



COVID-19 Case Data for Italy Stratified by Age Class

DATA PAPER

GIUSEPPE CALAFIORE

GIULIA FRACASTORO

*Author affiliations can be found in the back matter of this article

]u[ubiquity press

ABSTRACT

The dataset described in this paper contains daily data about COVID-19 cases that occurred in Italy over the period from Jan. 29, 2020 to October 15, 2021, divided into ten age classes of the population, the first class being 0–9 years, the tenth class being >90 years. The dataset contains eight columns, namely: date (day), age class, number of new cases, number of newly hospitalized patients, number of patients entering intensive care, number of deceased patients, number of recovered patients, number of active infected patients. This data has been officially released for research purposes by the Italian authority for COVID-19 epidemiologic surveillance (*Istituto Superiore di Sanità* – ISS), upon formal request by the authors, in accordance with the Ordinance of the Chief of the Civil Protection Department n. 691 dated Aug. 4, 2020. A separate file contains the numerosity of the population in each age class, according to the *National Institute of Statistics*’ (ISTAT) data of the resident population of Italy as of Jan. 2020. This data has potential use, for instance, in epidemiologic studies of the effects of the COVID-19 contagion in Italy, in mortality analysis by age class, and in the development and testing of dynamical models of the contagion.

CORRESPONDING AUTHOR:

Giuseppe Calafiore

Dipartimento di Elettronica e Telecomunicazioni, Politecnico di Torino, IT

giuseppe.calafiore@polito.it

KEYWORDS:

COVID-19; Age classes; COVID Italy; Epidemiologic data

TO CITE THIS ARTICLE:

Calafiore G and Fracastoro G. 2022. COVID-19 Case Data for Italy Stratified by Age Class. *Open Health Data*, 9: 1. DOI: <https://doi.org/10.5334/ohd.34>

1. OVERVIEW

INTRODUCTION

The age structure of the population appears to play a key role in determining the severity of symptoms and the mortality of the disease caused by the SARS-CoV-2 infection. The importance of the demographic structure in determining the pandemic's progression and impact has indeed been well recognized by researchers, see, e.g., [1, 2]. Also, a clearer understanding of the contagion's interaction dynamics among age classes appears to be fundamental for devising effective containment measures and for establishing priorities for the vaccination campaigns. Despite the importance of age-related COVID-19 data, and despite the fact that calls for countries to provide this data have been repeatedly made (see, e.g., [2–4]) this type of data has been to date essentially unavailable to the public, and even to researchers. This fact motivated us to formally request specific age-related COVID-19 data to Italian authorities in charge of the COVID-19 surveillance (*Istituto Superiore di Sanità* – ISS), so to make them available to the public for research purposes.

2. CONTEXT AND METHODS

SPATIAL AND TEMPORAL COVERAGE

The data refers to the population of Italy and covers the period from Jan. 29, 2020 to October 15, 2021, with daily frequency. Data relative to the early phase of the contagion (i.e., previous to March 2020) have several missing values for some age classes.

METHODOLOGY AND QUALITY CONTROL

The data reported in the file are the data present in the Italian COVID-19 surveillance system, updated to the extraction date of October 15, 2021. The data represents aggregations of positive cases for SARS-CoV-2 derived from the Integrated Covid-19 Surveillance coordinated by the ISS (Ordonnance no. 640 of February 27, 2020). The Integrated Surveillance data is updated daily by each Region, both with new cases and with the addition of new information on cases already communicated previously, as they become available. In addition, the constant quality control of the data also seldom highlights the need, on the part of the Regions, to cancel some cases that are mistakenly duplicated.

The data collected is in a continuous phase of consolidation and, as expected in an emergency situation, some information is incomplete. In particular, the possibility of a delay of a few days between the execution of the swab for diagnosis and reporting on the dedicated platform is noted. Therefore, the number of cases observed in the most recent days, compared to the extraction date, must be interpreted as provisional and incomplete. The same applies to reporting hospitalization and death.

PRIVACY

The data reported are disaggregated in a manner that guarantees compliance with the privacy legislation. In particular, it should be noted that for frequency values between 1 and 4 the value is expressed as “<5”.

