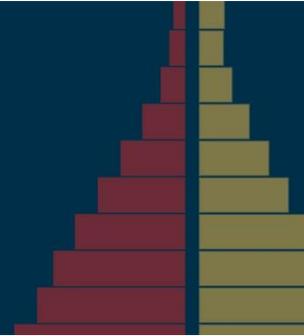


BEYOND THE NUMBERS: COVID-19 and the Philippine Population

Research Brief No. 4 | 8 May 2020



Reliable data needed to address COVID-19

Background

The University of the Philippines Population Institute (UPPI) and the Demographic Research and Development Foundation (DRDF), intending to contribute to the understanding of the on-going COVID-19 pandemic in the Philippine context, are compiling information coming from the Department of Health (DOH). Using this and data from the UPPI and DRDF surveys conducted in the Philippines, along with data from the Philippine Statistics Authority (PSA) that are publicly available, we can have a grasp of how the pandemic will affect Filipinos. Analyses are released through this series of research briefs we call “Beyond the Numbers: COVID19 and the Philippine Population”.

On 15 March 2020, Metro Manila was put under community quarantine. On 17 March, the community quarantine was extended to the rest of Luzon. Beginning 18 March, UPPI and DRDF started monitoring the daily DOH updates, manually copying data from official DOH sources to adapt them into a format that can easily be analyzed by our UPPI-DRDF Team.

Data sources

Initially, data came from the DOH ncovtracker¹. This website housed a dashboard using Esri’s ArcGIS, which is also used by Johns Hopkins University in the US² and Italy’s *Dipartimento della Protezione Civile*³. We were able to copy information from each case, particularly:

1. Age
2. Sex
3. Nationality
4. Province of residence
5. Travel history
6. Exposure to known COVID-19 cases
7. Date of lab confirmation
8. Facility of admission/consultation
9. Epidemiologic link
10. Status of condition (Mild, Asymptomatic, Severe, Critical, Expired, Recovered)

Information was also shared through the DOH website, its official Facebook and Twitter pages, the DOH PH COVID-19 Viber community, and the PH Coronavirus Updates Telegram channel. Documents from a shared Google Drive⁴ that compiled communication materials were also used to verify information and add details if available. Through these other channels, we were able to add the following information to our database:

1. Symptoms
2. Onset of symptoms

¹ <https://ncovtracker.doh.gov.ph/>

² <https://coronavirus.jhu.edu/map.html>

³ <http://opendatadpc.maps.arcgis.com/apps/opsdashboard/index.html#/b0c68bce2cce478eaac82fe38d4138b1>

⁴ <http://bit.ly/COVIDPH>

3. Date of admission
4. Date swabbed for testing
5. Date of discharge (if applicable)
6. Date of death (if applicable)
7. Cause of death (if applicable)
8. Comorbidities

However, at some point, more and more cases had details on symptoms, dates, epidemiologic link, and status of condition marked “For verification” or “For validation” until those variables were eventually removed. Since 2 April 2020, we have relied on news reports and bulletins from regional DOH offices to update our database. The DOH regional bulletins do not have a uniform format.

On 13 April 2020, DOH launched a Tableau-based dashboard⁵ and on 14 April 2020, data used for this dashboard was released through a publicly accessible Google Drive⁶ with an invitation to “Go forth and research!”. This COVID-19 Data Drop is currently the primary source of our research briefs.

COVID-19 Data Drop

The file structure of the Data Drop has gone through many iterations since its launch. As of this writing, it has two folders and 1 PDF (Policy and Confidentiality Statement). The two folders are (1) COVID-19 DATA, which have sub-folders that group data releases by month and by date, and (2) COVID-19 SITUATIONER, which contains PDFs of the “Beat COVID-19 Today Situationer” launched on 28 April 2020.

File formats have changed as well. From the initial Google Sheets and CSV combination, the present iteration delivers 1 Excel file containing different types of data per sheet and the CSV versions of these sheets.

