

# COVID-19 Case Surveillance Public Use Data with Geography Utility Summary

Users should consider the level of completeness, including suppression levels when planning their analyses and use of public datasets. Privacy protections will suppress field values to reduce reidentification risks. Completeness varies by jurisdiction (i.e., state, local, and territorial) and time period. Variables are consistently coded to the value “Unknown” when jurisdictions specify in the case data submitted to CDC that the value is unknown, the value “Missing” when jurisdictions do not provide a value, and the value “NA” when the value is suppressed as part of privacy protections.

Dataset version: 2021-08-03

Total records in dataset: 27,879,433

## Quick Summary

	all_fields	quasi_fields
total_fields	19	8
total_records	27,879,433	27,879,433
total_cells	529,709,227	223,035,464
missing_fields	111,554,840	13,255,484
missing_pct	21%	6%
complete_fields	418,154,387	209,779,980
complete_pct	79%	94%
unknown_fields	25,585,971	13,272,930
unknown_pct	5%	6%
suppressed_fields	16,555,445	15,002,459
suppressed_pct	3%	7%
available_fields	376,012,971	181,504,591
available_pct	71%	81%

## Field-level Utility Summary

	suppressed	suppressed_percent	missing	missing_percent
case_month	0	0.0%	0	0.0%
res_state	1,067	0.0%	0	0.0%
res_county	1,551,919	5.6%	0	0.0%
age_group	336,062	1.2%	185,083	0.7%
sex	879,455	3.2%	32,835	0.1%
race	4,973,923	17.8%	2,176,676	7.8%
ethnicity	5,878,345	21.1%	1,639,315	5.9%
death_yn	1,381,688	5.0%	9,221,575	33.1%
records_with_any_field	7,881,138	28.3%	10,383,332	37.2%