3. DATASET DESCRIPTION

OBJECT NAME

Two files are provided. The first file is the main COVID-19 data file named “*covid_ageclass_Italy.csv*” while the second file named “*ageclass_pop.csv*” is an ancillary file that contains the population cardinality for each age class.

FORMAT NAMES AND VERSIONS

File format is textual comma separated values (CSV).

DATASET CREATORS

The dataset was extracted from the national official database, upon request from the authors, by Dr. Patrizio Pezzotti from the Epidemiology, Biostatistics and Mathematical Models Department of the ISS.

LICENCE

The data is provided under the CC0-Public Domain Dedication waiver license.

REPOSITORY LOCATION

dataverse.harvard.edu: <https://doi.org/10.7910/DVN/VSS4CO>

Publication date: Nov. 22, 2021

DATA DESCRIPTION

The data file “*covid_ageclass_Italy.csv*” contains 6069 rows (plus the headings row) and eight columns. The columns contain the following data:

1. “date” contains the date indicating the day to which the data in the other columns refers. It is the date of the confirmed diagnosis of microbiological SARS-CoV-2 infection, or the date of hospitalization, the date of recovery, the date of death, etc.
2. “age_class” is the age class, in a ten-year range. In some rare cases it can be “unknown.”
3. “cases” contains the number of confirmed positive SARS-CoV-2 infected cases for that day in the given age class.
4. “hospitalized” contains the number of patients hospitalized (due to COVID) in that day in the given age class.
5. “intensive_care” contains the number of patients that entered intensive care (due to COVID) in that day in the given age class.
6. “deceased” contains the number of deceased persons (with death ascribed to COVID) in that day in the given age class.

- 7. “recovered” contains the number of persons that recovered (from COVID) in that day in the given age class.
- 8. “active_infected” contains the total number of persons that are active and infected with SARS-CoV-2 on the given day in the given age class.

The ancillary file data file “ageclass_pop.csv” contains 10 rows (plus the headings row) and two columns. The first column “age_class” contains the age class, the second column “population” contains the number of individuals resident in Italy for that age class, as of Jan. 2020.

DATA OVERVIEW

A cumulative summary of part of the data is shown in **Table 1**. Mortality is here computed simply as the ratio between deceased individuals in a given age class and the population of that class. Lethality is computed as the ratio between deceased individuals in a given age class and the infected individuals (cases) in that class. Values reported as “<5” in the data are imputed a default value of 2. **Figure 1** shows a pie chart of the deaths by age. **Figure 2** shows an example of time-series data representing the daily cases for the 50–59 age class; the regular spikes in the plot correspond to Sundays. **Figure 3** shows the time-

| AGE CLASS | POPULATION | CASES | INTENSIVE CARE | DECEASED | %MORTALITY | %LETHALITY |
|------------|-----------------|----------------|----------------|---------------|---------------|---------------|
| 0–9 | 4892494 | 273207 | 166 | 34 | 0.0007 | 0.0124 |
| 10–19 | 5706116 | 491752 | 233 | 44 | 0.0008 | 0.0089 |
| 20–29 | 6084382 | 585913 | 540 | 148 | 0.0024 | 0.0253 |
| 30–39 | 6854632 | 595830 | 1258 | 399 | 0.0058 | 0.0670 |
| 40–49 | 8937229 | 747765 | 3493 | 1215 | 0.0136 | 0.1625 |
| 50–59 | 9414195 | 793382 | 9353 | 4848 | 0.0515 | 0.6111 |
| 60–69 | 7364364 | 500624 | 15637 | 13825 | 0.1877 | 2.7616 |
| 70–79 | 5968373 | 360470 | 18035 | 33309 | 0.5581 | 9.2404 |
| 80–89 | 3628160 | 267017 | 8970 | 53126 | 1.4643 | 19.8961 |
| >=90 | 791543 | 91373 | 1606 | 25578 | 3.2314 | 27.9930 |
| ALL | 59641488 | 4707333 | 59291 | 132526 | 0.2222 | 2.8153 |

Table 1 Summary table.