The Excel worksheets are as follows:

1. Metadata – Sheets (details of the sheets in the Excel file)
2. Metadata – Fields (details of each field/variable in the Excel file)
3. Case Information (details of each case)
4. Daily Report (information from hospitals and infirmaries related to capacity)
5. Weekly Report (information from hospitals and infirmaries related to supplies)
6. Testing Aggregates (information from laboratories)

Detailed descriptions of items 3 through 6 are in the Technical Notes PDF included in each date sub-folder. For the database being used in the research briefs, we focus on the third sheet of the Excel file. As of this writing, this sheet contains the following information⁷:

Variable	Description
CaseCode	Random code assigned for labeling cases; does not equate to the unique case number assigned by DOH
Age	Age
AgeGroup	Five-year age group
Sex	Sex
DateRepConf	Date publicly announced as confirmed case

⁵ <https://www.doh.gov.ph/covid19tracker>

⁶ It was initially a Sheet at <http://bit.ly/DOHDataDrop> but was migrated to a full Drive at <http://bit.ly/dohcovid19data> on 15 April 2020 due to heavy traffic

⁷ Descriptions are from Sheet2 (Metadata – Fields) of the Excel file

Variable	Description
DateRecover	Date recovered
DateDied	Date died
RemovalType	Type of removal (recovery or death)
DateRepRem	Date publicly announced as removed
Admitted	Binary variable indicating patient has been admitted to hospital
RegionRes	Region of residence
ProvCityRes	Province of residence
RegionPSGC	Philippine Standard Geographic Code of Region of Residence
ProvPSGC	Philippine Standard Geographic Code of Province of Residence
MunCityPSGC	Philippine Standard Geographic Code of Municipality or City of Residence
HealthStatus	Known current health status of patient (asymptomatic, mild, severe, critical, died, recovered)
Quarantined	Ever been home quarantined, not necessarily currently in home quarantine

Variables `RegionPSGC`, `ProvPSGC`, `MunCityPSGC` were added on 23 April. `HealthStatus` was added on 24 April, while `Quarantined` was added on 29 April 2020. Data is shared between 4:30 PM and 9:00 PM daily.

Remarks on the COVID-19 Data Drop

The following are our observations on DOH's daily deliveries since 14 April 2020 which are merged (herewith referred to as version 2) with the original database constructed from the `ncovtracker` and other DOH channels (which we refer to as version 1).

1. *Variables no longer available in the Drop*

The main observation is that variables we were able to source from different DOH releases are not offered in the Drop. These variables are:

- Nationality
- Travel history
- Exposure to known COVID-19 cases/Epidemiologic link
- Date swabbed for testing
- Date of lab confirmation
- Facility of admission/consultation
- Symptoms
- Onset of symptoms
- Date of admission
- Cause of death
- Comorbidities

These variables would allow researchers numerous avenues of enquiry, for example: the connection between epidemiologic link and travel history to infections, recovery, and death. It would also allow mapping areas with increased risk of person-to-person spread⁸ due to population density compounded by number of cases in that area, the virus' incubation period,

⁸ How COVID-19 Spreads. (2020, April 13). Retrieved from <https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/how-covid-spreads.html>

outcomes of patients with comorbidities^{9,10}, contribution to knowledge about deaths attributable to COVID-19¹¹, and many others. It would be more useful if these variables will be included again in the data that DOH releases regularly.

2. Changing date formats

Initially, dates under `DateRecover` and `DateDied` were encoded in M/D/YYYY format, sometimes showing unrealistic information. In the 21 April 2020 Drop, for example, there were several recovery dates and dates of deaths that were later than April. On 22 April, the format was changed to DD-MMM-YYYY.

3. Mismatched Case Codes

Using Case Codes as unique identifiers for each case and under the assumption that these will not change once assigned, there have been two dates so far when the codes did not line up correctly with the cases: Data Drops on 17 April and 5 May 2020. Consequently, Database v.2 was not updated on these dates.

The succeeding comparisons exclude these two dates.

