

# UP experts spot 'alarming errors' in DOH's COVID-19 patient data

*Ratziel San Juan*

UP experts spot 'alarming errors' in DOH's COVID-19 patient data

[Ratziel San Juan](#) (Philstar.com) - May 12, 2020 - 4:07pm

MANILA, Philippines — The University of the Philippines COVID-19 Pandemic Response Team in their policy [note](#) released Tuesday pointed out errors and discrepancies in the Department of Health's data on COVID-19 patients.

Comparing official data published only a day apart on April 24 and 25, experts from the state university found that 45 cases had their recorded sex changed (male to female or vice-versa), 75 patients got older or younger by as much as decades, 516 had their residence moved to another city or a "completely imaginary city" (like a barangay or district) and even a patient initially reported dead being reclassified as alive the next day.

Figure 2. Example of errors in reported sex, age, and residence of confirmed cases as released by DOH on 24 April and 25 April 2020

Sex			Age			City/Municipality of Residence		
Case#	April 24	April 25	Case#	April 24	April 25	Case#	April 24	April 25
xx66	M	F	Cxxxx47	28	29	Cxxxx19	City of Manila	Sampaloc
xx61	M	F	Cxxxx22	70	76	Cxxxx29	City of Manila	Sampaloc
xx29	M	F	Cxxxx05	48	49	Cxxxx73	City of Manila	Santa Ana
xx57	F	M	Cxxxx92	36	35	Cxxxx51	City of Manila	Santa Cruz
xx27	F	M	Cxxxx82	8	1	Cxxxx10	City of Manila	Pandacan
xx12	F	M	Cxxxx16	31	28	Cxxxx78	Quezon City	City of San Juan
xx20	M	F	Cxxxx91	21	46	Cxxxx16	City of Manila	Sampaloc
xx70	F	M	Cxxxx61	28	70	Cxxxx23	City of Manila	Santa Cruz
xx25	M	F	Cxxxx96	39	34	Cxxxx58	City of Manila	Ermite
xx78	M	F	Cxxxx29	60	28	Cxxxx08	City of Manila	Sampaloc

Total erroneous cases: 45      Total erroneous cases: 75      Total erroneous cases: 516

Moreover, the health department logged conflicting location data with inconsistencies in the coding of the Philippine Geographic Code (PSGC), which is ironically supposed to provide "uniform standards and classification systems in the generation of government statistics to ensure harmonization and comparability of statistics in the country and at the international level," according to the Philippine Statistics Authority.

The UP team discovered overlapping codes used to describe the same region.

3. Data inconsistencies from DOH data drop on 6 May 2020

Row Labels	Count of DateRepConf	Row Labels	Count of DateRepConf
DR	6596	Region IV-B: MIMAROPA	28
I030000000	2	PH040000000	1
I040000000	5	PH170000000	26
I130000000	6581	(blank)	1
I140000000	1	Region IX: Zamboanga Peninsula	49
(blank)	7	PH090000000	46
Region III: Central Luzon	389	(blank)	3
I030000000	383	Region V: Bicol Region	55
I130000000	6	PH050000000	48
Region IV-A: ILAGAN LABARZON	1245	(blank)	7
I040000000	1237	Central Visayas	1180
I130000000	7	Region VII:	
(blank)	1	PH070000000	1179
		PH130000000	1
Grand Total: 9542*			

Similarly, DOH data reflect different data formats, adopting three different systems on April 14 (MM/DD/YYYY), April 22 (DD-MMM-YYYY) and May 5 (YYYY-MM-DD), respectively.

This "makes it difficult for automated systems of extracting and updating data from case information," read the report dated May 8.

Table 4. Inconsistency in reporting of date formats

Date Format	Date First used by DOH Data Drop
/DD/YYYY (e.g., 04/14/2020)	April 14, 2020
MMM-YYYY (e.g., 27-Apr-2020)	April 22, 2020 ☐ except DateRepRem (Date of report of removal, which is recovery or death), which is MM/DD/YYYY, ☐ all dates in similar format by April 24, 2020
Y-MM-DD (e.g., 2020-04-14)	May 5, 2020

COVID-19 data reporting protocols were also noted to have changed often since March, with certain variables appearing, disappearing and some later reappearing.

These include data reported (COVID-19 case count, PUI count, PUM count, demographics, health information, death information, date of recovery), case number format, and reporting platform (case summary in PDF format, data drop via Google Drive, nCoV Tracker individual case info).

Table 1. Timeline of changes in DOH's COVID-19 data reporting

	DATE									
	3/10 to 3/11	3/12	3/13	3/14	3/15 to 3/16	3/17 to 4/2	4/3 to 4/6	4/7 to 4/13	4/14 to 4/26	
I. Data Reported										
1. Count of COVID-19 Positive Cases										
2. Count of PUI										
3. Count of PUM										
4. Case Information										
a. Demographics										
- Age										
- Sex										
- Residence										
- Nationality										
b. Health Information										
- History of travel/exposure										
- Date of Onset										
- Date of Admission										
- Date of Laboratory Confirmation										
- Health Facility Admitted										
c. Death Information										
- Date of Death										
- Cause										
- Co-morbidity										
d. Date of Recovery										
II. Format of Case #										
1. PHx, x=1,2,...										
2. Cx, x is random										
III. Reporting Platform										
1. Summary of Case Information - pdf format										
2. nCoVtracker individual case information										
3. data drop - google drive										

“These lapses may seem small relative to the total size of data contained in the daily updates, but they have significant implications on the reliability of our scientific analyses on COVID-19,” wrote the UP team.

“As we have repeatedly pointed out since Day 1, any analysis is only as good as the data that we have...We do recognize the difficulties in the collection of detailed and timely records of COVID-19 data on a nationwide scale. However, no matter how difficult, these problems need to be addressed at the soonest possible time and should be of top priority.”

Aside from errors contained in the data itself, the team affirmed the observations of netizens that DOH data does not always match those of local government units, citing how the Health department reported 22 deaths and 65 recoveries fewer than the provincial tally of Laguna cases on May 3.

“Almost three months since we had our first confirmed case in the person of a Chinese national on 30 January, we have yet to reconcile differences in numbers between DOH and LGU sources,” read the report.

The team, however, recognized that the health department "has been open to feedback" and said that "this has been reflected in the improvement in reporting," citing the government's live COVID-19 [tracker](#) and habit of reporting cases at a uniform time daily (4:00 p.m.).

“However, there is still room for improving the quality of the data...It is important to make sure that correct data is captured as swiftly as possible to minimize uploading of erroneous and anomalous, if not missing, data.”