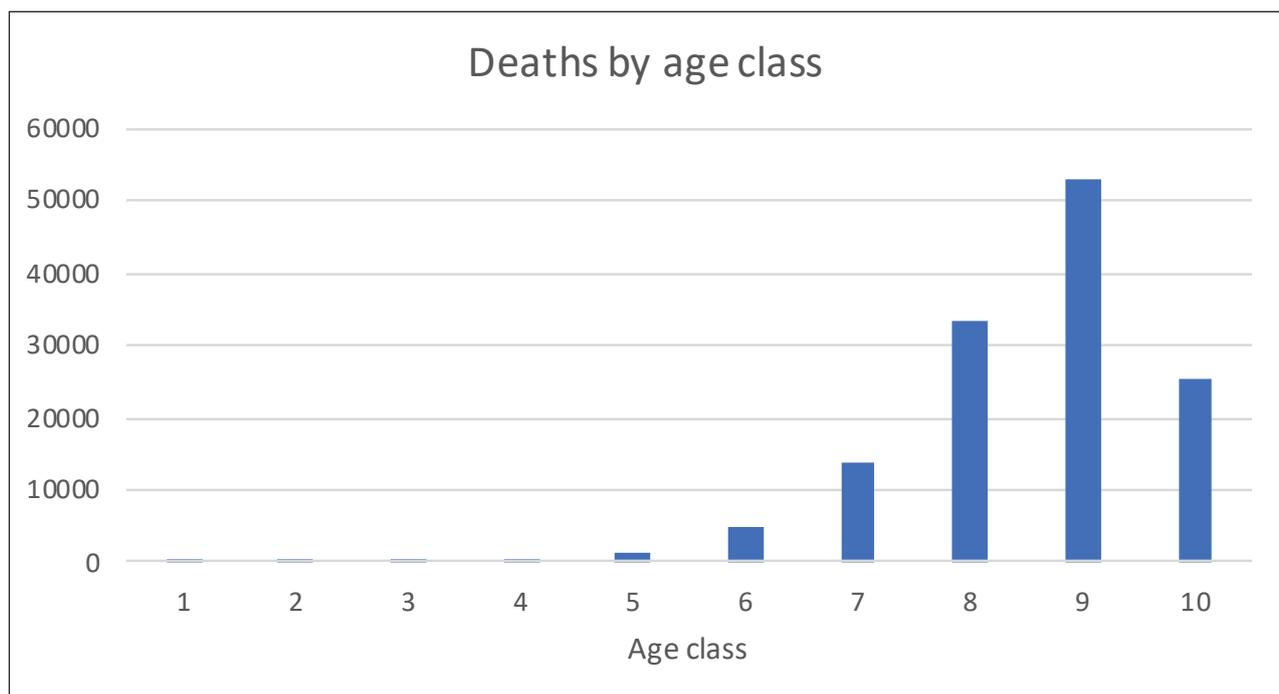


Figure 1 Deaths by age class.