4. Changes in the Age variable

Out of the 9,485 confirmed cases as of 4 May 2020, 218 (2.3%) had at least 1 change in age information. Excluding the apparent mistyping of age 456 for C569615, on 30 April (corrected to 45 years old the next day), changes in age range from 1 (unknown if this is a product of computing against the date of birth) to 68 years. Below are examples from data on April 25 and 26:

CaseCode	25-Apr	26-Apr	Diff
C653368	71	6	65
C774599	36	76	40
C621158	2	39	37
C418474	38	28	10
C596566	38	29	9

In addition, of the 7,192 confirmed cases as of 25 April 2020, 75 (or 1%) of these had changes in age information compared to data on 24 April. The changes range from 1 to 59 years (age 21 on 24 April and age 80 based on the 25 April data). A change of 59 years is not negligible, considering that those in the older age groups are more vulnerable to severe infection from the virus than others.

⁹ Yang, J., Zheng, Y., Gou, X., Pu, K., Chen, Z., Guo, Q., Ji, R., Wang, H., Wang, Y., & Zhou, Y. (2020). Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. *International Journal of Infectious Diseases*, 94, 91–95. doi: 10.1016/j.ijid.2020.03.017

¹⁰ Guan, W. J., Liang, W. H., Zhao, Y., Liang, H. R., Chen, Z. S., Li, Y. M., Liu, X. Q., Chen, R. C., Tang, C. L., Wang, T., Ou, C. Q., Li, L., Chen, P. Y., Sang, L., Wang, W., Li, J. F., Li, C. C., Ou, L. M., Cheng, B., ... He, J. X. (2020, March 26). Comorbidity and its impact on 1590 patients with Covid-19 in China: a nationwide analysis. *European Respiratory Journal*, 2000547. Advance online publication. doi: 10.1183/13993003.00547-2020

¹¹ Vincent, J-L., & Taccone, F.S. (2020, April 6). Understanding pathways to deaths in patients with COVID-19. *The Lancet Respiratory Medicine*, 8(5), 430-432. doi: 10.1016/S2213-2600(20)30165-X

5. Changes in the Sex variable

Out of the 9,485 confirmed cases as of 4 May 2020, 101 (1.1%) had problems with encoded information on sex. Majority of these cases involve a change Male to Female, or vice versa; below are examples where the information changed from Male to Female and back to Male:

CaseCode	Coded Male	Coded Female	Coded Male
C331958	24-Apr	25-Apr	26-Apr
C345588	15-Apr	21-Apr	24-Apr
C540600	15-Apr	18-Apr	19-Apr
C791729	22-Apr	25-Apr	29-Apr

Obtaining correct information about the patient's sex is important because there are significant differences in mortality and morbidity that are linked to sex and studies from other countries has so far shown that in the case of COVID-19, mortality is higher among males than females.

6. Changes in RemovalType variable

Two cases were initially declared "Died" but were changed to "Recovered". Lacking information from other DOH sources, their status was set to "Recovered".

CaseCode	Died	Recovered
C494273	15-Apr	21-Apr
C995502	15-Apr	22-Apr

Determining the correct date of death and matching this up with the correct number of cases is important, for example, in the computation of the ratio of deaths to infections and tracking this over time. The number of deaths per day inform projections on the spread of the virus, the doubling time, reproduction rate and the effects of the physical distancing measures that are in place¹².

7. Changes in place of residence variables

Out of the 9,485 confirmed cases as of 4 May 2020, only 8,488 have information on place of residence (89%). Of these, 280 had at least 1 change in information regarding place of residence (3.3%). Most are simply the in-and-out of the same place across different Drops, but the rest reflect changes in location. Some examples are:

- C735916: City of Manila on April 15 and 18, but no information on April 16 and on April 19 to 27; coded Quezon City from April 28 to May 3, but reverted to City of Manila on May 4
- C715421: Pasay City on April 15 and 18, City of Parañaque City from April 24
- C765730: Pasay City on April 15 and 18, Ilocos Norte from April 24
- C994912: Cavite on April 15 then no entry until April 23; Batangas on April 24; reverted to Cavite on April 25
- C211863: City of Makati on April 15; Pampanga on April 16 to 27; reverted to City of Makati from April 28

8. Reverting to no information

There are some cases where one Drop provides information on a Case and in the next, this information reverts to missing. This occurs particularly in `RemovalType`, `DateRecover`, `DateDied`, and in the place of residence.

