addressed to the establishment’s Data Protection Officer. Patients who objected to the use of their data were excluded before data were collected and extracted from the information system.

Dataset Validation

We performed an external validation of our dataset using data from the French National Hospital Database (PMSI). The PMSI is a medical-administrative database that gathers data on hospitalisation which are communicated on a monthly basis by all public and private hospitals in France [9]. Data are uploaded by each hospital to a secure national platform managed by the French Agence Technique de l’Information sur l’Hospitalisation (ATIH). A physician from AP-HM, responsible for medical information and evaluation, extracted data from the PMSI database for all patients treated for COVID-19 in our centre between 2 March and 31 December 2021. Using the IPP reference, we were able to merge our dataset with the data extracted from PMSI (Figure 3).

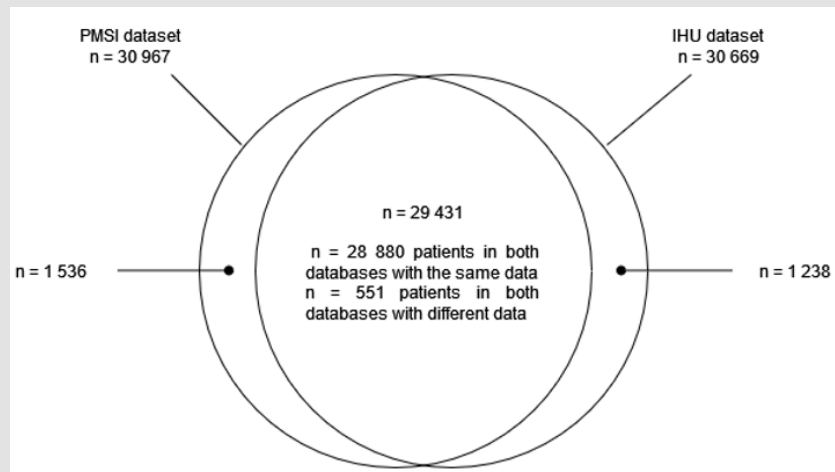


Figure 3: External dataset validation.

This allowed us to identify four groups of patients:

1. Patients in both datasets (same IPP) with the exact same data (age, sex, death status, ICU transfer, treatment data and patient management, n=28 880): no action required.
2. Patients in both datasets (same IPP) with different data (n=551): data were manually checked.
3. Patients who only appeared in the IHU dataset (n=1 238):

a) For outpatients (n=794): all data were manually checked; most patients (90%) were kept in our dataset.

b) For inpatients (n=444): we manually checked 5% of this group of patients (n=23, randomly selected). Twenty-two out of 23 patients were non-COVID-19 patients. We then decided to exclude most of these patients from our dataset (only four patients were kept).

4. Patients who only appeared in the PMSI dataset (n=1 536):

a) For outpatients (n=1 375): all data were manually checked; one patient out of two (52%) was kept in our dataset.

b) For inpatients (n=161): we manually checked 25% of this group of patients (n=41, randomly selected). Thirty-nine out of 41 patients were actual COVID-19 inpatients treated in our Institute. We decided to add most of these patients to our dataset (159 out of 161).

This external validation with PMSI data allowed us to consolidate our dataset (final dataset size n=30 423).

Retrospective Analysis of Published Paper

We then checked the robustness of the results published in our previous studies on patients treated for COVID-19 in our Institute [10-12]. We did not detect any statistically significant differences in the number of deaths and transfers to ICU compared to those published in these three studies (Table 1).

Table 1: Dataset validation - Comparison with previously published articles.

Article	Published data			IHU Verified dataset [†]				
	Sample Size	ICU transfer	Deaths	Sample size	ICU Transfer	p-value [‡]	Deaths	p-value [‡]
Outcomes of 3,737 COVID-19 patients treated with hydroxychloroquine/azithromycin and other regimens in Marseille, France: A retrospective analysis.	3 737	67 (1.8%)	35 (0.9%)	3 819	67 (1.8%)	0.899	48(1.3%) ^{††}	0.182
Outcomes of 2111 COVID-19 Hospitalized Patients Treated with Hydroxychloroquine/ Azithromycin and Other Regimens in Marseille, France, 2020: A Monocentric Retrospective Analysis	2 111	271 (12.8%)	239 (11.3%)	2 154	271 (12.6%)	0.802	243 (11.3%) ^{‡‡}	0.967
Early combination therapy with hydroxychloroquine and azithromycin reduces mortality in 10,429 COVID-19 outpatients (BDD HDJ 2020)	10 429	24 (0.23%)	16 (0.15%) ^{†††}	10 973	29 (0.26%)	0.615	18 (0.16%) ^{‡‡‡}	0.845

Note: †: External data validation with PMSI data + updated number of deaths from French National Death Registry

‡: Chi-square test versus "published data".