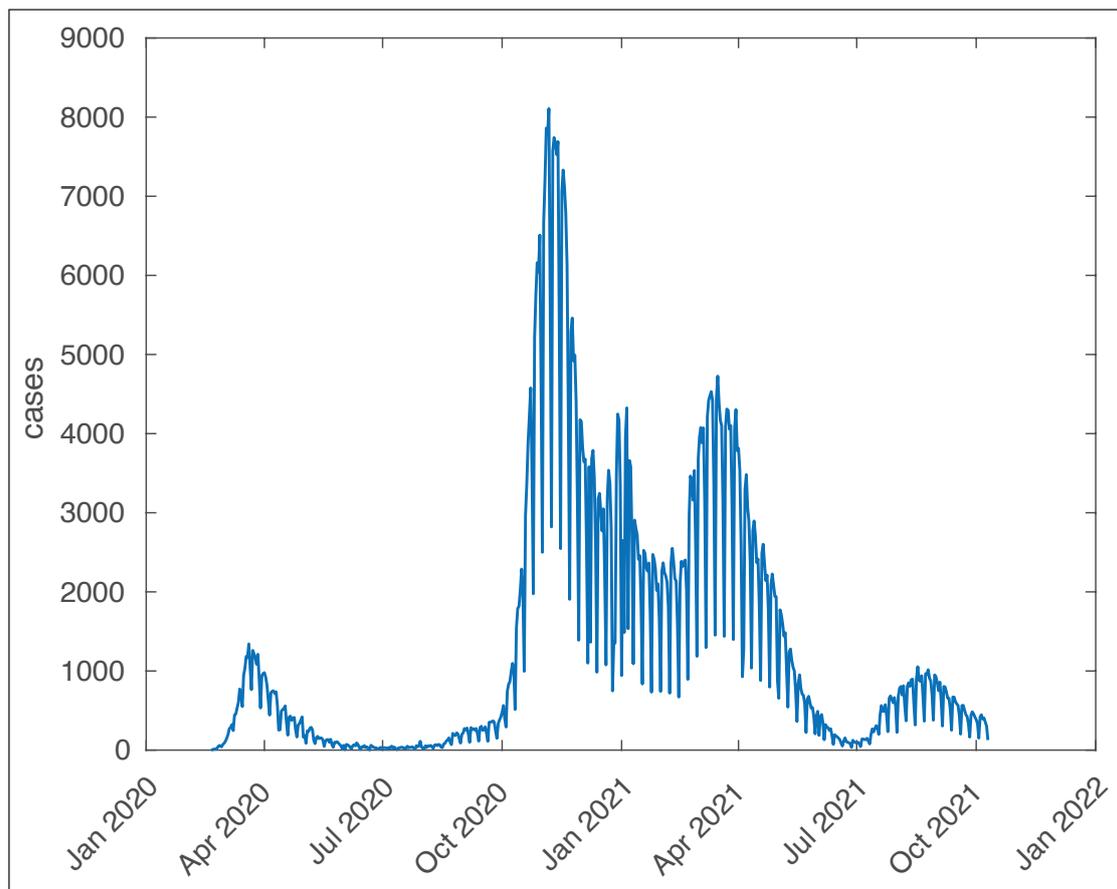


Figure 2 Cases for the 50-59 age class.