¹² Subbaraman, N. (2020, April 9). Why daily death tolls have become unusually important in understanding the coronavirus pandemic. Retrieved from <https://www.nature.com/articles/d41586-020-01008-1>

For example, DOH reported 631 closed cases on 15 April 2020 and 797 on 16 April. However, 68 cases were already reported dead or recovered on 15 April whose status went back to no information on 16 April. This implies that on 16 April, there should have been 865 reported closed cases as we assume that the data on 15 April is correct until otherwise corrected in future Drops.

RemovalType	15-Apr (A)	16-Apr (B)	Reverted* (C)	Cases in Database v. 2 for 16-Apr (B+C)
Recovered	295	435	66	501
Died	336	362	2	364
Total	631	797	68	865

* With status in 15-Apr but reverted to no information (blank) on 16-Apr

Out of the 1,938 declared closed as of 4 May 2020, 55 cases (2.8%) have shifted state in combinations of “Recovered”, “Dead”, and no information. The 55 cases include the two mentioned in the previous item.

Additionally, on April 22, 34 cases that already had information on `DateRecover` on April 20 were reverted to no information; information on April 21 was not used because it had date format issues.

Meanwhile, 6 cases have changes in `DateDied`:

CaseCode	16-Apr	18-Apr	22-Apr	28-Apr	29-Apr	30-Apr	1-May
C133008	24-Mar			25-Mar			
C228313					14-Apr	no info	
C494273	4-Apr	no info					
C612797	26-Mar		no info				
C711507		6-Apr				no info	26-Apr
C713217	28-Mar					no info	

Finally, 243 cases with information on place of residence on 25 April were reverted to blanks in the data a day later.

9. Wrong categorization or wrong date entry

One case (C727220) has `RemovalType` “Died” but the removal date (10 April) is under `DateRecover`. We are uncertain if the error is on the `RemovalType` or the variable in which the date was encoded.

10. Inconsistencies with previously released information

We were able to connect some Case Codes with their DOH case numbers. For the research briefs, we retained the original information. We assumed them to be more accurate particularly if they came from the detailed Case Bulletins. Some examples:

Ncovtracker	Data Drop	Data source apart from ncovtracker
PH 0031 – Recovered Mar 22	C501602 – Recovered Mar 24	Case Bulletin #011 (March 25)
PH 0145 – 32 M	C229116 – 30 M	Facebook: officialDOHCHDSOCCSKSARGEN
PH 0389 – 71 F	C583663 – 75 F	Case Bulletin #013 (March 27)
PH 1488 – Expired Mar 26 5:58 PM	C612797 – Died Mar 30	Case Bulletin #016 (March 30)
PH 1810 – 42 M	C637588 – 48 F	Facebook: dohdavao

While encoding data from `ncovtracker`, we already experienced changing information almost on a daily basis. For this reason, we adopted a standard procedure that we maintained even in the introduction of the Data Drop, that is, to retain information from detailed Case Bulletins and

other DOH websites unless otherwise verified. We also retain information from the most recent Data Drop when succeeding Data Drops revert to no information for certain variables of some cases.

Research briefs in context

The first three research briefs we released used version 1 of the database while the succeeding research briefs will be using version 2.

While recognizing the difficulty of collecting and collating data comprehensively, we would like to appeal to DOH to share to the public, all available information they have collected on COVID-19 cases and to do so on a timely basis.

As with all data, we practice caution in interpretation. Aside from the underlying undercount as an artifact of testing, factors including clerical error and those described above also affect the data we have for analysis. Until the pandemic is over and the data is cleaned, perhaps we cannot consider the details in the dataset as final.

As demographers, our analyses are only as good as the data we use.

This research brief was prepared by Klarriness P. Tanalgo, Micaella Lou-Andrea M. Garduce, Jodie Mae U. Penado, Mark Ryan B. Paguirigan, Maria Midea M. Kabamalan, Elma P. Laguna, and Maria Paz N. Marquez

Suggested citation: University of the Philippines Population Institute (UPPI) and Demographic Research and Development Foundation, Inc. (DRDF). (2020, May). *Reliable data needed to address COVID-19*. Retrieved from <https://www.uppi.upd.edu.ph/sites/default/files/pdf/Reliable-data-needed-to-address-COVID-19.pdf>.