††: 8 deaths added from French National Death Registry + 5 patients died after end of follow-up.

‡‡: 4 dead patients were added from the PMSI data.

†††: Only deaths attributed to Covid-19 were reported in the published data.

‡‡‡: 2 deaths non attributed to Covid-19 were added.

Quality Control by an Independent Adjudicator

The process to quality control the dataset was carried out by an independent adjudicator, who verified and attested to the presence of all the traceability elements and to the anonymisation process, guaranteeing the quality of the data in the dataset.

Disclosure Statement

The authors declare no support from any organisation for the submitted work, no financial relationships with any organisations that might have an interest in the submitted work in the previous three years, and no other relationships or activities that could appear to have influenced the submitted work.

Funding

This work was performed by academic doctors working in the IHU Méditerranée Infection. IHU Méditerranée Infection is funded

by the French government and received a grant from the Agence Nationale de la Recherche: ANR-15-CE36-0004-01 and the ANR "Investissements d'avenir", Méditerranée Infection 10-IAHU-03, and was also supported by the Région Provence-Alpes-Côte d'Azur.

Acknowledgements

The authors thank all participants involved in the construction of this database, the medical doctors who gave their time and provided the best treatment for their patients, the statisticians who gave their time for quality control and ad hoc analyses, and the adjudicators for the time spent observing the implementation of the quality of the database.

Author Contributions

DR and PB contributed to conception and design of the research, the acquisition of data and drafting and revision of the manuscript. Both authors read and approved the final manuscript.

Data Availability Statement

Raw data are publicly available online in two public open-access repositories. Science Data Bank [13] available at <https://doi.org/10.57760/sciencedb.07803> and DRYAD [14]. <https://doi.org/10.5061/dryad.ksn02v78v>. The conditions of reuse are covered by the Creative Commons Zero (CC0) license for both deposits. The SAS code is available upon request from the authors.

References

- (2023) John Hopkins University & Medicine. Johns Hopkins Coronavirus Resource Center.
- Doshi P, Godlee F, Abbasi K (2022) Covid-19 vaccines and treatments: we must have raw data, now. *BMJ* 376: o102.
- Godlee F (2020) Covid-19: The lost lessons of Tamiflu. *BMJ* 371: m4701.
- Mehra MR, Desai SS, Ruschitzka F, Patel AN (2020) RETRACTED: Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis. *Lancet*.
- Colson P, Fournier PE, Chaudet H, Delerce J, Giraud-Gatineau A, et al. (2022) Analysis of SARS-CoV-2 Variants From 24,181 Patients Exemplifies the Role of Globalization and Zoonosis in Pandemics. *Frontiers in Microbiology*.
- (1970) Fichiers des personnes décédées depuis 1970 | Insee [Internet]. [cited 2023 May 9].
- (2013) World Medical Association. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. *JAMA* 310(20): 2191-2194.
- (2017) Sciences C for IO of M. International ethical guidelines for health-related research involving humans. *International ethical guidelines for health-related research involving humans*.
- Boudemaghe T, Belhadj I (2017) Data Resource Profile: The French National Uniform Hospital Discharge Data Set Database (PMSI). *International Journal of Epidemiology* 46(2): 392-392d.
- Lagier JC, Million M, Gautret P, Colson P, Cortaredona S, et al. (2020) Outcomes of 3,737 COVID-19 patients treated with hydroxychloroquine/azithromycin and other regimens in Marseille, France: A retrospective analysis. *Travel Med Infect Dis* 36: 101791.
- Million M, Lagier JC, Tissot-Dupont H, Ravaux I, Dhiver C, et al. (2021) Early combination therapy with hydroxychloroquine and azithromycin reduces mortality in 10,429 COVID-19 outpatients. *Rev Cardiovasc Med* 22(3): 1063-1072.
- Lagier JC, Million M, Cortaredona S, Delorme L, Colson P, et al. (2020) Outcomes of 2111 COVID-19 Hospitalized Patients Treated with Hydroxychloroquine/Azithromycin and Other Regimens in Marseille, France, 2020: A Monocentric Retrospective Analysis. *Ther Clin Risk Manag* 18: 603-617.
- Monocentric Retrospective Cohort of 30,423 COVID-19 Patients [Internet].
- Dryad | Data -Monocentric retrospective cohort of 30,423 COVID-19 patients [Internet].

ISSN: 2574-1241

DOI: 10.26717/BJSTR.2023.52.008265

Philippe Brouqui. Biomed J Sci & Tech Res



This work is licensed under Creative Commons Attribution 4.0 License

Submission Link: <https://biomedres.us/submit-manuscript.php>



Assets of Publishing with us

- Global archiving of articles
- Immediate, unrestricted online access
- Rigorous Peer Review Process
- Authors Retain Copyrights
- Unique DOI for all articles

<https://biomedres.us/>