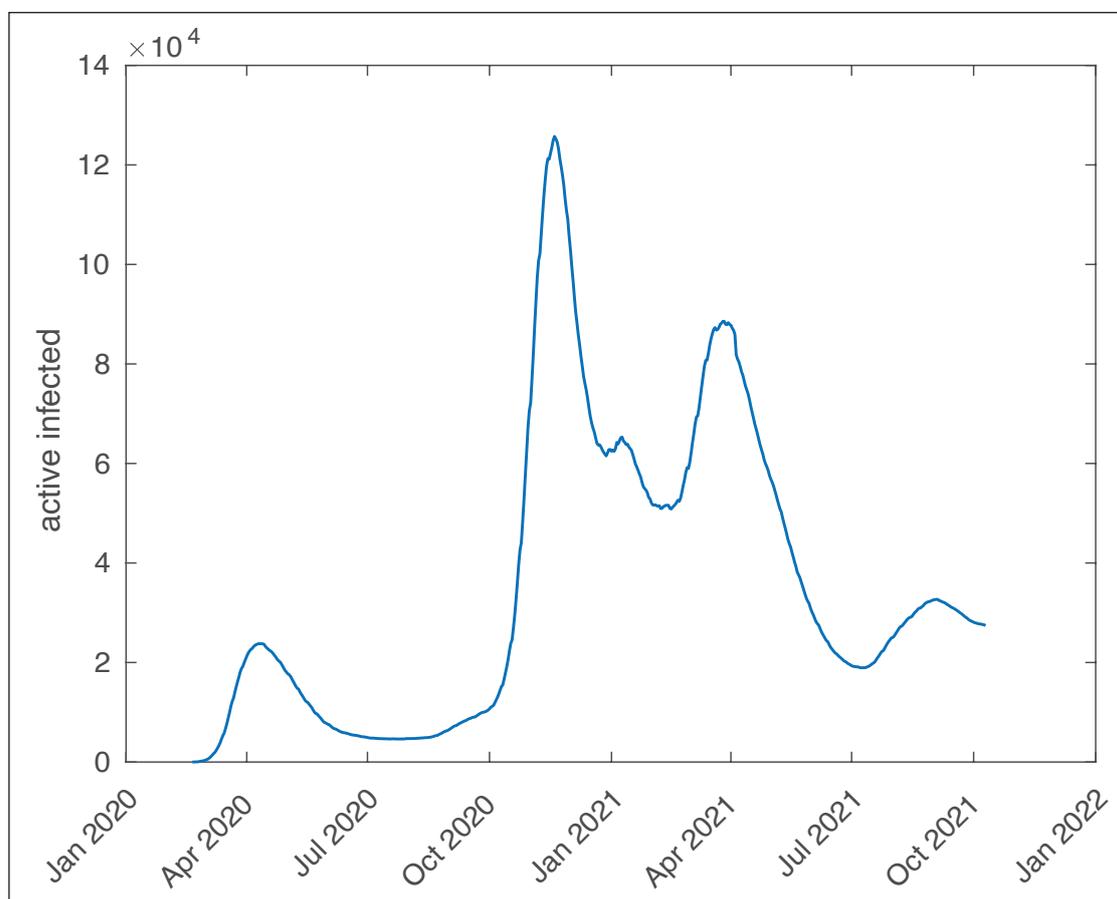


Figure 3 Active infected individuals for the 50-59 age class.

series of the active infected individuals for the 50–59 age class; three major infection peaks are visible, the first in mid-April 2020, the second in late November 2020, and the third in early April 2021.

4. REUSE POTENTIAL

The data can be used for research purposes, including aggregation, analysis, reference, model (e.g., SIRD) building and validation, teaching or collaboration.

DATA ACCESSIBILITY STATEMENT

This data paper is available as a preprint on arXiv at <https://arxiv.org/abs/2104.06199>.

ACKNOWLEDGEMENTS

We acknowledge the help of prof. Andrea Bianco, head of the Department of Electronics and Telecommunications Engineering of Politecnico di Torino, Italy, for his help in managing the formal data request to ISS.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Giuseppe Calafiore  orcid.org/0000-0002-6428-5653

Dipartimento di Elettronica e Telecomunicazioni, Politecnico di Torino, IT

Giulia Fracastoro  orcid.org/0000-0001-8495-1097

Researcher, Dipartimento di Elettronica e Telecomunicazioni, Politecnico di Torino, IT

REFERENCES

1. **Down JB**, et al. Demographic science aids in understanding the spread and fatality rates of COVID-19. *Proceedings of the National Academy of Sciences*. 2020; 117(18): 9696–9698. DOI: <https://doi.org/10.1073/pnas.2004911117>
2. **Omori R, Matsuyama R, Nakata Y**. The age distribution of mortality from novel coronavirus disease (COVID-19) suggests no large difference of susceptibility by age. *Scientific Reports*. 2020; 10(1): 1–9. DOI: <https://doi.org/10.1038/s41598-020-73777-8>
3. **Nepomuceno MR**, et al. Besides population age structure, health and other demographic factors can contribute to understanding the COVID-19 burden. *Proceedings of the National Academy of Sciences*. 2020; 117(25): 13881–13883. DOI: <https://doi.org/10.1073/pnas.2008760117>
4. **Calafiore G, Fracastoro G**. Age class structure in SIRD models for the COVID-19 - An analysis of Tennessee data. *Proceedings of the IEEE MED Conference*; 2021. DOI: <https://doi.org/10.1109/MED51440.2021.9480314>

TO CITE THIS ARTICLE:

Calafiore G and Fracastoro G. 2022. COVID-19 Case Data for Italy Stratified by Age Class. *Open Health Data*, 9: 1. DOI: <https://doi.org/10.5334/ohd.34>

Published: 26 April 2022

COPYRIGHT:

© 2022 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Open Health Data is a peer-reviewed open access journal published by Ubiquity Press.